PLANT PATHOGEN evolution in THREE DIMENSIONS

DAVID E. TORRES SÁNCHEZ

Propositions

1. Transposable elements are the most potent weapons of plant pathogens.
   (this thesis)

2. The natural selection of specific mutations is determined by the epigenetic landscape.
   (this thesis)

3. The complexity of microbial communities challenges the accuracy and interpretation of metagenome assemblies.

4. The use of artificial intelligence in plant breeding imposes a risk for food security.

5. Dutch academia flourishes at the expense of the global south.

6. The recognition of queer people in society perpetuates racist and sexist hierarchies.

Propositions belonging to the thesis, entitled

"Plant pathogen evolution in three dimensions"

David Eduardo
Torres Sánchez
Wageningen, 18 September
2023

Plant pathogen evolution in three dimensions

David E. Torres Sanchez

David E. Torres Sanchez
Plant pathogen evolution in three dimensions,
306 pages.

PhD thesis, Wageningen University, Wageningen, the Netherlands (2023)
With references, with summary in English

ISBN 978-94-6447-799-3
DOI https://doi.org/10.18174/634751

Plant pathogen evolution in three dimensions

David E. Torres Sanchez

Thesis

submitted in fulfilment of the requirements for the degree of doctor
at Wageningen University
by the authority of the Rector Magnificus,
Prof. Dr A.P.J. Mol,
in the presence of the
Thesis Committee appointed by the Academic Board
to be defended in public
on Monday 18 September 2023
at 11 a.m. in the Omnia Auditorium.

Thesis committee

Promotor
Prof. Dr B.P.H.J. Thomma
Professor of Phytopathology
Wageningen University & Research

Co-promotor
Dr M. F. Seidl
Associate professor, Theoretical Biology & Bioinformatics
Utrecht University

Other members
Prof. Dr E.H. Stukembrock, Max Planck Institute of Evolutionary Biology, Plön, Germany and Christian-Albrechts University of Kiel, Germany
Prof. Dr B.J. Zwaan, Wageningen University & Research
Dr R.A. Ohm, Utrecht University
Dr M. Bemer, Wageningen University & Research

This research was conducted under the auspices of the Graduate School Experimental Plant Sciences.

Table of contents

Summary

Despite the enormous amount of DNA that needs to fit in the eukaryotic nucleus, genomes are typically highly organized. First, genes and other genetic elements, such as transposable elements, are arranged in an organized manner on the linear DNA strand. Furthermore, histone proteins fold the DNA into chromatin that is organized into regions that are more compacted or more relaxed, thereby affecting accessibility of the DNA. Finally, chromatin is arranged in the three-dimensional (3D) space within the nucleus, with some chromatin regions (co)localizing in nuclear sub-regions. Thus, one can say that genomes are organized in three dimensions: linear DNA, chromatin, and 3D organization. The three dimensions and the interactions between them compose the epigenome. The work described in this thesis combines complementary analyses across these three dimensions to try and understand genome evolution of *Verticillium dahliae*, a pathogenic fungus that infects hundreds of plant species.

In **Chapter 1**, I introduce the interactions between plants and fungal plant pathogens that engage in co-evolutionary 'arms races'. Plants have evolved immune systems to detect pathogen intrusions and to mount defense responses, while fungal pathogens evolved complex molecular tool kits to disarm plants' defenses and support host colonization. To enable these cycles of adaptation and counter-adaptation, pathogen populations require sufficient genetic variation, which occurs in specific dynamic genomic compartments in many plant pathogens. In *V. dahliae*, these dynamic genomic compartments are embedded within the eight chromosomes that are typically observed in this species, and are enriched in genomic rearrangements, transposable elements, presence/absence polymorphisms, unique chromatin characteristics, and in *in planta*-induced genes that encode proteins mediating virulence during host infection. In *V. dahliae*, these dynamic genomic compartments are known as adaptive genomic regions (AGRs). I argue that the intimate relationship between DNA, gene expression, chromatin, the spatial organization of the genome, and therefore, the epigenome, contributes to *V. dahliae* genome evolution.

The 'two-speed' genome model has often been used to describe genome organizations of plant pathogens. This model describes the compartmentalization of genomes characterized by the presence of dynamic regions that evolve more rapidly when compared with conserved regions that compose the core genome. In **Chapter 2,** we critically review the two-speed model by discussing recent work on epigenetics, transposable element dynamics, and population genetics in filamentous plant pathogens. Numerous examples from different filamentous plant pathogens have suggested that effector genes, genes encoding secreted molecules that are involved in host colonization, localize in rapidly evolving genome compartments. However, recent evidence suggests that the two-speed genome is unlikely to have evolved to specifically benefit plant pathogenic lifestyles, but instead dynamic genomic regions occur in many branches across the tree of life. Thus,

we propose that fundamental drivers of eukaryotic genome evolution have shaped both pathogen and non-pathogen genomes alike.

The epigenome is composed of the earlier mentioned three genomic dimensions and their interactions. The first dimension of the epigenome considers the linear nucleotide sequence. In **Chapter 3**, we explore the contribution of transposable element (TE) dynamics to genomic variation as well as gene expression variation in *V. dahliae*. We use whole-genome sequencing data of 42 *V. dahliae* strains to identify polymorphic TEs. Polymorphic TEs are evolutionary younger and more active when compared with non-polymorphic TEs. Importantly, polymorphic TEs are restricted to AGRs and are associated with the occurrence of genomic structural variation. Additionally, polymorphic TEs are enriched near highly expressed pathogenicity-related genes. Collectively, our analyses demonstrate that TE dynamics in *V. dahliae* contributes to genomic variation, correlates with expression of pathogenicity-related genes, and potentially impacts the evolution of AGRs.

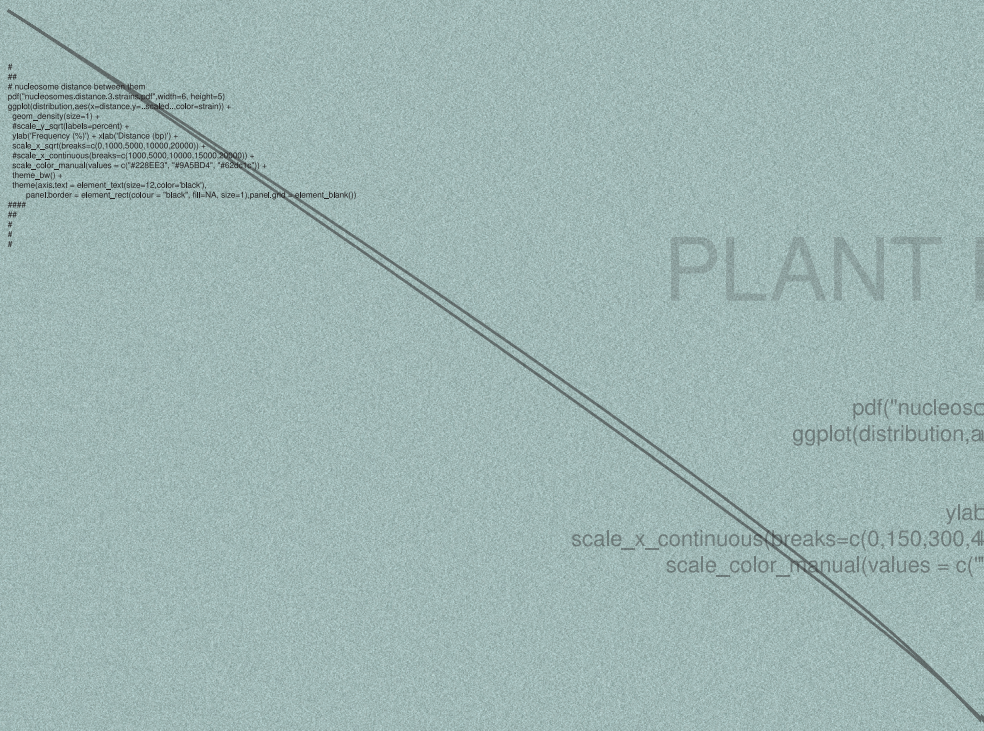
The second dimension of the epigenome focusses on the organization of the linear DNA into chromatin. In **Chapter 4**, we address the contribution of chromatin to intraspecific genomic variation and differential gene expression. Using two *V. dahliae* strains that belong to divergent genetic lineages, we show that gene expression differs for one third of the genes, and these differences are likely not caused by changes in the organization and sequence of *cis*-regulatory regions. We summarize different combinations of seven histone modifications in distinct chromatin states and a subset of those states can be linked to gene expression changes between strains. In eukaryotes, the histone modification H3K27me3 (tri-methylation of lysine 27 on histone 3) is associated with repressed gene expression, and we have previously demonstrated that H3K27me3 is enriched in *V. dahliae* AGRs. Intriguingly, H3K27me3 also occurs in chromatin states together with histone modifications typically associated with active gene expression such as H3K9ac, H3K4me3, or H3K36me3 (acetylation of lysine 9, tri-methylation of lysine 4, or tri-methylation of lysine 36 on histone 3, respectively). These chromatin states are associated with differential gene expression *in planta* and differential gene expression between *V. dahliae* strains. Particularly, genes with H3K27me3-associated chromatin states are enriched for pathogenicity-related gene functions that mostly reside in AGRs. Thus, H3K27me3-associated chromatin states contribute to the genome organization and the evolution of gene expression in *V. dahliae*.

The third dimension of the epigenome describes the spatial organization of chromatin within the nucleus. In **Chapter 5**, we utilize Hi-C data to uncover the 3D genome organization of *V. dahliae*. Locally, we reveal the presence of topologically associating domains (TADs) with gene-rich and repeat-poor boundaries that show reduced levels of gene expression and of genomic variation. Remarkably, we find that TADs in the AGRs of *V. dahliae* are less well insulated than TADs in the core genome, indicating that TADs in AGRs are not as well established as those in the core. Intriguingly, we show that the physical interaction of the AGRs follow the pattern of segmental duplications in *V. dahliae* and we confirm a similar physical association across different *Verticillium* species, suggesting that

the 3D organization is conserved within the *Verticillium* genus. Collectively, our analysis links the 3D organization to genome function and evolution throughout the *Verticillium* genus.

In **Chapter 6**, we discuss the current knowledge on fungal genome organization in the 3D space within the nucleus. On a local scale, fungal genomes are regionally organized into TADs. However, the insulation of TADs is diverse, and the detailed molecular mechanisms behind TAD formation, organization, and function remain to be uncovered. We also discuss how chromatin organization impacts DNA-templated processes across the fungal genome. On a global scale, fungal genomes are organized within the nucleus in a so-called Rabl configuration, which is characterized by clustering of centromeres on one side of the nuclear envelope and chromosomal arms extending towards the opposing nuclear periphery where (sub)telomeres associate. Proteins known to be involved in formation of the Rabl configuration are widely present across the fungal kingdom, suggesting that the Rabl conformation is a common feature in fungi. Nevertheless, we argue that our current picture of fungal genome organizations is limited to only a few fungal taxa given the paucity of fungal Hi-C experiments. Thus, we advocate for exploring genome organization across diverse fungal lineages to understand the impact of the nuclear organization on fungal genome function and evolution.

Finally, in **Chapter 7**, I address the contribution of the three dimensions of the epigenome to discuss their implications in the genome evolution of *V. dahliae*. I provide an evolutionary hypothesis to explain the genome organization and the potential role in the evolution of dynamic genomic compartments, such as the AGRs in the plant pathogen *V. dahliae*. Since dynamic genomic regions do not only occur in plant pathogens, but have been similarly observed in non-pathogenic organisms (**Chapter 2**), I conclude that a holistic view of all the interactions of the three dimensions of the epigenome is needed to understand the origin and function of the dynamic genomic compartments through the evolution of the eukaryotic genomes.



PLANT

pdf("nucleosc
ggplot(distribution,a

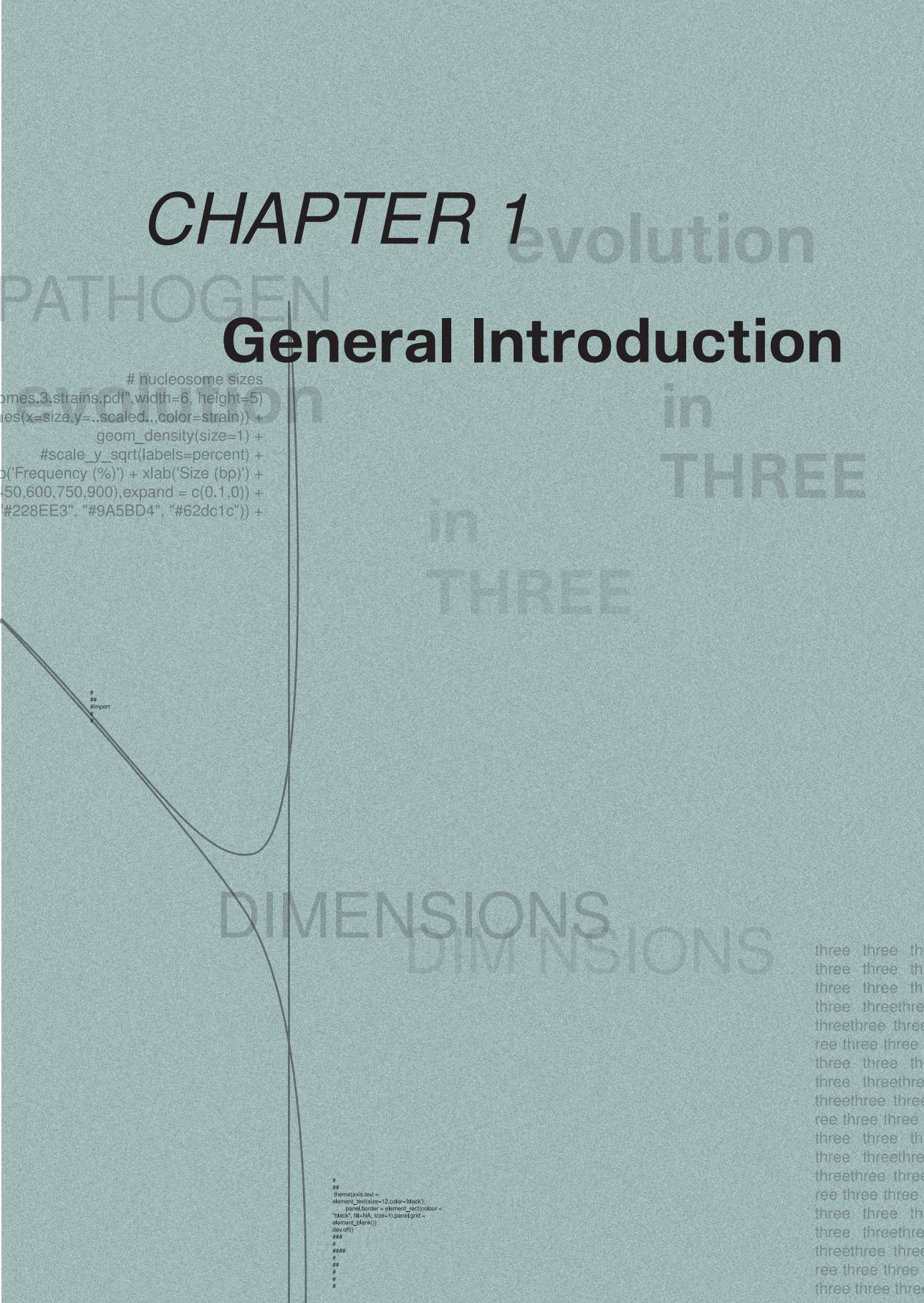
ylab
scale_x_continuous(breaks=c(0,150,300,4
scale_color_manual(values = c("

#
##
import#
#
#

```
# core genes
core.25ka<- read.table("~/Dropbox/rebuttal_raw/materi-
al/core.25k.a.bed",sep="\t",fill=T,
                       col.names=c("Chr_bnd","start_bnd","end_bnd","BND","Ins","X-
","Chr","start","end","strand","genes","ID",
                                   "baseMean","log2FoldChange","lfcSE","stat","pvalue","pad-
j","O","P","Q","R","S","W","Z"))%>%
  dplyr::select(-X,-O,-P,-Q,-R,-S,-W,-Z)%>%dplyr::mutate(region=rep("Core"))
ov <- core.25ka%>%group_by(Chr_bnd)%>%mutate(a=start-end_b-
nd,b=end-start_bnd)%>%mutate(ov=ifelse(a<0&b>0,"ov","no"))%>%filter(ov=="ov")
%>%
  dplyr::select(-a,-b)%>%mutate(dist=rep(0),cat=rep("0k"))
```

CHAPTER 1

General Introduction



Plant diseases challenge humankind

The domestication of plants played a crucial role in the rise, establishment, and development of human civilizations through a continuous supply of fibers, food and energy (Zeder 2015). The enormous growth in human population and the current climate change make securing crops as a resource, one of the crucial goals of our current and future generations (Uphoff 2012; OECD and FAO 2022). This challenge concerns from commodity crops such as wheat, or cotton (Lyson and Guptill 2004), to crops deeply rooted in our cultural identities as olives, chili peppers, or bananas (Wilson 2007).

Diseases that can effect crops have had catastrophic effects causing huge economic losses, starvation, reconfiguration of societal dynamics, and may lead to environmental changes (Cheatham et al. 2009; Zadoks 2017; Ristaino et al. 2021; Rizzo et al. 2021). The emergence of diseases, particularly those that have suddenly increased in incidence, severity, geographical or host range, imposes a huge challenge for humankind to overcome (Fisher et al. 2012; Rizzo et al. 2021). A notorious example is the emergence of the Panama disease of banana during the mid-1890's in Central America. This disease was caused by the fungus *Fusarium oxysporum* f.sp. *cubense* and led to immense economic losses (Ploetz 2015, 2000; Viljoen et al. 2020). The disease caused a reconfiguration of the banana industry, which led to worsened labor conditions for thousands of workers in plantations (Marquardt 2001; Ploetz 2015), which led to a major reshaping of the landscape in countries such as Costa Rica, Honduras, and Panama due to the dramatic deforestation of tropical rainforests by farmers in search of new productive land (Marquardt 2001). As a consequence, this plant disease epidemic caused not only huge economic losses, but changed the landscape in Central America, and the lives of thousands.

Plant diseases are the outcome of unbalances in the complex interactions between the plant host, the pathogen, and the environment (Scholthof 2007). For instance, the homogeneity of plant genotypes in monocultures has facilitated the proliferation of pathogens (Stukenbrock and McDonald 2008; McDonald and Stukenbrock 2016). Climate change can foster the emergence of plant pathogens in new geographical locations (Fisher et al. 2012). Historically, plant pathogens have been widely controlled by the application of agrochemicals (Agrios 2005). However, agrochemicals exert a high pressure on plant pathogen populations to overcome by developing resistance (Lamberth et al. 2021; Steinberg and Gurr 2020). Collectively, the unbalances between host, environment and pathogens have made plant diseases a challenge for securing crop resources for human societies.

Plant pathogenic organisms are diverse. The ability of pathogens to infect plant species has evolved independently several times across the tree of life, including diverse groups as bacteria, stramenopiles, nematodes, and by far the most important microbial plant disease agents: fungi (Fisher et al. 2012, 2020).

Fungal lifestyles and the plethora of plant-fungal interactions

Fungi are considered to be one of the most versatile eukaryotic kingdoms on earth, as they can thrive in conditions that involve extreme pressure, temperature, salinity, desiccation, pH, and pollution (Naranjo-Ortiz and Gabaldón 2019a, 2019b). Under such diverse environmental conditions, fungi can exploit diverse nutrient sources and establish complex symbiotic associations with other organisms. Some fungi can exploit the available nutrients from dead sources as saprotrophs (Rousk and Bååth 2011; Boddy and Hiscox 2016). Furthermore, some fungi can engage in symbiotic relationships that range in a continuum from mutualistic to parasitic (Bonfante and Anca 2009; Lo Presti et al. 2015; Selosse et al. 2018).

Parasitic plant-fungal interactions typically form a convoluted continuum that can be defined from necrotrophic to obligate biotrophic interactions (Jones and Dangl 2006; Zeilinger et al. 2016; Zanne et al. 2020; Thomma et al. 2011). On the one extreme end of the continuum, necrotrophic fungi interact with the host and induce cell death to retrieve nutrients from their hosts. Important pathogens such as *Botrytis cinerea* and *Sclerotinia sclerotiorum* are considered typical examples of necrotrophic fungi (van Kan et al. 2014; Kim et al. 2008). On the other extreme end, obligate biotrophic fungi interact with the host and do not cause cell death, but establish a long-lasting interaction to retrieve nutrients from the plant. Rust and powdery mildew fungi are among the economically most important groups of obligate biotrophic pathogens (Helfer 2014; Lorrain et al. 2019; Garnica et al. 2014). In between those two extremes, hemibiotrophic fungi display early a biotrophic and later a necrotrophic lifestyle during their life cycle. One of the most important hemibiotrophic fungi is *Magnaporthe oryzae* (Wilson and Talbot 2009; Cruz-Mireles et al. 2021). However, plant-fungal interactions have shown not to be static, and as such, intimately associated with fungal plant pathogen speciation (Stukenbrock 2013).

The everlasting co-evolution of plant-fungal interactions

The long-lasting interaction between plants and fungal pathogens exerts strong selection pressure on pathogens as well as on their hosts (Jones and Dangl 2006). Plants have evolved immune systems to detect pathogen intrusions and to mount defense responses, while fungal pathogens evolved complex molecular tool kits to support host colonization (Rovenich et al. 2014; Cook et al. 2015). The everlasting co-evolutionary 'arms race' is characterized by repeated cycles of adaptation and counter-adaptation of pathogens and their hosts (Strotz et al. 2018; McDonald and Stukenbrock 2016) (Fig. 1). A key component of the arms race are so-called effectors, which have been typically defined as proteins that are secreted by the pathogen to support host colonization (Lo Presti et al. 2015; Rovenich et al. 2014; Snelders et al. 2018). Also secondary metabolites, and small RNAs can act as effectors (Collemare

et al. 2019). Interestingly, genes encoding effectors or other proteins with roles in fungal-adaptation often co-localize in discrete gene-sparse regions in plant pathogen genomes, which contrasts with the conserved housekeeping genes often being depleted from those regions and instead localized in gene-rich regions (Haas et al. 2009; Raffaele et al. 2010). This observation gave rise to the 'two-speed' genome hypothesis (Raffaele et al. 2010; Raffaele and Kamoun 2012) (Fig. 1) that posits that genes localized in gene-sparse regions evolve faster, while genes located in gene-rich regions evolve at a slower pace (Raffaele and Kamoun 2012; Dong et al. 2015).

The 'two-speed' genome hypothesis conceptually states the presence of two distinctive genomic compartments, namely a 'core' compartment enriched for genes evolving at a low speed, and a 'dynamic' compartment enriched for genes evolving at a higher speed (Fig. 1). Following this thought, the dynamic compartment is often enriched for structural variations, copy-number variations, presence-absence variations, sequence polymorphisms, genomic rearrangements, and transposable element polymorphisms (Faino et al. 2016; Frantzeskakis et al. 2018; Raffaele et al. 2010; Torres et al. 2021; Thon et al. 2006; Plissonneau et al. 2018). In contrast, the core compartment is usually depleted in such genomic variations. Collectively, these observations have fueled the initial hypothesis that genes embedded in the dynamic compartments show a higher evolutionary speed due to the abundance of genomic variation.

The two-speed genome model has been very useful to explain the association of specific adaptive genes to discrete repeat-rich genomic regions in many plant pathogens (Fouché et al. 2018; Möller and Stukenbrock 2017; Frantzeskakis et al. 2019). For example, effector genes are associated to AT-rich isochores in *Lepthosphaeria maculans* (Rouxel et al. 2011) and *Epichlöe festucae* (Hassing et al. 2019). Similarly, the sub-telomeric regions of *Fusarium fujikuroi* (Chiara et al. 2015) and some species of *Colletotrichum* (Gan et al. 2021) as well as the accessory chromosomes of *Magnaporthe oryzae* and *Fusarium oxysporium* are enriched for effector genes and other environmentally-response genes (van Dam et al. 2017; Fokkens et al. 2018; Peng et al. 2019). These observations, as well as similar observations for other plant pathogenic fungi, collectively revealed the widespread occurrence of genomic compartmentalization, even though there are also clear exceptions that lack this organization such as the genomes of *Blumeria graminis* (Frantzeskakis et al. 2018) and *Ramularia collo-cygni* (Stam et al. 2018).

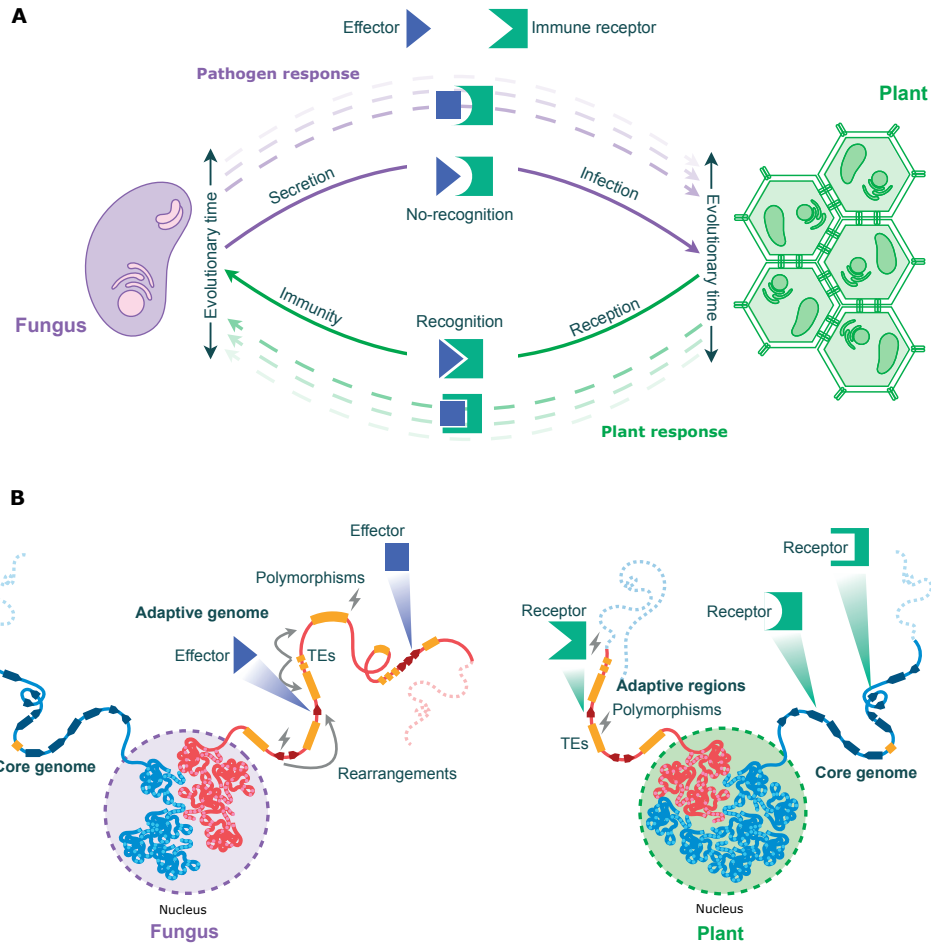


Figure 1. The 'two-speed genome' model and its role in plant-pathogen co-evolution.
(A) Co-evolution model between a plant and a fungal pathogen. The fungus secretes effectors that happen to be not recognized by the plant immune system, and consequently colonization progresses. The plant evolves immune receptors that are able to recognize effectors and establishes immunity. Consequently, other effectors could be unrecognized by the plant. This cycle of adaptation and counteradaptation repeats over and over in the everlasting plant-pathogen co-evolutionary arms races. **(B)** The genome according to the 'two-speed' model. The genome is compartmentalized into a 'core genome' and a 'dynamic genome'. The dynamic genome is enriched in transposable elements (TEs), genomic rearrangements, polymorphisms, nucleotide variations, and *in planta* expressed fungal genes. Possibly, plant genomes obey to a two-speed model as well, with immune receptor genes in the dynamic genome compartments.

Chromatin, the complex of DNA and its associated proteins, determines the organization of the genome within the nucleus (Olins and Olins 2003; Maeshima et al. 2021). Nowadays, chromatin characteristics are increasingly used to further detail the function, evolution, and relevance of the genomic compartments of the two-speed model (Seidl et al. 2016). For instance, in *Z. tritici*, *M. oryzae*, and *Fusarium* spp., the accessory

chromosomes have been associated with a specific chromatin profile, particularly enriched for the tri-methylation of the lysine 27 in the histone 3 (H3K27me3) (Schotanus et al. 2015; Fokkens et al. 2018; Möller and Stukenbrock 2017). The presence of H3K27me3 has been associated with silencing of gene expression (Zhang et al. 2021b), chromosome stability (Möller et al. 2019) and nucleotide variation (Feurtey et al. 2020; Tralamazza et al. 2022; Habig et al. 2021). These observations suggest an intimate relationship between DNA, gene expression, chromatin organization, and genome evolution of plant pathogens.

Genome evolution in three dimensions

One of the fundamental distinctions in biology is between the DNA, with the genes as carriers of information, and the proteins which generate the phenotype (Szathmáry and Smith 1995). The term 'epigenetics' was originally coined by Conrad Waddington in the '40s of the 20th century to explain the connection between genes and phenotype (Waddington 1942). He described it as 'the interactions of genes with their environment, which bring the phenotype into being' (Waddington 1942). Nowadays, a modern definition of epigenetics could be considered as 'the sum of all genetic and non-genetic factors acting upon cells to selectively control the gene expression needed in development and evolution' (Hallgrímsson and Hall 2011). Thus, the cooperativity, dynamics and organization of the genetic and non-genetic factors could be summarized in at least the following three dimensions, not only in fungi, but potentially in all eukaryotes.

The first dimension: on the linear thread

The first dimension considers the linear nucleotide sequence in the control and evolution of gene expression (Fig. 2). The organization of genes on the linear DNA and the gene sequence itself determine, at least partially, gene expression. Gene expression can be altered by changes in the gene sequence due to single nucleotide mutations (Shastry 2009), genomic rearrangements (Escaramís et al. 2015; Ho et al. 2020). These mutations may change amino acids or introduce stop codons (nonsynonymous mutations), or do not alter protein sequences (synonymous mutations) (Ohta and Ina 1995; Nei and Kumar 2000). Typically, nonsynonymous mutations are under strong negative selection because of their potential changes in phenotype (Ohta and Ina 1995; Nei and Kumar 2000; Conrad and Hurles 2007; Feulner and De-Kayne 2017). Conversely, synonymous mutations are considered mostly neutral (Ohta and Ina 1995; Kimura 1968; Nei and Kumar 2000), or at least, nearly neutral as they potentially could have slight effects on the phenotype (Lynch and Conery 2003; Lynch and Walsh 2007; Chamary et al. 2006; Ohno 2013; Bergthorsson et al. 2007).

The linear nucleotide genome sequence is composed of more than just protein-coding genes. Genes are flanked by regulatory sequences that modulate their expression timing and transcript levels (Wray et al. 2003; Wittkopp and Kalay 2011). Many types of

regulatory sequences have been described, including for example promoters, and enhancers (Wray et al. 2003; Gasch et al. 2004; Schoenfelder and Fraser 2019; Andersson and Sandelin 2020; Brázda et al. 2021). Promoters are regulatory sequences located in upstream proximity of a transcription start site (TSS), whereas enhancers are not necessarily located near a TSS and can activate multiple genes (Gagniuc and Ionescu-Tirgoviste 2012; Pennacchio et al. 2013). Regulatory sequences allow the binding of transcription factors (TFs) and promote the recruitment of the transcriptional machinery responsible to ensure timely and spatially organized gene expression. Regulatory sequences could be more prone to tolerate point mutations, deletions, or structural variants that can modify the binding affinity of TFs, and consequently modify gene expression (Schmidt et al. 2010; Schoenfelder and Fraser 2019; Villar et al. 2015; Berthelot et al. 2018, 2015; Young et al. 2015). Therefore, the plasticity of regulatory sequences is relevant for the evolution of gene expression.

The size of the linear eukaryotic genome does not necessarily correlate with the number of genes and regulatory sequences (Lynch 2006; Lynch and Conery 2003; Wells and Feschotte 2020). This is a consequence of the proliferation of non-coding DNA such as transposable elements (TEs). TEs are repetitive DNA sequences that are capable of transposing to diverse locations within the host genome, and thus are considered 'mobile genetic elements' (McClintock 1950; Wicker et al. 2007; Wells and Feschotte 2020; Feschotte 2008). TEs can impact genome organization, gene expression, and gene function by disrupting protein-coding regions, modifying gene regulatory sequences, or causing large-scale chromosomal rearrangements (Chuong et al. 2017; Wells and Feschotte 2020; Bourque et al. 2018; Muszewska et al. 2019). TEs can insert in regulatory sequences and induce changes in gene expression (Chuong et al. 2017; Kitano et al. 2018), or insert in coding regions and change the abundance of transcripts (Van't Hof et al. 2016). Therefore, host genomes deploy different defense mechanisms to suppress TE mobilization and decrease the detrimental effect of such mobilizations. This represents a delicate trade-off for the host genomes, as TEs have also been recognized to play important roles in adaptive evolution (Lynch et al. 2011b; Bourque et al. 2018; Wells and Feschotte 2020). For example, TEs have shown to play a role in the immune system of animals, some domestication traits in plants, or the environmental response of fungi (Sakamoto et al. 2004; Kawakatsu et al. 2016; Seidl et al. 2016; van Wersch and Li 2019). Therefore in the linear space of the DNA, the organization of genes, regulatory sequences, repeats and transposable elements is crucial to understand gene expression, and evolution.

The second dimension: on the knitted thread

The second dimension goes beyond the linear DNA sequences and considers the organization of the genomic DNA into chromatin (Fig. 2). Chromatin is defined as the sum of the eukaryotic DNA and all the protein complexes that allow the organization and regulation of the linear nucleotide strand inside the nucleus (Olins and Olins 2003; Maeshima et al. 2021). The linear DNA strand could be chemically modified by direct DNA modifications, mostly in the form

of methylations (Meng et al. 2015; Luo et al. 2015; Boulias and Greer 2022; Zhong 2016; Nai et al. 2020; He et al. 2020; Kumar et al. 2018). Additionally, in eukaryotes, DNA is wrapped around proteins to form a DNA-protein complex called a nucleosome (Luger et al. 1997), which is the basic unit of chromatin. Nucleosomes are composed of an octamer formed by two copies of each of the histone proteins H2A, H2B, H3, and H4 that wrap together around 150 bp of DNA. Each histone have chain of unstructured amino acid residues that can be post-translationally modified through methylation, acetylation, and phosphorylation (Zhao and Garcia 2015; Bannister and Kouzarides 2011; Zentner and Henikoff 2013). To date, hundreds of post-translational modifications have associated with nucleosome structures and with roles in transcriptional regulation (Bannister and Kouzarides 2011; Zentner and Henikoff 2013). In addition, each nucleosome can be modified by chaperones that change the nucleosome structure by interchanging histone variants (Talbert and Henikoff 2010, 2017), or by changing the nucleosome position in relation to the associated DNA (Saha et al. 2006; Nocetti and Whitehouse 2016).

Histone post-translational and DNA modifications establish a chromatin landscape that can be broadly divided into transcriptionally active euchromatin and repressive heterochromatin. Euchromatin is mostly characterized by high chromatin accessibility, depletion of DNA methylation, and enrichment of specific histone modifications such as H3K36me3 (tri-methylation of lysine 36 on the tail of histone 3) or H3K27ac (mono-acetylation of lysine 27 on histone 3) (Zentner and Henikoff 2013; Huang and Zhu 2018). Typically, euchromatic regions are associated with strong negative selection, and therefore genomic variation is depleted (Makova and Hardison 2015; Monroe et al. 2022b; Hazarika et al. 2022; Quiroz et al. 2022). In contrast, heterochromatic regions are typically characterized by poor accessibility, enrichment in DNA methylation, and enrichment of specific histone modifications such as H3K9me3 or H3K27me3 (Zentner and Henikoff 2013; Bannister and Kouzarides 2011; Grewal and Jia 2007; Sasaki et al. 2014; Schuster-Böckler and Lehner 2012; Sun et al. 2016). Heterochromatic regions are associated with relaxed selection and prone to accumulate TEs, SVs, and nucleotide variations (Lewis et al. 2009; Makova and Hardison 2015; Janssen et al. 2018; de la Peña et al. 2023; Habig et al. 2021; Yasuhara et al. 2005; Liu et al. 2020a). Thus, it has been hypothesized that in contrast to euchromatin, heterochromatin can serve as a cradle for genomic variability (Liu et al. 2020a; Yasuhara et al. 2005). Taken together, in the second dimension of the genome, the chromatin organization is important not only for gene expression regulation, but also for the evolution of eukaryotic organisms.

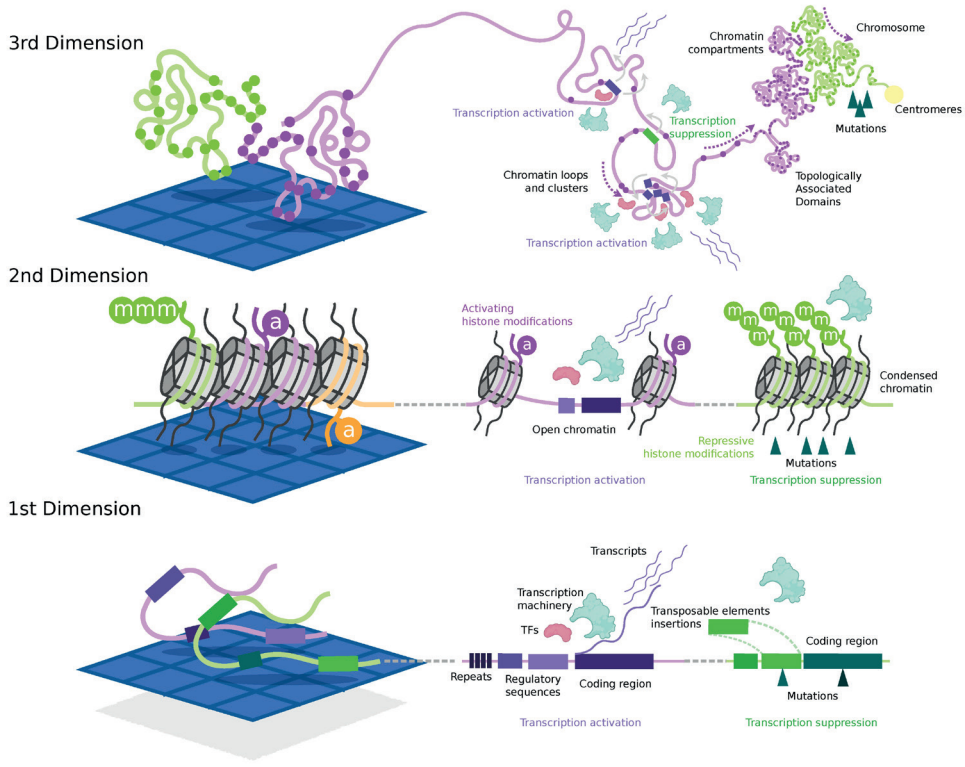


Figure 2. The genome in three dimensions.
The genome is organized in at least three dimensions. The interactions between these three dimensions compose the epigenome. **The first dimension covers the architecture and dynamics of the linear DNA sequence.** The genome is composed of coding, regulatory, and repetitive sequences. Regulatory sequences allow binding of transcription factors (TFs) to recruit the transcription machinery, thereby enabling the 'reading' of coding regions (i.e., genes). Changes in regulatory sequences due to transposable element (TE) insertions or mutations (green arrows) can lead to gene silencing or expression variation. **The second dimension covers the organization of the DNA into chromatin.** DNA is wrapped around histone proteins to form nucleosomes. Histone proteins can be modified on their tails (purple, orange balls = acetylations), and these modifications can modulate the DNA accesibility to TFs, thereby promoting transcriptional activation. Other histone modifications (green balls = methylation) can promote chromatin condensation, thereby suppressing transcription or the accumulation of mutations (green arrows). **The third dimension depicts the spatial and physical organization of the chromatin within the nucleus.** The chromatin strand can form loops to enable regulatory sequences (dark purple boxes) to get physically closer to target genes. Loops can form topologically associating domains (TADs). The acumulation of TADs can enable the formation of active or repressive chromatin compartments (purple and green chromatin strands). Compartments form the typical structure of chromosome domains, like centromeres (yellow) or chromosome arms.

The third dimension: on the folded knitted thread

The third dimension recognizes that chromatin is organized into three-dimensional (3D) structures that play a role in the control and evolution of gene expression within the eukaryotic nucleus (Fig. 2). The 3D folded chromatin strand allows the spatial association of genomic elements separated on the linear DNA strand (Lieberman-Aiden et al. 2009). The spatial organization ranges from small-scale chromatin loops of a few kilobases that could contribute to transcriptional regulation based on DNA-DNA contacts, to large-scale subdomains composed of hundreds of kilobases that arrange large chromosomal regions into active or silent chromatin regions in 'A' and 'B' subdomains, respectively (Rao et al. 2014; Holwerda and De Laat 2012; Rowley and Corces 2018; Eagen et al. 2017; Jerkovic and Cavalli 2021). Chromatin loops can form through the action of proteins that progressively extrude chromatin from a boundary element, till they encounter the next boundary element (Brackley et al. 2018; Banigan and Mirny 2020; Davidson and Peters 2021). These boundary elements are diverse in three of life. For example, in vertebrates, boundaries are enriched in CTCF (CCCTC binding factor) dimers (de Wit et al. 2015; Sanborn et al. 2015). However, different boundary elements than CTCF dimers have been observed in plants (Wang et al. 2015b; Dong et al. 2017; Xie et al. 2019; Li et al. 2019), *D. melanogaster* (Ramírez et al. 2018; Arzate-Mejía et al. 2020), *C. elegans* (Anderson et al. 2019), *S. pombe* (Mizuguchi et al. 2014), *S. cerevisiae* (Schalbetter et al. 2019), and *N. crassa* (Galazka et al. 2016; Rodriguez et al. 2022). The accumulation of loops shapes the chromosome structure into discrete regions called topologically associating domains (TADs) in mammals and plants, or 'globules' in fungi (Dong et al. 2017; Mizuguchi et al. 2014; Dixon et al. 2012). While named differently, these in principle refer to homologous topological structures. Although the function of TADs is controversial and still under debate (Ghavi-Helm et al. 2019; Kaushal et al. 2021; Cavalheiro et al. 2021; Arzate-Mejía et al. 2020), many studies have associated the biological function of TADs to transcriptional regulation (Dixon et al. 2012; Rao et al. 2014; Cavalheiro et al. 2021) and genome replication (Eser et al. 2017; Yang et al. 2019).

Chromatin loops, TADs and large scale domains are usually characterized by the presence of distinct chromatin patterns within the nucleus. For example, heterochromatic domains are usually tethered in the nuclear envelope in the form of lamina-associated domains (Guelen et al. 2008; Rao et al. 2014), in centromere-specific domains (Mekhail and Moazed 2010), or by the interaction of heterochromatic marks with specific nuclear envelope proteins (Buchwalter et al. 2019; Briand and Collas 2020; Mekhail and Moazed 2010). Changes in the organization of these wide-scale domains have been associated with multiple effects on the phenotype, including diseases in humans (Szabo et al. 2019; Niu et al. 2021). Therefore, studies into various groups of related organisms suggest an evolutionary conservation of specific 3D genome structures (Harmston et al. 2017; Fudenberg and Pollard 2019; Golicz et al. 2020; Krefting et al. 2018; Liu et al. 2017; McArthur and Capra 2021; Rowley et al. 2017; Rao et al. 2014; Yang et al. 2019; Huynh and Hormozdiari 2019). As such, the 3D genome organization has been associated with the evolution of specific traits in

vertebrates and plants (Lonfat and Duboule 2015; Deschamps 2016; Yang et al. 2019; Golicz et al. 2020; Corbo et al. 2022; Hoencamp et al. 2021). Collectively, the conservation of the 3D organization of the genome suggests an association of the regulation of gene expression with chromatin organization, which plays a significant role in evolution of eukaryotes.

In summary, the three dimensions are tightly linked to a modern definition of epigenetics. The three dimensions are part of a single epigenetic unit that is subject to evolutionary forces: the 'epigenome'. The 'epigenome' could be considered as a unit composed of all the genetic and non-genetic factors that collectively determines the phenotype. Thus, theoretically, the epigenome cover the three dimensions and the interactions between that directly or indirectly regulate gene expression and organization under the lens of evolution.

Verticillium dahliae genome in three dimensions

The ascomycete fungus *Verticillium dahliae* is a plant pathogen that causes Verticillium wilt, a devastating vascular disease (Fradin and Thomma 2006). The species *V. dahliae* causes wilting disease in hundreds of plant species, including economically important ones such as tomato and cotton (Fradin and Thomma 2006; Klosterman et al. 2009). *Verticillium dahliae* infects host plants through the roots, and colonizes the xylem vessels where it produces conidiospores. Upon colonization, the fungus proliferates in the vascular tissue resulting in wilting of the plants (Reusche et al. 2014; Buhtz et al. 2017; Vallad and Subbarao 2008). Furthermore, *V. dahliae* forms melanized resting structures called microsclerotia that maintain dormant for several years in the soil, making that soil unsuitable for the growing of many agricultural important plant species. Therefore, *V. dahliae* is difficult to control utilizing crop rotation and fungicides (Fradin and Thomma 2006; Klosterman et al. 2009). Only a few genetic resistance sources to Verticillium wilt have been identified thus far in few plant species only, making the control of Verticillium wilt in most crops a challenge (Song et al. 2020; Shaban et al. 2018; Usami et al. 2017; Chen et al. 2021a).

The Lineage-Specific Regions: the first dimension

During the infection process, *V. dahliae* makes use of different effectors to successfully colonize the host (Inderbitzin and Subbarao 2014). To date, the activity of several effectors secreted by *V. dahliae* has been experimentally verified (de Jonge et al. 2012; Kombrink et al. 2017; Chavarro-Carrero et al. 2021; Snelders et al. 2023, 2020). For instance, functional analysis revealed that Ave1 (Avirulence on *Ve1* tomato) acts as a major virulence factor on tomato plants lacking the *Ve1* immune receptor (de Jonge et al. 2012). The secreted small cysteine-rich protein Ave1 displays antimicrobial activity and facilitates colonization through the manipulation of the plant microbiome by suppressing antagonistic bacteria (Snelders et al. 2020). Interestingly, the gene coding for Ave1 is located in highly dynamic genomic

regions (de Jonge et al. 2012). Genomic comparisons between *V. dahliae* strains revealed that their genomes evolved through extensive large-scale chromosomal rearrangements (de Jonge et al. 2013; Faino et al. 2016). These rearrangements are associated with lineage-specific (LS) regions that are highly variable between *V. dahliae* strains (Klosterman et al. 2011; de Jonge et al. 2012, 2013; Faino et al. 2016; Depotter et al. 2019). These regions are enriched in relatively young and transcriptionally active TEs (Faino et al. 2016; Cook et al. 2020). Therefore, the presence of active TEs is associated with the formation and maintenance of these dynamic regions (Amyotte et al. 2012; Faino et al. 2016; Cook et al. 2020; Seidl et al. 2020). Thus, dynamics of the linear DNA strand has shown to be associated with the genome evolution of *V. dahliae*.

The Adaptive Genomic Regions: the second and third dimensions

Recent work on the chromatin landscape has revealed that the *V. dahliae* LS regions display a unique chromatin profile when compared with the core genome (Cook et al. 2020). The chromatin profile of LS regions is characterized by high chromatin accessibility, low DNA methylation, and enrichment of the histone modification H3K27me3 (Cook et al. 2020; Kramer et al. 2021, 2022). Notably, a substantial number of additional genomic regions share chromatin characteristics with the LS regions and, accordingly, the amount of LS DNA has been underestimated (Cook et al. 2020). The originally identified LS regions, together with the additional genomic regions that share the similar chromatin profile, are now collectively referred to as adaptive genomic regions (AGRs) (Cook et al. 2020). These AGRs are enriched not only in *in planta*-induced genes but also in genes that are differentially expressed between *in vitro* growth conditions (Klosterman et al. 2011; de Jonge et al. 2012, 2013; Kombrink et al. 2017; Cook et al. 2020; Chavarro-Carrero et al. 2021; Kramer et al. 2021, 2022; Gibriel et al. 2019). Additionally, the presence of a hallmark TE associated with unique H3K9me3 heterochromatin domains in centromeres correlates with the karyotype evolution of *Verticillium* spp. (Seidl et al. 2020). Thus, chromatin dynamics showed that the second dimension of the genome is important for gene expression regulation and genome evolution of *V. dahliae*. Further work is needed to understand not only the interplay of the observations in the first and second dimensions of the genome, but to integrate the third dimension of the genome that has not been explored thus far.

Research question and thesis outline

The multidimensional epigenome fuels the genomic variation needed for the evolution of filamentous fungal plant pathogens, such as *V. dahliae*. The epigenome is the collaborative dynamics of multiple dimensions. Previous analyses have shown the interplay of the linear DNA dimension and the chromatin dimension in *V. dahliae*. I aimed to investigate the role of the epigenome dynamics between strains of *V. dahliae*, and by comparing *V. dahliae* to all species of the *Verticillium* genus. Therefore, the main objective of my doctoral research is to explore the epigenome dynamics to understand its implications in the genome evolution of the soil-borne fungal plant pathogen *V. dahliae*.

In **Chapter 2**, we critically analyze the 'two-speed' model by examining recent studies on epigenetics, transposable element dynamics, and population genetics in filamentous plant pathogens. Emerging evidence suggests that the 'two-speed' genome model is not exclusive to plant pathogens, but rather dynamic regions are present across diverse organisms. Consequently, we propose that fundamental factors influencing eukaryotic genome evolution have shaped both pathogen and non-pathogen genomes alike.

In **Chapter 3**, we explore the contribution of transposable element (TE) dynamics to the genomic and gene expression variation in *V. dahliae*. We use whole-genome sequencing data of 42 *V. dahliae* strains to identify polymorphic TEs associated with genomic variation. Polymorphic TEs are evolutionary younger and more active, compared to the non-polymorphic TEs. Additionally, polymorphic TEs occur mostly in adaptive genomic regions (AGRs) in nearby association with highly expressed pathogenicity-related genes.

In **Chapter 4**, we investigate how chromatin influences genomic variation and differential gene expression. Comparison of two strains from different genetic lineages shows that one-third of the genes are differentially expressed between both strains. These differences are likely not associated with changes in the *cis*-regulatory regions. Instead, by analyzing different combinations of seven histone modifications, we identified chromatin states associated with differential gene expression. H3K27me3-rich chromatin states are associated with differential gene expression of pathogenicity-related genes *in planta* and differential gene expression between *V. dahliae* strains.

In **Chapter 5**, we further use proximity ligation to uncover the three-dimensional (3D) genome organization of *V. dahliae*. We reveal the presence of topologically associating domains (TADs). We find that TADs in AGRs are not as well established as those in the core genome. Additionally, we show that the physical interaction of AGRs follows the pattern of segmental duplications in *V. dahliae* and we confirmed a similar physical association across different *Verticillium* species, suggesting that the 3D organization is conserved within the *Verticillium* genus.

In **Chapter 6**, we analyze the current knowledge on fungal genome organization in the 3D space within the nucleus. We discuss how chromatin organization impacts DNA-

templated processes across the fungal genome. We cover the fungal genome organization from the local scale, in the regionally organized TADs, to the global scale, in the so-called Rabl configuration. We further show that proteins known to be involved in the genome organization in the 3D space within the nucleus are widely present across the fungal kingdom.

In **Chapter 7**, I address the contribution of the epigenome to discuss their implications in the genome evolution of *V. dahliae*. Furthermore, I discuss an evolutionary hypothesis for the potential role of the epigenome in the genome evolution of the plant pathogen *V. dahliae*, and potentially for the evolution of similar eukaryotic genomes.

PLANT

```
# core genes
core.25ka<- read.table("~/Dropbox/rebuttal_raw/materi-
al/core.25k.a.bed",sep="\t",fill=T,
                    col.names=c("Chr_bnd","start_bnd","end_bnd","BND","Ins","X-
"start","end","strand","genes","ID",
                    "baseMean","log2FoldChange","lfcSE","stat","pvalue","pad-
j","O","P","Q","R","S","W","Z"))%>%
  dplyr::select(X,O,P,Q,R,S,W,Z)%>%dplyr::mutate(region=rep("Core"))
ov <- core.25ka%>%group_by(Chr_bnd)%>%mutate(a=start-end_b-
nd,b=end-start_bnd)%>%mutate(ov=ifelse(a<0&b>0,"ov","no"))%>%filter(ov=="ov")
%>%
  dplyr::select(-a,-b)%>%mutate(dist=rep(0),cat=rep("0k"))
```

```
#
##
# nucleosome distance between them
pdf("nucleosomes.distance.3.strains.pdf",width=6, height=5)
ggplot(distribution,aes(x=distance,y=..scaled..,color=strain)) +
  geom_density(size=1) +
  #scale_y_sqrt(labels=percent) +
  ylab('Frequency (%)') + xlab('Distance (bp)') +
  scale_x_sqrt(breaks=c(0,1000,5000,10000,20000)) +
  #scale_x_continuous(breaks=c(1000,5000,10000,15000,20000)) +
  scale_color_manual(values = c("#228EE3", "#9A5BD4", "#62dc1c")) +
  theme_bw() +
  theme(axis.text = element_text(size=12,color='black'),
     panel.border = element_rect(colour = "black", fill=NA, size=1),panel.grid = element_blank())
####
##
#
#
#
```

CHAPTER 2

Genome evolution in fungal plant pathogens: looking beyond the two-speed genome model

David E. Torres[1,2], Ursula Oggenfuss[3], Daniel Croll[3], and Michael F. Seidl[2]

[1]Wageningen University & Research, NL
[2]Utrecht University, NL
[3]Neuchâtel University, CH

Modified version published in *Fungal Biology Reviews*, 34(3), 136-143. (2020).

Abstract

The interaction of pathogens with their hosts creates strong reciprocal selection pressures. Pathogens often deploy an arsenal of small proteins called effectors that manipulate the plant immune system and promote disease. In the post-genomics era, a major interest has been to understand what shapes the localization of effector genes in pathogen genomes. The two-speed genome model originated with the discovery of repeat-rich and gene-sparse genome compartments with an over-representation of effector-like genes in a subset of plant pathogens. These highly polymorphic genome compartments are thought to create unique niches for effector genes and facilitate rapid adaptation. Research over the past decade has revealed a number of twists to the two-speed genome model and raised questions about the universality among plant pathogens. Here, we critically review the foundations of the two-speed model by presenting recent work on epigenetics, transposable element dynamics, and population genetics. Numerous examples have demonstrated that the location of effector genes in rapidly evolving compartments has created key adaptations. However, recent evidence suggests that the two-speed genome has unlikely evolved to benefit specifically the plant pathogen lifestyle. We propose that fundamental drivers of eukaryotic genome evolution have shaped both pathogen and non-pathogen genomes alike. An evolutionary genomics perspective on the two-speed genome model will open up fruitful new research avenues.

Introduction

Eukaryotic microorganisms are crucial components of worldwide ecosystems. Many eukaryotic microorganisms engage in symbiotic relationships that can range from mutualistic to pathogenic. Within this continuum, tight interactions between pathogens and their hosts exert strong selection pressures (Papkou et al. 2019). Hosts evolved immune systems to detect pathogen intrusions and to mount defense responses while pathogens evolved molecular tools to support host colonization (Cook et al. 2015), and these everlasting co-evolutionary 'arms-races' are thus characterized by repeated cycles of adaptation and counteradaptation (Strotz et al. 2018; Papkou et al. 2019).

The genomic era has provided unprecedented insights into the genomic signatures of adaptation for many eukaryotic microorganisms that form symbiotic relationships with plants or animals. One of the key observations was that effectors, genes or genomic loci encoding secreted proteins as well as non-proteinaceous effector molecules such as secondary metabolites or small RNAs that can collectively support ecological niche colonization (Snelders et al. 2018; Rovenich et al. 2014; Lo Presti et al. 2015; Collemare et al. 2019), and other genes with roles in adaptation often co-localize within the genome (Seidl and Thomma 2014, 2017; Sánchez-Vallet et al. 2018; Dong et al. 2015; Raffaele and Kamoun 2012). Genes localized in repeat-rich and gene-sparse regions tend to evolve faster, in contrast to slow-evolving housekeeping genes localized in repeat-poor and gene-rich regions (Haas et al. 2009; Dong et al. 2015; Raffaele et al. 2010; Raffaele and Kamoun 2012). The observation of this particular genome organization gave rise to the popular 'two-speed' genome model that is commonly used to describe genome organization and evolution in plant-pathogenic fungi and oomycetes (Dong et al. 2015; Raffaele and Kamoun 2012; Frantzeskakis et al. 2019). Here we discuss some of the key assumptions underlying the two-speed hypothesis, highlight recent work shedding light on the evolution of this peculiar genome organization, and provide some critical re-evaluation in the post-genomics era.

Evolution at two speeds, a misleading concept?

A key assumption of the two-speed hypothesis underscores the presence of two discrete types of genomic regions that supposedly evolve at different speeds (Dong et al. 2015). *Fast*-evolving regions in compartmentalized genomes are typically defined to be rich in repeats such as transposable elements (TEs) and to contain effector genes as has been observed in AT-isochores, sub-telomeric regions, and repeat islands embedded within chromosomes or accessory chromosomes (Fig. 1) (Thon et al. 2006; Rouxel et al. 2011; de Jonge et al. 2012; Wang et al. 2017; Dutheil et al. 2016; Faino et al. 2016; de Jonge et al. 2013; Plissonneau et al. 2018; Croll and McDonald 2012; Goodwin et al. 2011; Akagi et al. 2009; Plissonneau et al. 2016; Ma et al. 2010; van Dam et al. 2017; Fokkens et al. 2018).

However, recent genomics analyses of diverse plant pathogens have revealed that effectors do not always cluster, that effectors and TEs do not necessarily co-localize (Frantzeskakis et al. 2018; Schwessinger et al. 2018, 2020; Dutheil et al. 2016; Wyka et al. 2020; Stam et al. 2018), and that more than two discrete types of genomic regions can occur (Fokkens et al. 2018). These observations blur the binary notion of the original two-speed hypothesis that had been initially founded on observations in just few plant pathogens (Frantzeskakis et al. 2019; Raffaele et al. 2010; Raffaele and Kamoun 2012), and highlight that a clear two-speed genome compartmentalization is not rule for all plant pathogens. It is conceivable that the plethora of different genome organizations reflects the tremendous genomic diversity in filamentous microbes, and is not *per se* tied to the *speed* assumption underling the two-speed genome hypothesis (Frantzeskakis et al. 2019).

A second key assumption of the 'two-speed' hypothesis is that genes embedded in repeat-rich regions are assumed to evolve *faster* compared with genes residing in the core genome. However, what *speed* implies is often glossed over. Does speed imply a higher mutation rate per generation? Or does speed imply stronger positive selection? Repeat-rich genomic regions in plant pathogens are often enriched for structural variations, copy-number variations, or sequence polymorphisms (Raffaele et al. 2010; Faino et al. 2016; Plissonneau et al. 2018; Frantzeskakis et al. 2018; Thon et al. 2006; Schirawski et al. 2010). Furthermore, *fast*-evolving genes such as effectors are often identified by genome-wide scans for positive selection (Aguileta et al. 2009; Sánchez-Vallet et al. 2018). Nevertheless, not all effectors evolve under positive selection or even show signs of increased sequence polymorphisms. For example, *in planta* expressed effector genes in the vascular wilt pathogen *Verticillium dahliae* display few sequence polymorphisms and are not evolving under positive selection but spur frequent gene presence/absence polymorphisms between different *V. dahliae* strains (de Jonge et al. 2013, 2012; Kombrink et al. 2017; Faino et al. 2016; Depotter et al. 2019). The common observation of genes under positive selection and high degree of polymorphisms in repeat-rich genomic regions are the result of selection and thus are not *per se* an indication for speed. To our knowledge, genomic mutation rates have not yet been rigorously assessed in plant pathogens. Such experiments would require carefully controlled conditions and well-designed sequencing strategies to avoid false positives. Furthermore, mutation rates cannot easily be estimated from population sequencing data. Consequently, we know little about whether mutation rates truly differ between genome compartments, and how they differ to genomes lacking clear genome compartmentalization. Most observed differences in polymorphisms in repeat-rich regions can be explained by mutation accumulation due to relaxed selection. In line with Frantzeskakis and colleagues (Frantzeskakis et al. 2019), we therefore propose to refer to these regions as 'dynamic compartments' rather than '*fast*-evolving' regions.

Apart from plant pathogens, compartmentalized genomes have been found in many bacteria and eukaryotes where co-expressed genes, genes that operate in the same biological process, or genes involved in adaptation cluster together (Gokcumen et al.

2011; Lynch and Conery 2003; Hurst et al. 2004; Batada and Hurst 2007; Yeaman 2013) (Fig. 1). In plants and animals, for example, genes with roles in immunity often cluster and are located in dynamic regions (Roach et al. 2005; Seidl and Thomma 2017; Kawakatsu et al. 2016; Mascher et al. 2017; Leister 2004; Gokcumen et al. 2011; Laun et al. 2006). Notably, irrespective in which species genome compartmentalization has been observed, this phenomenon has nearly always been linked to the occurrence of repetitive genomic elements.

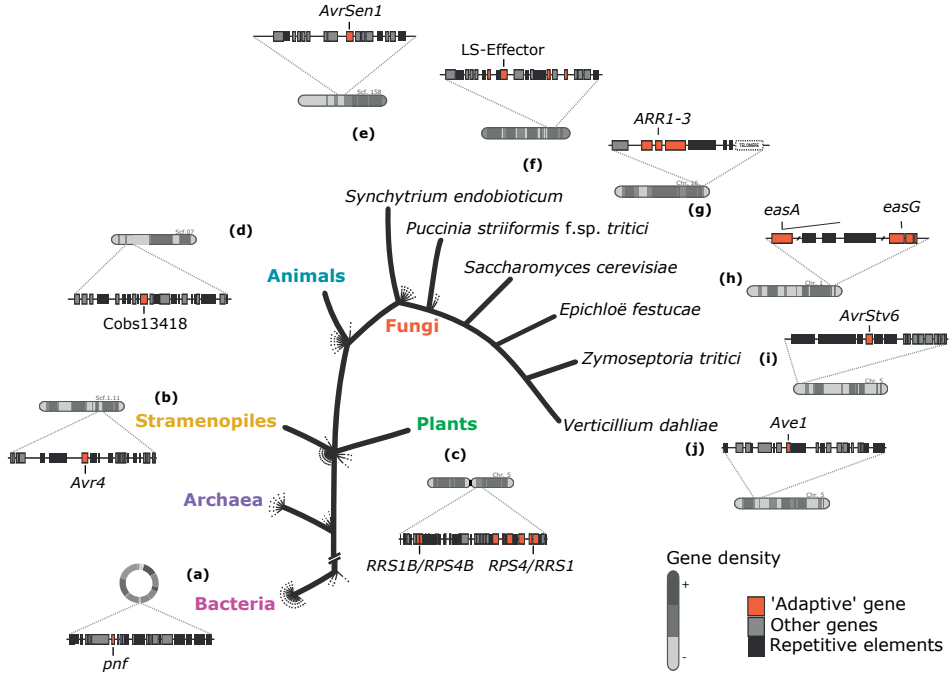


Figure 1. Dynamic genomic compartments occur in many branches of the tree of life.
Illustration of the phylogenetic tree depicting the relationship between bacteria, archaea, and eukaryotes highlights few **(a-d)** major non-fungal taxa (bacteria, Stramenopiles, animals, plants) and members of the fungal taxa **(e)** Chytridiomycetes, **(f)** Basidiomycetes, and **(g-j)** Ascomycetes. These diverse taxonomic groups contain species that have adapted to different ecological niches and lifestyles. The species differ in their precise genome organizations, but dynamic compartments are enriched to be encode genes will roles in virulence **(a-b)** (Heermann and Fuchs 2008; Waterfield et al. 2002; Duchaud et al. 2003; Haas et al. 2009; Raffaele et al. 2010), immunity **(c)** (Kawakatsu et al. 2016; van Wersch and Li 2019), or environmental responses **(d)** (Schrader et al. 2017). A spectrum of diverse dynamic genomic compartments have been reported in diverse fungi **(e-j)** ranging from genomes without clear genomic compartments **(e-f)** (van de Vossenberg et al. 2019a, 2019b; Schwessinger et al. 2018; Xia et al. 2018) to clearly defined genome compartments **(g-j)** (Peter et al. 2018; Yue et al. 2017; Brown et al. 2010; Fleetwood et al. 2011; Kema et al. 2018; Plissonneau et al. 2018; Zhong et al. 2017; de Jonge et al. 2012, 2013; Faino et al. 2016; Winter et al. 2018). Importantly, irrespective of the details of the genome compartmentalization, effector genes and other genes with roles in adaptation in fungal plant pathogens and also other species typically co-localize with transposable elements **(b-j)**.

Do transposable element insertion dynamics create compartmentalized genomes?

Repetitive sequences such as TEs are ubiquitous in eukaryotes (Wicker et al. 2007), although TE content and distribution can remarkably differ between species. For example, the TE content in fungi can differ between very streamlined genomes as found in the yeast *Saccharomyces cerevisiae* (~3.3%) (Carr et al. 2012) and genomes with a TE content above 80% such as the ectomycorrhiza *Cenococcum geophilum* (Peter et al. 2016) or the wheat powdery mildew *Blumeria graminis* (Muller et al. 2019b). Genomes are colonized by actively transposing TEs, but these evolutionary young TEs are not evenly distributed in fungal genomes (Faino et al. 2016; Muszewska et al. 2019; Cook et al. 2020). Disruptive impacts of new TE insertions trigger purifying selection, while insertions into non-coding regions should have a less dramatic impact on host fitness (Lynch 2007). Thus, TE insertions are likely under relaxed selection in gene-poor and repeat-rich regions. The number of TE insertions in gene-poor regions may slowly increase, while deleterious TE insertions into genes or regulatory regions will be purged very rapidly from the population (Fig. 2). Furthermore, random processes can act on the population frequency of inserted TEs. If a pathogen population undergoes a bottleneck or colonizes a new habitat, chance events can drastically increase or decrease the frequency of a TE at a specific locus within the population.

TE insertions tend to accumulate in accessory chromosomes or accessory compartments in core chromosomes (Croll and McDonald 2012; Sánchez-Vallet et al. 2018). In the fungal pathogens *Fusarium oxysporum*, *Leptosphaeria maculans*, and *Zymoseptoria tritici*, accessory chromosomes or accessory compartments are generally characterized by low gene density, low gene expression, facultative heterochromatin, and enrichment of species-specific genes (Fokkens et al. 2018; Soyer et al. 2014; Feurtey et al. 2019; Plissonneau et al. 2018). Relaxed purifying selection on accessory genes could have allowed TEs to accumulate. Frantzeskakis and colleagues (2019) argued purifying selection should remain strong on housekeeping genes regardless of their localization (Frantzeskakis et al. 2019) and accessory genes should universally experience more relaxed selection. However, how did effectors end up or emerge in close proximity to TEs? Is it a consequence of relaxed selection against TE insertions or selection favoring the localization of effectors in dynamic regions? The challenge will be to disentangle these two fundamentally linked phenomena in the future.

Dynamic chromatin for dynamic compartments

Active TEs have a strong mutagenic potential, and consequently TE-rich regions are often highly condensed into heterochromatin to suppress TE activity (Grewal and Jia 2007). The TE-rich dynamic compartments in filamentous microbes are typically associated with condensed heterochromatin (Connolly et al. 2013; Schotanus et al. 2015; Moller et al. 2019; Janevska et al. 2018; Wang et al. 2017; Cook et al. 2020). However, facultative heterochromatin can rapidly switch to an open euchromatic state depending on developmental stage or environmental cues, and therefore act as an important modulator of effector gene expression (Soyer et al. 2014, 2019), secondary metabolite gene clusters (Chujo and Scott 2014; Chujo et al. 2019; Pfannenstiel and Keller 2019; Collemare and Seidl 2019), or genes encoding proteins with various roles in carbohydrate metabolism (Fokkens et al. 2018).

Next to playing an important role in modulating gene expression, chromatin has been hypothesized to influence genome evolution (Seidl et al. 2016). Typically, heterochromatin is thought to suppress recombination (Grewal and Jia 2007). However, this observation is inconsistent with data from pathogen genomes where heterochromatic regions are often unstable or enriched in copy number variations (Hastings et al. 2009; Schotanus et al. 2015; Seidl et al. 2016; Plissonneau et al. 2018; Möller et al. 2019). Heterochromatic regions are considered to be more prone to DNA breakages (Sasaki et al. 2014) and heterochromatic regions often co-localize within the nucleus (Galazka et al. 2016), and it is conceivable that chromatin structure can also impact genetic variation. Furthermore, heterochromatic regions are often associated with higher densities of single nucleotide polymorphisms compared with euchromatic regions (Fokkens et al. 2018; Makova and Hardison 2015; Wang et al. 2017; Schuster-Böckler and Lehner 2012), which is likely due to altered accessibility for DNA repair machineries (Sun et al. 2016). Thus, it has been hypothesized that heterochromatin serves as cradle for genomic variability (Yasuhara et al. 2005; Liu et al. 2020a; Seidl et al. 2016). However, the roles of chromatin in the formation and evolution of dynamic compartments is still elusive. Is the association between dynamic compartments and heterochromatin a consequence of heterochromatin formation to suppress and silence recent TE insertions (Stuart et al. 2016; Rebollo et al. 2011; Sentmanat and Elgin 2012)? Alternatively, heterochromatin and higher degrees of polymorphism could directly drive the formation of the dynamic regions (Fig. 2), for instance by the spread of heterochromatic regions (Wang et al. 2015a, 2014; Dixon et al. 2012) or by co-opting heterochromatic genomic regions such as sub-telomeres (Hocher and Taddei 2020; Juarez-Reyes and Castano 2019). These hypotheses are not mutually exclusive, yet it remains challenging to experimentally elucidate how chromatin states affect genome compartmentalization dynamics and the embedded genes.

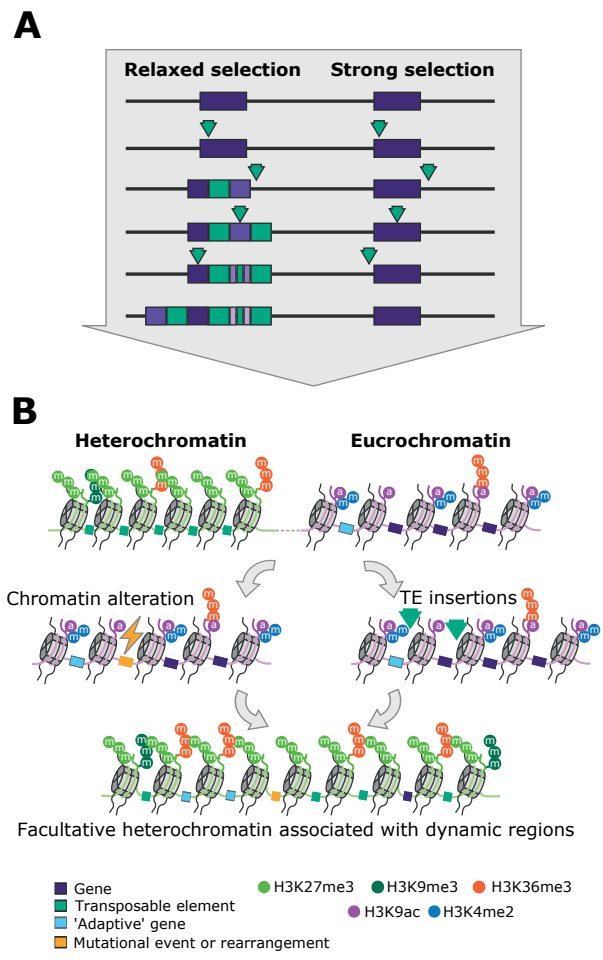


Figure 2. Origin and evolution of dynamic compartments in filamentous plant pathogens.
(A) Transposable elements (TEs) insertions events (green arrows) occur in theory randomly throughout the genome. In gene-poor regions and under relaxed selection these insertions would be tolerated. Over evolutionary time, the number of TE insertions will increase as additional TE insertions would be tolerated, and consequently genes (purple boxes) located in these regions would be 'trapped' in proximity to TEs. The TE accumulation could over longer time contribute to the formation of dynamic compartments. In gene-rich regions, strong selection is expected to purge TE insertions and TE content would maintain constant over time. **(B)** It is conceivable that the chromatin landscape contributes to the formation of dynamic compartments. Genomic rearrangements, e.g. mediated by TEs, could modify the chromatin landscape and induce the spread of heterochromatin, thereby altering the boundaries of different chromatin conformations (Wang et al. 2015a, 2014). Furthermore, TE insertions could induce chromatin-based silencing mechanisms aimed to further suppress their spreading (Rebollo et al. 2011; Sentmanat and Elgin 2012), and thereby induce heterochromatin formation and spread.

The chicken and egg problem: dynamic compartments and genes underlying rapid adaptation

Effector genes and other genes with roles in rapid adaptation are often located in dynamic compartments (Dong et al. 2015; Raffaele and Kamoun 2012) (Fig. 1). In line with the two-speed genome hypothesis, the presence of adaptive genes in TE-rich dynamic compartments is considered beneficial (Seidl and Thomma 2017). However, the evolutionary transitions towards a compartmentalized genome architecture each have to be advantageous and favored by selection over less compartmentalized genome architectures. At what stage did a pathogen species harbor different variants of a genome architecture and selection could favor one over the other? Did a compartmentalized genome architecture emerge through random events (*e.g.* massive TE proliferation) in some ancestral pathogen

species and selection favored the pathogen species carrying two-speed genomes? Is this compartmentalization reversible? Resolving such questions will clarify our understanding of why a compartmentalized genome is thought to be beneficial. Understanding mechanisms driving genome evolution can be fruitful to answer these major questions. TEs could influence effectors through leaky TE defense mechanisms such as repeat-induced point mutation (Galagan and Selker 2004; Fudal et al. 2009), by altering the expression of the effector (Chuong et al. 2016; Omrane et al. 2017), through disturbance through ectopic recombination (Devos et al. 2002), by leaking of silencing (Hollister and Gaut 2009), or even by creating effector genes *de novo* (Nottensteiner et al. 2018; Sabelleck and Panstruga 2018). Deletion of a TE-rich region carrying recognized effectors enables plant pathogens such as *Z. tritici* or *V. dahliae* to increase virulence on specific host genotype (de Jonge et al. 2012; Hartmann et al. 2017). Conceivably, these processes could potentially also reduce fitness, e.g. by deleting, mutating, or silencing effector genes. Thus, what are the selective advantages of the association between effector genes or other genes with adaptive roles and TE-rich dynamic compartments, and how did dynamic compartments emerge in the first place?

TE-rich compartmentalized genomes are commonly observed in nature (Lynch and Conery 2003; Lynch 2007) and their maintenance (e.g. replication and transcription) could be evolutionary costly (Oliver and Greene 2009; Schrader and Schmitz 2019; Bennetzen and Wang 2014), and thus it is conceivable that TE-rich compartments directly or indirectly contribute to higher pathogen fitness. It has been proposed that over evolutionary time, compartmentalized genomes can evolve from a random genome through TE-mediated rearrangements, which increases the probability of rapid adaptation to novel environments (Crombach and Hogeweg 2007). Thus, over evolutionary timescales these TE-rich dynamic compartments can emerge and are maintained as they positively contribute to evolvability (Kidwell and Lisch 2001; Slotkin and Martienssen 2007; Fouche et al. 2020; Cuypers and Hogeweg 2012; Fablet and Vieira 2011). Due their increased evolvability, these regions can serve as cradles for adaptive genome evolution.

Conclusions

Plant pathogenic fungi provide some of the most fascinating examples of how genome evolution is linked to phenotypic trait evolution. Effector genes encoded in fast evolving genome compartments can undergo rapid rearrangements, silencing, and activations due to the unique genomic environment. However, recent genome analyses across plant pathogenic fungi suggest that no unifying rules exist how effector genes are localized in the genome. We suggest that the evolutionary history, mode of reproduction, and population structure have strongly influenced characteristics of the pathogen genome. Such features include chromatin landscape along chromosomes, the dynamics of TE insertions, how

selection can act on beneficial mutations, and how random processes such as founder effects structure polymorphism within a species. A major unanswered question is how associations of effector genes and fast evolving genome compartments have evolved in some lineages but not others. A crucial missing link is convincing evidence for the stepwise re-localization of effector genes to fast-evolving genome compartments or for the emergence (*i.e.* birth) of effector genes in fast evolving genome compartments. Suitable pathogen systems with multiple closely related species and high-quality genome and population resources will be critical to address these questions in the future.

Acknowledgments

DET acknowledges funding by the Consejo Nacional de Ciencia y Tecnología, México. DC acknowledges funding by the Swiss National Science Foundation and Pierre Mercier pour la science.

REBUTTAL
# libraries
# Base environment
library(Rsamtools)
library(GenomicAlignments)
library(BiocParallel)
library(Biostrings)
library(ggsci)
library(data.table)
library(tidyverse)
library(tidyselect)
library(plyr)
library(dplyr)
library(IRanges)
library(GenomicRanges)
library(ShortRead)
library(doParallel)
library(ggplot2)
library(magrittr)
library(rtracklayer)
library(ComplexHeatmap)
library(cluster)
library(viridis)
library(gplots)
library(scales)
library(preprocessCore)
library(tidyr)
library(tidyselect)
library(tidyverse)
library(tibble)
library(ggridges)
library(plyranges)
library(genomation)
library(EnrichedHeatmap)
library(circlize)
library(magick)
library(strawr)
library(hicrep)
library(colorspace)
library(GenomicRanges)
library(InteractionSet)
library(ggpubr)
library(FactoMineR)
library(factoextra)
library(ggpointdensity)
library(pals)
library(ggridges)
library(ggpubr)
library(stringr)
library(colorspace)
# change for the one in my personal computer
setwd("~/Dropbox/rebuttal_raw/material")
# core genes
core.25ka<- read.table("~/Dropbox/rebuttal_raw/material/core.25k.a.bed",sep="\t",fill=T,
col.names=c("Chr_bnd","start_bnd","end_bnd","BND","Ins","X","Chr","start","end","strand","genes","ID",
"baseMean","log2FoldChange","lfcSE","stat","pvalue","padj","O","P","Q","R","S","W","Z"))%>%
dplyr::select(-X,-O,-P,-Q,-R,-S,-W,-Z)%>%dplyr::mutate(region=rep("Core"))
ov <- core.25ka%>%group_by(Chr_bnd)%>%mutate(a=start-end_bnd,b=end-start_bnd)%>%mutate(ov=ifelse(a<0&b>0,"ov","no"))%>%filter(ov=="ov")%>%
dplyr::select(-a,-b)%>%mutate(dist=rep(0),cat=rep("0k"))
ov2 <- core.25ka%>%group_by(Chr_bnd)%>%mutate(a=start-end_bnd,b=end-start_bnd)%>%mutate(ov=ifelse(a<0&b>0,"ov","no"))%>%filter(ov=="no")

CHAPTER 3

Transposable elements contribute to genome dynamics and gene expression variation in the fungal plant pathogen *Verticillium dahliae*

David E. Torres[1,2], Bart P.H.J. Thomma[2,3], and Michael F. Seidl[1]

[1]Utrecht University, NL
[2]Wageningen University and Research, NL
[3]University of Cologne, DE

Published in *Genome Biology and Evolution* 13,7 (2021): evab135

Abstract

Transposable elements (TEs) are a major source of genetic and regulatory variation in their host genome and are consequently thought to play important roles in evolution. Many fungal and oomycete plant pathogens have evolved dynamic and TE-rich genomic regions containing genes that are implicated in host colonization and adaptation. TEs embedded in these regions have typically been thought to accelerate the evolution of these genomic compartments, but little is known about their dynamics in strains that harbor them. Here, we used whole-genome sequencing data of 42 strains of the fungal plant pathogen *Verticillium dahliae* to systematically identify polymorphic TEs that may be implicated in genomic as well as in gene expression variation. We identified 2,523 TE polymorphisms and characterize a subset of 8% of the TEs as dynamic elements that are evolutionary younger, less methylated, and more highly expressed when compared with the remaining 92% of the TE complement. As expected, the polyrmorphic TEs are enriched in the adaptive genomic regions. Besides, we observed an association of dynamic TEs with pathogenicity-related genes that localize nearby and that display high expression levels. Collectively, our analyses demonstrate that TE dynamics in *V. dahliae* contributes to genomic variation, correlates with expression of pathogenicity-related genes, and potentially impacts the evolution of adaptive genomic regions.

Introduction

Repetitive sequences such as transposable elements (TEs) are ubiquitous components of prokaryote and eukaryote genomes (Bowen and Jordan 2002; Wicker et al. 2007). In eukaryotes, the repeat content can differ between species and range from only ~5.6% in the fish *Tetraodon nigroviridis* (Jaillon et al. 2004) to up to 85% in the maize genome (Schnable et al. 2009; Jiao et al. 2017). TEs can transpose to different locations within the host genome, and thus are often considered 'mobile genetic elements' (McClintock 1950). TEs have been traditionally classified into two classes based on their mode of transposition: class I retrotransposons transpose via a copy-and-paste mechanism, while class II DNA transposons mainly mobilize via a cut-and-paste system (Wicker et al. 2007; Piégu et al. 2015). TEs can impact host genome organization and gene function by disrupting protein-coding regions, modifying gene expression, or causing large-scale chromosomal rearrangements (Feschotte 2008; Eichler and Sankoff 2003; Le et al. 2015; Mita and Boeke 2016; Bennetzen and Wang 2014). Consequently, TEs and their activities have been considered to generally decrease fitness, and TEs and TE-induced mutations are typically rapidly purged from a population. Furthermore, TE activities are generally suppressed by host defense processes. These processes include DNA methylation (Selker et al. 2003; Zemach et al. 2010), histone modifications associated with highly condensed heterochromatin (Slotkin and Martienssen 2007; Hocher et al. 2018; Hocher and Taddei 2020), targeted processes such as RNA-silencing (Lippman and Martienssen 2004; Nicolas et al. 2013; Yadav et al. 2018b), or repeat-induced point (RIP) mutation, a fungal specific process in in which duplicated or repetitive sequences are targeted by C to T mutations (Cambareri et al. 1989; Galagan and Selker 2004; Wang et al. 2020b; Fudal et al. 2009).

Over the last decades, besides their detrimental effects, TEs have been increasingly recognized to play important roles in adaptive genome evolution (Venner et al. 2009; Lynch et al. 2011a; Bourque et al. 2018; Hua-Van et al. 2011). For example, TEs can be co-opted as a source of *cis*-regulatory elements (Chuong et al. 2016; Kitano et al. 2018) and mediate swift changes in the transcriptional landscape under biotic and abiotic stresses (Muszewska et al. 2019; Sinzelle et al. 2009; Bourque et al. 2018). A classic example is the dark pigmentation of the peppered moth in response to industrialization that was caused by a single TE insertion into the first intron of a cell-cycle regulator gene, changing the transcript abundance of this gene and allowing the insect to camouflage on polluted surfaces (Cook and Saccheri 2013; Van't Hof et al. 2016). Changes in mutation rates in genes in proximity TEs can be caused by leakage of TE silencing mechanisms (e.g., RIP) or DNA repair following TE insertion (Rouxel et al. 2011; Yoshida et al. 2016; Wang et al. 2020b; Fudal et al. 2009). Furthermore, abundant TEs and other repetitive elements can act as an ectopic substrate for double-strand break repair pathways, thereby fostering chromosomal rearrangements (structural variants; SVs), such as chromosomal translocations and inversions as well as large-scale duplications and deletions (Chan and Kolodner 2011; Bourque et al. 2018; Faino et al. 2016).

For example, TE expansions in crops such as tomato, rice, wheat and soybean, have been shown to foster an increased accumulation of SVs associated with important traits during domestication (Fuentes et al. 2019; Alonge et al. 2020; Liu et al. 2020b; Walkowiak et al. 2020). Interestingly, as TEs and SVs co-localize, they tend to form clusters in discrete regions in many eukaryotic organisms (Hastings et al. 2009; Lin and Gokcumen 2019; Marand et al. 2019). This co-localization, and its consequences for genome variation and organization, are important to understand the genome evolution of eukaryotic organisms (Kidwell and Lisch 2001; Wright and Finnegan 2001; Lynch and Conery 2003; Lynch 2007; Hurst et al. 2004). Remarkably, these TE-rich regions are often enriched in environmentally responsive genes, for example in genes associated with roles in immunity or in secondary metabolism in plants (Kawakatsu et al. 2016; van Wersch and Li 2019; Sakamoto et al. 2004; Seidl et al. 2016). Therefore, TE-rich compartments are considered relevant for contribution to adaptation to various abiotic and biotic stresses (Crombach and Hogeweg 2007; Knibbe et al. 2007).

TE-rich compartments play important roles in the co-evolutionary 'arms-race' between pathogens and their hosts. Pathogens secrete molecules known as effectors to establish a successful parasitic relationship on their hosts (Rovenich et al. 2014; Sánchez-Vallet et al. 2018). In turn, hosts evolve receptors to activate immunity and halt pathogen colonization (Cook et al. 2015; Han 2019). Thus, in order to overcome recognition by host immune systems, pathogens have to purge or diversify effector genes once their gene products become recognized (Fouche et al. 2018; Badet and Croll 2020; Moller and Stukenbrock 2017). In many filamentous pathogens, effector genes are enriched in TE-rich and gene-sparse compartments, while they are typically underrepresented in TE-poor and gene-dense genomic regions that typically harbor housekeeping genes (Dong et al. 2015; Seidl and Thomma 2014; Frantzeskakis et al. 2020). TE-rich compartments are often characterized by increased substitution rates and increased occurrence of SVs and presence/absence polymorphisms (Raffaele et al. 2010; Croll and McDonald 2012; de Jonge et al. 2013, 2012; Dutheil et al. 2016; Fokkens et al. 2018; Plissonneau et al. 2018; Wang et al. 2017; Gan et al. 2020; Grandaubert et al. 2019; Hartmann and Croll 2017; Hartmann et al. 2017; Wyka et al. 2020). Notably, a similar association of TEs with genes involved in immune responses has also been observed in plant hosts (Kawakatsu et al. 2016; Seidl and Thomma 2017; Leister 2004; Mascher et al. 2017). Therefore, TE-rich compartments have been suggested to facilitate the generation of variation that is necessary to fuel the 'arms races' between pathogens and their hosts (Frantzeskakis et al. 2020, 2019; Schrader and Schmitz 2019; Torres et al. 2020; Seidl and Thomma 2017).

Verticillium dahliae is an asexual soil-borne fungal plant pathogen that colonizes the xylem of hundreds of susceptible host plants (Klosterman et al. 2009, 2011; Fradin and Thomma 2006; Inderbitzin and Subbarao 2014). Genomic comparisons between *V. dahliae* strains revealed that their genomes evolved by extensive chromosomal rearrangements (Faino et al. 2016; de Jonge et al. 2013). These rearrangements include chromosomal translocations, inversions, large-scale deletions, and duplications that collectively contributed

to the formation of adaptive genomic compartments that originally have been referred to as lineage-specific (LS) regions, as these are variable between *V. dahliae* strains (Faino et al. 2016; de Jonge et al. 2013, 2012; Depotter et al. 2019; Klosterman et al. 2011). Initially, LS regions were solely defined by presence/absence polymorphisms among sequenced strains (de Jonge et al. 2013; Faino et al. 2016). However, subsequent work on the chromatin landscape in *V. dahliae* has revealed that these LS regions display a unique chromatin profile (Cook et al. 2020). Moreover, it was shown that a substantial number of additional genomic regions share chromatin characteristics with LS regions and, accordingly, that the amount of LS DNA has been underestimated (Cook et al. 2020). The originally identified LS regions, together with the additional genomic regions that share a similar chromatin profile, are now referred to as adaptive genomic regions (Cook et al. 2020). Importantly, these adaptive genomic regions are enriched for *in planta* induced effector genes that contribute to host colonization (de Jonge et al. 2013, 2012; Klosterman et al. 2011; Kombrink et al. 2017; Gibriel et al. 2019; Chavarro-Carrero et al. 2021; Cook et al. 2020). Furthermore, these regions are enriched in relatively young and transcriptionally active TEs (Faino et al. 2016; Cook et al. 2020). While a key role of TEs to drive formation and maintenance of adaptive genomic regions has been revealed (Amyotte et al. 2012; Cook et al. 2020; Faino et al. 2015, 2016; Shi-Kunne et al. 2018; Depotter et al. 2019; Seidl et al. 2020), it remains unclear what the impact of TE variation is on the biology of individual strains of *V. dahliae*. Changes in the TE landscape between *V. dahliae* strains, involving novel insertions as well as deletions, can cause structural as well as gene expression variation. Here, we exploited a collection of *V. dahliae* strains to analyze their TE landscape and systematically identify TE polymorphisms. Furthermore, we analyzed the polymorphic TEs in detail to study their impact on genome function and organization. Our analysis extends previous hypotheses on the relevance of TE dynamics in adaptive genomic regions (Faino et al. 2016; Cook et al. 2020), and further characterizes a TE subset implicated in genomic variation and gene expression in *V. dahliae*.

Results

Structural variants are common in *V. dahliae*

Previous genomic comparisons among a narrow selection of *V. dahliae* strains have revealed extensive chromosomal rearrangements, which were proposed to have contributed to the establishment of adaptive genomic regions (Faino et al. 2016; de Jonge et al. 2013). To systematically assess their prevalence and abundance in a considerably larger collection of *V. dahliae* strains, we used paired-end sequencing reads for *de novo* prediction of SVs in 42 *V. dahliae* strains (Chavarro-Carrero et al. 2021; Cook et al. 2020) together with the gapless genome assembly of strain JR2 as a reference (Faino et al. 2015). With a combination of four commonly used SV-callers (Supplementary Fig.1), 919 high-confidence SVs were identified in total, comprising 410 deletions, 49 inversions, 24 duplications, and 436 translocations

(Fig. 1A, Supplementary Fig.1). Importantly, we successfully recalled all 19 chromosomal rearrangements that were previously manually identified between *V. dahliae* strains JR2 and VdLs17 (Faino et al. 2016). To investigate the occurrence of the 919 SVs across the *V. dahliae* strains, we calculated the frequency of each SV over all the strains included in the analysis. We observed that 87% (*n*=803) of the SVs is shared by <20% of the strains, i.e., they occur in less than eight strains (Fig. 1B). Among the SVs, deletions occur at the highest frequency (Figure 1A, B; $p<2\times10^{-16}$, Kruskal-Wallis test). Interestingly, the number of SVs per strain does not correspond with the phylogenetic relationship among the *V. dahliae* strains (Fig. 1C, D). Collectively, our data confirm that SVs occur commonly in *V. dahliae,* but that individual SVs are typically only shared by a small subset of strains.

Previously, we have demonstrated that SVs cluster at adaptive genomic regions in *V. dahliae* (de Jonge et al. 2013; Faino et al. 2016). Here, we similarly observed distinct genomic regions comprising multiple non-randomly distributed SVs ($p<2\times10^{-16}$, Kolmogorov-Smirnov test, Supplementary Fig. 2), enriched in 17 regions (Pignatelli et al. 2009) (hypergeometric test; $p<0.05$ after Benjamini-Hochberg correction; Supplementary Table 2), and co-localized with the previously identified adaptive genomic regions (one-sided Fisher´s exact test, $p = 0.00049$; Supplementary Fig. 3). Interestingly, SVs occur largely independently of nucleotide mutations in adaptive genomic regions (Supplementary Table 3). We observed that only 0.06% of the nucleotides in adaptive genomic regions are affected by single nucleotide variations while 14.8% are affected by SVs. Thus, we conclude that SVs are a more likely phenomenon to increase variation in adaptive genomic regions than nucleotide mutations, especially when considering the previously reported high levels of sequence conservation in adaptive genomic regions (de Jonge et al. 2013; Depotter et al. 2019; Faino et al. 2016; Shi-Kunne et al. 2018).

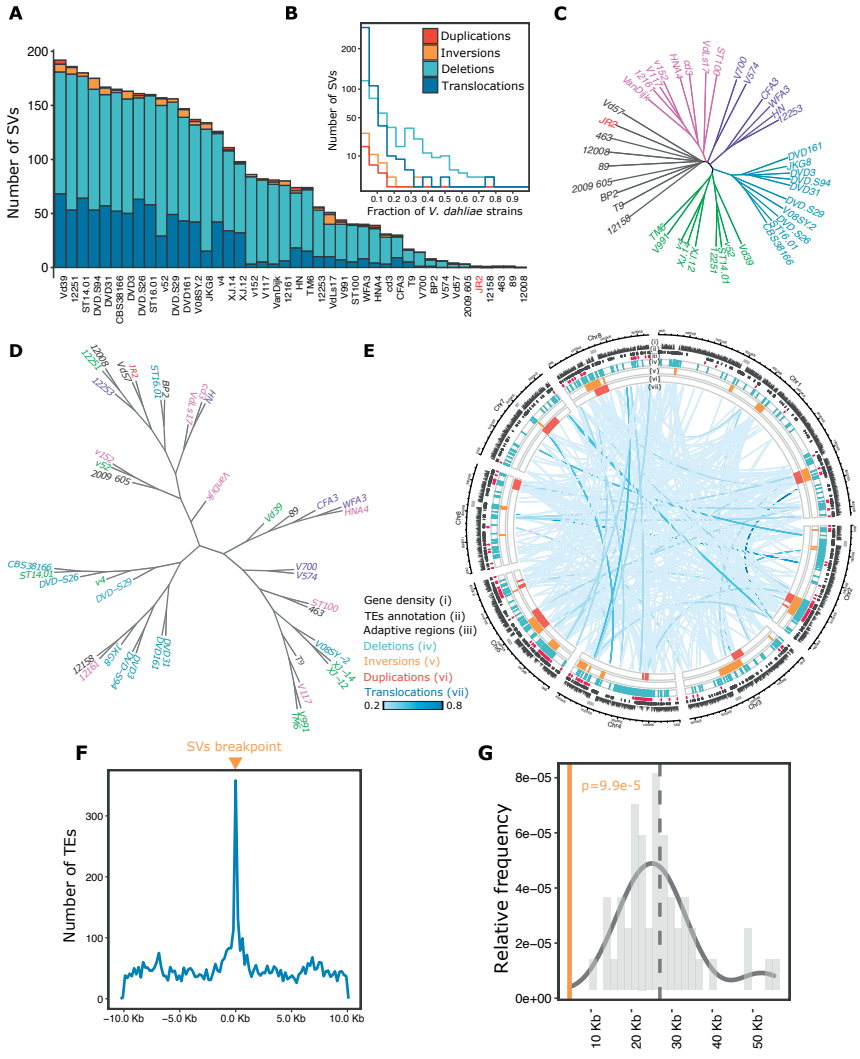


Figure 1. Structural variants (SVs) in a collection of 42 *Verticillium dahliae* strains.
(A) The number of SVs is depicted per type for each *V. dahliae* strain included in the analysis; the color-code for each variant type as indicated in panel **(B)**; **(B)** Frequency distribution of the 919 SVs in the collection of *V. dahliae* strains; **(C)** Unrooted neighbor-joining tree based on the presence-absence of 919 high confident SVs relative to the reference genome assembly of *V. dahliae* strain JR2 (red), color code depicts different groups; **(D)** Unrooted neighbor-joining tree of the collection of *V. dahliae* strains based on 287,251 high-confident SNVs relative to the reference genome assembly of *V. dahliae* strain JR2 (red); the color of the strain names corresponds to the colors represented in **(C)**; **(E)** Circular plot displaying the genomic distribution of the 919 SVs along the eight chromosomes of the *V. dahliae* strain JR2 reference genome assembly. The tracks are shown in the following order from outside to inside: centromeric regions (Seidl et al. 2020), gene density in 10 kb windows, TE distribution, adaptive genomic regions distribution (Cook et al. 2020), 410 deletions (light blue), 49 inversions (yellow), 24 duplications (red), and 436 translocations (blue lines). The color intensity of the lines for an individual translocation is based on its frequency in the *V. dahliae* collection; **(F)** Number of TEs around SV breakpoints is displayed in 10 kb windows; **(G)** The distribution of 10,000 random permutations of distances between SVs and TEs in adaptive genomic regions. The vertical dashed line represents the mean number of random occurrences, the yellow line indicates the average distance at 4,681.85 bp, significance p= 9.99×10^{-5}, z-score=-2.32.

A subset of transposable elements in *V. dahliae* is polymorphic

TEs and other repetitive elements are often considered to drive the formation of SVs (Bourque et al. 2018). In *V. dahliae*, TEs have been proposed to mediate genomic rearrangements (Faino et al. 2016), and TEs and repetitive elements encompass 12.3% of the *V. dahliae* strain JR2 genome (Faino et al. 2015). Class I LTR retroelements are the most abundant, while Class II DNA is less abundant (Supplementary Table 4). In line with previous observations (Faino et al. 2015, 2016; de Jonge et al. 2013; Klosterman et al. 2011; Cook et al. 2020), TEs are significantly enriched (3.5-fold) in adaptive genomic regions when compared with the core genome and co-localize with SVs (Fig. 1F-G; Supplementary Fig. 4). As TE dynamics may promote the formation of SVs, insertions, and deletions of TEs (i.e., polymorphic TEs) may directly impact host genome organization. We identified TE insertion and deletion polymorphisms using TEPID (Stuart et al. 2016) by querying paired-end sequencing data for 42 *V. dahliae* strains relative to the *V. dahliae* strain JR2 reference genome; deletions are defined as TEs that are present in the reference genome but absent in other strains, while insertions are those present in at least one other strain and absent from the syntenic region in the reference genome. Our approach uses a single *V. dahliae* strain as a reference, and thus we will be limited by the reference TEs annotation when describing TE polymorphisms. For example, two DNA/Tc1-Mariner elements annotated in *V. dahliae* strain VdLs17 (Amyotte, et al. 2012) are absent in the reference strain JR2 (Faino, et al. 2015), and consequently these two elements would have not been considered in our analysis. In total, we identified 135 and 30 unique loci that underwent TE deletions and insertions, respectively; these loci are responsible for a total of 2,387 deletions and 136 insertions we identified across the 42 *V. dahliae* strains (Fig. 2B). In the case of insertions, we decided to keep only elements that belong to designated TE superfamilies, removing twelve 'unspecified' elements that do not show association to a specific TE superfamily or the reference sequence length is incomplete. Only around 8% of all TEs are polymorphic, i.e., they show insertion or deletion events in different *V. dahliae* strains compared with the reference strain JR2 (Fig. 2A). We observed that polymorphic TEs are significatively larger than non-polymorphic TEs (Supplementary Figure 6). To further analyze the polymorphic TEs, we first calculated their frequency in the *V. dahliae* strains relative to the reference strain JR2. This revealed that TE deletions occurred more frequently than insertions ($p = 0.0349$, one-sided Fisher´s exact test; Fig. 2C); a similar frequency distribution was also observed for SV deletions (Fig. 1C). Only five TEs display both deletions and insertions, of which one belongs to the LTR/Gypsy and four to the LTR/Copia superfamily. We observed that 33 polymorphic TEs (37% of which belong to the LTR/Copia superfamily) overlap protein-coding genes, 22 of which overlap introns and 11 exons. Thus, only a limited subset of TEs in the *V. dahliae* genome is polymorphic, while the vast majority is shared by all strains analyzed.

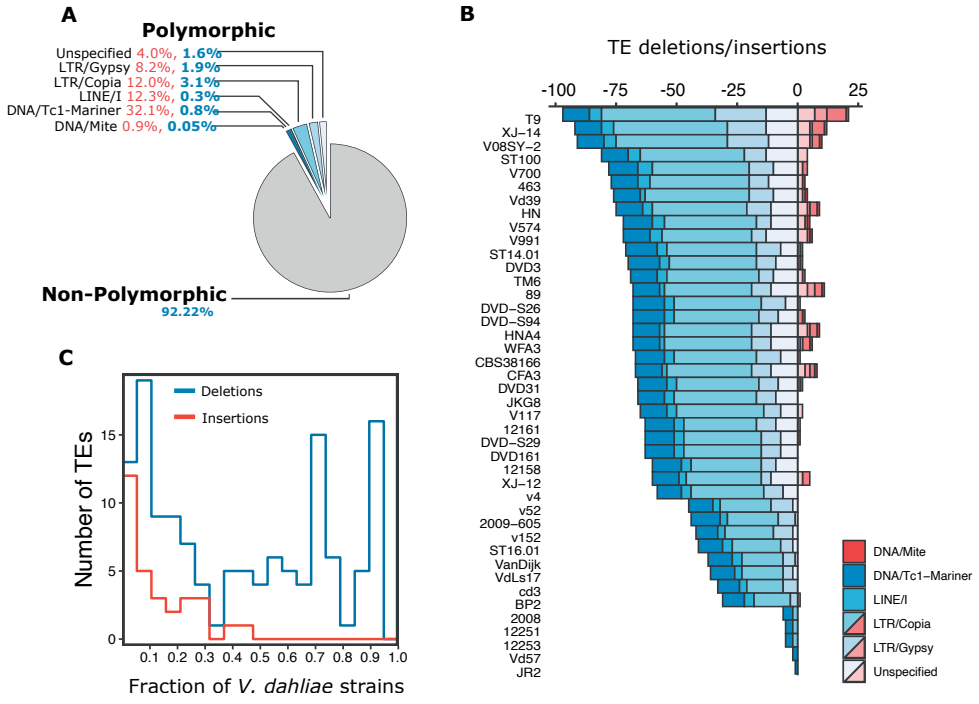


Figure 2. Transposable element polymorphisms in a collection of 42 *Verticillium dahliae* strains.
(A) The percentage of polymorphic (shades of blue) and non-polymorphic (grey) TEs in *V. dahliae* strain JR2 is depicted as a pie chart. The percentages of polymorphic TEs separated by superfamily relative to the TE annotation of the *V. dahliae* strain JR2 is shown in blue, and the percentage of polymorphic TEs within each superfamily is shown in red; **(B)** The proportion of identified TE polymorphisms is shown for every *V. dahliae* strain, depicting deletions (blue) and insertions (red) per superfamily**; (C)** Frequency distribution of 165 polymorphic TEs in the *V. dahliae* strain collection, depicted by TE variant type.

Polymorphic TEs are dynamic in *V. dahliae*

Fungi evolved a range of mechanisms to suppress TE activity and avoid TE expansions (Castanera et al. 2016). However, we previously observed that a fraction of TEs is transcriptionally active in *V. dahliae*, suggesting it comprises TEs that are potentially able to multiply and accumulate in the *V. dahliae* genome (Faino et al. 2016; Cook et al. 2020). Therefore, we hypothesized that polymorphic TEs are not suppressed and still able to multiply. To assess if polymorphic TEs lack typical signatures of suppression when compared with non-polymorphic TEs in *V. dahliae* strain JR2, we assessed DNA methylation levels (in the three methylation contexts; mCG, mCHG, and mCHH), RIP signature (Composite Repeat-Induced-Mutations Index; CRI), Kimura distance (divergence from the TE consensus sequence), GC content, and expression levels *in vitro* (counts per million; CPM) by quantifying the abundance of TE-derived reads across all different TE copies in the genome using TEtranscripts (Jin and Hammell 2018; Jin et al. 2015). Dimensional reduction by principal component analysis (PCA) revealed that polymorphic TEs cluster and are largely separated from the non-polymorphic TEs (Fig.

3A). On the first dimension of the PCA, explaining 31.6% of the variation, polymorphic and non-polymorphic TEs are separated based on CRI, DNA methylation, and GC content (Supplementary Fig. 5). This observation can be explained by the association of TEs with strictly heterochromatic regions, in which TEs are characterized by high CRI and DNA methylation levels as well as by low GC content (Cook et al. 2020). The second dimension of the PCA, explaining 14.5% of the variation, revealed a more distinct separation between polymorphic and non-polymorphic TEs mainly based on Kimura distance, DNA methylation, and transcriptional activity *in vitro* (Supplementary Table 5). Specifically, we observed that the sequence of polymorphic TEs is more similar to the consensus sequence of the TE family (Kimura distance) and is characterized by a higher GC content when compared with non-polymorphic TEs, indicating that polymorphic TEs are evolutionary younger when compared with non-polymorphic TEs (Fig. 3B). Notably, the CRI of polymorphic TEs is lower than non-polymorphic TEs and resembles that of genomic regions located in the core genome (Supplementary Fig. 6). Together with lower CG methylation levels in polymorphic TEs, our data suggest that polymorphic TEs are not silenced and are likely more active, as proxied by an increased transcriptional activity when compared with other TEs (Fig. 3B). As expected, polymorphic TEs are expressed at higher levels *in vitro* and *in planta* when compared with non-polymorphic TEs (Fig. 3B; Supplementary Fig. 6). Collectively, our analyses identified a subset of TEs as polymorphic among strains, and these tend to be evolutionary younger, less suppressed by RIP and methylation, as well as more highly expressed in *V. dahliae.*

Adaptive genomic regions in *V. dahliae* have been previously shown to be enriched for transcriptionally active TEs (Supplementary Fig. 4) (Faino et al. 2016; Cook et al. 2020). We hypothesized adaptive genomic regions are enriched for polymorphic TEs. Therefore, we analyzed the distribution of polymorphic TEs in *V. dahliae* strain JR2 (Fig. 3D), leading to the identification of clusters of polymorphic TEs that co-localize with adaptive genomic regions (p =0.00024, one-sided Fisher´s exact test; Supplementary Table 6; Fig. 3C). Interestingly, polymorphic TEs do not localize closer to SVs than non-polymorphic TEs, suggesting that both types of TEs can contribute to the formation of SVs. While deleted TEs were enriched in adaptive genomic regions, we could not observe a similar enrichment for TE insertions in the JR2 strain (Supplementary Fig. 7). Polymorphic TEs belonging to LTR/Copia superfamilies or unspecified repetitive elements are highly enriched in adaptive regions (p =8.24×10^{-9} and p =1.36×10^{-5} respectively, one-sided Fisher´s exact test), and for instance, LTR/Copia represents 51.1% of all polymorphic TEs in adaptive regions (Supplementary Table 7). Therefore, our results suggest that the LTR superfamily is an important component of the dynamic TE landscape.

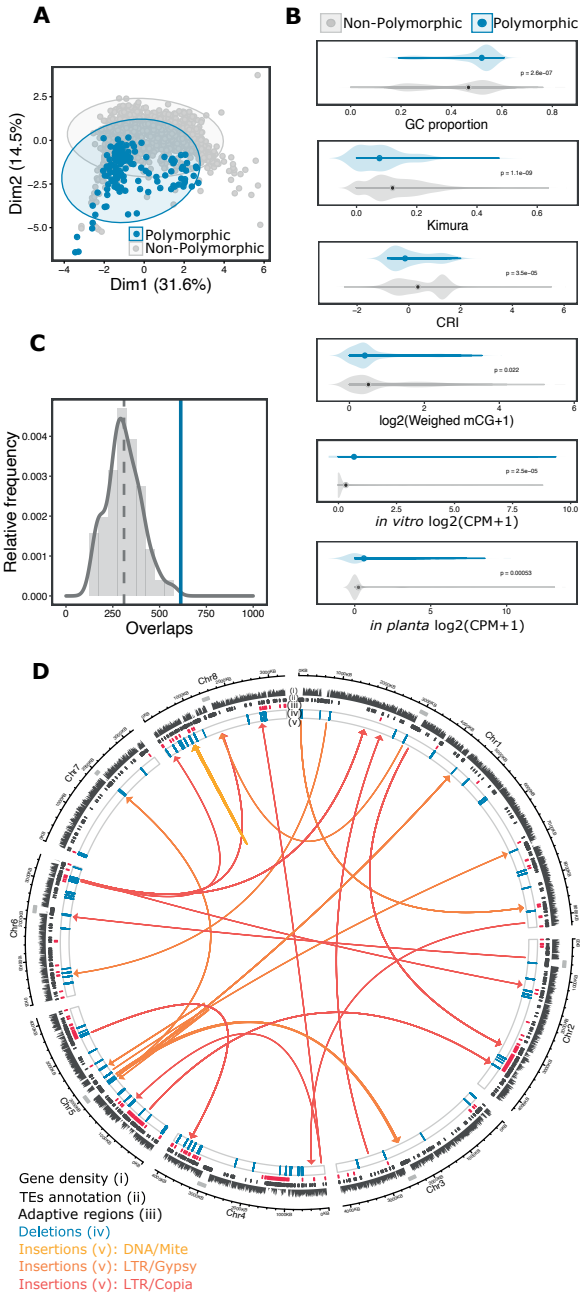


Figure 3. The TE landscape across the *V. dahliae* genome is dynamic.
(A) Principal Component Analysis of TEs in *V. dahliae* strain JR2 based on methylation (mCG, mCHG, mCHH), GC proportion, Kimura distance, Composite Repeat Induced Point Mutation index (CRI), frequency, and *in vitro* expression. Each point represents a single TE and colored ellipses represent the confidence interval for the polymorphic and non-polymorphic TEs; **(B)** Polymorphic (n=165) and non-polymorphic TEs (n=1,956) were compared based on GC proportion, the Kimura distance to the TE family consensus sequence, CRI, cytosine methylation, and *in vitro* expression (CPM; Counts per Million). Statistical significance was assessed using one-sided Wilcoxon rank-sum test; **(C)** The association between polymorphic TEs and adaptive genomic regions was tested using a permutation test. The vertical dashed line shows the mean number of polymorphic TEs overlaps expected at 10,000 random permutations and the blue line indicates the empirical number of overlaps, significance $p = 9.99\times10^{-5}$, z-score=3.2927; **(D)** Circular plot displays the genomic distribution of 165 polymorphic TEs along the eight chromosomes of *V. dahliae* strain JR2. The tracks display the centromeric regions, gene density (in 10 kb windows), TE annotation, adaptive genomic regions, TE deletions, and TE insertions (with arrows indicating the insertion direction) from outside to inside.

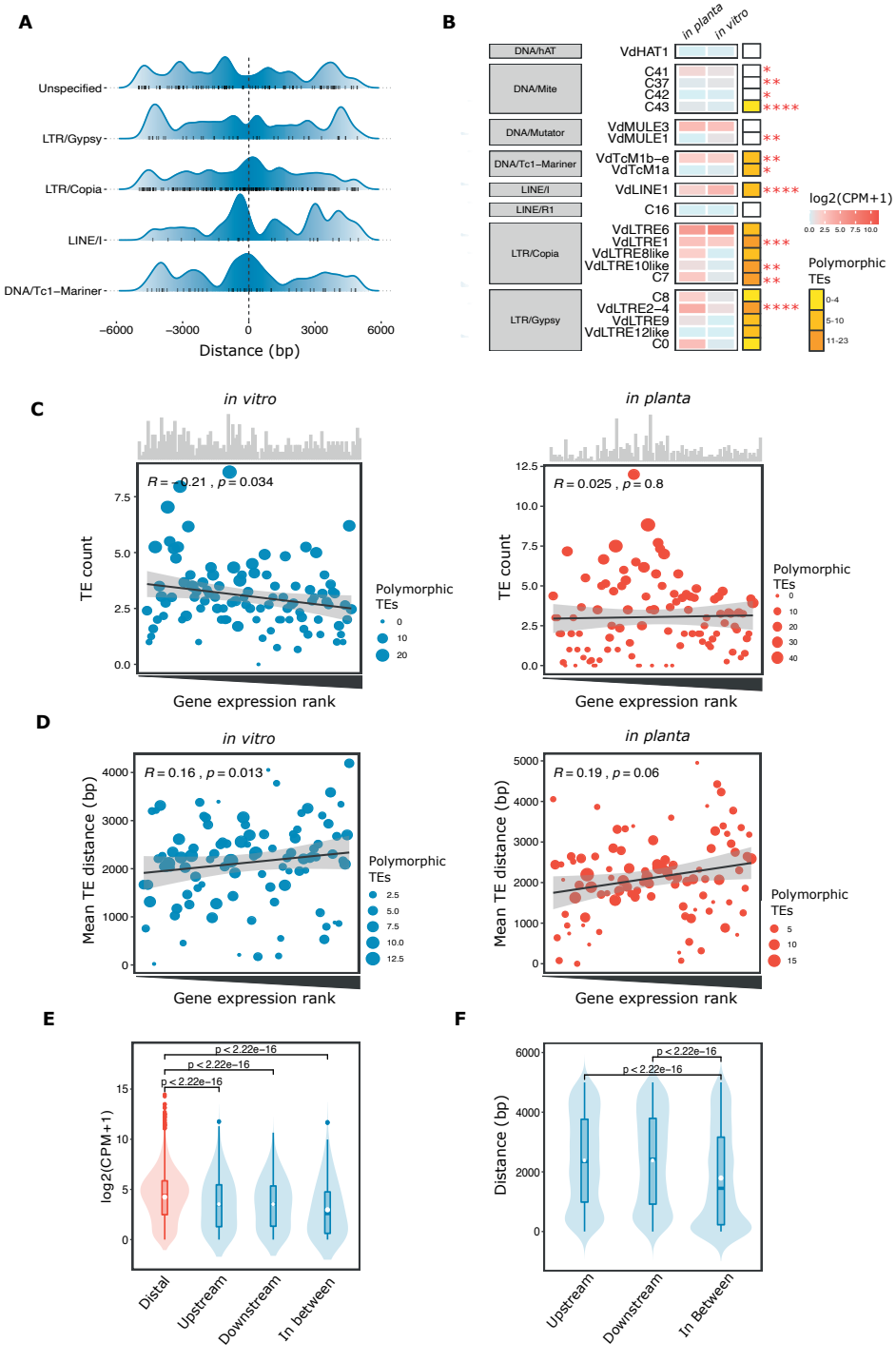
Polymorphic TEs impact the functional genome of *V. dahliae*

Next to their association with SVs, TE presence and activity have been suggested to directly or indirectly impact the functional genome, for example by modifying the expression of genes in their vicinity or by inducing changes in gene structure (Bourque et al. 2018; Schrader and Schmitz 2019). To establish to which extent TEs may affect gene expression in *V. dahliae* strain JR2, we first summarized the occurrence of TEs up to 5 kb upstream and downstream of all protein-coding genes (Supplementary Fig. 8). The majority of annotated TEs are located within 5 kb range of a gene (67% of total TEs, n=1,418; p =0.0094, one-sided Fisher's exact test), with an overrepresentation of polymorphic TEs belonging to the LTR/Copia, LINE/I, and DNA/Tc-1-Mariner superfamilies (Fig. 4A; Supplementary Fig. 8). By assessing the expression of TEs nearby genes (n=1,418) as a proxy for their activity, we observed that many TE families are expressed *in vitro* and typically at higher levels during infection (p<0.05, Fig. 4B; Supplementary Fig. 9). As expected, these families are abundant in polymorphic TEs (Fig. 4B), suggesting that environmental changes induce the expression of TEs.

>>

Figure 4. The presence of polymorphic transposable elements is correlated with expression of genes that localize nearby.

(A) Ridgeline shows the density distribution of polymorphic TEs, classified by TE superfamily, upstream and downstream of genes; **(B)** Unsupervised clustering of TE expression per TE family and superfamily, color-coded based on the mean-per-family expression value (log2(CPM+1)). The column depicts the abundance of polymorphic TEs within each group (yellow coded). Statistical significance for comparisons between *in vitro* (growth in MS media) and *in planta* (colonization of *Arabidopsis thaliana* at 28 days post inoculation) was assessed using a one-sided Wilcoxon rank-sum test * p<=0.05, ** p<=0.01, *** p<=0.001, **** p<=0.0001; **(C)** Relationship of gene expression ranks *in vitro* (blue) and *in planta* (red) and the number of TEs. Low rank = low expression and high rank = high expression. The plots display linear regression (dark line) and confidence interval (light grey) as well as the R and P values after linear regression. The histograms on top of the graph show the number of polymorphic TEs per expression rank and the dot-size reflects the proportion of polymorphic TEs relative to the total number of TEs per expression rank; **(D)** Relationship of gene expression rank and mean distance to TEs (bp) with annotations as for **(C)**; **(E)** Gene expression in relation to the TE context (TEs within 5 kb); no TE in proximity = Distal (n=8,742), or TEs flanked upstream (n=1,098), downstream (n=1,051), or in between (n=546; blue). Significant differences were assessed using one-sided Wilcoxon rank-sum test; **(F)** TE distance to genes in 5 kb windows; TEs in upstream (n=1,609), downstream (n=1,511) or in between (n=1,582). Statistical differences were assessed using one-sided Wilcoxon rank-sum test.


A
Unspecified
LTR/Gypsy
LTR/Copia
LINE/I
DNA/Tc1−Mariner
−6000
−3000
0
3000
6000
Distance (bp)
B
in planta
in vitro
DNA/hAT
VdHAT1
DNA/Mite
C41
C37
C42
C43
DNA/Mutator
VdMULE3
VdMULE1
DNA/Tc1−Mariner
VdTcM1b−e
VdTcM1a
LINE/I
VdLINE1
LINE/R1
C16
LTR/Copia
VdLTRE6
VdLTRE1
VdLTRE8like
VdLTRE10like
C7
LTR/Gypsy
C8
VdLTRE2−4
VdLTRE9
VdLTRE12like
C0
log2(CPM+1)
Polymorphic TEs
0-4
5-10
11-23
C
in vitro
in planta
R = −0.21 , p = 0.034
R = 0.025 , p = 0.8
TE count
Gene expression rank
Polymorphic TEs
D
in vitro
in planta
R = 0.16 , p = 0.013
R = 0.19 , p = 0.06
Mean TE distance (bp)
Gene expression rank
Polymorphic TEs
E
p < 2.22e−16
log2(CPM+1)
Distal
Upstream
Downstream
In between
F
p < 2.22e−16
Distance (bp)
Upstream
Downstream
In Between

To further assess the correlation between the proximity of TEs and the expression of nearby genes, we compared the expression of genes located within 5 kb windows from TEs under *in vitro* and *in planta* conditions. We observed that 23% of all annotated genes reside within 5 kb of a TE (*n*=2,696). Genes in proximity to TEs are transcribed at significantly lower levels when compared with genes that do not reside in close proximity to TEs ($p < 2\times10^{-16}$, *in vitro*, $p = 1.9\times10^{-9}$ *in planta*; Kruskal-Wallis test; Fig. 4E, Supplementary Fig.10). Therefore, we hypothesized that the presence of polymorphic TEs correlates with expression in *V. dahliae*. To further investigate this, we aggregated genes and TEs within 5 kb by their expression rank and obtained the number of TEs in each rank. We observed only a moderate negative correlation between the number of polymorphic TEs and gene expression ranks *in vitro* (Fig. 4C; R = –0.21, $p = 0.034$ linear regression). Conversely, the distance between genes and polymorphic TEs is moderately positively correlated with higher expression ranks (Fig. 4D; R = 0.16, $p = 0.013$ linear regression). These results suggest that polymorphic TEs are associated, albeit weakly, with genes expressed at low levels under *in vitro* conditions. However, we observed a different pattern when considering host colonization, as gene expression ranks *in planta* did not correlate with TE density (Fig. 4C; R = 0.025, $p = 0.8$ linear regression) whereas, similar to *in vitro* conditions, TE distance and expression are moderately positively correlated (Fig. 4D; R = 0.19, $p = 0.06$ linear regression). Collectively, our data suggest the polymorphic TE density nearby protein-coding genes correlates only weakly with gene expression in different conditions.

To further assess the relationship between TEs and gene expression, we categorized genes into four categories based on the relative position of TEs; genes with no TEs in proximity (*n*=8,742), genes with TEs localized upstream (*n*=1,098), downstream (*n*=1,051), or genes flanked by TEs on both sides (*n*=546). The distance between TEs and genes is significantly shorter when genes are flanked by TEs, when compared with genes having TEs only upstream or downstream (Fig. 4F). Importantly, genes flanked by TEs show reduced expression *in vitro* when compared with genes in the proximity of TEs located either upstream or downstream (Fig. 4E), a pattern we similarly observed during host colonization. We similarly observed this pattern when only considering the subset of polymorphic TEs (Supplementary Fig. 10, 11). Interestingly, during host colonization genes in proximity to polymorphic TEs do not show a significant decrease in expression (Supplementary Fig. 10). Thus, genes in proximity to non-polymorphic TEs showed lower expression levels when compared with genes in proximity to polymorphic TEs.

Pathogenicity-related genes co-localize with polymorphic TEs

To understand which types of genes are enriched in proximity to TEs, we performed functional annotation of the *V. dahliae* strain JR2 gene catalog using a Cluster for Orthologs Groups (COG) approach and subsequently tested for enrichment of genes with specific COG categories. No significant enrichment of genes in proximity to TEs could be observed for most of the categories belonging to 'cellular processes' or to 'information processing'. In contrast, genes annotated as 'metabolism' and 'poorly characterized' are significantly enriched (Fig. 5A). Especially, genes belonging to the sub-categories 'defense mechanisms' or 'carbohydrate metabolism' are highly enriched (Fig. 5A). A total of 39% of genes that reside in proximity to TEs fall in the category 'function unknown' (*n*=579) or 'not categorized' (*n*=482), which refers to orthologs present in other organisms but without known function. Furthermore, we could not find orthologs for 14.9% of the genes that reside in proximity to TEs (*n*=402), described in this work in the 'not recognized orthologs' category, which can probably be explained as gene content specific to the *V. dahliae* lineage (Fig. 5A; Supplementary Table 8). Interestingly, most of these gene categories are associated with TEs (Fig. 5A) and overlap with various SVs (Fig. 5B). Subsequently, we annotated all protein-coding genes for functions that have previously been associated with pathogenicity, such as genes encoding secreted proteins (*n*=672), carbohydrate-active enzymes (CAZymes, *n*=498), secondary metabolites (*n*=25), or effector candidates (*n*=193). These genes can be broadly classified as 'pathogenicity-related' and are enriched in the 'poorly characterized' functional category, especially secreted proteins and effector candidates ($p = 3.49 \times 10^{-15}$ and *n*=390, $p = 2.85 \times 10^{-6}$ and *n*=127, respectively, one-sided Fisher's exact test; Supplementary Table 9). Genome-wide, we observed that 40.72% (*n*=401) of predicted pathogenicity-related genes have a TE localized within 5 kb (Fig. 5E). Polymorphic TEs are enriched in proximity (5 kb) to effector candidates (Z-score=3.4838, $p = 0.0001$; 5,000 random permutations), even though we did not observe that these genes and other pathogenicity-related genes are significantly closer to polymorphic TEs (Fig. 5G). Thus, commonly considered pathogenicity-associated genes reside in proximity to TEs, yet only some genes are in proximity to polymorphic TEs.

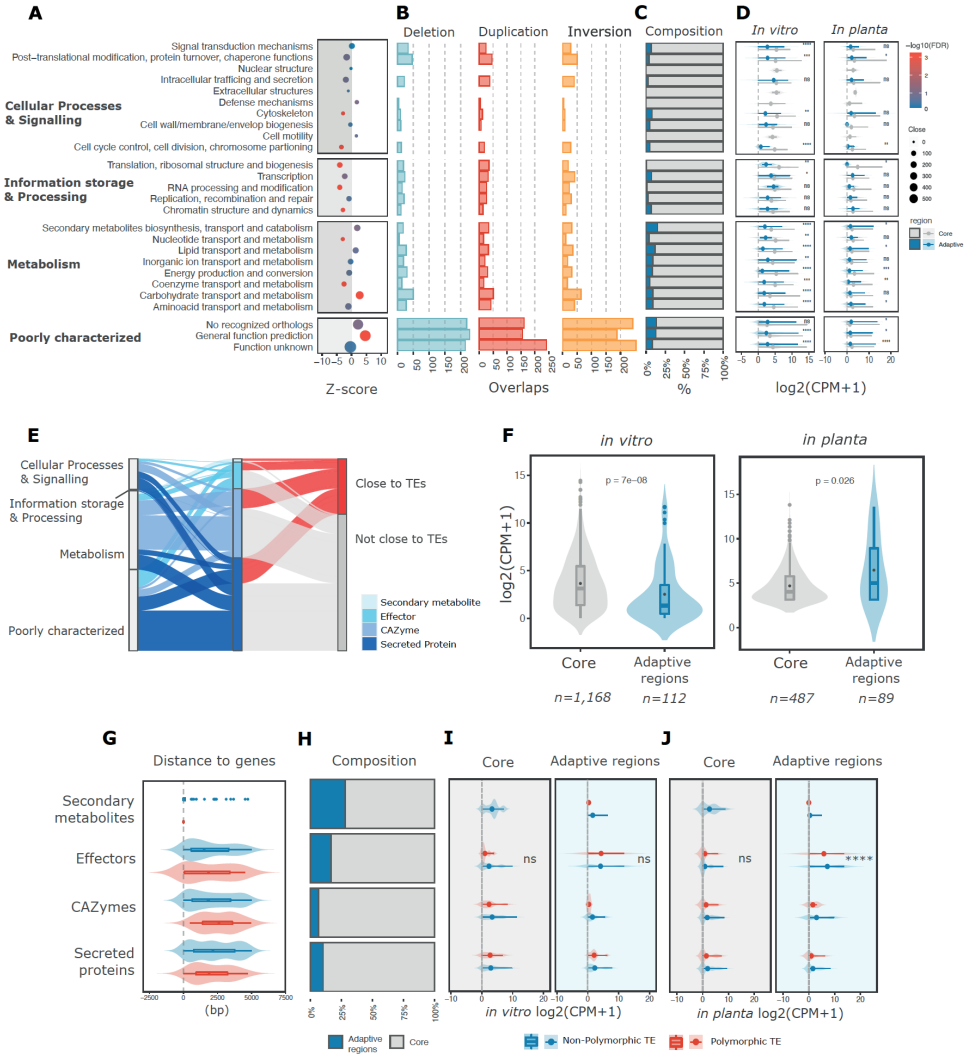


Figure 5. Polymorphic transposable elements correlate with expression of a subset of genes.
(A) Enrichment of TEs in proximity (within 5 kb range) to genes in different functional categories. Functional annotation was based on Clustering for Orthologous Groups (COGs) categories. The y-axis shows the different functional categories and the x-axis the z-score after 5,000 random permutations; the grey shadow represents values with a negative z-score indicating depletion rather than enrichment, and the -log10(*p-value*) after Benjamini-Hochberg correction for false-discovery rate is color-coded. The size is relative to the number of genes associated with TEs in each category; **(B)** The number of genes overlapping deletions (blue) and duplications (red), and inversions (yellow) ; **(C)** The relative proportions of genes localized in the adaptive genomic regions (blue) and core (grey) genomic regions; **(D)** Gene expression distribution in two conditions (*in vitro* and *in planta*; log2(CPM+1)) and separated by core and adaptive genomic regions; **(E)** Predicted pathogenicity-related genes (*n*=1,388) overlapping with COG categories and depicting the proportion of genes with TEs in proximity (5 kb); **(F)** Gene expression of pathogenicity-related genes in two conditions (*in vitro* and *in planta*; log2(CPM+1)) and separated by core and adaptive genomic regions; **(G)** Distance between genes and TEs separated by polymorphic and non-polymorphic TEs. Statistical significance was assessed using Wilcoxon rank-sum test; **(H)** The composition of pathogenicity-related gene categories separated in adaptive (blue) and core (grey) genomic regions; Pathogenicity-related gene expression *in vitro* **(I)** and *in planta* **(J)** separated by polymorphic and non- polymorphic TEs. Statistical significance was assessed by a Kruskall-Wallis rank-sum test; **** *p*<=0.0001.

When compared with the core genome, adaptive genomic regions in *V. dahliae* are relatively gene-poor (Fig. 5C) and SV-rich (Supplementary Fig. 3), as well as enriched for *in planta*-expressed genes (de Jonge et al. 2013, 2012; Kombrink et al. 2017; Chavarro-Carrero et al. 2021; Cook et al. 2020) (Fig. 5D). While a broad range of functional categories can be observed in adaptive genomic regions (Fig. 5C), these are enriched for genes encoding secreted proteins and candidate effector genes (Fig. 5H; $p = 0.032$ and $p = 0.014$, respectively; one-sided Fisher's exact test) (Cook et al. 2020). Since we have shown that gene expression levels correlate with the presence of both polymorphic as well as non-polymorphic TEs (Fig. 4), we queried to which extent polymorphic TEs in proximity of pathogenicity-related genes correlate with gene expression patterns in the adaptive genomic regions. We observed a significant enrichment of polymorphic TEs co-localizing with highly expressed genes *in vitro* and *in planta* ($p = 0.0116$ and $p = 0.0010$ for 50% and 75% upper quartiles for expression, respectively, after one-sided Fisher's exact test; Supplementary Table 10). Under *in vitro* conditions, pathogenicity-related genes located in the core genome are expressed significantly higher than genes located in adaptive genomic regions (Fig. 5F). As expected, we observed the opposite trend *in planta* (Figure 5F), where pathogenicity-related genes that are localized in adaptive genomic regions are highly expressed during host colonization (de Jonge et al. 2013, 2012). Furthermore, we observed an enrichment of polymorphic TEs located within 5 kb of pathogenicity-related genes in adaptive genomic regions ($p = 0.0062$ after one-sided Fisher's exact test). As expected, a higher proportion of these highly expressed genes are effector candidates (Fig. 5I, J). These results suggest that polymorphic TEs occur more likely in close proximity to highly expressed pathogenicity-related genes in adaptive genomic regions.

Discussion

Transposable elements (TEs) can contribute to genomic and transcriptomic variation (Bourque et al., 2018) and, consequently, TEs have been often considered to play crucial roles in genome evolution and function (Feschotte 2008). Here, we analyzed the TE landscape in the plant pathogen *V. dahliae*. We show that the presence of TEs is associated with abundant SVs that emerged independently in individual *V. dahliae* strains. The SVs form discrete clusters and co-localize within adaptive genomic regions. Importantly, we demonstrate that these SVs co-locate with TEs, many of them polymorphic, and typically evolutionary younger, less silenced, and higher expressed. Furthermore, we show that the presence of polymorphic TEs is associated with gene expression of genes that can be assigned to diverse functional categories, particularly with host-pathogen interaction genes. Collectively, our results provide evidence for the hypothesis that polymorphic TEs contribute to increased genomic diversity, correlate with the expression of pathogenicity-related genes, and play important roles in the evolution of adaptive genomic regions.

TEs occur commonly in eukaryote genomes, yet there is an enormous variation in the TE content and diversity between different organisms (Wicker et al. 2007; Huang et al. 2012; Dietrich et al. 2013). Typically, TE expansions are suppressed by various genome defense mechanisms, but some TE copies can escape suppression and remain active and mobile. Polymorphic TEs have been associated with recent transposition activity (Huang et al. 2012), and in *V. dahliae*, a subset of TEs shares these characteristics. Only a subset of all predicted TEs is young and dynamic, which is in contrast to some other fungi in which there is a considerable amount of young TEs associated with recent TE expansions, such as in *Zymoseptoria tritici* (Badet et al. 2020; Oggenfuss et al. 2020; Lorrain et al. 2020), *Leptosphaeria maculans* (Grandaubert et al. 2014), *Blumeria graminis* (Frantzeskakis et al. 2018) and *Magnaporthe oryzae* (Kang et al. 2016). In many plants and fungi, TE expansions are mainly driven by Long Terminal Repeat (LTR) retrotransposons (Vitte and Bennetzen 2006; Muszewska et al. 2011; Huang et al. 2012; Lorrain et al. 2020; Galindo-Gonzalez et al. 2017; Tsukahara et al. 2009; Donnart et al. 2017; Amselem et al. 2015). We observed an enrichment of LTRs in facultative heterochromatic regions and at the centromeres in *V. dahliae* (Cook et al. 2020; Seidl et al. 2020). Similarly, LTRs are an important fraction of the polymorphic TEs in *V. dahliae* that are localized in the adaptive genomic regions. TE expansions have occurred at least twice during the evolution of *V. dahliae*, once during *Verticillium* diversification, and then right after *V. dahliae* speciation (Faino et al. 2016). The active polymorphic TEs we identified are likely the product of the most recent TE expansion in *V. dahliae*, and likely contributed to the formation of the adaptive genomic compartments.

Polymorphic TEs are a small fraction of the total TE repertoire, which suggests that most TEs are suppressed. Three main processes have been associated with TE suppression in fungi: repeat-induced point mutation (RIP), DNA methylation, and RNA-mediated silencing. RIP is usually associated with sexual fungal organisms (Galagan and Selker 2004), yet RIP mutations have been also observed in *V. dahliae* that is considered asexual (Clutterbuck 2011; Amyotte et al. 2012; Klosterman et al. 2011; Cook et al. 2020). RIP mutations have been associated with meiosis, but alternative mechanisms during vegetative propagation could lead to RIP-like mutations (Clutterbuck 2011). This has been recently observed in *Z. tritici* and *N. crassa* in which propagation through mitotic divisions generate RIP-like mutations (Möller et al. 2020; Wang et al. 2020b). Intriguingly, polymorphic TEs are depleted in RIP mutations. Consequently, we consider different RIP scenarios, such as that RIP activity was lost during an evolutionary transition to asexuality in *V. dahliae*. In such a scenario, the mutations we observed in most TEs are remains of ancestral RIP activity. A second scenario is that *V. dahliae* is not strictly asexual, but rare sexual activity occurs in the population. In this scenario, the RIP machinery may be active during the occasional meiosis. Thirdly, a RIP-like mechanism could act outside the typical meiotic cycles and might be active in *V. dahliae*. These hypotheses are not mutually exclusive, and it remains challenging to elucidate the mechanistic origin of RIP mutations in *V. dahliae*.

Next to RIP, DNA methylation targets repetitive elements in many fungi (Nakayashiki et al. 1999; Montanini et al. 2014; Zemach et al. 2010; Bewick et al. 2019). We have previously demonstrated that the majority of TEs in *V. dahliae* is methylated (Cook et al. 2020; Seidl et al. 2020). We now show that polymorphic TEs are typically less methylated and not affected by RIP, collectively indicating that these elements are not suppressed. In the model plant *Arabidopsis* and the fruit-fly *Drosophila*, polymorphic and active TEs occur in discrete regions where novel insertions are not efficiently suppressed, while TE insertions outside of these regions are mostly suppressed by methylation (Stuart et al. 2016; Lee and Karpen 2017; Hollister and Gaut 2009; Tsukahara et al. 2009). We observed that TEs in adaptive genomic regions display low methylation levels (Cook et al. 2020). Whereas effective TE suppression outside adaptive regions might be an important mechanism to restrict TE expansions in the core genome, the activity TEs in adaptive regions suggests that that these regions are prone to TE-mediated variation.

Environmental changes can alter suppression mechanisms and drive a de-repression of TEs (Bouvet et al. 2008; Carr et al. 2010; Fouche et al. 2020). We observed TE expression upon changes in environmental conditions, i.e., different *in vitro* growth media as well as *in planta* growth. In addition, TE expression has been observed previously in different abiotic conditions in *V. dahliae* (Amyotte et al. 2012; Faino et al. 2016; Cook et al. 2020). For example, some Tc1/Mariner and Mutator elements are induced upon heat stress (Amyotte et al. 2012), while we observed that some LTR/Copia and LTR/Gypsy families highly expressed *in planta*. However, the function of TE de-repression under changing environmental conditions remains unknown (Slotkin and Martienssen 2007; Galhardo et al. 2007; Koonin and Wolf 2009). Modifications in suppression patterns induced by changes in the environment could imply a trade-off between the functional TE de-repression and increased probability of TE expansions and TE mobilization on nearby genes, or even changes in the genome structure (Seidl et al. 2016). In *Schizosaccharomyces pombe,* for instance, changes in temperature and nutrient availability induce TE insertions in promoters of stress-dependent genes, and these insertions have been associated with higher expression of the targeted genes (Maxwell 2020; Feng et al. 2012). In the plant-pathogen *Z. tritici*, changes in nutrient composition led to increased TE expression, and a TE insertion in proximity of an effector gene contribute to reduced expression and virulence (Fouche et al. 2020). Although the mechanisms in which the environment and TE de-repression influence genome function and structure remain unclear, the expression of polymorphic TEs induced by changes in the environment could be similarly an important driver of the genome dynamics in *V. dahliae.*

Changes in the TE landscape have often been associated with gene expression changes (Goubert et al. 2020). Insertion of TEs within or near genes usually reduce gene expression, for example by spreading of silencing marks, such as DNA methylation or inducing changes in chromatin conformation (Castanera et al. 2016; Winter et al. 2018). TE insertions are associated with gene silencing and mostly have negative effects, yet also examples for an adaptive advantage exist. For example, in *Z. tritici* the insertion of a TE in the promoter

of a melanin biosynthesis gene causes reduced melanin synthesis and quicker growth, which would be beneficial for colonization (Krishnan et al. 2018). We observed that genes in proximity to TEs were typically less expressed during *in vitro* and *in planta* conditions, which was particularly clear for genes flanked by TEs. However, we observed that genes located in TE-rich regions are highly expressed *in planta*, raising a conundrum to link these opposing observations. Studies in adaptive genomic regions in *V. dahliae* and analogous regions in other fungi revealed that chromatin represents an additional layer of genomic regulation that creates a chromatin state allowing for accessibility to the transcriptional machinery (Schotanus et al. 2015; Soyer et al. 2019; Fokkens et al. 2018; Cook et al. 2020). We observed an enrichment of polymorphic TEs in close proximity of these highly expressed genes in adaptive regions. This observation raises the possibility that polymorphic TEs form part of the regulation of genes in the vicinity as *cis*-regulatory sequences or as a consequence of changes in the chromatin conformation (Le Rouzic et al. 2007; Hollister and Gaut 2009; Huang et al. 2012; Szitenberg et al. 2016; Choi and Lee 2020). For instance, TEs and TE-derived sequences can be co-opted in promoter sequences and contribute to the control of gene expression (Cosby et al. 2021; Sundaram and Wysocka 2020), as for example Tf1 retrotransposon insertions in *S. pombe* occur in promoters and induce consistently higher expression of neighboring genes (Maxwell 2020; Guo and Levin 2010; Leem et al. 2008). Furthermore, this could explain the maintenance of polymorphic TEs in *V. dahliae*, as the TE suppression of these elements could negatively affect the expression levels of neighboring genes, favoring their co-localization.

Genome structure can differ significantly between species and even between strains of the same species (Lynch 2007). Here, we demonstrated that even closely related *V. dahliae* strains are characterized by their unique SV repertoire and co-localization of polymorphic TEs and SVs in adaptive regions. As previously reported, we similarly observed high levels of sequence conservation in adaptive regions (Faino et al. 2016; de Jonge et al. 2013; Shi-Kunne et al. 2018; Depotter et al. 2019). Recently, adaptive regions were characterized by an intermediate chromatin state between heterochromatin and euchromatin (Cook et al. 2020), which could indicate that this unique chromatin organization can allow for lower mutation rates in these adaptive regions (Prendergast et al. 2007). We observed that SVs emerge rapidly due to their association with polymorphic TEs, and they represent a relevant mechanism to generate variation, rather than single-nucleotide changes in adaptive regions. The co-localization of TEs and SVs in distinct regions has been reported previously associated with genome compartmentalization in diverse fungi (Plissonneau et al. 2018, 2016; Schotanus et al. 2015; Faino et al. 2016; Shi-Kunne et al. 2018; Ola et al. 2020; Torres et al. 2020). Relaxed selection has been proposed to explain the association of variation and TEs in genomic compartments for example, the TE-rich accessory chromosomes in *Z. tritici* (Croll and McDonald 2012; Grandaubert et al. 2019; Hartmann et al. 2017). Conversely, adaptive regions in *V. dahliae* do not show signs of relaxed or strong positive selection (Depotter et al. 2019). Since *V. dahliae* is an asexual fungus and consequently has a low

effective population size, genetic drift could represent an important evolutionary force that maintains SVs in population (Kelkar and Ochman 2012; Cvijovic et al. 2015), and these can therefore contribute to genome dynamics.

Compartmentalized genomes have been observed in diverse organisms (Torres et al. 2020). In filamentous plant pathogens, this genome organization is described in the 'two-speed' genome model (Raffaele et al. 2010) in which dynamic genome compartments serve as cradles of variation (Croll and McDonald 2012; Frantzeskakis et al. 2019; Torres et al. 2020). Additionally, changes in chromatin organization have been shown to be relevant for compartmentalized genomes to further increase or facilitate genomic variation (Moller et al. 2019; Schotanus et al. 2015; Cook et al. 2020). Thus, further insights into chromatin organization and TE dynamics are necessary to further disentangle the impact of chromatin modifications on the emergence of genetic variation and unravel how these facilitate the evolution of virulence in plant pathogens.

Material and Methods

Verticillium dahliae JR2 genome, repetitive element, and functional gene annotation

Repetitive element annotation of the chromosome-level genome assembly of *V. dahliae* strain JR2 (Faino et al. 2015) was based on available annotation (Faino et al. 2015; Cook et al. 2020; Seidl et al. 2020). Briefly, repetitive elements were annotated by using a combination of LTRharvest (Ellinghaus et al. 2008), LTRdigest (Steinbiss et al. 2009), and RepeatModeler. The repeat predictions were further curated and classified (Wicker et al. 2007) by a combination of PASTEC (Hoede et al. 2014) and sequence similarity to known transposable elements. The genome-wide occurrence of repeats was determined with RepeatMasker v 4.0.9, and the output was further processed using 'one code to find them all' (Bailly-Bechet et al. 2014). Simple repeats and low-complexity regions were excluded (Cook et al. 2020; Seidl et al. 2020). GC-content, Kimura distance from a consensus TE family, and weighted DNA methylation data were previously summarized (Cook et al. 2020; Seidl et al. 2020). The composite RIP index (CRI) per transposable element was calculated by obtaining the RIP substrate and the RIP product index as defined by nucleotide frequencies: RIP product index = TpA/ApT and the RIP substrate index= (CpA + TpG)/(ApC+GpT). We also calculated the genome-wide RIP index (CRI) using RIPper (van Wyk et al. 2019) using sliding windows (1 kb, 500 bp slide).

Functional gene annotation was based on the available *V. dahliae* JR2 gene annotation (Faino et al., 2015) and was performed using eggNOG 5.0 (Huerta-Cepas et al. 2019) based on an hierarchical non-supervised orthology search, restricted to a fungal database with e-values >=0.001 and query coverage >=50%. To define the *V. dahliae* secretome, we predicted N-terminal signal peptides using SignalP v.4.1 (Nielsen 2017a). Subsequently, we

predicted putative effectors with EffectorP v.2.0 (Sperschneider et al. 2018), using default parameters. Secondary metabolite biosynthetic genes were previously predicted using antiSMASH fungal version 4.0.2 (Weber et al. 2015; Shi-Kunne et al. 2018). Similarly, we used previously predicted CAZymes that cover signatures of glycoside hydrolases, polysaccharide lyases, carbohydrate esterases, glycosyltransferases, and carbohydrate-binding molecules (Shi-Kunne et al. 2018; Seidl et al. 2015). For further analysis, we considered the groups of 'secreted proteins' as the secreted proteomes excluding candidate effectors, secondary metabolite genes, and CAZymes.

Gene and transposable element RNA-sequencing analysis

Transcriptional activity of genes and repetitive elements in *V. dahliae* strain JR2 was assessed using previously generated transcriptome data (Cook et al. 2020) of three *in vitro* conditions, (Potato Dextrose Broth, Murashige-Skoog and Czapeck-Dox media) and of *in planta* colonization (*Arabidopsis thaliana*; 28 dpi). For further analysis, we considered only MS media for *in vitro* and *in planta* comparisons. Single-end sequencing reads of three biological replicates per condition were mapped to *V. dahliae* strain JR2 genome assembly (Faino et al. 2015) using STAR v.2.4.2.a, allowing multiple mapped reads with the following settings: --outFilterMultimapNmax 100 --winAnchorMultimapNmax 200 --outSAMtype BAM Unsorted --outFilterMismatchNmax 3 (Jin and Hammell 2018; Jin et al. 2015; Fouche et al. 2020). The bam files were sorted by read name with samtools v.1.2 and the transcriptional activity level was quantified using TEcount from the TEtrancripts package 2.2.1 (Jin et al. 2015), with the following parameters: --stranded no --mode multi, --iteration 1000. TEcount considers multiple-mapped reads aligned to genes and TE regions to determine transcript abundance per condition/replicate. Furthermore, TEcount only considers reads that map in its entirety to TEs and reads mapping to only a fraction were discarded (Jin and Hammell 2018; Jin et al. 2015).

Sequencing reads summarized over TEs and genes were normalized between three replicates in the different conditions using R/Bioconductor package EdgeR v.3.8 (Robinson et al. 2010; Robinson and Oshlack 2010). For normalization, we only considered genes and TEs >=1 reads in all samples (Anders et al. 2013). TEs with read count 0 were assumed to have no transcriptional activity. Libraries were normalized with TMM method (Robinson and Oshlack 2010), and converted to counts per million (CPM) mapped reads using R/Bioconductor package EdgeR v.3.8 (Robinson et al. 2010). Comparisons between transcriptional levels were computed in R 3.6.3 (Team, R 2019)

Single nucleotide variant detection and analysis

Single nucleotide variants were detected using paired-end sequencing reads of each 42 previously sequenced *V. dahliae* strains (Supplementary Table 1). Each strain was mapped independently to the reference genome *V. dahliae* JR2 using BWA -MEM v.0.7 with default options (Li and Durbin 2010). Library artifacts were marked and removed using Picard tools

v.2.18 with -MarkDuplicates followed by -SortSam to sort the reads (http://broadinstitute.github.io/picard/). Single nucleotide variants were identified using the -HaplotypeCaller of the Genome Analysis Toolkit (GATK) v.4.0 (Poplin et al. 2018). Variations were detected individually per strain, using -ploidy 1 and -emitRefConfidence GVCF. Then, all strains were joined by -Jointgenotyping with -maxAltAlleles 2, and all non-SNVs were removed using SelectVariants -selectType SNP. To obtain high-quality SNVs, we applied the following cut-off filters: QUAL<250, MQ<40, QD<20, FS>60, SOR>3.0, ReadPosRankSum<-5.0, MQRankSum between -2 and 2. Finally, we excluded variants with missing genotype calls in 10% of the strains. To explore the similarity between strains, an unrooted phylogenetic tree was constructed using a Neighbor-Joining approach based on the final single nucleotide variance set, using distance function in R 3.6.3 (Team, R 2019).

To further assess the sequence diversity between *V. dahliae* strains, we calculated the nucleotide diversity (π) based on (Nei and Jin 1989) using a sliding window of 1 kb (500 bp sliding) as implemented in the PopGenome package (Pfeifer et al. 2014) in R. Single nucleotide variants were annotated using SNPeff v.3.2 (Cingolani et al. 2012) using the refined annotation of *V. dahliae* JR2 strain. Based on the SNPeff prediction, we calculated the number of synonymous and non-synonymous mutations.

Structural variant calling and analysis

To predict structural variants (SVs), we used the 'sv-callers' workflow with few modifications that enabled parallel execution of multiple SV callers (Kuzniar et al. 2020), an approach that is considered optimal as it exploits often complementary information to predict SVs (Goerner-Potvin and Bourque 2018; Cameron et al. 2019). Briefly, the workflow takes every strain and maps the genomic reads to the reference genome *V. dahliae* JR2 using BWA-MEM v.0.7 with default options (Li and Durbin 2010). Then, library artifacts were marked and removed using Picard tools v.2.18 with -MarkDuplicates followed by -SortSam to sort the reads (http://broadinstitute.github.io/picard/). We used four different callers: DELLY v.0.8.1 (Rausch et al. 2012), LUMPY v.2.13 (Layer et al. 2014), Manta v.1.6.0 (Chen et al. 2016), and Wham v.1.0 (Kronenberg et al. 2015) in single-sample mode, using each 42 previously sequenced *V. dahliae* strains independently (Supplementary Table 1). These callers incorporate a diverse range of information for SV detection such as split reads (DELLY, LUMPY, Wham), discordant read pairs (DELLY, LUMPY, Wham), read depth signal (LUMPY, Wham), short-read assembly (Manta, Wham), and soft-clipping detection (Wham). The four SV callers were used with default parameters, except for DELLY in which we used >1 as minimum quality for further processing. All outputs were first filtered using bcftools -filter v.1.3.2 with default settings used in the 'sv-callers' workflow (Li 2011), except for Manta since the score model for post-processing assumes a diploid genome (Chen et al. 2016). Therefore, we used the unscored Manta predictions for further processing.

For final filtering, the 'sv-callers' workflow post-processes the vcfs. Briefly, the results of the four different callers per strain were merged using SURVIVOR v.1.0.6 (Jeffares

et al. 2017), keeping SVs predicted by at least three callers, allowing 1000 bp as the maximum distance from breakpoints predicted by the different tools and considering only same SV types. Subsequently, only SVs with minimum size >50 bp and maximum size 1 Mb as well as localization outside of a low-quality region, defined as MQ=0 and read support <10 for every strain, were kept. Finally, the 42 independent datasets we combined into a single vcf file using SURVIVOR by merging SVs up to 1000 bp apart of the same SV type.

Transposable element polymorphism prediction and analysis

Transposable element presence/absence polymorphisms were analyzed using TEPID v.2.0 (Stuart et al. 2016), using paired-end short reads of 42 previously sequenced *V. dahliae* strains (Supplementary Table 1). TEPID integrates split and discordant read mapping, mapping quality, sequencing breakpoints, and local variations in coverage to identify variants with respect to a reference TE annotation (Stuart et al. 2016). We mapped each of the 42 *V. dahliae* strains individually to *V. dahliae* JR2 reference genome, using tepid-map (average insert size -s 500), which uses Bowtie2 v.2.2.5 (Langmead and Salzberg 2012). TE variant discovery was computed using tepid-discover, considering insertion and deletion prediction and a conservative discovery through –strict option. We subsequently refined the discovered variants using tepid-refine to reduce false-negative calls within the group of 42 strains.

Statistical analysis and visualization

The distribution of SVs over the *V. dahliae* strain JR2 genome was determined considering the breakpoints (±1 bp) overlapping in 10 kb non-overlapping windows. We predicted an expected probability distribution assuming a Poisson distribution, using λ=0.77 (mean number of SVs breakpoints per 10 kb windows). To investigate if SVs co-localize, we performed a clustering analysis using CROC (Pignatelli et al. 2009) based on a hypergeometric distribution test and posterior multiple-testing correction using Benjamini-Hochberg; CROC scans every chromosome with a sliding window (SV-breakpoints per window=10, slide=1, >3 SV-breakpoints as a minimum). We computed the same test for TE clustering using TEs per window=10, slide=1, >3 polymorphic TEs as a minimum. Permutation tests were computed using R/Bioconductor regionR v.1.18.1 package (Gel et al. 2016) and performed 10,000 iterations, using the mean distance to evaluate the closest relationship (bp distance) between SV-breakpoints and TE elements, and circular randomization to maintain the order and distance of the regions in the chromosomes. We assumed the same parameters to evaluate the association of TEs and adaptive regions but considering the number of overlaps instead of distance. Finally, we computed enrichment association tests using bedtools -fisher v.2.25.0 based on a one-sided Fisher's test (Quinlan and Hall 2010) and random association permutation tests using R/Bioconductor region v.1.18.1 (Gel et al. 2016), performing 5,000 iterations with a resampling randomization.

Principle component analyses were performed in R v.3.6.3, using packages FactoMineR v.1.42 and factoextra v.1.0.5 (Lê et al. 2008). The used variables for each TE were methylation (mCG, mCHG, mCHH), GC content, Kimura distance, CRI, Frequency, and *in vitro* (MS media) expression. To further investigate the relationship of TEs with gene expression, we summarized TEs in 5kb windows upstream and downstream genes, using bedtools -window (Quinlan and Hall 2010). We classified genes in 'upstream' or 'downstream' if >=1 TE is located in the position in relation to the gene. We considered genes to be located in 'between' TEs if these had >=1 TEs in both positions (upstream and downstream), and we excluded them from the other categories. We ranked each gene near TE based on their expression profile and counted the number of TEs associated with each independent rank. These rank counts were then binned to obtain the distribution of ranks and fit a linear model for the data and calculated the R^2 and *p* value for the fit of the model in R v.3.6.3 (Team, R 2019). All statistical analyses and comparison tests were performed in R v.3.6.3 (Team, R 2019), and visualization with R packages ggplot2 and circlize (Wickham 2016; Gu et al. 2014).

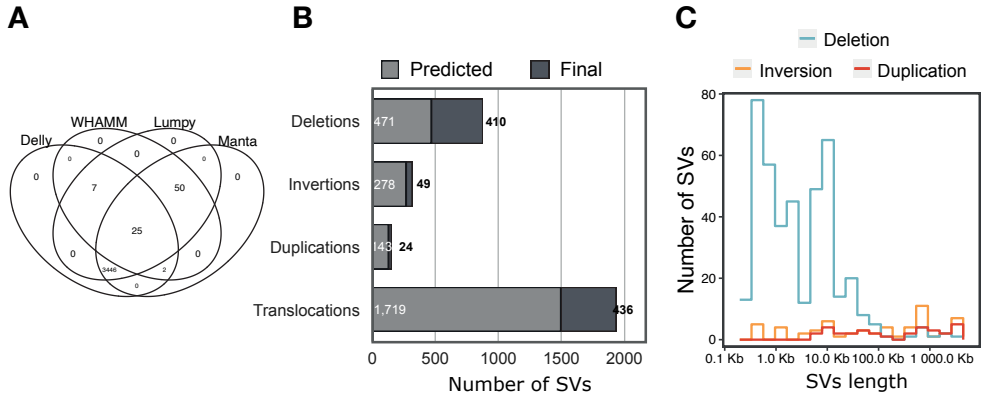
Acknowledgments

This work was supported by the Consejo Nacional de Ciencia y Tecnología, México and the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) under Germany´s Excellence Strategy – EXC 2048/1 – Project ID: 390686111.

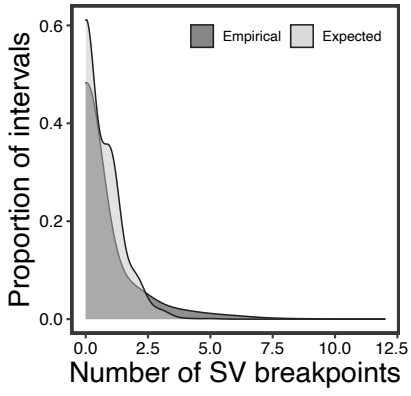
Data availability statement

The data underlying this article are accessible at the National Center for Biotechnology Information (NCBI) Sequence Read Archive (SRA) under BioProject PRJNA592220 (Supplementary Table 1).

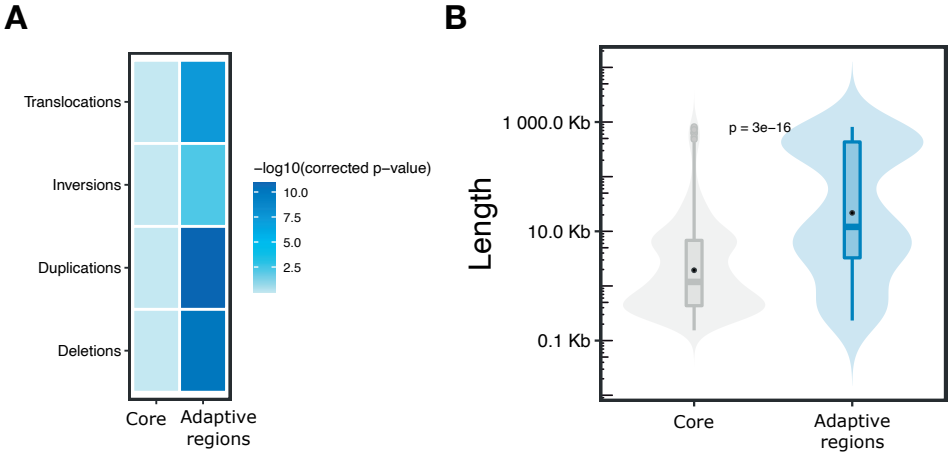
Supplementary data



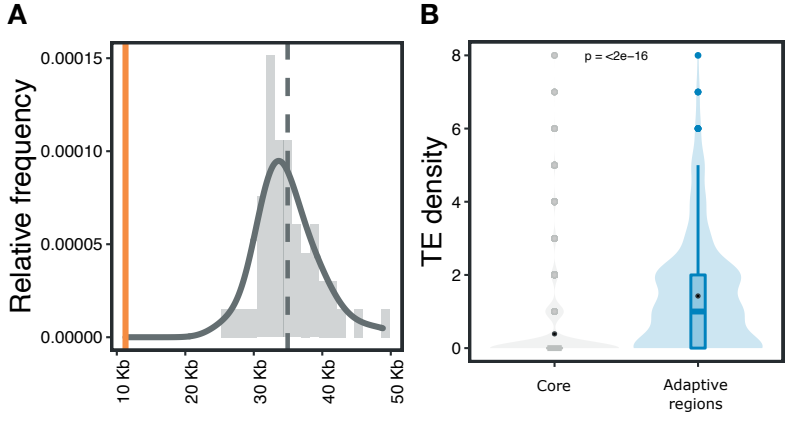
Supplementary Figure 1. Structural variant (SV) calling methods used in this study.
(A) The total number of SV predictions from the four different SV callers and their overlap is shown as a Venn diagram; **(B)** Overview of the 3,530 SVs identified in the 42 *V. dahliae* isolates after combining the results with SURVIVOR. For the 'final set', we only considered SVs predicted by at least three different callers (see Methods) and removed SVs due to mapping in low-quality regions (MQ=0) and to size (>1 Mb) (see Methods). The 'final set' was considered for all further analyses; **(C)** Length distribution of the final set of 505 SVs, here excluding translocations, used for all further analysis.



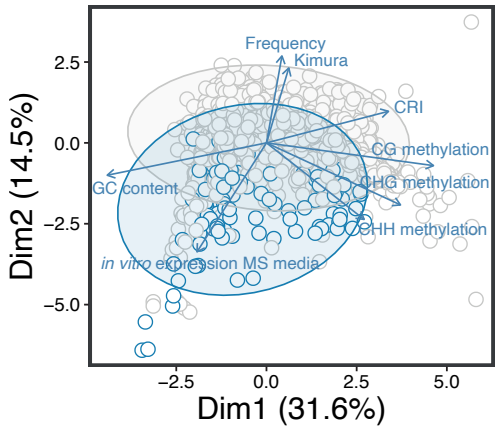
Supplementary Figure 2. Distribution of structural variant breakpoints in the *Verticillium dahliae* genome.
Density plot overlapping the empirical and expected probability distribution in 10 kb non-overlapping windows across the genome of *V. dahliae* strain JR2. The empirical distribution considers the breakpoints of all 941 SVs. The expected probability distribution was estimated assuming a Poisson distribution. Statistical difference between the two distributions was assessed using a Kolmogorov-Smirnov test (p<2.2e-16).



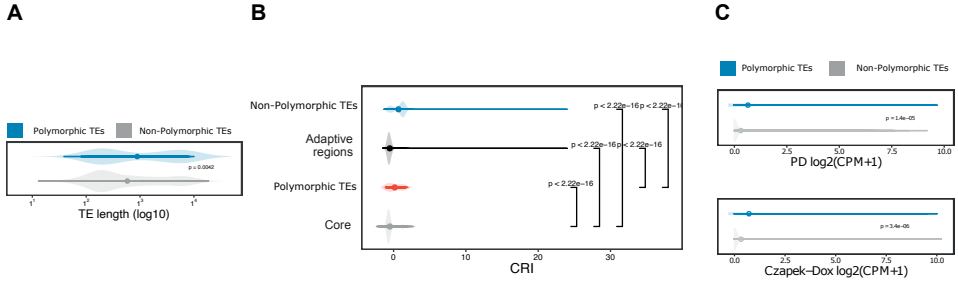
Supplementary Figure 3. Adaptive genomic regions in *Verticillium dahliae* are enriched in structural variants (SVs). **(A)** Enrichment of different classes of SVs in core and adaptive genomic regions. The heatmap displays the posterior false-discovery rate corrected p-value derived from a Fisher´s exact test; **(B)** Length (log10) distributions of SVs overlapping the core and adaptive genomic regions. The statistical significance was assessed using a one-sided Wilcoxon rank-sum test.



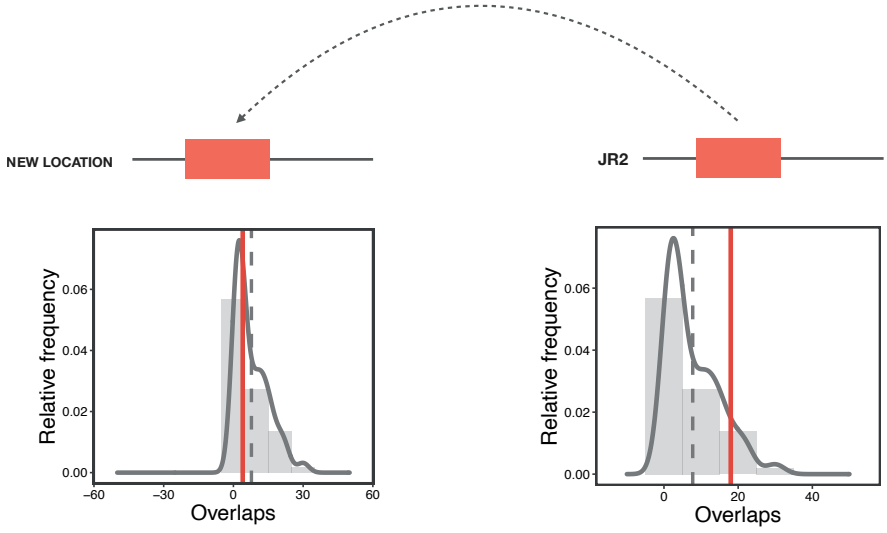
Supplementary Figure 4. Transposable elements are close to SVs.
(A) Permutation test to assess the distance between the full set of TEs and SV breakpoints. The vertical dashed line represents the mean distance of overlaps expected at 10,000 random permutations, the yellow line indicates the mean distance observed at a closer distance (11,284 bp), significance p = 9.9e-5, z-score=-5.43. TEs showed to be more closed to SV breakpoints than expected by chance; **(B)** The distribution of TE density (TE counts) summarized in 10 kb non-overlapping windows is shown. TEs at centromeric regions were removed. The statistical significance was assessed using a Wilcoxon rank sum test.



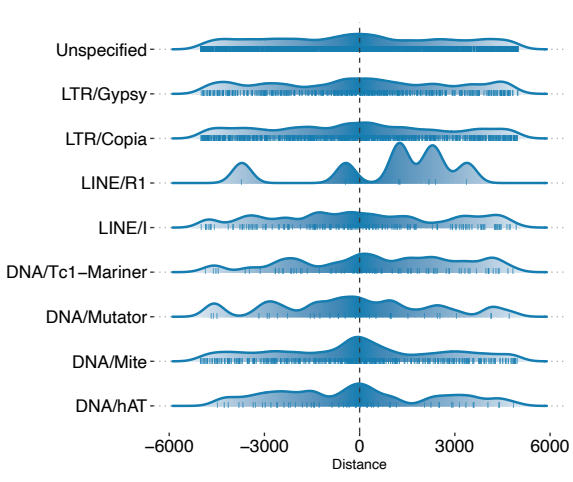
Supplementary Figure 5. Principle component analysis of TEs in *V. dahliae*.
Decomposition of the principal component analysis for eight variables summarized for each TE, excluding centromeric regions. Each vector represents one variable, with the length indicating the importance of the variable for this dimension, for angles <90º, the two variables are correlated, while >90º indicates that the variables are negatively correlated. The coordinates and contribution for each variable are further detailed in Supplementary Table 4.



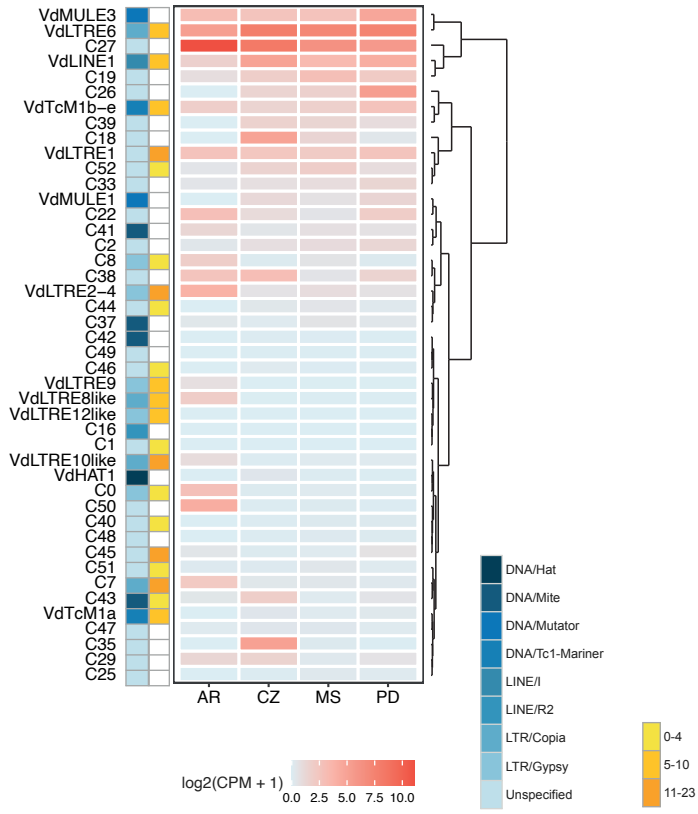
Supplementary Figure 6. Polymorphic transposable elements in *V. dahliae* strain JR2.
(A) Length distribution of polymorphic TEs and non-polymorphic TEs; the statistical differences were assessed by one-sided Wilcoxon rank-sum tests; **(B)** Composite RIP (repeat-induced point mutations) index comparison (window size 1 kb, 500 bp slide) in two genomic compartments separated by polymorphic and non-polymorphic TEs; statistical differences were assessed by one-sided Wilcoxon rank-sum test; **(C)** Transcript levels of *V. dahliae* strain JR2 TEs in two *in vitro* conditions. TE expression is displayed for two *in vitro* conditions, PD (Potato Dextrose broth) and CZ (Czapek-Dox) growth media. Expression values are log2(CPM+1) transformed. *p-value* after one-sided Wilcoxon rank-sum test.



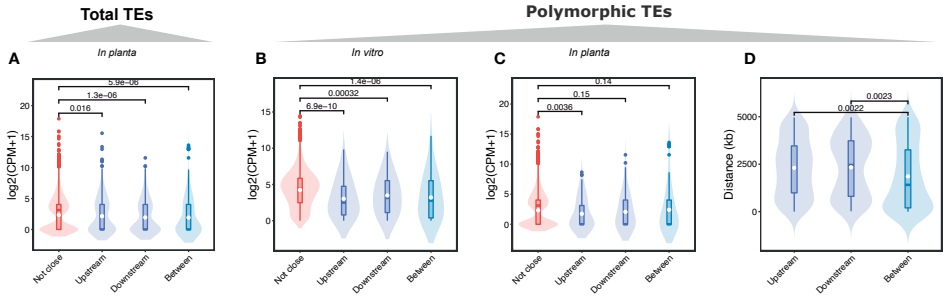
Supplementary Figure 7. Transposable element insertions are not enriched in adaptive genomic regions in *V. dahliae* strain JR2.
(A) Left, permutation test of overlaps between new TE insertions and adaptive genomic regions. The vertical dashed line represents the mean number of overlaps expected at 10,000 random permutations and the red line indicates the number of overlaps observed (7), significance p = 0.451, z-score=-0.534; **(B)** Right, permutation test of overlaps between original coordinates in JR2 of TE insertions and adaptive genomic regions. The vertical dashed line highlights the mean number of overlaps expected at 10,000 random permutations and the red line indicates the number of overlaps observed (18), significance p = 0.294, z-score=0.7862.



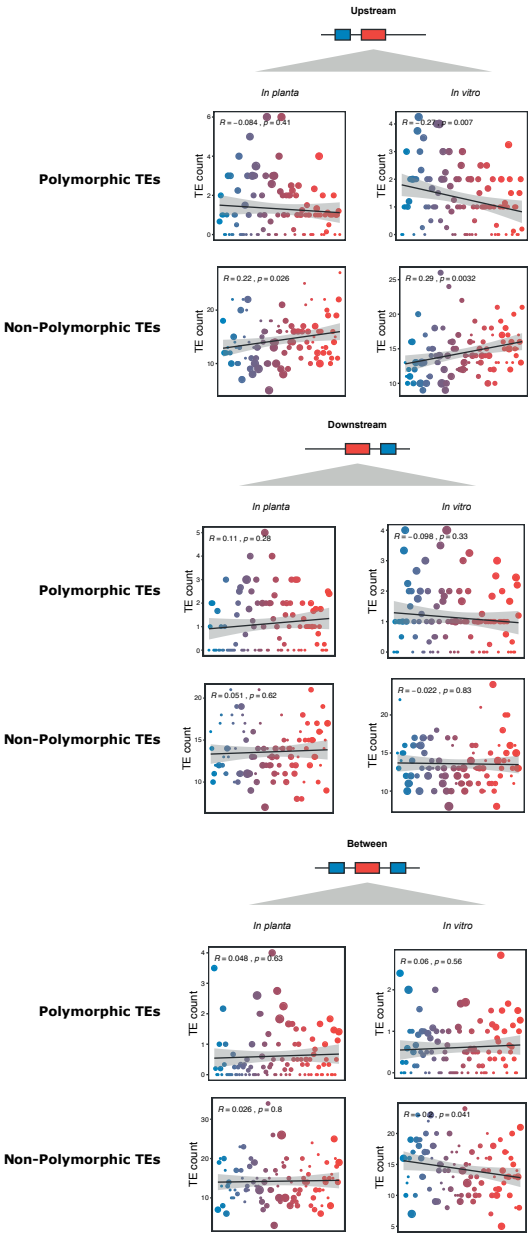
Supplementary Figure 8. Transposable element distribution relative to protein-coding genes in *V. dahliae*.
Density distribution of non-polymorphic TEs depicted by superfamilies relative to the location of protein-coding genes (distance=0), showing a random distribution within 5 kb windows upstream and downstream gene sequences.



Supplementary Figure 9. Transposable element expression in different conditions.
Unsupervised clustering of TE expression monitored under different *in vitro* conditions (Czapek-Dox, CZ; potato dextrose broth, PD; Murashige-Skoog, MS) and *in planta* condition (*Arabidopsis thaliana* 28 dpi; AR). The mean transcript abundance (counts per million; log2(CPM+1) transformed) is depicted in the heatmap. The blue-coded left column details the superfamily, and the yellow-coded column summarizes the abundance of polymorphic TEs in each of the TE family.



Supplementary Figure 10. Gene expression correlates with transposable elements occurrence
(A) *In planta* (*Arabidopsis thaliana* 28 dpi) gene expression of genes located upstream (*n*=1,098), downstream (*n*=1,051), between (*n*=546) TEs (5 kb proximity), or genes without TEs in proximity = Not close (5 kb; *n*=8,742) are shown; **(B)** *In vitro* and (**C**) *in planta* gene expression of genes upstream (*n*=153), downstream (*n*=134), and between (*n*=118) dynamic TEs; **(D)** Comparison of distances between polymorphic TEs to genes (within 5 kb windows); in all cases, *p-value* depicts one-sided Wilcoxon rank-sum test.



Supplementary Figure 11. Polymorphic transposable elements correlate with gene expression depending on their position on the genome.

Relationship between gene expression ranks (percentile windows, from low (blue) to high (red) expression; *in vitro* and *in planta*) and TE density, depending on relative TE positions (within 5 kb). The linear regression (dark line) and confidence interval (light grey) are shown as well as the R and *p-value* after linear regression. The upper panels depict the polymorphic TEs and the lower panels the non-polymorphic TEs.

Supplementary Table 1. ***Verticillium dahliae*** **strains used for this work**

Strain	Reference	Sequencing platform	bioProject	Origin	Location	Aligned reads*	Mean read length	Genome coverage**	Depth coverage
12008	Fan et al., 2018	Illumina HiSeq200	PRJNA344737	Strawberry	United Kingdom	9,471,730	299.47	98.7%	78.5
12158	Fan et al., 2018	Illumina HiSeq200	PRJNA352681	Strawberry	United Kingdom	13,504,404	207.82	88.8%	77.7
12161	Fan et al., 2018	Illumina HiSeq200	PRJNA352681	Strawberry	United Kingdom	9,986,230	218.03	89.0%	60.2
12251	Fan et al., 2018	Illumina HiSeq200	PRJNA352681	Strawberry	United Kingdom	13,052,198	172.52	94.5%	62.3
12253	Fan et al., 2018	Illumina HiSeq200	PRJNA352681	Strawberry	United Kingdom	7,201,832	221.43	93.3%	44.1
2009-605	Li et al., 2018	Illumina HiSeq200	PRJNA576642	Bell pepper	Ukraine	11,420,000	100.00	96.2%	31.6
463	Li et al., 2018	Illumina HiSeq200	PRJNA576642	Cotton	Mexico	11,050,000	100.00	91.6%	30.6
89	Chavarro-Carrero et al., 2021	Illumina HiSeq200	PRJNA639910	Sunflower	China	27,462,910	150.00	92.4%	114.0
BP2	Zhang et al., 2012	Illumina HiSeq200	PRJNA693498	Cotton	China	17,897,332	101.00	97.7%	50.0
CBS38166	De Jonge et al., 2012	Illumina HiSeq200	PRJNA169154	Tomato	Canada	11,023,740	100.00	93.7%	30.5
CFA3	Chavarro-Carrero et al., 2021	Illumina HiSeq200	PRJNA639910	Sunflower	China	27,261,526	150.00	92.5%	113.1
DVD-S26	De Jonge et al., 2012	Illumina HiSeq200	PRJNA169154	Soil	Canada	12,000,000	100.00	95.1%	33.2
DVD-S29	De Jonge et al., 2012	Illumina HiSeq200	PRJNA169154	Soil	Canada	11,000,000	100.00	92.2%	30.4
DVD-S94	De Jonge et al., 2012	Illumina HiSeq200	PRJNA169154	Soil	Canada	11,602,874	100.00	94.7%	32.1
DVD-161	De Jonge et al., 2012	Illumina HiSeq200	PRJNA169154	Tomato	Canada	11,000,000	100.00	93.5%	30.4
DVD-3	De Jonge et al., 2012	Illumina HiSeq200	PRJNA169154	Potato	Canada	11,000,000	100.00	94.9%	30.4
DVD-31	De Jonge et al., 2012	Illumina HiSeq200	PRJNA169154	Tomato	Canada	11,302,980	100.00	93.2%	31.3
HN	Xu et al., 2012	Illumina HiSeq200	PRJNA693498	Cotton	China	23,411,782	101.00	95.0%	65.4
HNA4	Chavarro-Carrero et al., 2021	Illumina HiSeq200	PRJNA639910	Sunflower	China	27,445,394	150.00	92.3%	113.9
JKG8	Kombrink et al., 2017	Illumina HiSeq200	PRJNA576642	Potato	The Netherlands	11,440,000	100.00	92.8%	31.7
JR2	Faino et al., 2015	Illumina HiSeq200	PRJNA169154	Tomato	Canada	11,700,862	100.00	99.6%	32.4
ST100	De Jonge et al., 2012	Illumina HiSeq200	PRJNA169154	Soil	Belgium	12,272,662	100.00	93.5%	34.0

Strain	Reference	Sequencing platform	bioProject	Origin	Location	Aligned reads*	Mean read length	Genome coverage**	Depth coverage
ST14.01	De Jonge et al., 2012	Illumina HiSeq200	PRJNA169154	Pistachio	USA	12,940,760	100.00	94.4%	35.8
ST16.01	Li et al., 2018	Illumina HiSeq200	PRJNA576642	Cotton	Syria	11,510,000	101.00	97.1%	32.2
T9	Keykhasaber, 2017	Illumina HiSeq200	PRJNA693498	Cotton	USA	24,582,176	101.00	90.7%	68.7
TM6	Keykhasaber, 2017	Illumina HiSeq200	PRJNA693498	Cotton	China	20,372,150	101.00	91.3%	56.9
V08SY-2	Chavarro-Carrero et al., 2021	Illumina HiSeq200	PRJNA639910	Sunflower	China	47,503,846	150.00	92.9%	197.2
V117	Keykhasaber, 2017	Illumina HiSeq200	PRJNA693498	Olive	Spain	6,155,058	100.00	92.1%	17.0
V574	Milgroom et al., 2014	Illumina HiSeq200	PRJNA693498	Artichoke	Spain	8,121,868	150.00	92.3%	33.7
V700	Milgroom et al., 2014	Illumina HiSeq200	PRJNA693498	Artichoke	Spain	8,082,488	150.00	92.2%	33.5
V991	Zhang et al., 2012	Illumina HiSeq200	PRJNA302216	Cotton	China	22,086,912	101.00	91.4%	61.7
VanDijk	Kombrink et al., 2017	Illumina HiSeq200	PRJNA576642	Chrysanthemum	The Netherlands	9,078,396	100.00	93.6%	25.1
Vd39	Li et al., 2018	Illumina HiSeq200	PRJNA576642	Sunflower	Germany	11,040,000	100.00	94.7%	30.5
Vd57	Li et al., 2018	Illumina HiSeq200	PRJNA693498	Strawberry	Germany	11,080,000	100.00	98.9%	30.7
VdLs17	Faino et al., 2015	Illumina HiSeq200	PRJNA276625	Lettuce	USA	11,420,000	100.00	96.8%	31.6
WFA3	Chavarro-Carrero et al., 2021	Illumina HiSeq200	PRJNA639910	Sunflower	China	27,219,230	150.00	92.3%	113.0
XJ-12	Chavarro-Carrero et al., 2021	Illumina HiSeq200	PRJNA639910	Sunflower	China	45,673,316	150.00	92.6%	189.6
XJ-14	Chavarro-Carrero et al., 2021	Illumina HiSeq200	PRJNA639910	Sunflower	China	40,753,298	150.00	92.8%	169.1
cd3	Xu et al., 2012	Illumina HiSeq200	PRJNA693498	Cotton	China	16,739,900	101.00	93.4%	46.8
v152	Kombrink et al., 2017	Illumina HiSeq200	PRJNA576642	Oak	Hungary	11,090,000	100.00	95.8%	30.7
v4	Keykhasaber, 2017	Illumina HiSeq200	PRJNA693498	Olive	Spain	10,878,440	100.00	94.6%	30.1
v52	Kombrink et al., 2017	Illumina HiSeq200	PRJNA576642	Pepper	Autria	11,060,000	100.00	95.7%	30.6

* Total number or reads after removing duplicates

** Percentage of total covered bases to reference strain JR2

Supplementary Table 2. Clustering test for SVs breakpoints in *Verticillium dahliae* JR2

Chromosome	Start	End	Cluster size (bp)	*p-value*
Chr1	60001	330000	269,999	0.0456345
Chr1	4250001	4470000	219,999	0.00705146
Chr1	7090001	7180000	89,999	0.0456345
Chr3	210001	260000	49,999	0.0456345
Chr3	**2370001**	**2430000**	**59,999**	**0.0456345**
Chr3	2750001	2860000	109,999	0.0456345
Chr4	840001	890000	49,999	0.0456345
Chr4	1170001	1260000	89,999	0.00705146
Chr4	**2990001**	**3360000**	**369,999**	**0.00705146**
Chr5	1300001	1570000	269,999	0.0456345
Chr5	**1890001**	**2030000**	**139,999**	**0.00061621**
Chr5	3460001	3630000	169,999	0.00061621
Chr6	**2210001**	**2330000**	**119,999**	**0.00705146**
Chr6	3170001	3230000	59,999	0.0456345
Chr7	**1980001**	**2050000**	**69,999**	**0.00705146**
Chr8	1060001	1140000	79,999	0.00705146
Chr8	**1430001**	**1560000**	**129,999**	**0.00705146**

Significance is indicated based on a hypergeometric distribution test, multiple-testing was addressed by Benjamini-Hochberg correction; clusters overlapping centromeres are highlighted in bold.

Supplementary Table 3. Comparison between SNVs and SVs in core or adaptive regions in 42 *Verticillium dahliae* strains

Variable	Core	Adaptive	Significance
SNVs	14.2	12.1	$p<2\times10^{-16}$
π	0.0025	0.00015	$p<2\times10^{-16}$
Non-synonymous	0.0176	0.0127	$p<1.3\times10^{-5}$
Synonymous	0.0264	0.0144	$p<2\times10^{-16}$
Size (bp)	32,813,998	3,328,633	
Total of nucleotide changes (SNVs bp)*	263,975	23,276	
% genome***	0.73%	0.06%	
Total of nucleotide changes (SVs bp)**	13,527,461	5,235,157	
Deletions	3,654,762	1,977,686	
Duplications	3,999,565	570,326	
Inversions	5,872,396	2,687,011	
Translocations	738	134	
Polyrmorphic TEs	318,041	111,289	
% genome***	37.43%	14.48%	

Median values are reported (1 kb windows) and statistical significance was assessed with a one-sided Wilcoxon rank sum test
*Sum of SNVs per genomic compartment
**Sum of SV length overlaps (deletions, duplications, inversions, translocations) per genomic compartment, do not include polymorphic TEs; considering >= 50% total length overlap
*** Percentage with respect to the total number of bases per genomic compartment

Supplementary Table 4. Repetitive elements in *Verticillium dahliae* JR2

Class	Superfamily	Families	Elements	Mean Size
DNA-TIR	hAT	1	13	394
DNA-TIR	Mite	4	116	290
DNA-TIR	Mutator	2	9	940
DNA-TIR	Tc1-Mariner	2	53	1567
LINE	I	1	57	1888
LINE	R1	1	3	859
LTR	Copia	5	545	2348
LTR	Gypsy	5	500	4101
Unspecified	Unspecified	22	817	559

Total number of annotated repetitive elements used for the TE polymorphism analysis

Supplementary Table 5. Principal Component Analysis of transposable elements

	Coordinates of variables as correlations					Contribution of variables				
Variable	**Dim.1**	**Dim.2**	**Dim.3**	**Dim.4**	**Dim.5**	**Dim.1**	**Dim.2**	**Dim.3**	**Dim.4**	**Dim.5**
Methylation CG	0.84	-0.13	0.07	0.00	0.15	27.56	1.38	0.42	0.00	2.97
Methylation CHG	0.67	-0.35	0.18	-0.04	-0.09	17.75	10.31	3.16	0.18	1.00
Methylation CHH	0.49	-0.43	0.41	-0.08	-0.39	9.43	15.71	15.63	0.70	19.88
GC content	-0.80	-0.18	0.19	0.00	-0.24	25.06	2.80	3.39	0.00	7.62
Kimura distance	0.11	0.42	0.73	0.33	0.37	0.49	15.03	49.44	11.52	17.51
CRI*	0.61	0.18	-0.49	-0.05	0.17	14.71	2.67	22.28	0.26	3.77
Frequency	0.08	0.48	0.23	-0.83	-0.09	0.24	20.26	5.06	72.92	1.00
in vitro expression MS media	-0.35	-0.61	0.08	-0.37	0.59	4.77	31.83	0.62	14.42	46.25

Coordinates for each variable as proxy for correlations and the contributions of each variable (%) for the first five dimensions

*Composite Repeat Index

Supplementary Table 6. Clustering of polymorphic transposable elements in *Verticillium dahliae* JR2

Chromosome	Start	End	Cluster size (bp)	*p-value*
Chr1	53070	75768	22698	0.00640506
Chr1	2902101	2919752	17651	0.0472553
Chr1	3728458	3751806	23348	0.0297937
Chr1	7726623	7731761	5138	0.0156587
Chr2	3306675	3312299	5624	0.0156587
Chr3	1730692	1738069	7377	0.0112741
Chr4	581484	605820	24336	0.0297937
Chr4	877261	878231	970	0.0156587
Chr4	3383031	3389191	6160	0.0472553
Chr5	22170	55414	33244	2.13E-06
Chr5	560562	561866	1304	0.0112741
Chr5	1932698	1972808	40110	0.00432344
Chr5	2136723	2145914	9191	0.000860712
Chr5	2360598	2390674	30076	0.0112741
Chr5	3620143	3629611	9468	0.0472553
Chr6	344418	359784	15366	0.00548821
Chr6	2796059	2817652	21593	0.000402971
Chr6	3179711	3185903	6192	0.00548821
Chr6	3212527	3229593	17066	0.000402971
Chr8	27028	32235	5207	0.0156587
Chr8	470207	478533	8326	0.00548821
Chr8	985415	1006393	20978	0.0297937
Chr8	2753583	2775625	22042	0.0297937

Clustered regions of polymorphic TEs after hypergeometric distribution test, *p-values* after Benjamini-Hochberg correction

Supplementary Table 7. Distribution of transposable elements over different genomic compartments

	Total amount of TEs							Polymorphic TEs							
TE superfamily	**Centromere**		**Core**		**Adaptive**		**Total**	**Centromere**		**Core**		**Adaptive**		**Total**	**Total (%)**
DNA/hAT	0	0.00%	7	0.55%	6	1.23%	13	0	0.00%	0	0.00%	0	0.00%	0	0.00%
DNA/Mite	0	0.00%	110	8.63%	6	1.23%	116	0	0.00%	1	0.97%	0	0.00%	1	0.86%
DNA/Mutator	0	0.00%	1	0.08%	8	1.64%	9	0	0.00%	0	0.00%	0	0.00%	0	0.00%
DNA/Tc1-Mariner	4	1.11%	28	2.20%	21	4.31%	53	1	5.88%	12	11.65%	4	8.89%	17	32.08%
LINE/I	0	0.00%	13	1.02%	44	9.03%	57	0	0.00%	5	4.85%	2	4.44%	7	12.28%
LINE/R1	0	0.00%	1	0.08%	2	0.41%	3	0	0.00%	0	0.00%	0	0.00%	0	0.00%
LTR/Copia	90	25.00%	360	28.26%	99	20.33%	549	3	17.65%	40	38.83%	23	51.11%	66	12.02%
LTR/Gypsy	261	72.50%	197	15.46%	43	8.83%	501	13	76.47%	24	23.30%	4	8.89%	41	8.18%
Unspecified	5	1.39%	557	43.72%	258	52.98%	820	0	0.00%	21	20.39%	12	26.67%	33	4.02%
Total	**360**	**100.00%**	**1274**	**100.00%**	**487**	**100.00%**	**2121**	**17**	**100.00%**	**103**	**100.00%**	**45**	**100.00%**	**165**	

Left column depicts the total number of TEs per genomic compartment and right column the percentage of them relative to the total counts per compartment

Supplementary Table 8. Functional gene annotation and association with TEs

COG	Genes close TE	Genes not close to TE	Genes total	Z-score	*p-value*
Cell cycle control, cell division, chromosome portioning	25	140	165	-3.453	0.00057766
Cell wall/membrane/envelop biogenesis	20	67	87	-0.399	0.41699993
Cell motility	2	1	3	1.624	0.16665556
Post-translational modification, protein turnover, chaperone functions	107	436	543	-1.673	0.13343998
Signal transduction mechanisms	71	218	289	0.102	0.60079184
Intracellular trafficing and secretion	76	319	395	-1.921	0.13343998
Defense mechanisms	18	29	47	1.790	0.07718456
Extracellular structures	0	3	3	-1.099	0.39173035
Nuclear structure	1	3	4	-0.188	0.65406919
Cytoskeleton	11	80	91	-2.877	0.00057766
RNA processing and modification	40	253	293	-3.993	0.00057766
Chromatin structure and dynamics	21	115	136	-2.909	0.00057766
Translation, ribosomal structure and biogenesis	57	308	365	-3.970	0.00057766
Transcription	62	256	318	-2.293	0.07718456
Replication, recombination and repair	50	176	226	-0.861	0.24981004
Energy production and conversion	72	253	325	-0.799	0.21042107
Aminoacid transport and metabolism	101	365	466	-1.079	0.16665556
Nucleotide transport and metabolism	14	83	97	-2.962	0.00057766
Carbohydrate transport and metabolism	200	482	682	2.770	0.00057766
Coenzyme transport and metabolism	40	183	223	-2.542	0.00057766
Lipid transport and metabolism	85	219	304	1.426	0.16665556
Inorganic ion transport and metabolism	61	200	261	-0.315	0.38100951
Secondary metabolites biosynthesis, transport and catabolism	99	254	353	1.924	0.07718456
Function unknown	579	1902	2481	-0.392	0.39173035
General function prediction	482	1245	1727	4.657	0.00057766
No recognized orthologs	402	1152	1554	2.201	0.07718456

Genes in orthologs groups close to TEs (within 5 kb window) or not close (>5 kb). *p-value* after 5,000 random permutations and Benjamini-Hochberg correction; Z-score with respect to permutation test

Supplementary Table 9. Pathogenicity-related genes relative to COG categories

COG category	Pathogenicity-related	Genes	Total %	Category %
Cellular Processes & Signaling	CAZ	74	0.64%	4.54%
Cellular Processes & Signaling	Effector	17	0.15%	1.04%
Cellular Processes & Signaling	No pathogenicity	1407	12.26%	86.32%
Cellular Processes & Signaling	Secreted	132	1.15%	8.10%
Information Storage & Processing	CAZ	2	0.02%	0.15%
Information Storage & Processing	Effector	3	0.03%	0.22%
Information Storage & Processing	No pathogenicity	1327	11.57%	99.18%
Information Storage & Processing	Secreted	6	0.05%	0.45%
Metabolism	CAZ	354	3.09%	12.91%
Metabolism	Effector	46	0.40%	1.68%
Metabolism	No pathogenicity	2176	18.96%	79.33%
Metabolism	Secondary	23	0.20%	0.84%
Metabolism	Secreted	144	1.26%	5.25%
Poorly characterized	CAZ	68	0.59%	1.18%
Poorly characterized	Effector	127	1.11%	2.20%
Poorly characterized	No pathogenicity	5176	45.11%	89.81%
Poorly characterized	Secondary	2	0.02%	0.03%
Poorly characterized	Secreted	390	3.40%	6.77%

Number of pathogenicity genes per COG category used for enrichment analysis

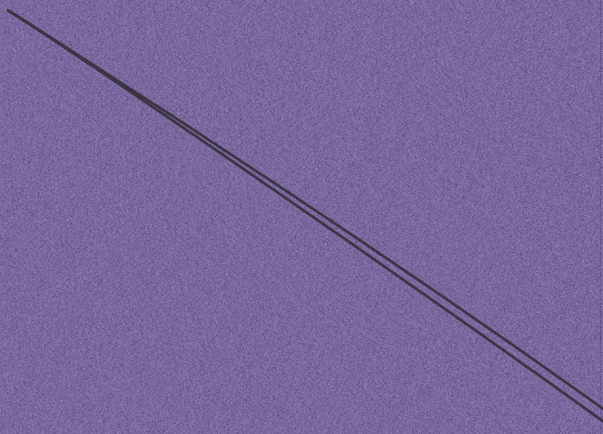
Supplementary Table 10. Polymorphic TEs associated with highly expressed genes

Pathogenesis related	TE type	Rank	Region	Number Genes	Number TEs
Secreted	Polymorphic TE	Q1	Core	4	4
Secreted	Polymorphic TE	Q1	Adaptive	1	1
Secreted	Polymorphic TE	Q2	Core	4	4
Secreted	Polymorphic TE	Q2	Adaptive	6	14
Secreted	Polymorphic TE	Q3	Core	6	8
Secreted	Polymorphic TE	Q3	Adaptive	2	2
Secreted	Polymorphic TE	Q4	Core	4	6
Secreted	Polymorphic TE	Q4	Adaptive	1	1
Secreted	TE	Q1	Core	40	0
Secreted	TE	Q1	Adaptive	12	0
Secreted	TE	Q2	Core	53	0
Secreted	TE	Q2	Adaptive	30	0
Secreted	TE	Q3	Core	57	6
Secreted	TE	Q3	Adaptive	7	2
Secreted	TE	Q4	Core	52	4
Secreted	TE	Q4	Adaptive	13	2
CAZ	Polymorphic TE	Q1	Core	6	6
CAZ	Polymorphic TE	Q1	Adaptive	2	4
CAZ	Polymorphic TE	Q2	Core	6	8
CAZ	Polymorphic TE	Q2	Adaptive	2	2
CAZ	Polymorphic TE	Q3	Core	8	16
CAZ	Polymorphic TE	Q4	Core	3	3
CAZ	TE	Q1	Core	39	10
CAZ	TE	Q1	Adaptive	14	0
CAZ	TE	Q2	Core	43	4
CAZ	TE	Q2	Adaptive	13	2
CAZ	TE	Q3	Core	60	3
CAZ	TE	Q3	Adaptive	5	0
CAZ	TE	Q4	Core	74	8
CAZ	TE	Q4	Adaptive	5	0
Effector	Polymorphic TE	Q1	Core	5	5
Effector	Polymorphic TE	Q1	Adaptive	4	6
Effector	Polymorphic TE	Q2	Core	5	9
Effector	Polymorphic TE	Q2	Adaptive	3	3
Effector	Polymorphic TE	Q3	Core	2	2
Effector	Polymorphic TE	Q4	Adaptive	4	4

Effector	TE	Q1	Core	32	12
Effector	TE	Q1	Adaptive	16	5
Effector	TE	Q2	Core	12	0
Effector	TE	Q2	Adaptive	10	4
Effector	TE	Q3	Core	20	0
Effector	TE	Q3	Adaptive	3	0
Effector	TE	Q4	Core	18	0
Effector	TE	Q4	Adaptive	15	15
Secondary	Polymorphic TE	Q2	Adaptive	2	2
Secondary	TE	Q1	Core	8	0
Secondary	TE	Q1	Adaptive	1	0
Secondary	TE	Q2	Core	4	0
Secondary	TE	Q2	Adaptive	15	15
Secondary	TE	Q3	Core	17	0
Secondary	TE	Q3	Adaptive	9	0
Secondary	TE	Q4	Core	5	0
Secondary	TE	Q4	Adaptive	2	0
			Total	**784**	**202**
			Core	**587**	**118**
			Adaptive	**197**	**84**

Quartile ranks based on expression data in vitro, Q1=25%, Q2=50%, Q3=75%, Q4=100%

```
#
##
 theme(axis.text = element_text(-
size=12,color='black'),
        panel.border = element_rect(colour
= "black", fill=NA, si ze=1),panel.grid =
element_blank())
dev.off()
###
#
####
#
##
#
#
#
```



CHAPTER 4

H3K27me3-linked chromatin states are associated with gene expression variation between *Verticillium dahliae* isolates

David E. Torres[1,2], Gabriel L. Fiorin[1], Anton Kraege[3], Edgar A. Chavarro-Carrero[1,3], Bart P.H.J. Thomma[3], and Michael F. Seidl[2]

[1]Wageningen University and Research, NL
[2]Utrecht University, NL
[3]University of Cologne, DE

Abstract

Gene expression is crucial to determine the phenotype of an organism. Variation in gene expression at the same genomic locus enables species to express different phenotypic traits facilitating adaptation, which is particularly relevant for pathogens that engage in host-pathogen 'arms races'. However, the relationship between genomic variation and gene expression differences in plant pathogens remains unclear. Here, we studied gene expression variation in *Verticillium dahliae*, a fungal plant pathogen known to have adaptive genomic regions (AGRs) that are enriched in genome rearrangements, genes with dynamic expression between conditions, accessible DNA, and the presence of the facultative heterochromatic H3 Lys-27 methylated histones (H3K27me3). Using two *V. dahliae* strains that belong to divergent genetic lineages, we show that gene expression differs for one-third of the genes, and these differences are likely not caused by changes in the organization and sequence of *cis*-regulatory regions. We summarize different combinations of seven histone modifications into distinct chromatin states and a subset of those states can be linked to gene expression changes between strains at AGRs. Intriguingly, H3K27me3 occurs in chromatin states together with histone modifications typically associated with active gene expression such as H3K9ac, H3K4me3, or H3K36me3. The H3K27me3-associated chromatin states correlate with reduced sequence conservation, differential gene expression *in planta*, and differential gene expression between *V. dahliae* strains. Particularly, genes with H3K27me3-associated chromatin states are enriched for pathogenicity-related gene functions. Thus, our analyses show that H3K27me3-associated chromatin states correlate with the evolution of gene expression and genome organization in *V. dahliae*.

Introduction

Gene expression determines the phenotype of an organism (Lewontin 2008; Nachtomy et al. 2007). Gene expression levels are typically conserved between organisms throughout evolution and are associated with nucleotide conservation (Martin and Fraser 2018; Bedford and Hartl 2009; Josephs et al. 2015). Gene expression is tightly regulated and allows organisms to respond dynamically to different stimuli (Hill et al. 2021), and consequently the regulation of gene expression is critical to determine the phenotype (Price et al. 2022; Ba et al. 2022). Despite gene expression conservation, a species can diverge multiple phenotypic traits by varying the expression of the same underlying loci (Mank 2017; Parsch and Ellegren 2013). Gene expression can vary through changes in the DNA, due to silent mutations in the coding sequence, through genomic rearrangements, or through mutations in regulatory sequences (Wittkopp and Kalay 2011; Zhou et al. 2016). Regulatory sequences upstream or as part of genes can modulate gene expression timing and levels (Wray et al. 2003; Wittkopp and Kalay 2011). Regions directly upstream of the coding sequences that typically contain transcription factor (TF) binding sites and regulate gene expression in *cis* are called *Cis*-regulatory REgions (CREs) (Wittkopp and Kalay 2011; Rodríguez-Trelles et al. 2003; Stern and Frankel 2013).

The DNA in the nucleus is wrapped around an octamer formed by two copies of the histone proteins H2A, H2B, H3, and H4 to form nucleosomes (Luger et al. 1997). Post-translational chemical modifications of histone proteins are associated with expression regulation (Dai et al. 2020; Villaseñor and Baubec 2021), and to date more than 300 histone modifications have been described (Zhao and Garcia 2015; Bannister and Kouzarides 2011; Zentner and Henikoff 2013). Additionally, DNA methylation and non-coding RNAs have been associated with expression regulation (Mattick and Makunin 2006; Yao et al. 2019; Zemach et al. 2010; Zemach and Zilberman 2010). Post-translational histone modifications and DNA methylation establish a chromatin landscape that can be broadly divided into transcriptionally active euchromatin and transcriptionally repressive heterochromatin. Thus, the chromatin landscape collectively plays an important role in gene expression regulation.

Euchromatin is characterized by high chromatin accessibility, depletion of DNA methylation, and enrichment in specific histone modifications such as tri-methylation of the lysine 4 or acetylation of the lysine 9 on H3 (H3K4me3 and H3K9ac, respectively) (Zentner and Henikoff 2013; Huang and Zhu 2018). Euchromatin is typically associated with strong negative selection, and therefore genomic variation is depleted (Makova and Hardison 2015; Quiroz et al. 2022; Monroe et al. 2022b; Hazarika et al. 2022). In contrast, heterochromatic regions are typically poorly accessible, enriched for DNA methylation, and for specific histone modifications such as tri-methylation of the lysine 9 or 27 on H3 (H3K9me3 and H3K27me3, respectively) (Zentner and Henikoff 2013; Bannister and Kouzarides 2011; Grewal and Jia 2007; Sasaki et al. 2014; Schuster-Böckler and Lehner 2012; Sun et al. 2016). Furthermore, heterochromatic regions are associated with relaxed selection, and as such

they are prone to accumulate transposable elements (TEs), structural variations (SVs), and nucleotide variations (Lewis et al. 2009; Makova and Hardison 2015; Janssen et al. 2018; de la Peña et al. 2023; Habig et al. 2021; Yasuhara et al. 2005; Liu et al. 2020a). Therefore, together with the gene sequences and CREs, the chromatin landscape has been implicated in expression variation within and between species and plays important roles in adaptation.

Plants and fungal pathogens are thought to engage in co-evolutionary 'arms races' (Koch et al. 2014; Upson et al. 2018; Frantzeskakis et al. 2019), where plants have evolved immune systems to detect pathogen intrusions and mount defense responses, while fungal pathogens evolved a complex molecular tool kit to support host colonization (Rovenich et al. 2014; Cook et al. 2015). Such co-evolutionary 'arms races' are characterized by repeated cycles of adaptation and counter-adaptation of pathogens and their hosts (Strotz et al. 2018; McDonald and Stukenbrock 2016). To enable these cycles of adaptation and counter-adaptation, genetic variation in pathogens is needed, which in many cases is restricted to specific genomic compartments (Dong et al. 2015; Möller and Stukenbrock 2017; Upson et al. 2018; Torres et al. 2020; Seidl and Thomma 2017). DNA methylation and histone modifications play a significant role in the formation and maintenance of these genomic compartments (Erlendson et al. 2017; Seidl et al. 2016), such as the accessory chromosomes of *Zymoseptoria tritici* (Schotanus et al. 2015; Habig et al. 2021; Feurtey et al. 2020), or the adaptive genomic regions (AGRs) in *Verticillium dahliae* (Cook et al. 2020; Torres et al. 2021; Kramer et al. 2022). Similarly, the expression of *in planta*-induced and other conditionally responsive genes that are thought to contribute to environmental adaptation has been associated with changes in histone modifications in many plant pathogens (Sánchez-Vallet et al. 2018; Seidl et al. 2016; Torres et al. 2020). Moreover, for *Phytophthora sojae* it was demonstrated that gene expression variation, mediated by changes in the chromatin landscape, can contribute to avoiding recognition by the plant immune system (Wang et al. 2020a; Pais et al. 2018). It is therefore conceivable that gene expression variation, like genetic variation, enables co-evolutionary arms-races between the host and the pathogen.

Verticillium dahliae is an asexual soil-borne fungal plant pathogen that infects hundreds of plant species (Fradin and Thomma 2006; Klosterman et al. 2011, 2009; Inderbitzin and Subbarao 2014). In *V. dahliae*, AGRs are embedded in the eight chromosomes that are typically observed in this species and are enriched in large-scale chromosomal rearrangements such as translocations, segmental duplications, and gene losses, as well as in active TEs (de Jonge et al. 2012, 2013; Faino et al. 2016; Torres et al. 2021; Cook et al. 2020). Moreover, AGRs show an unique chromatin profile that is characterized by enrichment of H3K27me3, depletion of DNA methylation, and the occurrence of accessible chromatin and genes with dynamic transcription between different conditions (Torres et al. 2021; Cook et al. 2020; Kramer et al. 2022, 2021). Previous observations revealed that *V. dahliae* strains can be divided into at least three genetically distinct lineages (de Jonge et al. 2013; Chavarro-Carrero et al. 2021; Torres et al. 2021). However, it presently remains unclear whether genomic variation correlates with gene expression differences between

strains of *V. dahliae*. Here, we use *in vitro* and *in planta* transcriptomic data to analyze gene expression variation between two *V. dahliae* strains. Furthermore, we use chromatin immunoprecipitation followed by sequencing (ChIP-seq) of seven activating or repressing histone modifications to understand the association of the chromatin landscape with gene expression variation in *V. dahliae*.

Results

Differential gene expression correlates with variable sequence conservation in *Verticillium dahliae*

Based on the genomic variation in *V. dahliae* (de Jonge et al. 2013; Faino et al. 2016; Torres et al. 2021), we hypothesized that gene expression levels vary between strains. We here focused on *V. dahliae* strains JR2 and CQ2, as these strains occur in two divergent lineages (Fig. 1A, Supplementary Table 1), and high-quality genome assemblies are available (Faino et al. 2015; Depotter et al. 2019). In line with previous observations (Depotter et al. 2019), whole-genome alignments show that 89.10% of the genome is syntenic between JR2 and CQ2 (aligned >=10 kb blocks; Fig. 1B; Supplementary Fig. 1A-C). Nevertheless, 97% (*n*=11,067) of genes occur in both JR2 and CQ2, while only 362 JR2-specific and 347 CQ2-specific genes are identified that mostly locate in AGRs (Fig. 1B; z-score=2.27, *p*=0.00019 JR2; z-score=6.69, *p*=0.00019 CQ2).

To assess whether genes that occur in both *V. dahliae* strains are differentially expressed between these strains, we generated RNA sequencing data for conidiospores harvested from *V. dahliae* grown on potato dextrose agar (PDA). We observe that homologous genes are similarly expressed in conidiospores of the two strains (Fig. 1C; $p<2.2\times10^{-16}$, one-sided Wilcoxon-rank sum test). Homologous genes vary in their sequence conservation across the *Verticillium* genus (Fig. 1D). While we observe a positive correlation between gene expression levels of homologs and their sequence conservation across the *Verticillium* genus (Fig. 1D; Supplementary Fig. 1F,G), reduced levels of sequence conservation positively correlate with increased differences in gene expression levels between both *V. dahliae* strains (R=0.87, $p=3.1\times10^{-7}$; Fig. 2A; Supplementary Fig. 2).

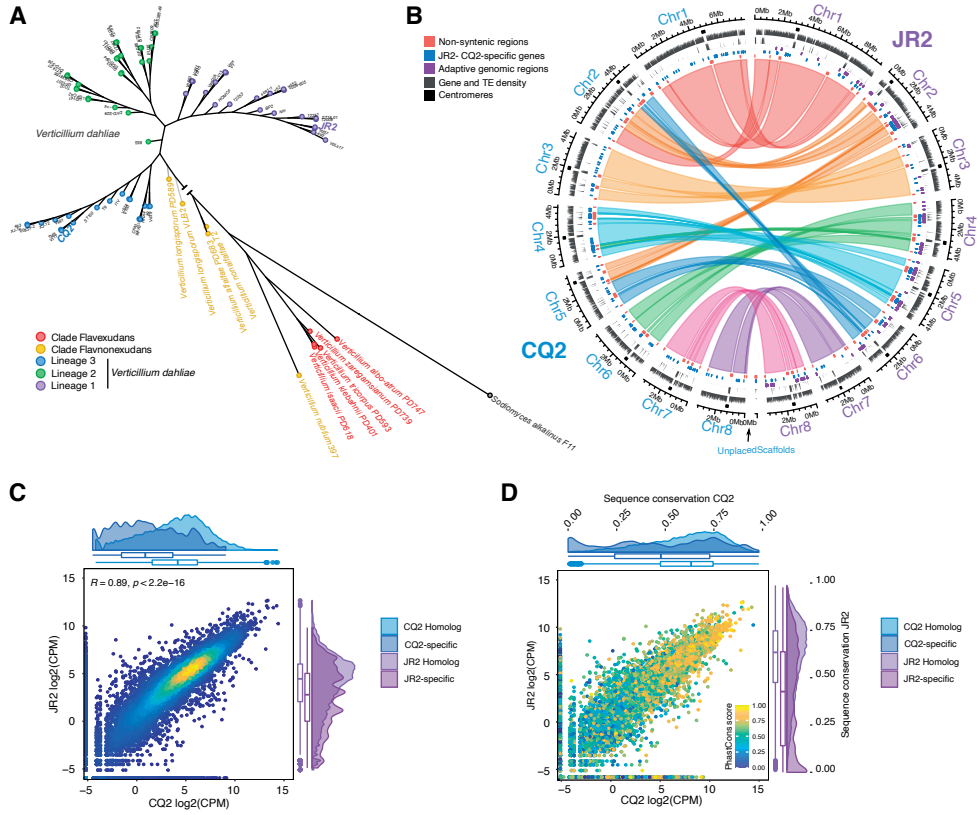


Figure 1. Gene expression correlates with sequence conservation in two divergent *Verticillium dahliae* strains. **(A)** Unrooted phylogenetic tree, based on genome comparisons and 58,364 single nucleotide variants (SNVs), shows the known relationships between the ten species of the *Verticillium* genus, and display three *Verticillium dahliae* lineages (blue, green, and purple). **(B)** Genome comparison between *V. dahliae* strains CQ2 and JR2 uncovers genetic variation; from outer to inner track: centromere locations, gene density (10 kb windows), repeat density (10 kb windows), adaptive genomic regions (AGRs; (Cook *et al.* 2020)), JR2- or CQ2-specific genes, and non-syntenic regions. The ribbons depict syntenic regions (10 kb windows). **(C)** *In vitro* gene expression levels in conidiospores harvested from *V. dahliae* strains JR2 and CQ2 grown on potato dextrose agar (PDA) are highly correlated. Pearson's correlation (R) and associated *p*-values are shown. Density plots along the axes show gene expression levels (log2(counts per million; CPM)) for homologous gene pairs (JR2 and CQ2) and JR2- or CQ2-specific genes. **(D)** Gene expression levels in conidiospores of PDA-grown *V. dahliae* correlate with sequence conservation. Color gradient as well as density plots at the axes depict the PhastCons scores for homologous gene pairs and for JR2- or CQ2-specific genes.

Next, we classified genes into those with conserved or differential expression patterns between the two *V. dahliae* strains based on the spread of gene expression values by using the coefficient of interquartile variation (CQV) between both strains as a quantitative measure for gene expression variation (Bonett 2006). Gene expression in conidiospores of JR2 and CQ2 is conserved for two thirds of the genes (n=7,581), while one third of the genes is differentially expressed (n=2,101) (Fig. 2B; Supplementary Fig. 2B-D). Genes with conserved expression have 1.42 times higher sequence conservation than differentially expressed genes (Supplementary Fig. 2D) and are enriched in GO terms associated with cell cycle regulation (Supplementary Fig. 2E, p<0.05). Conversely, differentially expressed genes are enriched in GO terms associated with conditionally responsive functions, such as metabolism (carbohydrates, lipids, or carboxylic acid metabolism), development (filamentous growth, cell development, or regulation), and cellular sensing (symbiotic interaction, response to external stimulus, cation transport, or oxo-reduction processes) (Fig. 2C, p<0.05). Differentially expressed genes between conditions are enriched in AGRs (Kramer et al. 2022), and similarly, differentially expressed genes between strains mostly localize in AGRs (Permutation test: z-score=5.35, p=0.00019).

Evolution of *cis*-regulatory regions is not associated with expression variation

Changes in CRE sequences can lead to gene expression variation (Wray 2007; Stone and Wray 2001). We here define CREs to be between ten and 450 bp upstream of coding regions, and we consider the coding sequence to comprise the sequence between the transcription start and end site (Supplementary Fig. 3A,B). One third (36%) of CREs is located between two coding regions in opposite direction within a single strain (n=3,983 and n=3,917 for CQ2 and JR2, respectively; Fig. 3A), and these bi-directional CREs are associated with 40% of genes for which the gene expression is conserved between strains (Fig. 3A,B). In contrast, most of the differentially expressed genes are associated with uni-directional CREs (Fig. 3A), yet differentially expressed genes with high expression levels also have bi-directional CREs (Fig. 3B). Similarly, 35% of the highly expressed JR2- and 24% of the highly expressed CQ2-specific genes have bi-directional CREs (n=122 and n=82, respectively; Fig. 3A,B).

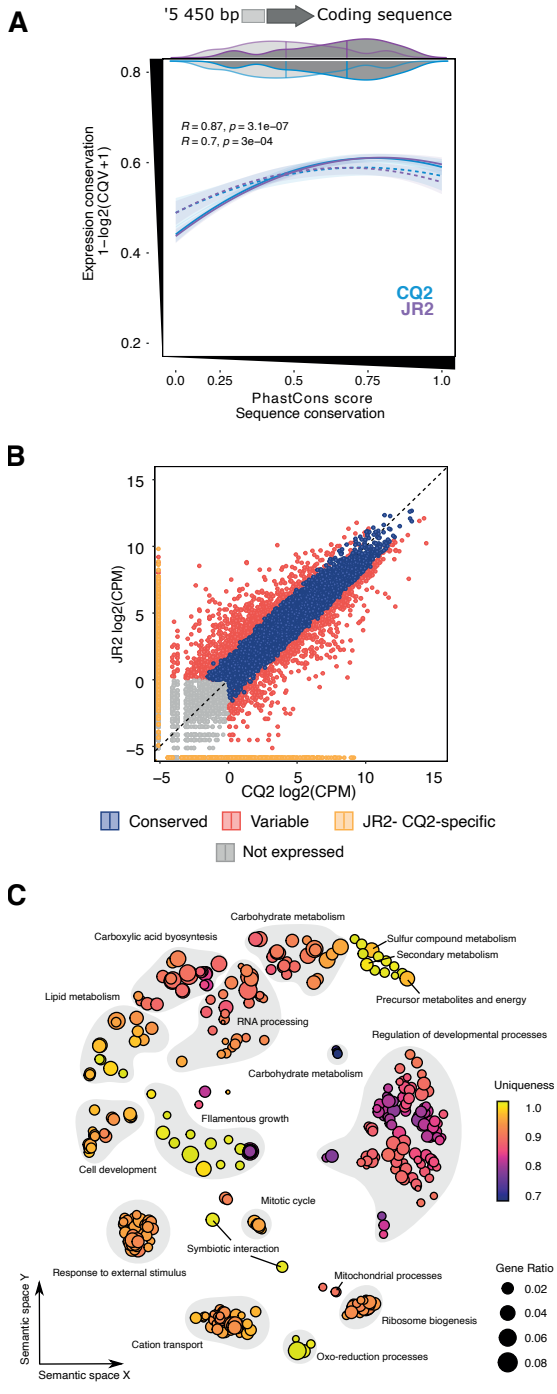


Figure 2. One third of *Verticillium dahliae* genes is differentially expressed between strains and are associated with variation in sequence conservation.

(A) Gene expression variation correlates with sequence conservation. Solid lines depict coding sequences, while dashed lines depict 5' regions up to 450 bp upstream of the coding sequences. We defined 'coding sequence' as the region between the transcription start and end sites, including introns. Bimodal density plots (x-axis) depict the PhastCons distribution in coding sequences (right) and the '5-upstream regions (left) for *V. dahliae* strains CQ2 (blue) and JR2 (purple). Spearman's correlations (R) and associated *p*-values are shown. **(B)** Genes can be classified based on the gene expression variation between two *V. dahliae* strains. Coefficient of interquartile variation (CQV) is used as a proxy for gene expression variation, and genes in the 4th upper quartile are classified as differentially expressed genes. Non-expressed genes (gray) are defined as genes with counts per million (CPM)<1. **(C)** Semantic similarity of GO terms enriched in the set of genes differentially expressed between the two *V. dahliae* strains. The circle size depicts the ratio of genes with a given GO term when compared with the total set of annotated genes. The color scale depicts the uniqueness score value; lower values (purple) indicate common GO terms in the full gene data set, while high values (yellow) show rare GO terms in the full gene dataset.

Changes in CRE organization by rearranging bi- or uni-directional CREs might lead to gene expression changes. However, genome comparisons between *V. dahliae* strain JR2 and CQ2 show that both bi-directional and uni-directional CREs are largely conserved (Fig. 3C). Genes with rearranged CREs can show expression changes between strains (Fig. 3D), yet we also observe that homologs with rearranged CREs show gene expression conservation. Moreover, genes with syntenic CREs do not always show similar expression (Fig. 3D), and thus, while CRE reorganization can lead to differences in gene expression between strains, most expression changes occur despite conserved CRE organization.

We observe reduced sequence conservation in CREs when compared with coding regions (Fig. 2E; Supplementary Fig. 3C). Thus, we tested if nucleotide changes in coding sequence or CREs might lead to differential gene expression between *V. dahliae* strains. We performed hierarchical clustering of genes, based on the sequence conservation of coding sequences and CREs together with gene expression conservation, and recover six clusters that capture different combinations of these features (Fig. 3F, Supplementary Fig. 3D). We observe that 85.5% (*n*=6,205) of the genes with conserved expression also has high levels of sequence conservation of their coding sequences and CREs (Fig. 3F). Interestingly, reduced sequence conservation of coding sequences and CREs is not observed at all differentially expressed genes, as for example 30.8% of the differentially expressed genes between *V. dahliae* strains (*n*=617) and most JR2- and CQ2-specific genes (*n*=190 and *n*=197, JR2 and CQ2 respectively) also display increased sequence conservation (Fig. 3F). Conversely, 14.5% (*n*=1,050) of genes with conserved expression also display reduced sequence conservation, and thus we conclude that sequence conservation of CREs and changes in their organization cannot fully account for the observed variation in gene expression between *V. dahliae* strains.

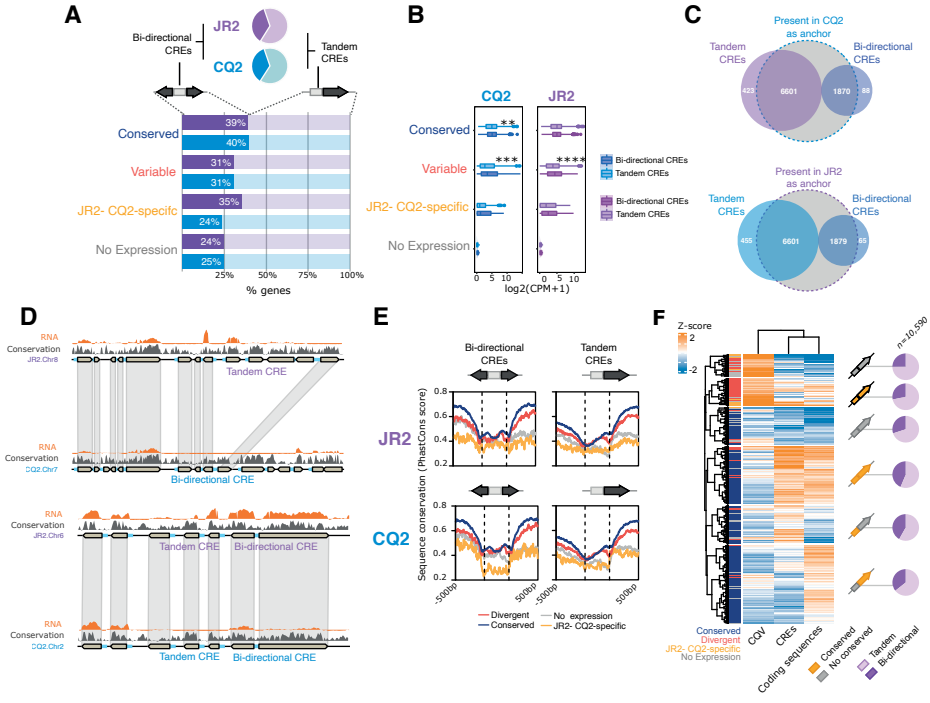


Figure 3. Changes in gene expression are not associated with changes of *cis*-regulatory regions.
(A) Bi-directional or uni-directional *Cis*-regulatory REgions (CREs) are associated with expression conservation in *V. dahliae* strains JR2 (purple) and CQ2 (blue). Pie plots depict the total number of genes with a bi-directional (dark colored) or uni-directional (light colored) CREs. Bar plots depict the percentage of bi-directional (dark colored) or uni-directional (light colored) CREs, separated based on the expression conservation of the associated gene (see Fig. 2). **(B)** Conidial gene expression is higher in genes associated with bi-directional CREs; *p-value*, one-sided Wilcoxon rank sum test. **(C)** The organization of most genes with bi-directional or uni-directional CREs is conserved between *V. dahliae* strains JR2 and CQ2. Overlapping areas indicate the number of JR2 CREs that can be identified in the CQ2 genome (upper Venn diagram) or CQ2 CREs in the JR2 genome (bottom Venn diagram). Non-overlapping areas depict genes with non-syntenic CREs. **(D)** A syntenic locus in *V. dahliae* strains JR2 and CQ2. Changes in CRE organization can lead to changes in gene expression (top). Conserved CRE organization can also show changes in gene expression (bottom). Ribbons link homologous genes between the two *V. dahliae* strains. From top to bottom: RNA *in vitro* (in conidiospores harvested from *V. dahliae* strains JR2 and CQ2 grown on PDA), sequence conservation (PhastCons score), gene models, CREs (light blue). **(E)** Sequence conservation (PhastCons score) differs between bi-directional or uni-directional CREs based on the expression category of the associated genes. **(F)** Genes are clustered by their expression conservation (CQV) as well as sequence conservation of the coding region and CRE. Diagram representation of the six clusters on the right. Pie plots depict the proportion of uni-directional or bi-directional CREs in *V. dahliae* strain JR2 per cluster (dark and light purple).

Histone modifications correlate with gene expression variation

Next to the sequence and organization of CREs, chemical modifications of the DNA and histones are associated with transcriptional regulation (Smolle and Workman 2013; Lawrence et al. 2016). To determine if specific histone modifications are associated with gene expression in *V. dahliae*, we performed *in vitro* chromatin immunoprecipitation followed by sequencing (ChIP-seq) for *V. dahliae* strains CQ2 and JR2 in two biological replicates that were significantly correlated (Supplementary Fig. 4). We targeted six modifications on histone H3

(H3K4me3, H3K4ac, H3K9ac, H3K9me3, H3K36me3, and H3K27me3) and one modification on histone H4 (H4K16ac) (Fig. 4A). H3K4me3, H3K4ac, H3K9ac, H3K36me3, and H4K16ac are associated with gene-rich regions (Fig. 4A; Supplementary Fig. 5,6) and, conversely, regions enriched in H3K9me3 are depleted in genes and associated with repetitive elements such as the regional centromeres of *V. dahliae* (Fig. 4A; Supplementary Fig. 5,6) (Seidl et al. 2020). As we have demonstrated previously (Cook et al. 2020), the facultative heterochromatic mark H3K27me3 is enriched in AGRs (Fig. 4A; z-score=1.86, p=0.0386 and z-score=2.77, p=0.0002 based on permutation test for *V. dahliae* strains JR2 and CQ2, respectively). Together with H3K27me3, we observe that also H3K36me3 is enriched in AGRs (Fig. 4A; z-score=2.16, p=0.00019 and z-score=5.099, p=0.0002 based on permutation test for *V. dahliae* strains JR2 and CQ2, respectively). Principle component analyses based on the genome-wide distribution of histone modifications in *V. dahliae* strains JR2 and CQ2 show that the same histone modifications cluster irrespective of the strain (Fig. 4B; Supplementary Fig. 6A,C), confirming the expectation that the genome-wide histone landscape defined by these seven histone modifications in *V. dahliae* is largely conserved between lineages.

To determine how the seven histone modifications associate with gene expression in *V. dahliae*, we summarized their coverage over coding sequences as well as their surrounding upstream and downstream regions, and further separated genes based on their expression levels (Fig. 4C). We observe that highly expressed genes are enriched in H3K4me3, H3K9ac, and H3K4ac close to the transcription start site (TSS) and at the 5'flanking regions, which contain CREs, while H3K36me3 is enriched at coding sequences (Fig. 4C; Supplementary Fig. 7A,B). Highly expressed genes are also enriched in H4K16ac at the TSS and transcription end site (TES) (Fig. 4C; Supplementary Fig. 7A,B). In contrast, lowly expressed genes are enriched in H3K27me3 or H3K9me3 (Fig. 4C), which is in line with our previous observations (Cook et al. 2020; Kramer et al. 2022). Interestingly, the presence of H3K36me3 positively correlates with the occurrence of H3K27me3, yet both highly and lowly expressed genes are enriched in H3K36me3 (Fig. 4C,D).

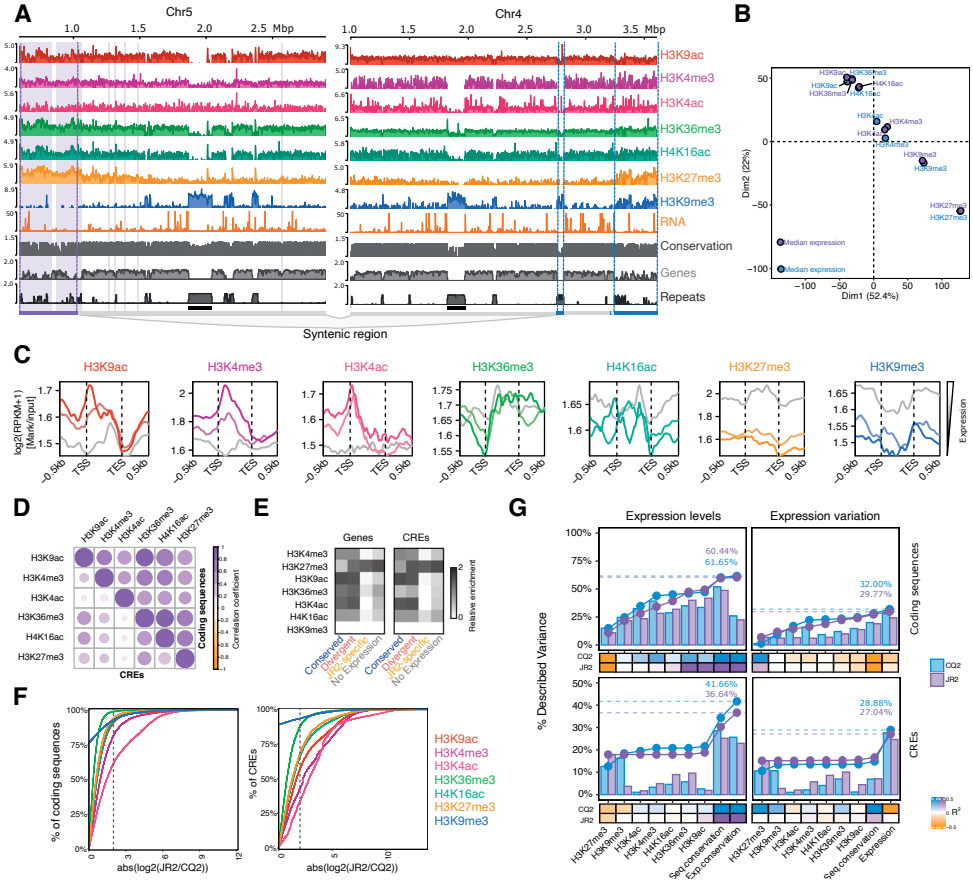
Next, we determined if changes in histone modifications between *V. dahliae* strains are associated with differential gene expression. More than 75% of coding sequences do not display marked differences in histone modification levels when compared between *V. dahliae* strains JR2 and CQ2 (Fig. 4F). While CREs display larger differences in histone modifications (Fig. 4F), we do not observe that genes with associated CREs or coding sequences that differ in histone modifications levels also show differentially expression between strains (Supplementary Table 2). Still a small fraction of coding sequences and CREs shows difference in histone modification levels between both strains. As expected, only H3K27me3 is significantly enriched at differentially expressed, CQ2- or JR2-specific genes, as well as at non-expressed genes, while depleted from genes with conserved expression (Fig. 4E). Thus, our results suggest that H3K27me-associated genes, which mainly reside in AGRs, are enriched for genes that show differential expression not only *in planta* (Cook et al. 2020)

and between growth conditions (Kramer et al. 2022), but that these are also differentially expressed between *V. dahliae* strains.

To quantify the contributions of the seven histone modifications to gene expression levels and gene expression variation, we built sparse multiple regression models (Koenker and Ng 2003) for each of the two strains (Fig. 4G; Supplementary Table 3). Sparse multiple regression accounts for the inherent sparse enrichment of histone modifications and the large number of coding sequences or CREs to describe the relationships between histone modifications, expression levels, and expression variation. Considering only coding sequences, a model that combines all histone modifications account for 60.4% and 32.0% of the variation in gene expression levels and gene expression variation, respectively (Fig. 4G). Importantly, H3K27me3 alone can account for at most 15% and 12% of the variation in gene expression levels and gene expression variation, respectively (Fig. 4G; Supplementary Table 3), while combining H3K27me3 with the other six histone modifications describes more variation in gene expression levels and gene expression variation, a trend we similarly observed when only considering CREs (Fig. 4G; Supplementary Table 3). Thus, as previously shown, H3K27me3 cannot account for *in vitro* gene expression levels (Kramer et al. 2022), and we here also demonstrated that H3K27me3 alone cannot account for the differential expression between two *V. dahliae* strains.

>>

Figure 4. Histone modifications in *Verticillium dahliae* are associated with distinctive gene expression patterns. **(A)** Overview of the distribution of seven histone modifications over a large syntenic region in *V. dahliae* strains JR2 (left) and CQ2 (right). *In vitro* histone modifications in conidiospores harvested from *V. dahliae* strains JR2 and CQ2 grown on potato dextrose agar (PDA) are shown; histone modifications were normalized using a micrococcal nuclease digestion control. Syntenic (bottom grey) and non-syntenic blocks in *V. dahliae* strains JR2 (purple) and CQ2 (blue), as well as AGRs (purple and blue boxes) are shown. **(B)** Histone modification distribution is conserved between *V. dahliae* strains JR2 and CQ2. Principal component analysis of homologous regions in *V. dahliae* strains JR2 (purple) and CQ2 (blue) based on 10 kb windows that summarize the seven histone modifications and gene expression levels. **(C)** The occurrence of specific histone modifications at and around genes correlates with differences in gene expression levels. The relative coverage of histone modifications is calculated over coding regions and ±500 bp flanking sequence of the transcription start site (TSS) and transcription end site (TES), respectively. Genes are grouped into low expression (1st quartile), mid-expression (2nd and 3rd quartile), and high expression (4th quartile) based on gene expression levels when grown on PDA. Grey color indicates low expression, fade color mid-expression, and bright color high expression. **(D)** Spearman's correlation for histone modifications at coding sequences (top) and CREs (bottom) in *V. dahliae* strain JR2; color code and circle size depict the direction and the absolute strength of the correlation, respectively. **(E)** Relative enrichment of histone modifications in *V. dahliae* strain JR2 over coding sequences (left) and CREs (right) and grouped based on their expression conservation (see Fig. 2B). **(F)** Differential enrichment of histone modifications over coding sequences and CREs in *V. dahliae* strains JR2 and CQ2; the dashed line depicts a 2.0-fold enrichment. **(G)** Percentage of described variance by seven histone modifications for expression levels and expression variation using a sparse multiple regression model. Bar plots depict the variance described by a single histone modification, the sequence conservation, the gene expression levels and gene expression conservation between *V. dahliae* strains JR2 and CQ2 along the x-axis. The line plots depict the cumulative variance described by each histone modification indicated on the x-axis from left to right. Heatmaps at the bottom depict the R^2 value indicating a positive or negative correlation of each histone modification in the full multiple regression model.



Seven histone modifications define distinct chromatin states in *Verticillium dahliae*

Given that histone modifications often co-occur genome-wide (Fig. 4E,D) and single histone modifications can only partially explain gene expression levels and conservation (Fig. 4G), we reasoned that discrete combinations of histone modifications might be relevant for gene expression. To define combinations of the seven histone modifications, we used a multivariate Hidden Markov Model (Ernst and Kellis 2012, 2017), which yielded 20 combinations of histone modifications in each *V. dahliae* strain, referred to as chromatin states (Baker 2011) (Fig. 5A; Supplementary Fig. 8,9). We subsequently annotated these states based on their association with transcription (Fig. 4; Fig. 5C, e.g. 'Promoter activating', 'Transcription', and 'Repressive') or by their shared histone modification (e.g. 'H4K16ac', or 'H3K27me3'). Fifteen chromatin states are found in both strains and are here referred to as 'shared' states (Fig. 5A; Supplementary Fig. 9), while we also observe five unique chromatin states in *V. dahliae* strain JR2 and CQ2, referred to as 'strain-specific' states (Fig. 5A; Supplementary

Fig. 9). These strain-specific states collectively cover 17.23% and 19.17% of the JR2 and CQ2 genome, respectively (Fig. 5B). Interestingly, the fifteen shared chromatin states are also similar in *V. dahliae* strain JR2 and CQ2 (Fig. 5A) as they are enriched in the same genomic compartments and occur at homologous genes in both strains (Fig. 5B; Supplementary Fig. 9). For example, the shared chromatin states annotated as 'Promoter activating', 'Transcription', or 'H4K16ac' are enriched in core regions, whereas the 'Repressive' state is associated with centromeres and TEs (Fig. 5B). Five of the fifteen shared chromatin states contain H3K27me3 in combination with H3K9ac, H3K4me3, H3K36me3, or H4K16ac (Fig. 5B). As to be expected, H3K27me3-associated shared and strain-specific chromatin states are enriched in AGRs (Fig. 5B) (Cook et al. 2020; Kramer et al. 2022), and our results thus indicate that discrete combinations of histone modifications are associated with genome compartmentalization in *V. dahliae*.

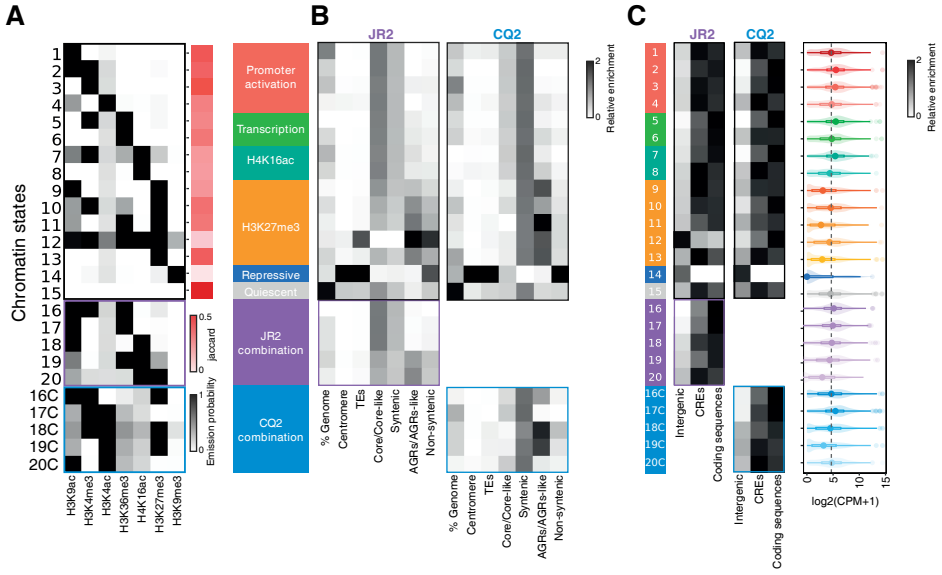
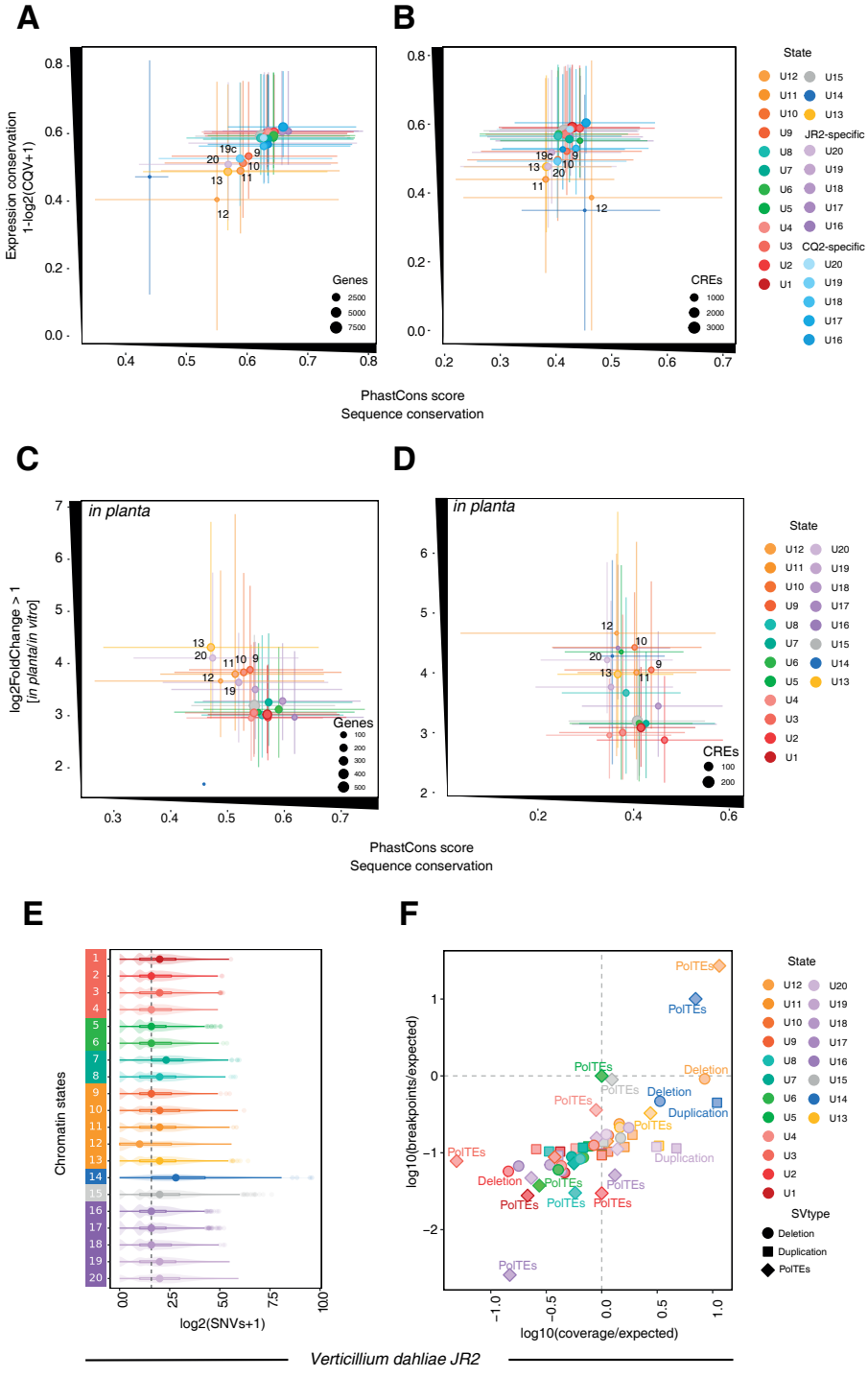


Figure 5. Distinct chromatin states are associated with genome compartmentalization in *Verticillium dahliae*. **(A)** The combinations of 15 shared chromatin states of *V. dahliae* strains JR2 and CQ2 and five CQ2- or JR2-specific chromatin states. The heatmap shows the probability of each histone modification to belong to a named chromatin state (emission probability). On the right, the heatmap depicts the Jaccard's similarity (red) based on the emission probabilities of each shared chromatin state and the presence at homologous genes. **(B)** Specific chromatin states are associated with genome compartmentalization into core regions, centromeres, and AGRs in two *V. dahliae* strains. **(C)** Chromatin states associated with gene expression in conidiospores harvested from PDA-grown *V. dahliae* are differentially enriched at intergenic regions, CREs, and coding sequences in two *V. dahliae* strains. The dotted line depicts the genome-wide gene expression average as a reference.

Chromatin states link gene expression and sequence variation

Specific chromatin states are associated with increased or decreased expression levels in conidiospores (Fig. 5C). For example, state 13 (H3K27me3-alone) is associated with low gene expression levels, while state 10 (H3K27me3-H3K4me3) occurs at genes that are expressed to higher levels. We therefore reasoned that specific chromatin states might also be linked to sequence and expression variation (Fig. 1). High gene expression and sequence conservation is associated with chromatin states without H3K27me3, but with activating histone modifications such as H3K4me3, H3K9ac, or H3K36me3 (Fig. 6A,B). Conversely, gene expression and sequence conservation are reduced for states associated with H3K27me3 (Fig. 6A,B). Similarly, CQ2- and JR2-specific genes are almost exclusively associated with H3K27me3 states (Supplementary Fig. 10A), as expected, since these genes typically occur in AGRs that are enriched in H3K27me3 (Cook et al. 2020). Notably, not only genes associated with state 13 (H3K27me3-alone), but also chromatin states in which H3K27me3 occurs in combination with other histone modifications, such as chromatin states 9 (H3K9ac-H3K27me3), 10 (H3K4me3-H3K27me3), and 11 (H3K36me3-H3K27me3), show reduced gene expression and sequence conservation in both strains (Supplementary Fig. 10B,C). Importantly, genes enriched in the H3K27me3-associated states 9, 10 and 11 *in vitro* are also differentially upregulated during *Arabidopsis thaliana* infection in both strains (Fig. 6C,D; Supplementary Fig. 10F; Supplementary Table 4; $p<0.05$), thus indicating that H3K27me3-associated chromatin states correlate with reduced gene expression and sequence conservation, as well as differential gene expression upon plant colonization in both *V. dahliae* strains.

As AGRs are enriched in SVs and large-scale genomic rearrangements (de Jonge et al. 2013; Faino et al. 2016; Torres et al. 2021), we expected that H3K27me3-associated chromatin states would also be enriched in genomic variation, while other states would be depleted of variation, possibly due to differences in selection (Fudenberg and Pollard 2019). To this end, we used the previously generated data for a set of 42 *V. dahliae* strains (Torres et al. 2021), to query the distribution of single nucleotide variants (SNVs) and structural variations (SVs) over the chromatin states of *V. dahliae* JR2 (Supplementary Table 5). As expected, we observe that chromatin states associated with active transcription (Fig. 5D) are depleted in SNVs and SVs, whereas H3K27me3- and H3K9me3-associated states are enriched for SVs and SNVs (Fig. 6E,F; Supplementary Fig. 10D,E). Even though we observe small differences between H3K27me3-associated chromatin states, our results indicate that all H3K27me3-associated states show high genomic variation in *V. dahliae*.

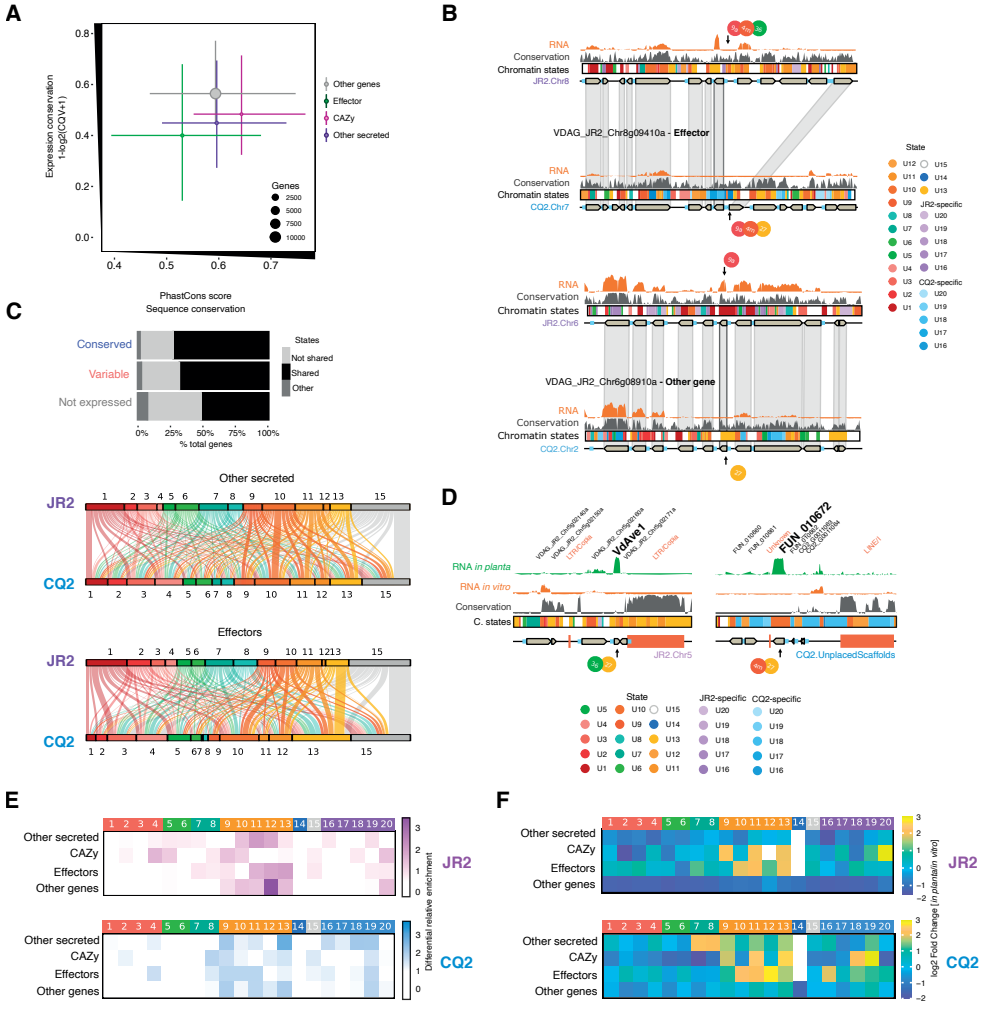

A
B
C
D
E
F
Expression conservation
1-log2(CQV+1)
PhastCons score
Sequence conservation
Genes
CREs
State
JR2-specific
CQ2-specific
in planta
log2FoldChange > 1
[in planta/in vitro]
Chromatin states
log2(SNVs+1)
log10(breakpoints/expected)
log10(coverage/expected)
SVtype
Deletion
Duplication
PolTEs
Verticillium dahliae JR2

<<

Figure 6. Chromatin states link expression and sequence conservation in *Verticillium dahliae*.
(A) Gene expression conservation (coefficient of interquartile variation) and sequence conservation (PhastCons score) is shown for genes grouped by their associated chromatin state. Genes with H3K27me3-associated chromatin states have lower gene expression conservation and sequence conservation. The individual points summarize the mean values for conservation and expression conservation, the point size depicts the total number of genes associated with the chromatin state, and the line indicates the minimum and maximum value. **(B)** Similar as **(A)** for chromatin states at CREs. **(C)** Differential gene expression *in planta* (*Arabidopsis thaliana*, 21 dpi) compared with *in vitro* and sequence conservation (PhastCons score) in *V. dahliae* strain JR2 is shown for genes grouped by their associated chromatin state. Similar results for *V. dahliae* strain CQ2 are shown in Fig. S9. Gene associated with H3K27me3-states have a higher, positive $\log_2$ fold-change (i.e., induced gene expression *in planta* compared with *in vitro*; $p<0.05$) and lower sequence conservation. The individual points summarize the mean values for sequence conservation and $\log_2$ fold-change, while the point size depicts the total number of associated genes with the chromatin state, and the line indicates the minimum and maximum value. **(D)** Similar as **(C)**, for chromatin states at CREs. **(E)** Single nucleotide variants (SNVs) are depleted in chromatin states associated with active transcription. The dotted line shows the genome-wide average number of SNVs. **(F)** Structural variants (SVs) are depleted in chromatin states associated with gene expression. Only deletions, duplications, and polymorphic TEs (polTEs) are shown, the full set of SVs is depicted in Supplementary Fig. 9.

Chromatin states are associated with expression and sequence conservation of conditionally responsive genes

The histone modification H3K27me3 is associated with conditionally responsive genes in *V. dahliae* (Kramer et al. 2022; Cook et al. 2020). It is thought that some of those conditionally responsive genes contribute to host colonization and environmental adaptation (Kramer et al. 2021, 2022; de Jonge et al. 2013; Torres et al. 2021; Cook et al. 2020; Faino et al. 2016). We therefore investigated if differentially expressed genes between *V. dahliae* strains JR2 and CQ2 are enriched in typical pathogenicity-related gene functions (Cook et al. 2020; Kramer et al. 2022). We defined pathogenicity-related gene functions to include predicted effectors (n=320 and 354; JR2 and CQ2, respectively), secreted carbohydrate active enzymes (CAZy) (n=229 and 210; JR2 and CQ2, respectively), and other secreted proteins (n=364 and 335; JR2 and CQ2, respectively) (Supplementary Fig. 11A). The AGRs are enriched for effectors (z-score=4.71, p=0.0001 and z-score=1.62, p=0.0385 based on permutation test for JR2 and CQ2, respectively), CAZy (z-score=0.61, p=0.0482 and z-score=0.96, p=0.0346 based on permutation test for JR2 and CQ2), and other secreted proteins (z-score=2.67, p=0.0001 and z-score=4.02, p=0.0001 based on permutation test for JR2 and CQ2, respectively). As expected, when compared with other genes, most pathogenicity-related genes have higher sequence and expression variation, especially for effectors and other secreted proteins (Fig. 7A; Supplementary Fig. 11B). Consequently, differentially expressed genes between strains are enriched in pathogenicity-related gene functions (Supplementary Fig. 11B,C).


A
Expression conservation
1-log2(CQV+1)
Other genes
Effector
CAZy
Other secreted
Genes
B
RNA
Conservation
Chromatin states
JR2.Chr8
VDAG_JR2_Chr8g09410a - Effector
CQ2.Chr7
VDAG_JR2_Chr6g08910a - Other gene
JR2.Chr6
CQ2.Chr2
State
JR2-specific
CQ2-specific
C
PhastCons score
Sequence conservation
Conserved
Variable
Not expressed
States
Not shared
Shared
Other
% total genes
Other secreted
Effectors
JR2
CQ2
D
RNA in planta
RNA in vitro
Conservation
C. states
JR2.Chr5
CQ2.UnplacedScaffolds
E
Other secreted
CAZy
Effectors
Other genes
Differential relative enrichment
F
log2 Fold Change (in planta/in vitro)

<<

Figure 7. H3K27me3-associated chromatin states correlate with gene expression variation and sequence variation of pathogenicity-related genes.

(A) Predicted pathogenicity-related genes are associated with high gene expression variation (coefficient of interquartile variation) and low sequence conservation (PhastCons score). The individual points summarize the average values for sequence conservation and gene expression variation, the point size depicts the total number of genes, and the line indicates the minimum and maximum value. Genes coding for: CAZy = carbohydrate active enzymes, effectors, and other secreted proteins. 'Other genes' encode non-pathogenicity-related functions. **(B)** A syntenic locus in *V. dahliae* strains JR2 and CQ2, as in Fig. 3D. Changes in chromatin states can lead to changes in gene expression. Ribbons link homologous genes between the two strains. From top to bottom: RNA *in vitro* (from conidiospores harvested from *V. dahliae* strains JR2 and CQ2 grown on PDA) sequence conservation (PhastCons score), gene models, and CREs (light blue). The color code for the chromatin state is shown at the bottom. **(C)** Changes in chromatin states between *V. dahliae* strains JR2 and CQ2 are associated with differentially expressed and non-expressed genes (only homologous genes are considered). Homologous genes in JR2 and CQ2 with more than one similar chromatin state are considered to be 'conserved'. Genes with chromatin states 16-20 are considered in the 'other' category. The two bottom plots show homologous effectors and other secreted protein coding genes overlapping with conserved and different chromatin states *V. dahliae* strains JR2 and CQ2. **(D)** *In planta* expressed effectors are enriched in H3K27me3-associated chromatin states. From top to bottom: RNA *in planta* track depicts *V. dahliae* strain JR2 (*Arabidopsis thaliana*, 21 dpi) and CQ2 (*Arabidopsis thaliana*, 15 dpi), left and right, respectively; RNA *in vitro* depicts *V. dahliae* JR2 and CQ2 (in conidiospores harvested from *V. dahliae* grown on PDA); sequence conservation (PhastCons score); chromatin states distribution is color coded; gene models (grey), CREs (light blue), and TEs (red). **(E)** Differential expression *in vitro* between strains in pathogenicity-related genes is associated with H3K27me3-rich chromatin states. Differential enrichment of 20 chromatin states for differentially expressed genes *versus* genes with conserved expression in each category for *V. dahliae* strains JR2 (purple) and CQ2 (blue). Black dots show significant enrichment of a given state at the pathogenicity-related differentially expressed genes. Statistical significance was assessed by Chi-square test, and *p-value* adjusted for false discovery rate <0.05.**(F)** Differential *in planta* expression of pathogenicity-related genes is associated with H3K27me3-rich chromatin states. Heatmap depicts the $\log_2$ fold-change (*in planta* compared with *in vitro*) for *V. dahliae* strains JR2 (*Arabidopsis thaliana*, 21 dpi) and CQ2 (*Arabidopsis thaliana*, 15 dpi).

Next, we reasoned that differential expression between *V. dahliae* strains might be associated with changes in the chromatin states between strains (Fig. 7B). Interestingly, while 69.39% of homologous genes (*n*=7677) shows conserved chromatin states in both strains (Fig. 7C), we observe that genes with differential gene expression also tend to differ in chromatin states (Fig. 7C; X^2=8.4; *p*=0.0037; Supplementary Table 6). We observe that only 27.34% (*n*=700) and 54.51% (*n*=538) of differentially and non-expressed genes, respectively, are associated with the same chromatin states in both strains (Supplementary Table 6). Interestingly, such conserved chromatin states are chromatin state 10 (H3K27me3-H3K4me3) for effector coding genes (X^2=13.5; *p*=0.00020), and state 11 (H3K27me3-H3K36me3) for genes encoding secreted proteins (X^2=5.8; *p*=0.0158), besides the expected presence of chromatin state 13 (H3K27me3-only) (Fig. 7C; X^2=5.56; *p*=0.018, X^2=5.09; *p*=0.024, effector and secreted proteins, respectively; Supplementary Table 7). Taken together, our results show that specific H3K27me3-associated chromatin states are conserved in pathogenicity-related genes in *V. dahliae.*

H3K27me3 is associated with *in planta*-induced genes in *V. dahliae* (Cook et al. 2020; Kramer et al. 2022). For example, the *in planta* highly expressed effector gene *Ave1* from *V. dahliae* strain JR2 (de Jonge et al. 2012; Snelders et al. 2020) or a putative effector of *V. dahliae* CQ2 are only lowly expressed in conidiospores harvested *in vitro* and are linked with the H3K27me3-associated chromatin states 10 (H3K27me3-H3K4me3) and

11 (H3K27me3-H3K36me3) *in vitro*, respectively (Fig. 7D). We observe that H3K27me3-associated chromatin states 9, 10, 11 and 12, but not 13 (H3K27me3-alone) are enriched for pathogenicity-related genes that are differentially expressed between the two *V. dahliae* strains *in vitro* (Fig. 7E; Supplementary Table 8), and genes differentially expressed *in planta* compared with *in vitro* (Fig. 7F). Even though we cannot observe strong differences between H3K27me3-associated chromatin states, these results suggest that a subset of genes with specific functions are differentially expressed *in planta* and associated with only a subset of H3K27me3-associated chromatin states in both *V. dahliae* strains.

Discussion

Gene expression variation plays an important role in adaptation (Bedford and Hartl 2009; Hill et al. 2021). Here, we studied gene expression variation in two genetically divergent strains of the fungal plant pathogen *V. dahliae*. We show that gene expression differs for one-third of the genes, and these differences are likely not caused by changes in the organization and sequence of CREs. We summarize different combinations of seven histone modifications in distinct chromatin states and a subset of those states can be linked to gene expression changes between strains. Intriguingly, H3K27me3 occurs in chromatin states together with histone modifications typically associated with active gene expression such as H3K9ac, H3K4me3, or H3K36me3. These chromatin states are associated with reduced sequence conservation, differential gene expression *in planta*, and differential gene expression between *V. dahliae* strains. Genes with H3K27me3-associated chromatin states are enriched for pathogenicity-related gene functions that mostly reside in AGRs. Thus, H3K27me3-associated chromatin states correlate with the evolution of gene expression and genome organization in *V. dahliae*.

To identify the combinations of histone modifications, so-called chromatin states, we use a diverse set of seven histone modifications determined in conidiospores. We define 20 combinations of these seven histone modifications in each strain of *V. dahliae*. We recover similar combinations in fifteen shared chromatin states, a similar genome-wide coverage of those states, and many homologous genes with conserved chromatin states in two *V. dahliae* strains. Next to the conservation of fifteen chromatin states, we also observe five CQ2- or JR2-specific states. Unique chromatin states have been widely associated with differential gene expression between individuals, sexes, and cell types in animals and plants (Kasowski et al. 2013; Pellacani et al. 2016; Roadmap Epigenomics et al. 2015; Brown and Bachtrog 2014; Vu and Ernst 2022; Fu et al. 2022; Zhao et al. 2020; Liu et al. 2018). Thus, possibly, the CQ2- or JR2-specific chromatins states could be explained by the genetic variation between both strains. In a similar fashion, intraspecific comparisons in *Aspergillus fumigatus*, *Saccharomyces cerevisiae*, and the oomycete *P. sojae* showed differences in the presence of histone modifications between strains (Filleton et al. 2015; Colabardini et al.

2022; Wang et al. 2020a). However, we presently cannot determine if the strain-specific states in *V. dahliae* represent biologically meaningful chromatin states, as we observe little differences in the levels of histone modifications associated with differential gene expression between strains. Alternatively, variability within the population of conidiospores, hence cell heterogeneity, may be the source of those strain-specific chromatin states. In this respect it is interesting to note that in Aspergillus conidiospores show high expression variation with conidial age or slight changes in the environment (Wang et al. 2021a; Takahashi-Nakaguchi et al. 2018). However, in our study *V. dahliae* strains were cultivated simultaneously and under the same controlled conditions, thus suggesting a minimal effect on strain-specific states due to cell heterogeneity. While it is important to mention that our biological replicates and our conservative approach that consider only peaks present in both replicates to determine the chromatin states yielded similar results, further experiments should be performed to validate the presence of the strain-specific chromatin states. We here analyze two genetically divergent strains, thus, further analyses of strains that belong to the same genetic lineage could confirm similar CQ2- and JR2-specific chromatins states. Additionally, given the strain-specific histone modification combinations, a further confirmation could be possible by measuring the abundance and correlation of such histone modifications at specific loci by ChIP-qRT-PCR (Mukhopadhyay et al. 2008; Kim and Dekker 2018), or by a sequential ChIP of those histone modifications (Furlan-Magaril et al. 2009; Kinkley et al. 2016). Taken together, our analyses could recover fifteen chromatin states shared in two genetically divergent strains, but further analyses are needed to confirm the presence of CQ2- and JR2-specific chromatin states.

We previously demonstrated that H3K4me2, H3K27ac, H3K27me3, and H3K9me3 correlate with gene expression levels and genome organization in the mycelium of *V. dahliae* (Cook et al. 2020; Kramer et al. 2022; Torres et al. 2023). In *V. dahliae* and other fungi, the histone modification H3K27me3 plays a role in gene expression regulation (Weiner et al. 2015; Brosch et al. 2008; Lai et al. 2022; Freitag 2017). For example, removal of H3K27me3 by deletion of the gene encoding the Polycomb subunit Set7, which is responsible for the tri-methylation of H3K27, led to the expression of normally repressed genes that are, in wild-type, enriched in H3K27me3 (Connolly et al. 2013; Studt et al. 2017; Möller et al. 2019; Zhang et al. 2021b; Kramer et al. 2022; Jamieson et al. 2013; Lin et al. 2022; Chujo and Scott 2014). However, we previously demonstrated that H3K27me3 does not act as a global regulator of differential gene expression between growth conditions in *V. dahliae* (Kramer et al. 2022). Our results here indicate that H3K27me3 alone and in combination with other histone modifications in H3K27me3-associated chromatin states correlate with the observed differential gene expression between *V. dahliae* strains.

Histone modifications associated with activation or repression of gene expression are typically antagonistic in the same nucleosome (Bannister and Kouzarides 2011). However, some histone modifications can reside in different tails at the same nucleosome and thus coexist at the same locus (Macrae et al. 2023; Villaseñor et al. 2020; Bryan et al. 2021; Voigt

et al. 2012). The presence of chromatin states with typically repressive and active histone modifications is thought to regulate gene expression to allow for swiftly changes for future activation or repression depending on the cell's life stage (Macrae et al. 2023; Zhang et al. 2021ª; Zhao et al. 2020; Zhu et al. 2023; Zhao et al. 2021; Zeng et al. 2019; Blanco et al. 2020). While we currently cannot differentiate if modifications co-occur at tails at the same histone protein or at the same locus due to cell heterogeneity, our results show the presence of chromatin states defined by histone marks that are typically associated with either activation or repression of gene expression. Interestingly, we observe that H3K27me3 occurs not only alone, but also in combination with activating histone modifications such as H3K9ac, H3K4me3, or H3K36me3. We observe that genes associated with H3K27me3-associated chromatin states are lowly expressed *in vitro* and differentially expressed *in planta*, suggesting that the presence of H3K27me3-associated chromatin states could correlate with dynamic expression. Many *in planta*-induced as well as other conditionally responsive genes in diverse filamentous fungi reside in H3K27me3-rich regions (Kramer et al. 2022; Zhang et al. 2021b; Lukito et al. 2021; Connolly et al. 2013; Meile et al. 2020; Meng et al. 2021; Zhou et al. 2021; Xie et al. 2023; Chujo and Scott 2014). Additionally, observations in *Leptosphaeria maculans*, *Magnaporthe oryzae*, *Podospora anserina, Fusarium graminearum*, *F. fujikuroi*, and *Z. tritici*, also point towards the occurrence of H3K27me3 in combinations with H3K4me3 or H3K36me3, which is associated with low gene expression *in vitro* (Tralamazza et al. 2022; Zhang et al. 2021b; Connolly et al. 2013; Soyer et al. 2021; Janevska et al. 2018; Carlier et al. 2021; Studt et al. 2017; Möller et al. 2023). Moreover, the co-occurrence of H3K27me3 with other histone modifications has been associated with differential expression of secondary metabolites and other environmentally responsive genes (Janevska et al. 2018; Connolly et al. 2013; Zhang et al. 2021b; Möller et al. 2023). Thus, H3K27me3-associated chromatin states with combinations of H3K9ac, H3K4me3, or H3K36me3 might be relevant for the dynamic expression of conditionally responsive genes.

As expected, AGRs are enriched in H3K27me3-associated chromatin states in both *V. dahliae* strains. Interestingly, in plants and animals, the presence of H3K27me3 in combination with activating histone modifications (e.g. H3K4me3) is associated with the absence of DNA methylation, high chromatin accessibility, and physical co-localization into different nuclear compartments than just euchromatic or H3K27me3-only chromatin states (Fu et al. 2022; Roadmap Epigenomics et al. 2015; Macrae et al. 2023; Yuan et al. 2022; Li et al. 2019; Zhu et al. 2023; Ikeda et al. 2017). Thus, the presence of H3K27me3-associated chromatin states could correspond to the poor DNA methylation, high chromatin accessibility and physical co-localization observed in the AGRs (Torres et al. 2023; Seidl et al. 2020; Cook et al. 2020; Kramer et al. 2022, 2021). Such unique chromatin characteristics, together with segmental duplications, reciprocal gene losses, and active transposable elements define AGRs in *V. dahliae* (de Jonge et al. 2013; Faino et al. 2016; Cook et al. 2020; Kramer et al. 2022; Torres et al. 2021, 2023). Thus, H3K27me3-associated chromatin

states in combination with activating histone modifications could be involved in the genome organization beyond differential gene expression.

The histone modification H3K36me3 is associated with euchromatin in many eukaryotes (Weiner et al. 2015; Kolasinska-Zwierz et al. 2009). However, H3K36me3 establishes facultative heterochromatin at sub-telomeres in some filamentous fungi (Möller et al. 2023; Janevska et al. 2018; Bicocca et al. 2018; Connolly et al. 2013). In *Neurospora crassa*, *F. fujikuroi*, and *Z. tritici*, lowly-expressed genes that are enriched in H3K27me3 are also enriched for Ash1-specific H3K36me2/3 (Bicocca et al. 2018; Janevska et al. 2018; Möller et al. 2023). Similarly, we here observe that lowly-expressed genes are enriched in H3K36me3 and H3K27me3 in the facultative heterochromatic AGRs, and that H3K36me3 is also present at highly expressed genes in the core genome. Interestingly, the presence of H3K36me3 has been associated with the repair of DNA double-strand breaks in facultative heterochromatic regions (Sun et al. 2020; Sims and Reinberg 2009; Li et al. 2013; Musselman et al. 2012). We previously hypothesized that the physical colocalization of homologous sequences in AGRs may lead to decreased DNA separation efficiency during mitosis and increased DNA double-strand breaks that might explain the abundant genomic rearrangements in AGRs (Huang and Cook 2022; de Jonge et al. 2013; Seidl and Thomma 2017; Cook et al. 2020; Faino et al. 2016; Torres et al. 2021, 2023). Here, we similarly observe an enrichment of structural variations in H3K27me3-associated chromatin states, and thus we speculate that specifically H3K27me3-H3K36me3 chromatin states might be relevant for the formation and maintenance of AGRs in *V. dahliae*.

We show that H3K27me3-associated chromatin states correlate with sequence variation across the *Verticillium* genus at genes and CREs. The increased variation of CREs has been shown relevant for gene expression variation between species (Villar et al. 2015; Berthelot et al. 2018; Marand et al. 2023; Gasch et al. 2004). Unexpectedly, the increase in sequence variation of CREs cannot account for the total number of differentially expressed genes between *V. dahliae* strains. Instead, the presence of H3K27me3-associated chromatin states correlates with differential gene expression and sequence variation. Especially, it is thought that the presence of H3K27me3 marks poorly conserved genes and CREs, potentially as a mechanism of temporary gene silencing of 'new' genes and CREs that have been only recently incorporated into the genome, e.g., by horizontal gene transfer (Connolly et al. 2013; Jamieson et al. 2013; Cook et al. 2020; Habig et al. 2021; Tralamazza et al. 2022; Joubert and Krasileva 2023; Xie et al. 2023). Additionally, in *N. crassa* and *Z. tritici*, the presence of H3K27me3 is associated with high mutation rates, compared to regions depleted in H3K27me3 (de la Peña et al. 2023; Möller et al. 2021; Habig et al. 2021). Thus, it is tempting to speculate that H3K27me3 in combination with an euchromatic mark may have a protective role to avoid a complete suppression of genes with increased sequence variation. However, it remains unclear whether H3K27me3-associated chromatin states result from sequence variation, or whether H3K27me3-associated states reflect lower functional constraints at poorly conserved genes that increase sequence variation.

We observe a positive correlation between gene expression variation and low sequence conservation, which has been also found in other species (Pál et al. 2001; Krylov et al. 2003; Rocha and Danchin 2004; Drummond et al. 2005a, 2005b; Ingvarsson 2007; Zhang and Yang 2015; Tralamazza et al. 2022). However, the evolutionary forces behind this observation remain controversial. It has been suggested that stabilizing selection is responsible for reducing gene expression divergence by removing mutations that drastically decrease or increase gene expression (Berthelot et al. 2018; Chen et al. 2019; Romero et al. 2012; Josephs et al. 2015). However, under natural selection, the purge of mutations mainly occurs through recombination and depends on large effective population sizes (Price et al. 2022; Vanhoenacker et al. 2018). Given that *V. dahliae* is an asexual fungus and consequently is expected to have a low effective population size, we consider that stabilizing selection likely plays only a minor role in gene expression conservation. Instead, gene expression variation could be associated with an accumulation of 'strain-specific' variation due to genetic drift (Fay and Wittkopp 2008; Groen et al. 2020; Price et al. 2022). We speculate that such gene expression variation could be facilitated by the H3K27me3-associated chromatin states. Increasing evidence suggests that variation mediated by the chromatin structure is important for evolution (Hu and Barrett 2017; Minow and Colasanti 2020; Ashe et al. 2021; Stajic and Jansen 2021). However, further insights into the mechanisms of chromatin formation are necessary to understand the role of chromatin states in the evolution of gene expression and genome organization in *V. dahliae*.

Materials and Methods

Phylogenetic analyses of *Verticillium* strains and species

To determine phylogenetic relationships within the *Verticillium* genus, we compared the Hi-C corrected genome assemblies of ten *Verticillium* species to a collection of 63 *V. dahliae* strains that have previously been sequenced with short- and long-read sequencing technologies (Seidl et al. 2020; Chavarro-Carrero et al. 2021; Torres et al. 2021). The phylogenetic tree was generated with Realphy (version 1.12) (Bertels et al. 2014) using Bowtie2 (version 2.2.6) (Langmead and Salzberg 2012) to map the genomic reads against the genome assembly of *V. dahliae* strain JR2 (Faino et al., 2015). A maximum-likelihood phylogenetic tree was inferred by using RaxML (version 8.2.8) (Stamatakis 2014); we used *Sodiomyces alkalinus* as an outgroup species to root the tree

Sequence conservation analysis

Whole-genome alignments of chromosomes of the complete genome assemblies of *V. dahliae* strains JR2 and CQ2 (Faino et al. 2015; Depotter et al. 2019) were performed using nucmer (--maxmatch), which is part of the MUMmer package (version 3.23) (Kurtz et al. 2004). Next, following the same approach (Kurtz et al. 2004), we performed whole-genome

alignments of *V. dahliae* strain CQ2 to itself to identify large-scale duplication events (>80% identity; Fig. S1). Based on the genome alignments of the *V. dahliae* strains JR2 and CQ2, we obtained the number of SNVs (single nucleotide variants) over homologous regions (Fig. S1).

Nucleotide conservation of *V. dahliae* was determined using the high-quality genome assemblies of ten *Verticillium* species including *V. dahliae* strains JR2, CQ2, and VdLs17. We performed pan-genome alignments of the genome assemblies using ProgressiveCactus with default settings (Armstrong et al. 2020), an approach that reduces biases of a single reference genome, which is highly relevant for abundant presence/absence variation. We focused on *V. dahliae* strains JR2 and CQ2 and obtained MAF alignments by using the HAL package (Hickey et al. 2013) to determine syntenic regions and to identify single nucleotide variants (SNVs). To measure the nucleotide conservation in these two *V. dahliae* strains, based on the sequence similarity to species in the *Verticillium* genus (Supplementary Fig. 1), we used the Hidden Markov Model approach implemented in PhastCons to estimate the probability of each nucleotide to be conserved (Siepel et al. 2005). Briefly, for each of the eight chromosomes in *V. dahliae* we assumed a neutral evolution model and a correction for insertions/deletions, an approach similar to the one previously applied in *Saccharomyces cerevisiae* (Siepel et al., 2015). We used the Viterbi algorithm implemented in PhastCons to determine significantly conserved nucleotides. We calculated the PhastCons score by using either *V. dahliae* JR2 or CQ2 as independent references, and both yielded highly correlated results (R=0.98, $p<2.2 \times 10^{-6}$ Pearson's correlation test; Supplementary Fig. 1).

Verticillium dahliae genome annotation

De novo protein-coding genes in *V. dahliae* strain CQ2 were annotated by combining two approaches. First, repetitive elements were identified using RepeatModeler2 and masked using RepeatMasker (Flynn et al. 2020). Then, gene prediction was performed with the funannotate pipeline (Palmer and Stajich 2020) using a combination of *ab initio* gene predictors and consensus predictions obtained using Evidencemodeler (v1.1.1). Subsequently, gene prediction parameters were estimated using *in vitro* RNA-seq data, as well with JR2 *in vitro* RNA-seq for robustness. The final gene predictions were subsequently adjusted using information from the RNA-seq data and the previously estimated parameters. Second, we used homology to previously generated gene annotation for *V. dahliae* strain JR2 (Faino et al. 2015). This second annotation was performed exploiting the previously generated ProgressiveCactus alignment, using the Comparative-Annotation-Toolkit (v2.2.1) (default settings, except for 'global-near-best'=1 and 'augustus-species'=verticillium_dahliae1) to transfer the annotation between *V. dahliae* strains (Fiddes et al. 2018). Genes predicted by the first approach yet without homology to *V. dahliae* strain JR2 genes were considered lineage-specific in CQ2 (*n*=347). Similarly, genes in JR2 without homologous genes in CQ2, were considered lineage-specific in JR2 (*n*=362).

Functional gene annotation was performed independently for *V. dahliae* JR2 and CQ2. First, gene ontology term (GO term) annotation was performed by the parallel

execution of multiple GO terms annotation tools, an approach we considered optimal as it exploits the complementary information provided by each tool. Briefly, we used Interproscan (v.5.55_88.0 ; --goterms –iprlookup, default database) (Jones et al. 2014), Goanna (v.2.00; default settings, Agbase-Uniprot database) (McCarthy et al. 2007), blastp (v. 2.12.0; --max_target_seqs 1, uniprot fungi database) (Camacho et al. 2009), PANNZER2 (v. 0.3; --m 'Pannzer', default settings and database) (Medlar et al. 2018), ggnog (v.5.0; e-values³0.001, query coverage ³ 50%) (Huerta-Cepas et al. 2019) to retrieve the Goterms associated with the gene annotation of JR2 and CQ2. Additionally, for lineage-specific genes we predicted corresponding protein structures using Alphafold2 (v.2.2.0; -m 'monomer', -t 2020-01-01) (Jumper et al. 2021), retained the structures with pLDDT>50, and then we used DeepFRI (v. 1.2.0; default settings) (Gligorijević et al. 2021) to retrieve Goterms with scores ³ 0.15 based on protein structure. The results of these annotation tools were merged and filtered for spurious Goterms for further analysis. To define the *V. dahliae* secretome, we predicted N-terminal signal peptides using SignalP v.4.1 (Nielsen 2017b). We removed predicted proteins with transmembrane domains using DeepTMHMM (v.1.0.20; default settings) (Hallgren et al. 2022). We predicted putative carbohydrate active enzymes (CAZy) using dbCAN2 covering signatures of glycoside hydrolases, polysaccharide lyases, carbohydrate esterases, glycosil transferases, and carbohydrate-binding molecules (default settings, retained if predicted by >1 tool) (Zhang et al. 2018); and predicted the putative effectors with EffectorP v.3.0 (Sperschneider and Dodds 2022). Furthermore, putative antimicrobial proteins (AMPs) were predicted by the parallel execution of three AMPs prediction tools. We used AMPEPpy (version 1.0; --seed 2012, default settings) (Lawrence et al. 2021), AmpGram (v.1.1; default settings) (Burdukiewicz et al. 2020) and AMP Scanner (v.2.0; default settings, model 'Feb2020') (Veltri et al. 2018) for AMP prediction, then retained AMPs with a probability score >0.5 and <200 amino acids length (Gómez-Pérez et al. 2022). We used the consensus of the three AMP prediction tools for further analyses. We considered the groups of 'other secreted proteins' as the secreted proteomes excluding candidate effectors, AMPs and CAZy; 'effectors' as the secreted proteomes including candidate AMPs and excluding CAZy; and 'CAZy' as the secreted proteome excluding 'effectors' and 'other secreted proteins' (Supplementary Fig. 11).

De novo repeat annotation was performed for *V. dahliae* strain JR2 and CQ2 for consistency. First, we identified and masked repeats using RepeatModeler2 and RepeatMasker (Flynn et al. 2020; Tempel 2012). Then, we combined homology-based, structural-based, and *de novo* approaches to annotate transposable elements (TEs) using the Extensive *de novo* TE annotator (EDTA) pipeline (Ou et al. 2019). To increase robustness and to be consistent with previous analysis of TEs in *Verticillium dahliae*, we used a manually curated library of TEs in *Verticillium dahliae* as reference and to lift-over TE names (Amyotte et al. 2012; Faino et al. 2016; Torres et al. 2021). We performed EDTA using 'sensitive', 'anno', 'evaluate' and 'curatedlib' options, together with the previously generated gene annotation to maximize TE discovery, integrity, and to avoid false positives. Finally, genome-

wide occurrences of the resulting repeat families were determined using RepeatMasker (v4.0.9) (Tempel 2012), and the output was post-processed using 'one code to find them all' to annotate full-length LTRs (Bailly-Bechet et al. 2014).

RNA sequencing and analysis

Total RNA extraction was performed *in vitro* and *in planta*. For *in vitro* RNA extraction, *Verticillium dahliae* strains JR2 and CQ2 were grown on potato dextrose agar for 14 days at 22ºC in the dark. Then, conidiospores were collected using sterile water. The spore suspension was filtered over two layers of Miracloth into 50 mL tubes on ice, and subsequently centrifuged at 3620xg for 15 min at 4ºC. The supernatant was removed and pellet resuspended in 4-6 mL of water, then centrifuged at 3620xg for 15min at 4ºC. Then, the supernatant was removed and the pellet was frozen in liquid nitrogen and stored at -80ºC. Total RNA extraction was carried out in triplicate using TRIzol (Thermo Fisher Scientific, Massachusets, USA) following the manufacturer's guidelines. Following RNA re-suspension, contaminating DNA was removed using the TURBO DNA-free kit (Ambion, ThermoFisher Scientific, Massachusets, USA) and RNA integrity was estimated by separating 2 mL of each sample on a 2% agarose gel and quantified using Nanodrop (ThermoFisher Scientific, California, USA). Sequencing library preparation and sequencing was carried out at BGI (Beijing Genomics Institute, Hong Kong, China) where RNA-sequencing libraries were constructed following end-repair and adapter ligation protocols and PCR amplification. Purified RNA fragments were paired-end sequenced (100 bp) on the DNBseq platform.

For *in planta* gene expression analyses, we used previously generated *Verticillium dahliae* JR2 data (*Arabidopsis thaliana* 28 dpi) (Cook et al. 2020). Additionally, we here generated *in planta* transcriptome data for *V. dahliae* CQ2. For *in planta* inoculations, *V. dahliae* CQ2 was plated on PDA and harvested after 10 days by pipetting 2 mL of sterile water onto the plate and releasing conidiospores by gentle agitation with a spreader. After washing twice, the spore concentration was determined using a Buerker-Tuerk counting chamber and brought to a final concentration of 5×10^6 conidiospores/mL in sterile water. The inoculation of 14-day-old *A. thaliana* Col-0 plants grown on potting soil was performed via root dipping. Individual *A. thaliana* plants were up-rooted carefully and the soil was removed from the roots as much as possible by gentle washing in water. Next, roots were placed in a Petri dish with the appropriate conidiospore suspension, or water as a control, while the upper plant parts remained outside the Petri dish. A total of six plants were used. After 10 minutes, the plants were taken from the Petri dishes and transferred to individual pots. After 15 days post inoculation, the above-ground parts of the plants were harvested without the leaves and flash frozen in liquid nitrogen. Next, in order to gain sufficient starting material, two plants were pooled, resulting in three replicates with two plants for *V. dahliae* CQ2. The material was ground to a fine powder using mortar and pestle, before performing RNA extractions with the Maxwell RSC Plant RNA Kit (Promega, Fitchburg, WI, USA) according to manufacturer's guidelines. RNA concentrations were measured with the

Qubit RNA BR Assay-Kit (Thermo Fisher Scientific, Waltham, MA, USA). Sequencing library preparation and sequencing was carried out at the Cologne Center for Genomics (Cologne, Germany) where RNA-sequencing libraries were constructed following end-repair and adapter ligation protocols and PCR amplification, and then sequenced on the NovaSeq 6000 platform.

Sequencing reads were mapped to the *V. dahliae* strain JR2 or CQ2 genome assemblies (Faino et al. 2015; Depotter et al. 2019), respectively, using STAR v.2.6.0 (Dobin et al. 2013), with settings –sjdbGTFfeatureExon exon, --alignIntronMax 400, --outFilterMismatchNmax 100, --outFilterMismatchNmax 3, --winAnchorMultimapNmax 200, –outSAMtype BAM Unsorted (Jin et al. 2015; Jin and Hammell 2018). To visualize gene expression, RNA-seq mapping data were transformed using deeptools bamCoverage, using CPM, --exactScaling -centerReads, and window size of 50kb. Finally, for *the V. dahliae* JR2 and CQ2 gene sets, RNA counts were extracted using htseq-count (Putri et al. 2022) with default settings, then counts were normalized within each strain using the R/Bioconductor package EdgeR v.3.8 (Robinson et al. 2010). For normalization, we considered only genes with >1 reads in all samples (Anders et al. 2013) and genes with read count =0 were assumed to be transcriptional inactive. Libraries were normalized with the TMM method (Robinson and Oshlack 2010), and converted to CPM-mapped reads using the R/Bioconductor package EdgeR v.3.8 (Robinson et al. 2010).

To investigate gene expression conservation between the two *V. dahliae* strains, we used the mean expression across both strains as a reference to mitigate possible confounding effects on the expression values. Thus, we considered gene expression values (CPM) for homologous gene pairs by using the median of ratio to the geometric means between the biological triplicates of the two strains to further calculations. This approach reduced the possible confounding effect due to differences in transcript level distributions in each *V. dahliae* strain. Next, we calculated the coefficient of interquartile variation (CQV) of each homolog gene as a measure of divergence by using the formula: CQV=(Q3-Q1)/(Q3+Q1) (Bonett 2006), where Q1 and Q3 refers to the 1st quartile and 3^{d} quartile of the CQV distribution, respectively. Expressed homolog genes (>1 CPM) over 1st-3^{d} CQV quartile were considered as genes with conserved expression patterns, whereas genes in the 4th CQV quartile were considered to be differentially expressed between strains (Supplementary Fig. 2).

Chromatin immunoprecipitation, sequencing, and analysis

Chromatin immunoprecipitation was performed as described previously, but using conidiospores harvested from PDA-grown *Verticillium dahliae* JR2 and CQ2 (Kramer et al. 2022; Cook et al. 2020). Approximately 200 mg of ground material was resuspended in 4 mL ChIP Lysis buffer (50 mM HEPES-KOH pH7.5, 140 mM NaCl, 1 mM EDTA, 1% Triton X-100, 0.1% NaDOC) and grinded in a 10 cm^3 glass tube with tight fitting pestle in a RZR50 homogenizer (Heidolph, Schwabach, Germany), followed by five rounds of 20 seconds sonication on ice with 40 seconds rest between rounds with a Soniprep 150 (MSE, London, UK). Samples were

redistributed to 2 mL tubes and pelleted by centrifugation for 2 min at maximum speed. The supernatants were combined, together with 5 µL 1M CaCl2 and 0.63 µL Mnase, and after 10 minutes of incubation in a 37°C water bath with regular manual shaking, 20 µL 0.5M EGTA was added, and tubes were put on ice. Samples were pre-cleared by adding 20 µL Protein A Magnetic Beads (New England Biolabs, MA, United States) and rotating at 4°C for 60 min, after which the beads were captured, 1 mL fractions of supernatant were transferred to new 2 mL tubes containing 5 µL of H3K9ac, H3K4me3, H3K4ac, H3K36me3, H4K16ac, H3K9me3 or H3K27me3 antibody (Abcam: ab10812 [H3K9ac], ab9050 [H3K36me3]; EMD Millipore: #07-329 [H4K16ac], #07-539 [H3K4ac], #04-745 [H3K4me3]; ActiveMotif: #39765 [H3K9me3], #39155 [H3K27me3]) respectively and incubated overnight with continuous rotation at 4ºC. One sample without antibody was kept as Mnase control. Subsequently, 20 µL protein-A magnetic beads were added and incubated for 3 hours at 4ºC, after which the beads were captured on a magnetic stand and subsequently washed with 1 mL wash buffer (50 mM Tris HCl pH 8, 1 mM EDTA, 1% Triton X-100, 100 mM NaCl), high-salt wash buffer (50 mM Tris HCl pH 8, 1 mM EDTA, 1% Triton X-100, 350 mM NaCl), LiCl wash buffer (10 mM Tris HCl pH8, 1 mM EDTA, 0.5% Triton X-100, 250 mM LiCl), TE buffer (10 mM Tris HCl pH 8, 1mM EDTA). Nucleosomes were eluted twice from beads by addition of 100µL pre-heated TES buffer (100 mM Tris HCl pH 8, 1% SDS, 10 mM EDTA, 50 mM NaCl) and 10 minutes incubation at 65ºC. 10 mg/mL 2µL Proteinase K (10 mg/mL) was added and incubated at 65ºC for 3 hours, followed by chloroform clean-up. DNA was precipitated by addition of 2 volumes 100% ethanol, 1/10th volume 3 M NaOAc pH 5.2 and 1/200th volume 120 mg/mL glycogen, and incubated overnight at -20ºC. Two ChIP replicates were performed for each antibody from independently grown samples. Sequencing libraries were prepared using the TruSeq ChIP Library Preparation Kit (Illumina, City, Country) following the manufacturer's guidelines, but without gel purification and with use of the Velocity DNA Polymerase (BioLine, Luckenwalde, Germany) for 25 cycles of amplification. Single-end (75 bp) sequencing was performed on the Illumina NetSeq500 platform at the Utrecht Sequencing Facility (Utrecht, the Netherlands).

Raw reads were trimmed using TrimGalore (v.0.6.9; default settings) (Krueger 2015) and mapped to the *V. dahliae* either strain JR2 or CQ2 reference genome (Faino et al. 2015; Depotter et al. 2019) using BWA -MEM v.0.7 with default options (Li and Durbin 2010). For the JR2 genome, three regions of the genome were masked due to aberrant mapping, (chr1:1-45000, chr2:3466000-3475000, chr3:1-4200). Library artifacts were marked and removed using Picard tools v.2.18 with -MarkDuplicates (http://broadinstitute.github.io/picard/). Mapped reads were CPM-normalized using deepTools v.3.1.3 bamCompare (Ramírez et al. 2014) with Mnase as input control, --scaleFactorsMethod 'none', --smoothLength 30, --binSize 10 –pseudocount 25 –extendedreads 100 –operation 'ratio' for further processing and –binSize 100 and –smoothLength 300 for visualization over the genome. Normalized output was transformed to bedgraph using UCSC-bigWigToBedGraph. We used deepTools

v.3.1.3 multiBigwigSummary (Ramírez et al. 2014) to determine the correlation and clustering of replicate ChIP samples (Supplementary Fig. 4).

Chromatin states characterization

To define chromatin states or combinations of H3K9ac, H3K4me3, H3K4ac, H3K36me3, H4K16ac, H3K27me3, an H3K9me3 that co-occur at discrete genomic locations in *Verticillium dahliae* JR2 and CQ2, we defined enrichment peaks for each ChIP replicate using HOMER v.4.11 makeTagDirectory and findPeaks –size 500 –minDist 0 –style 'histone' (Heinz et al. 2010). Additionally, we obtained the consensus peaks for each histone modification per strain by keeping peaks present in the two replicates. Next, we used a multivariate Hidden Markov Model to predict the significant combinations of histone modifications (Ernst and Kellis 2012, 2017). Replicates and consensus peaks of the seven histone modifications were transformed into a binary matrix using chromHMM v.1.4 Binarizebed –center (Ernst and Kellis 2012, 2017). Subsequently, we used chromHMM v.1.4 LearnModel (default settings) (Ernst and Kellis 2012, 2017) for segmentation and predict up to 30 possible chromatin states based on the seven histone modifications in the replicates and the consensus for each strain. We obtained the emission probabilities (probability that a histone mark is present in a named chromatin state) and the transition probabilities (probability that a named chromatin state is similar to other chromatin states) for each of the 30 states. We further compare and aimed to determine the accurate combination of chromatin states (Gorkin et al. 2020). First, for each replicate and the consensus peaks we used the chromHMM v.1.4. CompareModels (default settings) to calculate the Pearson's correlation of the emission probabilities of the 30th model to other 29 models. We obtained saturation at 20 chromatin states in each replicate and the consensus peaks for both strains (Supplementary Fig. 8, 9). To validate the optimal number of states we clustered the emission probabilities for the two replicates in each strain. We performed *k*-means 2-30, and look at the separation for each *k*, as the ratio of the between *k* sum squares and the total *k* sum of squares under 10,000 iterations (Supplementary Fig. 8, 9). To validate the reproducibility of the 20 states in each strain, we clustered the emission probabilities in the 20 states of the two replicates (distance=Euclidean, method=complete) (Supplementary Fig. 8, 9). As we observed similar chromatin states in each replicate and consensus, we decided to continue with the consensus dataset for further analyses (Supplementary Fig. 8, 9). To determine the presence of similar chromatin states between *Verticillium dahliae* JR2 and CQ2, we followed the same approach and clustered the emission probabilities in the 20 consensus chromatin states of strain JR2 and CQ2 (distance=Euclidean, method=complete) (Supplementary Fig. 9). Additionally, we performed principal component analysis (PCA) using the emission probabilities in the 20 states (Supplementary Fig. 9) and calculated the Jaccard's coefficient to determine similarity for each state. Relative enrichment of different genomic features over the 20 chromatin states was performed using chromHMM v.1.4 OverlapEnrichment (default settings).

Genomic variation dataset

Genomic variation (duplications, deletions, inversions and translocations; collectively SVs), single nucleotide variants (SNVs) and polymorphic TEs were previously identified using paired-end sequencing data for 42 previously sequenced *V. dahliae* strains (Torres et al. 2021). Briefly, structural variants (SVs) were predicted using the 'sv-callers' workflow with few modifications that enabled parallel execution of multiple SV callers, an approach that is considered optimal as it exploits complementary information to predict (Kuzniar et al. 2020). Single nucleotide variants were identified using the -HaplotypeCaller of the Genome Analysis Toolkit (GATK) v.4.0 (McKenna et al. 2010). The nucleotide diversity (π) was performed using a sliding window of 1 kb (500 bp sliding) as implemented in the PopGenome package (Pfeifer et al. 2014) in R v.3.6.3 (Team 2013). Transposable element polymorphisms was analyzed using TEPID v.2.0 (Stuart et al. 2016). To investigate if SVs and polymorphic TEs co-localize at specific chromatin states, we summarized the overlap of each set of variants by their breakpoint frequency (start or ends ±1 bp of the feature) and coverage (number of bases covered) across the genome of *V. dahliae* strain JR2 (Fudenberg and Pollard 2019). Similar to Fudenberg and Pollard (2019), we calculated the log10(observed/expected) of each feature to estimate the deviation from a uniform distribution across the genome, therefore accounting for the proportion of the genome covered by a specific genomic feature.

Statistical analysis and visualization

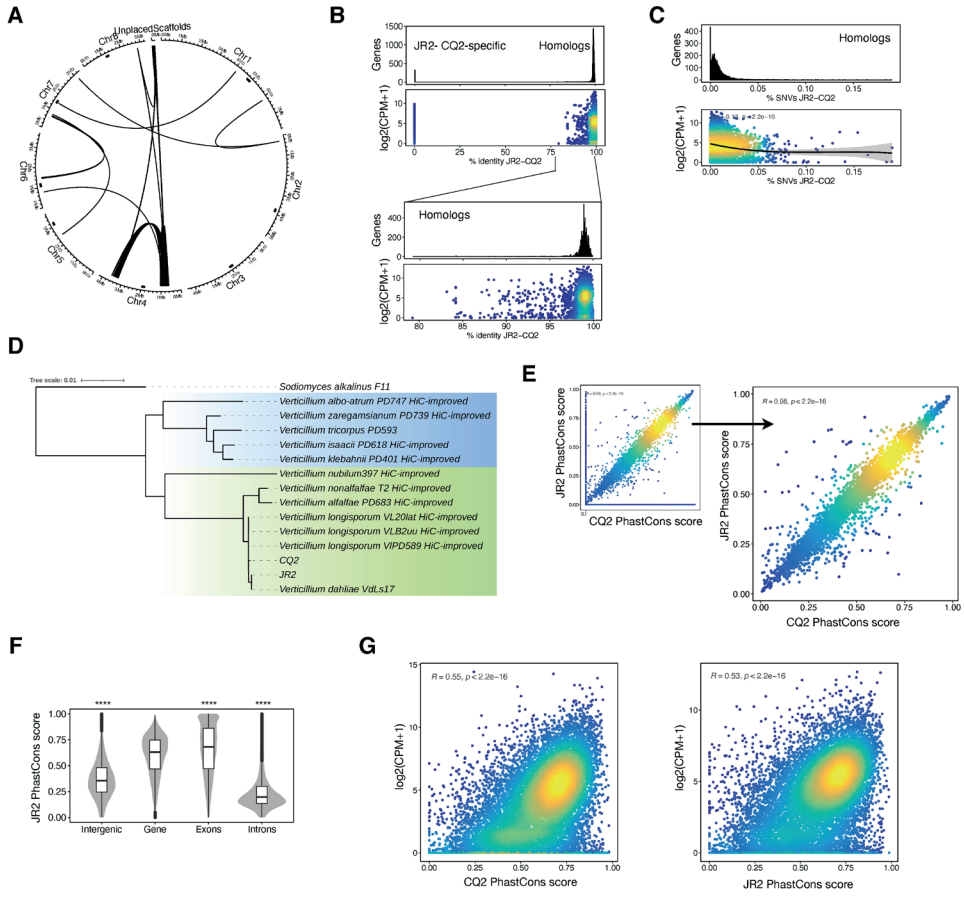
All statistical analyses and comparison tests were performed in R v.4.0.5 (Team 2013), and data visualized using circlize 0.4.15, gggenomes 0.9.5.9, ggalluvial 0.12.4, and ggplot2 3.3.6 (Gu et al. 2014; Wickham 2011). Principle component analyses were performed using FactoMineR v.1.42 and factoextra v.1.0.5 (Lê et al. 2008), and we used log2(H3K9ac), log2(H3K4me3), log2(H3K4ac), log2(H3K36me3), log2(H4K16ac), log2(H3K27me3), log2(H3K9me3) and *in vitro* (log2(CPM+1), PDA) expression over non-overlapping 10 kb windows as variables for *V. dahliae* strain CQ2 and JR2, respectively. Sparse multiple regression models were performed using the SparseM v.1.81 package (Koenker and Ng 2003) for *V. dahliae* strain CQ2 and JR2 independently. We performed multiple regression models with the variables in the following order: log2(H3K27me3+1), log2(H3K9me3+1), log2(H3K4ac+1), log2(H3K4me3+1), log2(H4K16ac+1), log2(H3K36me3+1), log2(H3K9ac+1), sequence conservation (PhastCons score), expression conservation (1-log2(CQV+1)) and/or expression (log2(CPM+1)). For the sparse multiple regression R, R^2 and Bonferroni corrected *p-value* were reported (Supplementary Table 4). Heatmap and enrichment visualization of histone modifications, PhastCons score, and SNVs, over genes or CREs were performed using the R/EnrichedHeatmap v1.2 package (Gu et al. 2018). Permutation tests were computed using R/Bioconductor regioneR v1.18.1 package (Gel et al. 2016) and performed with 10,000 iterations, using findOverlaps or meanDistance function and circular randomization to maintain the order and distance of the regions in the chromosomes. Functional Goterms

enrichment was performed using clusterProfiler v.3.18.1 (Yu et al. 2012), and then summarized the redundant GO terms with REVIGO (default settings) (Supek et al. 2011).

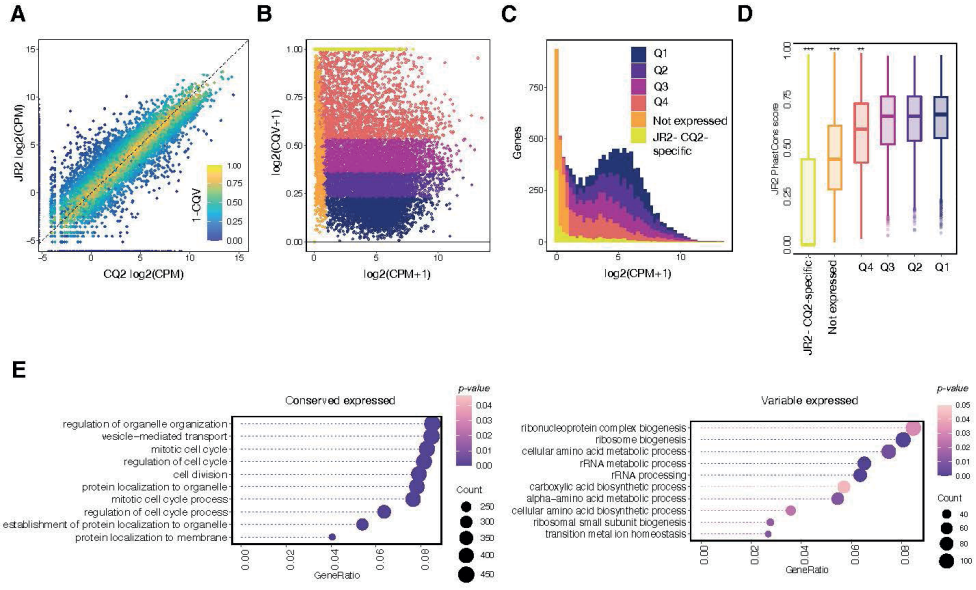
Acknowledgments

D.E.T and E.C-C acknowledge their PhD fellowship from the Consejo Nacional de Ciencia y Tecnología from México. G.L.F. acknowledges a PhD fellowship from the Coordination for the Improvement of Higher Education Personnel (CAPES) from the federal government from Brazil. Work in the laboratory of M.F.S is supported by the Research Council Earth and Life Sciences (ALW) of the Netherlands Organization of Scientific Research (I). B.P.H.J.T acknowledges funding by the Alexander von Humboldt Foundation in the framework of an Alexander von Humboldt Professorship endowed by the German Federal Ministry of Education and Research is furthermore supported by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) under Germany's Excellence Strategy – EXC 2048/1 Project ID: 390686111. The authors declare no conflict of interest exists.

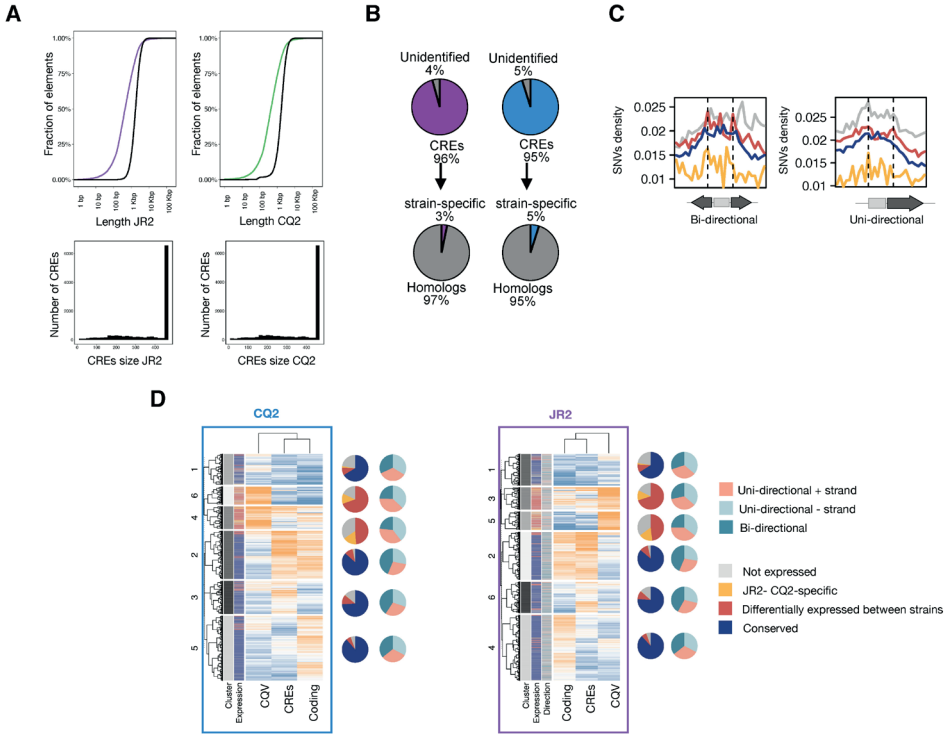
Supplementary data



Supplementary Figure 1. Sequence and expression conservation in *Verticillium dahliae*.
(A) Segmental duplications co-localize with adaptive genomic regions (AGRs) in *Verticillium dahliae* strain CQ2. **(B)** Genome comparisons between *V. dahliae* strains JR2 and CQ2 reveal homologous genes and strain-specific genes (0% similarity, left). The bottom two panels highlight homologous genes and show that highly similar genes are highly expressed. **(C)** Homologous genes are highly expressed and have a reduced number of single nucleotide variants (SNVs, from CQ2-JR2 comparison). The black line in the bottom panel depicts the Spearman's correlation (R= -0.13, *p-value* <0.05). **(D)** Phylogeny of the ten species in the *Verticillium* genus, with *Sodiomyces alkalinus* used to root the phylogeny. Flavexudans and Flavnonexudans clades are depicted in blue and green, respectively. The phylogenetic tree was used to guide the calculation of sequence conservation with PhastCons. **(E)** Sequence conservation (based on PhastCons) for *V. dahliae* strains CQ2 and JR2 is correlated. On the left, PhastCons scores for all genes in JR2 and CQ2 reveal many genes with score of 0 (347 and 367 for CQ2 and JR2, respectively), which represent JR2- or CQ2-specific genes. The color gradient shows the increased density around the mean sequence conservation. **(F)** Exons have higher conservation scores. Statistical significance was assessed using a one-sided Wilcoxon rank-sum test *p£0.05, **p£0.01, ***p£0.001, ****p£0.0001, when compared with 'Gene'. **(G)** Sequence conservation correlates with gene expression levels in *V. dahliae* strains CQ2 and JR2. R and *p-value* (<0.05) after Spearman's correlation.

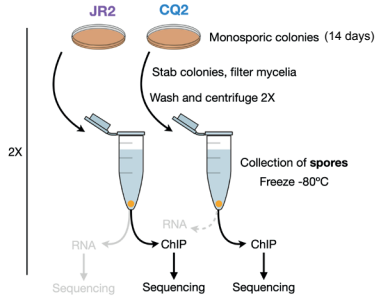


Supplementary Figure 2. Gene expression variation in *Verticillium dahliae*.
(**A)** The coefficient of interquartile variation is low (yellow) for similarly expressed homologous genes in *V. dahliae* strains JR2 and CQ2. **(B)** The upper quartile (Q4) of the coefficient of interquartile variation depicts the highest gene expression variation between *V. dahliae* strains JR2 and CQ2, i.e., these genes are differentially expressed between the two strains. Color code as in **(C). (C)** The majority of genes from quartile 1-3 (Q1-Q3) are highly expressed. Non-expressed genes (light orange) are defined as homologous genes with CPM<1. JR2- and CQ2-specific genes (yellow) are excluded from the coefficient of interquartile variation calculation. **(D)** The upper quartile (Q4) of the coefficient of interquartile variation shows less sequence conservation across the *Verticillium* genus (PhastCons score). Statistical significance was assessed using a one-sided Wilcoxon rank-sum test *p£0.05, **p£0.01, ***p£0.001, ****p£0.0001, when compared with the first quartile (Q1). **(E)** GO-term enrichment for genes with conserved and differential expression patterns between the two *V. dahliae* strains (left and right, respectively). The ten most significant GO-terms are shown. The color gradient depicts the *p-value* after Bonferroni correction and the dot size indicates the number of genes for each GO-term.



Supplementary Figure 3. *Cis*-regulatory regions in *Verticillium dahliae.*
(A) Distribution of intergenic length (black) when compared with the gene length in *V. dahliae* strain JR2 (purple) and CQ2 (green). Most intergenic distances between two genes in both *V. dahliae* strains are shorter than 450 bp. The lower panels depict the length distribution of the <450 bp and >10 bp '5 regions upstream of genes in CQ2 (*n*=10,958) and JR2 (*n*=10,941). **(B)** Regions of <450 bp and >10 bp '5 upstream of genes occur in >95% of genes in *V. dahliae* JR2 (purple) and CQ2 (blue). 'Unidentified' regions are the remaining regions that are <10 bp in size. **(C)** *Cis*-regulatory regions (CREs) are enriched in single nucleotide variants (SNVs). We used SNV data based on a collection of 42 *V. dahliae* strains with *V. dahliae* JR2 as reference (Torres et al., 2021). **(D)** Hierarchical clustering of genes based in their expression conservation (CQV) as well as sequence conservation (PhastCons score) in CREs and coding sequences. Six clusters are observed in both strains (numbers on the left of blue box, CQ2 and purple box, JR2, respectively). On the right of each heatmap (blue box, CQ2; purple box, JR2), the first pie plot column depicts per cluster the proportion of genes grouped by expression classification (see Fig. 2; Fig. S2). The second pie plot column depicts per cluster the proportion of genes grouped by the direction of the CREs.

A



B

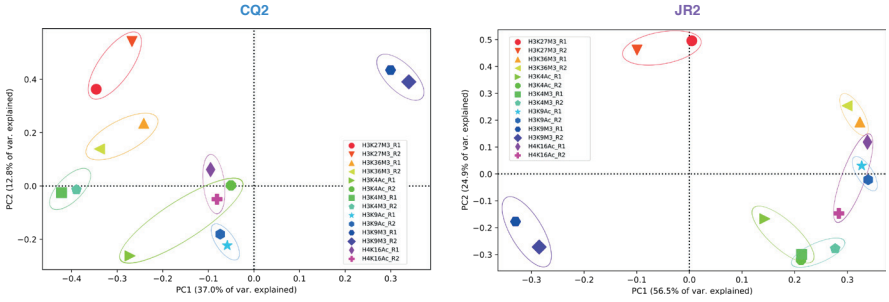
Mapped reads JR2

Sample	Total	Mappe	%
H4K16Ac_R2	10,162,465	9,782,199	96.3%
H4K16Ac_R1	10,189,904	9,867,224	96.8%
H3K9M3_R2	712,607	666,402	93.5%
H3K9M3_R1	8,881,532	8,006,836	90.2%
H3K9Ac_R2	9,570,565	9,344,179	97.6%
H3K9Ac_R1	24,673,390	23,730,345	96.2%
H3K4M3_R2	1,242,766	1,185,354	95.4%
H3K4M3_R1	7,653,246	6,698,200	87.5%
H3K4Ac_R2	1,472,989	1,383,467	93.9%
H3K4Ac_R1	1,912,008	1,630,677	85.3%
H3K36M3_R2	53,032,913	47,268,173	89.1%
H3K36M3_R1	18,102,912	17,601,891	97.2%
H3K27M3_R2	10,373,527	9,982,868	96.2%
H3K27M3_R1	10,344,727	8,725,581	84.3%

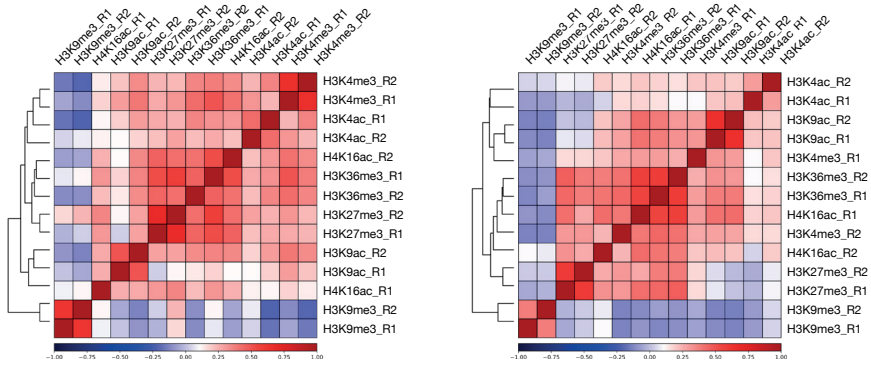
Mapped reads CQ2

Sample	Total	Mapped	%
H4K16Ac_R2	9,009,861	8,659,455	96.1%
H4K16Ac_R1	18,008,276	17,815,644	98.9%
H3K9M3_R2	3,170,274	3,023,635	95.4%
H3K9M3_R1	8,725,830	8,029,984	92.0%
H3K9Ac_R2	14,662,517	14,493,225	98.8%
H3K9Ac_R1	80,063,764	78,711,944	98.3%
H3K4M3_R2	7,570,492	7,391,179	97.6%
H3K4M3_R1	2,426,905	2,301,729	94.8%
H3K4Ac_R2	13,178,505	13,005,308	98.7%
H3K4Ac_R1	74,407,709	70,096,177	94.2%
H3K36M3_R2	45,403,733	43,678,974	96.2%
H3K36M3_R1	123,993,856	122,109,312	98.5%
H3K27M3_R2	14,786,203	14,201,775	96.0%
H3K27M3_R1	12,793,041	12,420,561	97.1%

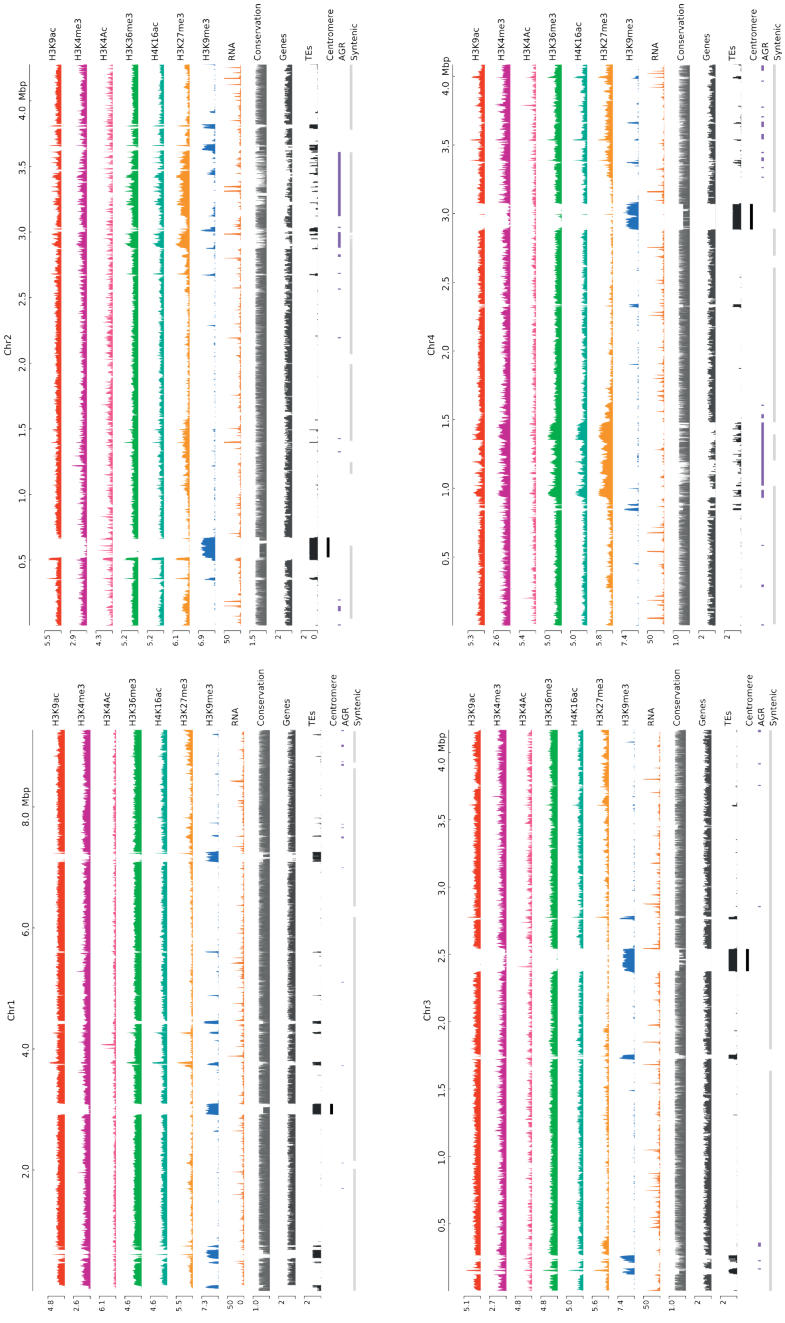
C

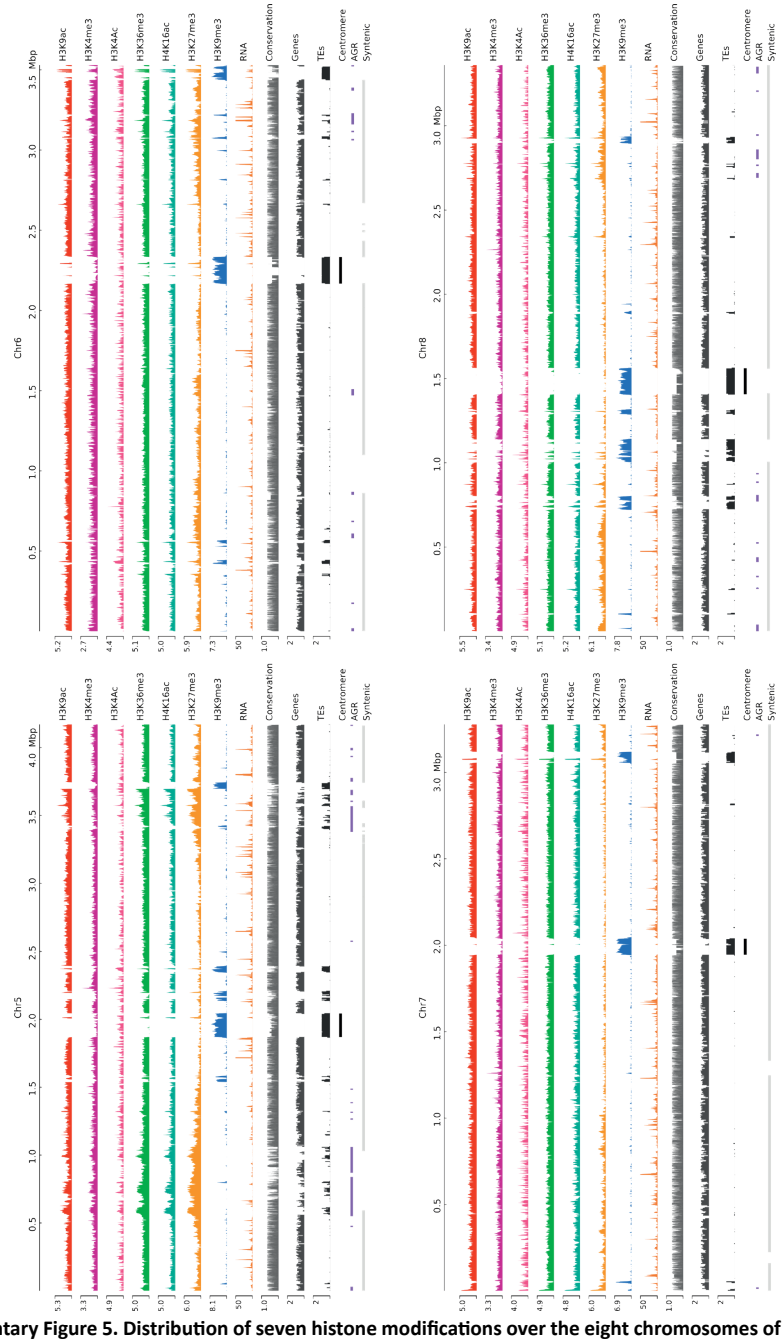


D



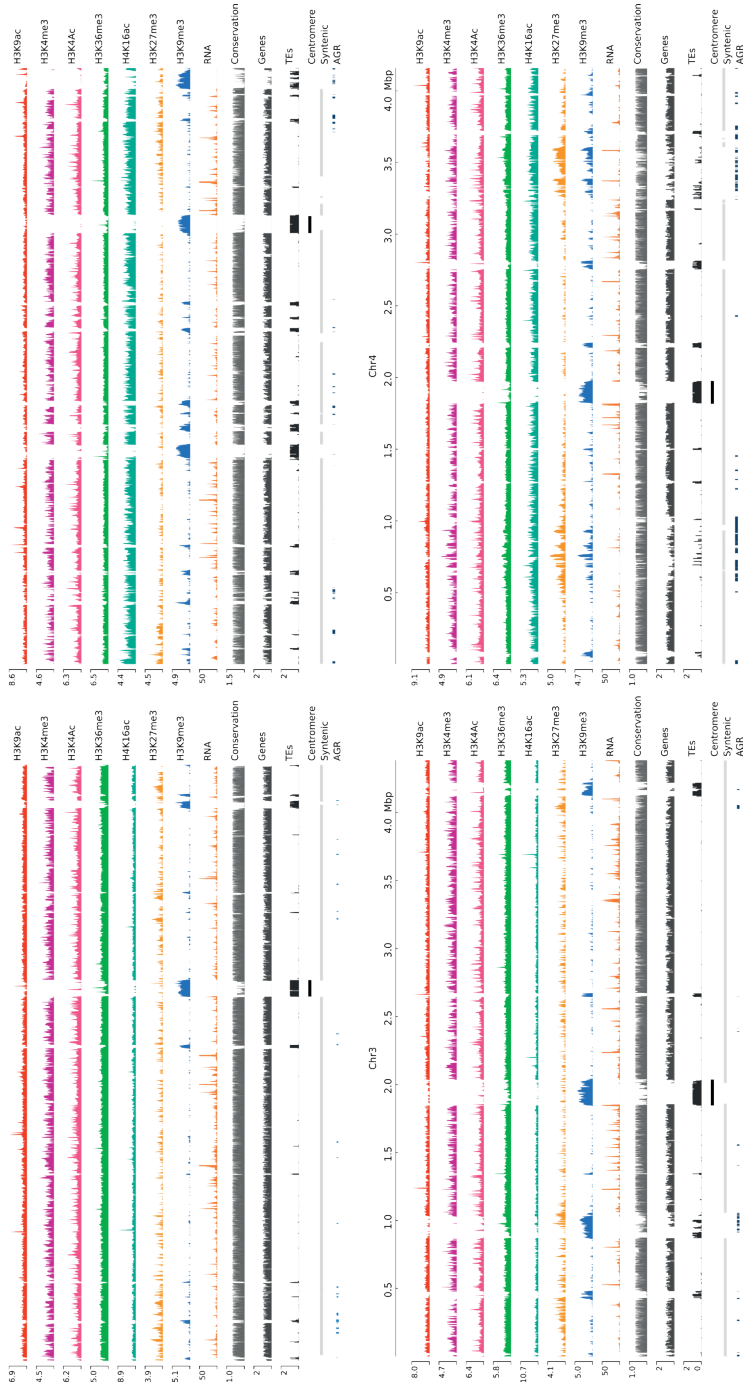
Supplementary Figure 4. Histone modifications in *Verticillium dahliae* JR2 and CQ2.
(A) Experimental approach for conidiospore isolation. Conidiospores were collected after 14 days for fungus grown on potato dextrose agar, washed, and frozen for ChIP-seq and RNA-seq analyses. (**B)** Most of the reads for each replicate are successfully mapped to the reference genomes. **(C)** Principal component analysis of histone modifications per replicate (raw read counts summarized in 10 kb windows) shows the clustering of biological replicates for each histone modification. **(D)** Heatmap display the Spearman's correlation between biological replicates of histone modifications (raw read counts summarized in 10 kb windows).

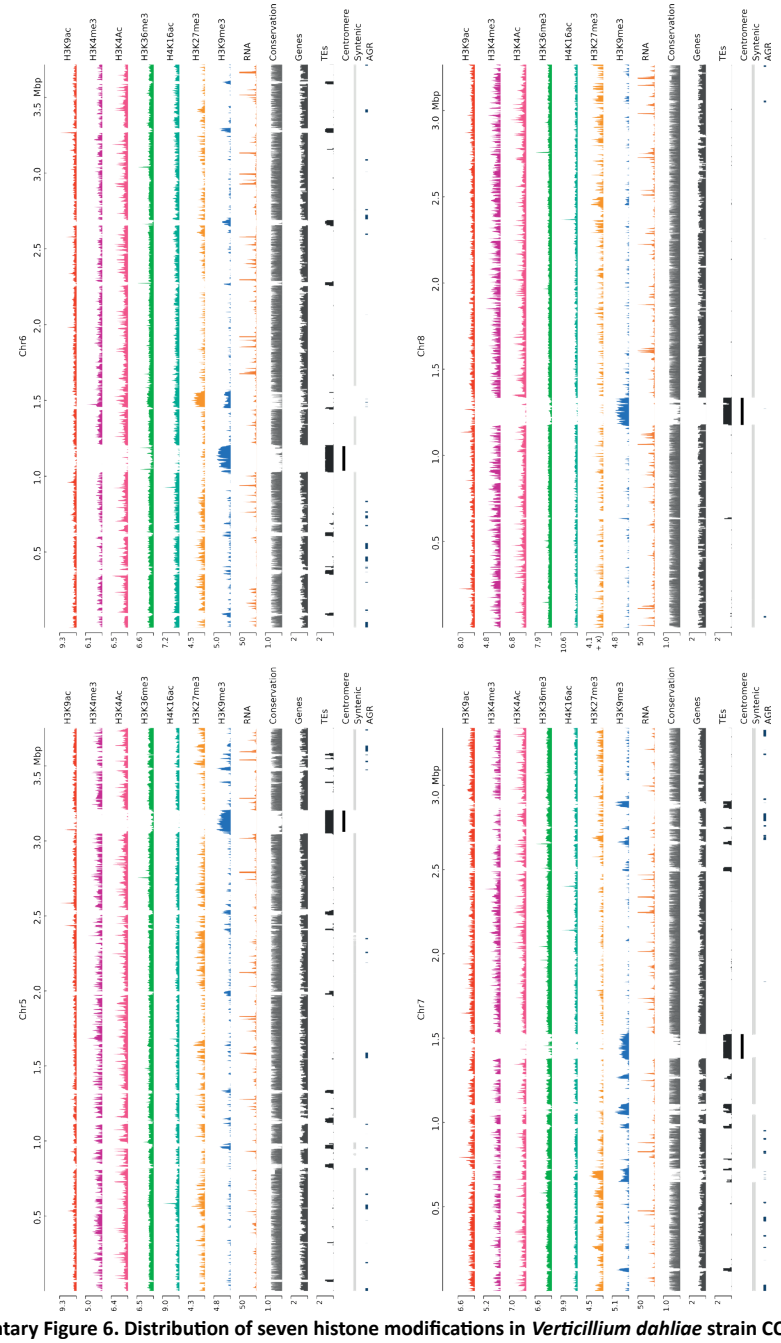

Chr1
Chr2
Chr3
Chr4
Mbp
H3K9ac
H3K4me3
H3K4Ac
H3K36me3
H4K16ac
H3K27me3
H3K9me3
RNA
Conservation
Genes
TEs
Centromere
AGR
Syntenic



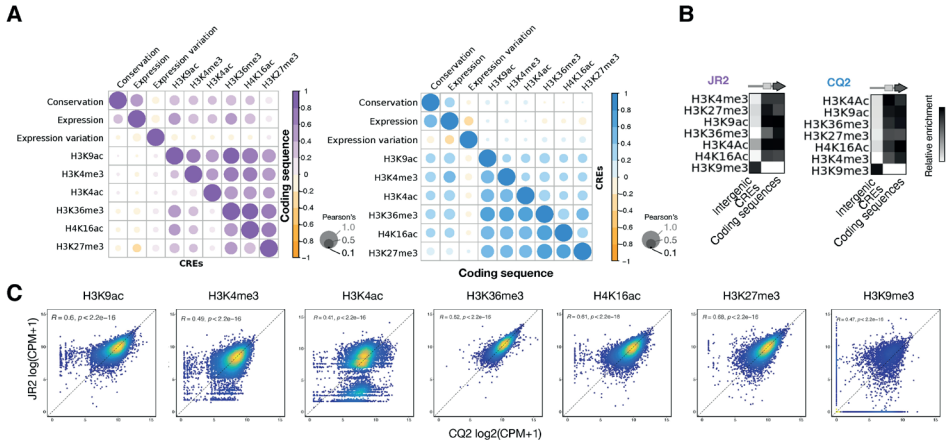
Supplementary Figure 5. Distribution of seven histone modifications over the eight chromosomes of *Verticillium dahliae* strain JR2

From left to right: histone modifications H3K9ac, H3K4me3, H3K4ac, H3K36me3, H4K16ac, H3K27me3, and H3K9me3 normalized using a micrococcal nuclease digestion control, transcriptomic data from *in vitro* RNA-seq on conidiospores harvested from *V. dahliae* strain JR2 grown on PDA, sequence conservation (PhastCons score), gene and transposable element (TE) densities in 10 kb non-overlapping windows, centromeres (Seidl et al. 2020), adaptive genomic regions (AGRs) (Cook et al. 2020), and syntenic regions between *V. dahliae* strains JR2 and CQ2. Chromosome sizes are displayed on the left.

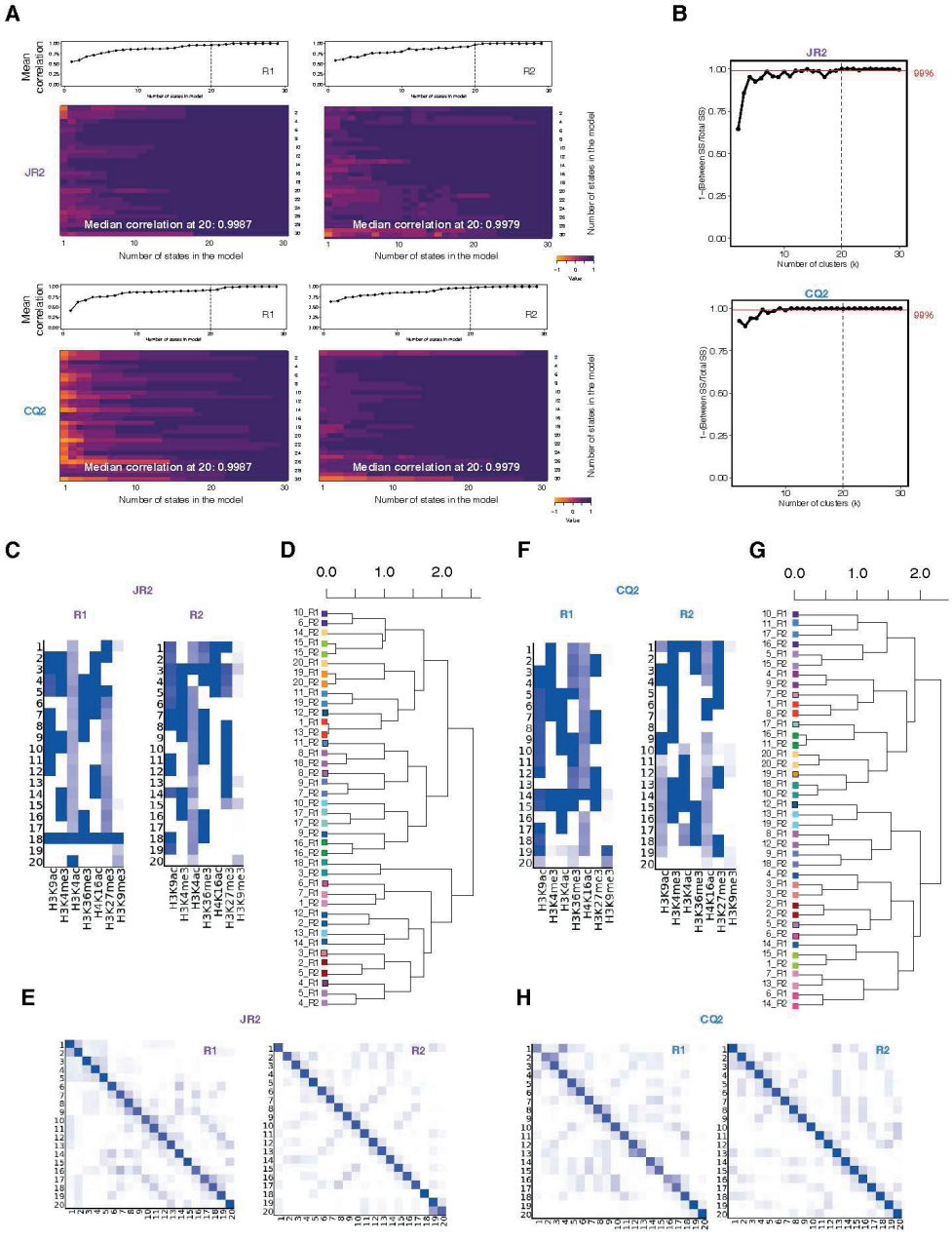

Chr1
Chr2
Chr3
Chr4
Mbp
H3K9ac
H3K4me3
H3K4Ac
H3K36me3
H4K16ac
H3K27me3
H3K9me3
RNA
Conservation
Genes
TEs
Centromere
Syntenic
AGR



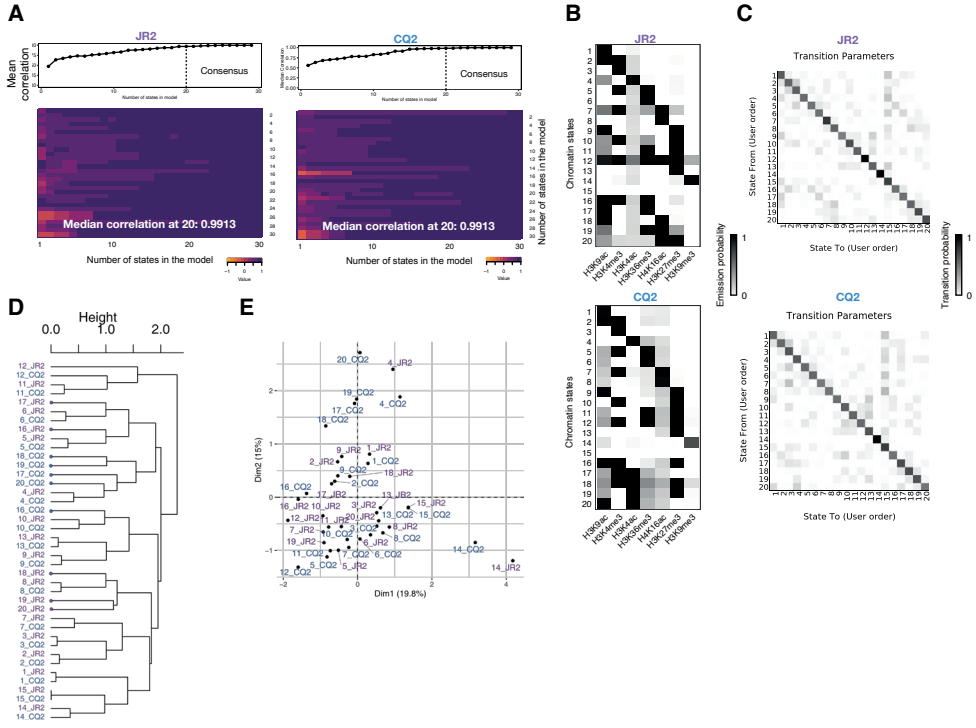
Supplementary Figure 6. Distribution of seven histone modifications in *Verticillium dahliae* strain CQ2.
From left to right: histone modifications H3K9ac, H3K4me3, H3K4ac, H3K36me3, H4K16ac, H3K27me3, and H3K9me3 normalized using a micrococcal nuclease digestion control, transcriptomic data from *in vitro* RNA-seq on conidiospores harvested from *V. dahliae* strain CQ2 grown on PDA, sequence conservation (PhastCons score), gene and transposable element (TE) densities in 10 kb non-overlapping windows, centromeres (Seidl et al. 2020), syntenic regions between *V. dahliae* strains CQ2 and JR2, and adaptive genomic regions (AGRs) (Cook et al. 2020) mapped to the CQ2 genome. Chromosome sizes are displayed on the left.



Supplementary Figure 7. Histone modifications in the coding sequences and CREs of *Verticillium dahliae*.
(A) Correlations between seven histone modifications, gene expression, expression conservation (CQV), and sequence conservation (PhastCons score). Histone modifications, gene expression (CPM+1), gene expression variation (CQV), and conservation (PhastCons) are summarized over coding sequences or CREs. The purple and blue gradient depict the direction of the correlation in *V. dahliae* strain JR2 and CQ2, respectively. The circle size shows the absolute strength of the correlation. **(B)** Relative enrichment of histone modifications at intergenic (without considering the CREs), *cis*-regulatory regions (CREs), and coding sequences in *V. dahliae* strains JR2 and CQ2. **(C)** Histone modification levels correlate for homologous genes in *V. dahliae* strains JR2 and CQ2. R and *p-value* after Spearman's correlation test.

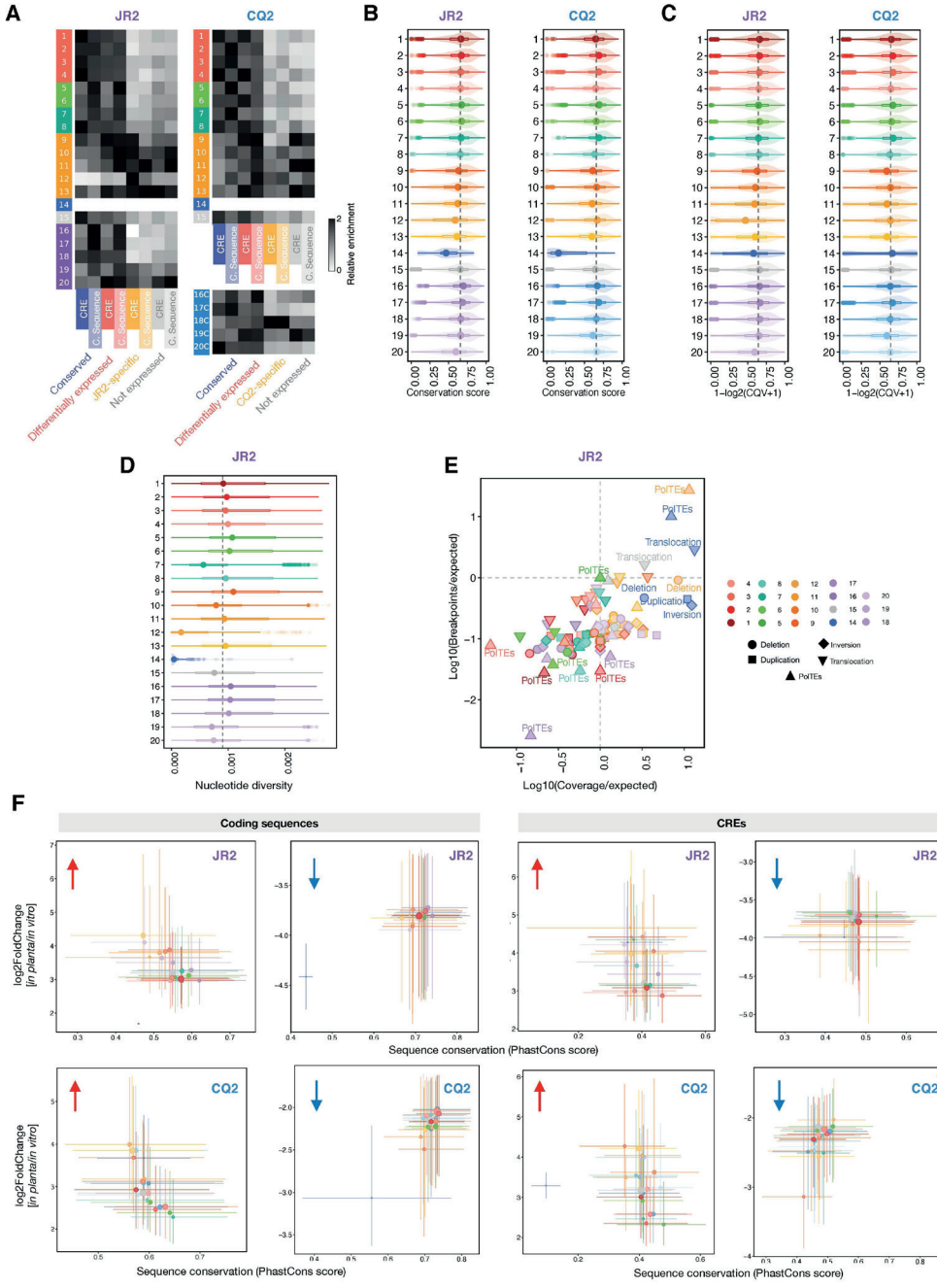

A
B
C
D
E
F
G
H
JR2
CQ2
R1
R2
Median correlation at 20: 0.9987
Median correlation at 20: 0.9979
Number of states in the model
Mean correlation
1−(Between SS/Total SS)
Number of clusters (k)
H3K9ac
H3K4me3
H3K4ac
H3K36me3
H4K16ac
H3K27me3
H3K9me3

Supplementary Figure 8. Based on seven histone modifications, twenty chromatin states are predicted in *Verticillium dahliae* strains JR2 and CQ2.
(A) We defined different combinations of histone modifications, so called chromatin states, with the use of a multivariate Hidden Markov Model. To define the optimal number of chromatin states, we varied the number of states between 2 to 30 (maximum tested model). Cumulative Pearson's correlation of the emission probabilities (the probability of each histone modification to belong to a chromatin state) against the maximum tested model with 30 states using the significant peaks of the histone modifications H3K9ac, H3K4me3, H3K4ac, H3K36me3, H4K16ac, H3K27me3, H3K9me3, for each replicate, is shown. Significant peaks, i.e., regions with more sequencing reads than expected by chance, were defined by Homer (Meers et al., 2019). The optimum number of states is reached when the median correlation for a given state reaches a plateau (dashed line), which is at 20 chromatin states. The heatmap depicts the raw correlation for each combination of states. Each analysis was performed for replicate 1 (R1) and replicate 2 (R2), and for *V. dahliae* strain JR2 (purple, top) and CQ2 (blue, bottom), respectively. **(B)** The optimal number of chromatin states is 20, based on the method of Gorkin et al., 2020. To validate the optimal number of states, we clustered the emission probabilities (the probability of each histone modification to belong to a named chromatin state) for the two replicates for each strain. Then, we performed *k*-means clustering with *k* ranging from 2 to 30 and considered the separation for each *k* as the ratio of the between sum squares and the total sum of squares based in 10,000 iterations. Perfectly separated clusters will have a ratio of 1. We considered the optimal number of clusters as those that would be shown at least at >=99.0% of the maximum ratio (red line). The dashed vertical line indicates 20 clusters as the optimal number of chromatin states. **(C)** The 20 chromatin states and the associated histone modifications are shown for each biological replicate (R1, R2) for *V. dahliae* strain JR2. The heatmap depicts the emission probabilities for each chromatin state and each histone modification. **(D)** Twenty chromatin states in *V. dahliae* strain JR2 are consistently found in replicate R1 and R2. Hierarchical clustering of the emission probabilities of the 20 chromatin states predicted for each replicate. Identical color boxes depict chromatin states consistently found for both replicates. **(E)** Transition probabilities from each chromatin state to each other of the 20 chromatin states in *V. dahliae* JR2 reveal high stability of chromatin states for each biological replicate (R1, R2). Dark blue color indicates a maximum transition probability of 1. **(F)** Chromatin states that were predicted in two biological replicates (R1, R2) of the *V. dahliae* strain CQ2. Similar as in **(C)**. **(G)** Twenty chromatin states in *V. dahliae* strain CQ2 are consistently found for R1 and R2. Similar as in **(D)**. **(H)** Transition probabilities of 20 chromatin states in *V. dahliae* strain CQ2 reveal high stability of chromatin states. Dar blue color indicates a maximum transition probability of 1, similar as in **(E).**



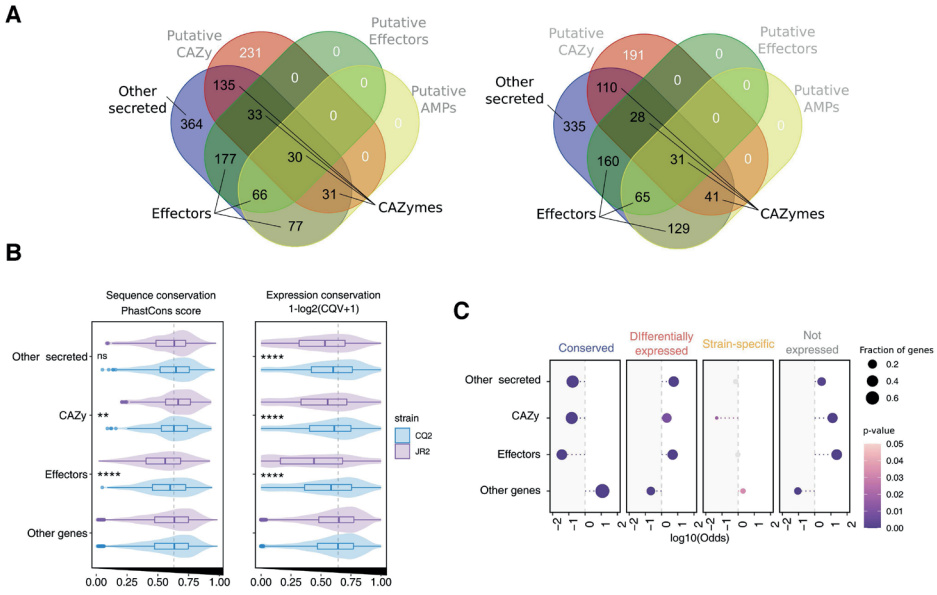
Supplementary Figure 9. Based on seven histone modifications, twenty chromatin states are predicted in *Verticillium dahliae* strains JR2 and CQ2,

(A) Twenty chromatin states were predicted based on the consensus histone modification peaks, i.e., peaks present in both replicates, for the histone modifications H3K9ac, H3K4me3, H3K4ac, H3K36me3, H4K16ac, H3K27me3, and H3K9me3 in *V. dahliae* strains JR2 and CQ2 . Histone modification peaks found in only one replicate were discarded. Similar to Figure S8, the optimal number of chromatin states is 20. Upper and lower panels as in Figure S8A**.** These 20 chromatin states were used for further analyses. **(B)** Emission probabilities for each state (the probability of each histone modification to belong to a named chromatin state) depict the combinations of seven histone modifications present at each state for *V. dahliae* strain JR2 (upper) and CQ2 (lower), respectively. **(C)** Transition probabilities for 20 chromatin states based on the consensus histone modification peaks. Dark black color indicates a maximum transition probability of 1. Low transition probability between states suggest that the 20 chromatin states are stable. **(D)** Hierarchical clustering of the chromatin states identified in *V. dahliae* strains CQ2 and JR2 based on the emission probabilities for each state. Clustering supports the presence of shared chromatin states (edged without dots) and strain-specific chromatin states (edges with dots). **(E)** Principal component analysis of chromatin states identified in *V. dahliae* strain JR2 and CQ2 based on the emission probabilities for each state.


A
JR2
CQ2
Relative enrichment
Conserved
Differentially expressed
JR2-specific
CQ2-specific
Not expressed
B
Conservation score
C
1−log2(CQV+1)
D
Nucleotide diversity
E
Log10(Breakpoints/expected)
Log10(Coverage/expected)
PolTEs
Translocation
Deletion
Duplication
Inversion
F
Coding sequences
CREs
log2FoldChange [in planta/in vitro]
Sequence conservation (PhastCons score)

<<

Supplementary Figure 10. Chromatin states correlate with gene expression and sequence conservation.
(A) Relative enrichment of 20 chromatin states at coding regions and their associated CREs classified based on their expression and sequence variation for *V. dahliae* strains JR2 and CQ2. **(B)** Sequence conservation of coding regions associated to each of the 20 chromatin states in *V. dahliae* strains JR2 (left) and CQ2 (right). The dashed line depicts the global mean value of sequence conservation. **(C)** Expression conservation of genes associated with the 20 chromatin states in *V. dahliae* strain JR2 (left) and CQ2 (right). The dashed line depicts the global mean value of expression conservation. **(D)** Nucleotide diversity in the 20 chromatin states in *V. dahliae* strain JR2. The nucleotide diversity was calculated based on a collection of 42 *V. dahliae* strains when compared with the genome of *V. dahliae* strain JR2 (Torres et al., 2021). The dashed line depicts the global mean value for nucleotide diversity. **(E)** H3K27me3-associated chromatin states are enriched in structural variants (SVs), as more SVs overlap (x-axis) and SVs breakpoints (y-axis) are present than expected by chance; the calculation is based on the approach by Fudenberg and Pollard (2019). **(F)** Differential gene expression comparing *in planta* and *in vitro* (potato dextrose agar) conditions for *V. dahliae* strains JR2 (top; 21 dpi *Arabidopsis thaliana*) and CQ2 (bottom; 15 dpi *Arabidopsis thaliana*). Differential gene expression *in planta* is higher in H3K27me3-associated chromatin states (states 9-13). The twenty chromatin states (points) at coding sequences or *cis*-regulatory regions (CREs) are shown based on their mean values for sequence conservation and gene expression change between *in planta* and *in vitro* conditions. The point size depicts the total number of genes associated with the specific chromatin state, and the line indicates the minimum and maximum values. Red and blue arrows highlight plots with genes that are differentially up- or down-regulated, respectively. The color code is similar as for **(B-E)**.



Supplementary Figure 11. Chromatin states correlate with the expression and sequence conservation of pathogenicity-related genes.

(A) We defined pathogenicity-related genes as effectors, CAZy (carbohydrate active enzymes), and other secreted proteins. For further analyses, we considered the groups of 'other secreted proteins' as the secreted proteome excluding putative effectors, AMPs, and CAZy; 'effectors' as the secreted proteome including putative AMPs and excluding CAZy; and 'CAZy' as the secreted proteome excluding 'effectors' and 'other secreted proteins'. The Venn diagram shows the overlap of the different gene categories for *V. dahliae* strains JR2 (left) and CQ2 (right). **(B)** Sequence conservation (left) and gene expression conservation (right) of genes that belong to pathogenicity-related functional categories. The dashed line depicts the global mean value. Data for *V. Dahliae* strains JR2 and CQ2 are shown in purple and blue, respectively. Statistical significance was assessed using a one-sided Wilcoxon rank-sum test *p£0.05, **p£0.01, ***p£0.001, ****p£0.0001, when compared with 'Other genes'. **(C)** Differentially expressed genes between *V. dahliae* strains JR2 and CQ2 are enriched for pathogenicity-related genes. Significance was assessed by a one-sided Fisher's test with Bonferroni correction for the *p-value*. The circle size depicts the fraction of genes in each category.

Supplementary Table 1. Overview of the Verticillium collection used in this study

Species	Isolate	Reference	bioProject
Verticillium dahliae	89	Chavarro-Carrero et al., 2021	PRJNA639910
Verticillium dahliae	463	Li, 2018	PRJNA576642
Verticillium dahliae	12008	Fan et al., 2018	PRJNA344737
Verticillium dahliae	12158	Fan et al., 2018	PRJNA352681
Verticillium dahliae	12161	Fan et al., 2018	PRJNA352681
Verticillium dahliae	12251	Fan et al., 2018	PRJNA352681
Verticillium dahliae	12253	Fan et al., 2018	PRJNA352681
Verticillium dahliae	VdLs17	Faino et al., 2015	PRJNA276625
Verticillium dahliae	2009-605	Li, 2018	PRJNA576642
Verticillium dahliae	85S	Depotter et al., 2018	PRLJ00000000
Verticillium dahliae	BP2	Zhang et al., 2012	PRJNA693498
Verticillium dahliae	CBS38166	de Jonge et al., 2012	PRJNA169154
Verticillium dahliae	CBS385.49	Chavarro-Carrero et al., 2021	PRJNA639910
Verticillium dahliae	CBS809,97	Chavarro-Carrero et al., 2021	PRJNA639910
Verticillium dahliae	cd3	Xu et al., 2012	PRJNA693498
Verticillium dahliae	CFA3	Chavarro-Carrero et al., 2021	PRJNA639910
Verticillium dahliae	CQ2	Depotter et al., 2018	PRLI00000000
Verticillium dahliae	DVD-S26	de Jonge et al., 2012	PRJNA169154
Verticillium dahliae	DVD-S29	de Jonge et al., 2012	PRJNA169154
Verticillium dahliae	DVD-S94	de Jonge et al., 2012	PRJNA169154
Verticillium dahliae	DVD161	de Jonge et al., 2012	PRJNA169154
Verticillium dahliae	DVD3	de Jonge et al., 2012	PRJNA169154
Verticillium dahliae	DVD31	de Jonge et al., 2012	PRJNA169154
Verticillium dahliae	EU35	Chavarro-Carrero et al., 2021	PRJNA639910
Verticillium dahliae	EU7	Chavarro-Carrero et al., 2021	PRJNA639910
Verticillium dahliae	GF1192	Chavarro-Carrero et al., 2021	PRJNA639910
Verticillium dahliae	GF1207	Chavarro-Carrero et al., 2021	PRJNA639910
Verticillium dahliae	GFCA2	Chavarro-Carrero et al., 2021	PRJNA639910
Verticillium dahliae	GFCB5	Chavarro-Carrero et al., 2021	PRJNA639910
Verticillium dahliae	HN	Xu et al., 2012	PRJNA693498
Verticillium dahliae	HNA4	Chavarro-Carrero et al., 2021	PRJNA639910
Verticillium dahliae	HOMCF	Chavarro-Carrero et al., 2021	PRJNA639910
Verticillium dahliae	I1V	Chavarro-Carrero et al., 2021	PRJNA639910
Verticillium dahliae	JKG8	Kombrink et al., 2017	PRJNA576642
Verticillium dahliae	JR2	Faino et al., 2015	PRJNA169154
Verticillium dahliae	Sick1-1	Unpublished	NA
Verticillium dahliae	ST100	de Jonge et al., 2012	PRJNA169154

Species	Isolate	Reference	bioProject
Verticillium dahliae	ST14.01	de Jonge et al., 2012	PRJNA169154
Verticillium dahliae	ST16.01	Li, 2018	PRJNA576642
Verticillium dahliae	T9	Keykhasaber, 2017	PRJNA693498
Verticillium dahliae	TM6	Keykhasaber, 2017	PRJNA693498
Verticillium dahliae	TO22	Chavarro-Carrero et al., 2021	PRJNA639910
Verticillium dahliae	UD141	Chavarro-Carrero et al., 2021	PRJNA639910
Verticillium dahliae	V08SY-2	Chavarro-Carrero et al., 2021	PRJNA639910
Verticillium dahliae	V1	Unpublished	NA
Verticillium dahliae	V117	Keykhasaber, 2017	PRJNA693498
Verticillium dahliae	V13	Unpublished	NA
Verticillium dahliae	v152	Kombrink et al., 2017	PRJNA576642
Verticillium dahliae	v4	Keykhasaber, 2017	PRJNA693498
Verticillium dahliae	V574	Milgroom et al., 2014	PRJNA693498
Verticillium dahliae	V700	Milgroom et al., 2014	PRJNA693498
Verticillium dahliae	V991	Zhang et al., 2012	PRJNA302216
Verticillium dahliae	VanDijk	Kombrink et al., 2017	PRJNA576642
Verticillium dahliae	Vd39	Li, 2018	PRJNA576642
Verticillium dahliae	vd52	Kombrink et al., 2017	PRJNA576642
Verticillium dahliae	Vd57	Li, 2018	PRJNA576642
Verticillium dahliae	Vd653	Unpublished	NA
Verticillium dahliae	VdH5	Unpublished	NA
Verticillium dahliae	VdR1	Unpublished	NA
Verticillium dahliae	Vmx	Unpublished	NA
Verticillium dahliae	VT2A	Chavarro-Carrero et al., 2021	PRJNA639910
Verticillium dahliae	WFA3	Chavarro-Carrero et al., 2021	PRJNA639910
Verticillium dahliae	XJ-12	Chavarro-Carrero et al., 2021	PRJNA639910
Verticillium dahliae	XJ-14	Chavarro-Carrero et al., 2021	PRJNA639910
Verticillium albo-atrum	PD747	Seidl et al., 2020	PRJNA641329
Verticillium alfalfae	PD683	Seidl et al., 2020	PRJNA641329
Verticillium isaacii	PD618	Seidl et al., 2020	PRJNA641329
Verticillium klebhanii	PD401	Seidl et al., 2020	PRJNA641329
Verticillium longisporum	PD589	Seidl et al., 2020	PRJNA641329
Verticillium longisporum	VLB2	Seidl et al., 2020	PRJNA641329
Verticillium nonalfalfae	T2	Seidl et al., 2020	PRJNA641329
Verticillium nubilum	397	Seidl et al., 2020	PRJNA641329
Verticillium tricorpus	PD593	Seidl et al., 2020	PRJNA641329
Verticillium zaregamsianum	PD739	Seidl et al., 2020	PRJNA641329
Sodiomyces alkalinus	F11	Grum-Grzhimaylo et al. 2018	PRJNA196044

Supplementary Table 2. Differential enrichment of histone modifications is associated with gene expression in the *Verticillium dahliae* strains JR2 and CQ2.

			Total number of	Total number of	One-sided Fisher's test	
Expression category	**Histone modification**	**Genetic element**	**differential enriched genes**	**similarly enriched genes**	***p-value****	**Odds value**
Stable	H3K9ac	Coding sequence	479	6800	1.00E+00	0.726
Variable	H3K9ac	Coding sequence	199	1809	7.01E-02	1.134
Strain-specific	H3K9ac	Coding sequence	0	0	1.00E+00	0.000
Not expressed	H3K9ac	Coding sequence	262	1077	2.20E-16	2.506
Stable	H3K4me3	Coding sequence	1150	6129	1.00E+00	0.831
Variable	H3K4me3	Coding sequence	370	1638	5.11E-01	1.000
Strain-specific	H3K4me3	Coding sequence	0	0	1.00E+00	0.000
Not expressed	H3K4me3	Coding sequence	438	901	2.20E-16	2.152
Stable	H3K4ac	Coding sequence	2579	4700	1.00E+00	0.875
Variable	H3K4ac	Coding sequence	797	1211	1.74E-01	1.049
Strain-specific	H3K4ac	Coding sequence	0	0	1.00E+00	0.000
Not expressed	H3K4ac	Coding sequence	720	619	2.20E-16	1.854
Stable	H3K36me3	Coding sequence	49	7230	9.97E-01	0.636
Variable	H3K36me3	Coding sequence	26	1982	2.00E-01	1.231
Strain-specific	H3K36me3	Coding sequence	0	0	1.00E+00	0.000
Not expressed	H3K36me3	Coding sequence	37	1302	2.01E-06	2.667
Stable	H3K16ac	Coding sequence	468	6811	1.00E+00	0.804
Variable	H3K16ac	Coding sequence	167	1841	2.65E-01	1.061
Strain-specific	H3K16ac	Coding sequence	0	0	1.00E+00	0.000
Not expressed	H3K16ac	Coding sequence	202	1137	2.20E-16	2.078
Stable	H3K27me3	Coding sequence	337	6942	1.00E+00	0.792
Variable	H3K27me3	Coding sequence	126	1882	2.06E-01	1.092
Strain-specific	H3K27me3	Coding sequence	0	0	1.00E+00	0.000

Supplementary Table 2. Continued

Expression category	Histone modification	Genetic element	Differential enriched genes	Similar enriched genes	*p-value**	Odds value
Not expressed	H3K27me3	Coding sequence	151	1188	6.14E-13	2.072
Stable	H3K9me3	Coding sequence	587	6692	3.72E-02	1.109
Variable	H3K9me3	Coding sequence	150	1858	2.56E-03	1.316
Strain-specific	H3K9me3	Coding sequence	0	0	1.00E+00	0.000
Not expressed	H3K9me3	Coding sequence	42	1297	1.00E+00	0.528
Stable	H3K9ac	CRE	3123	4138	4.57E-01	1.004
Variable	H3K9ac	CRE	852	1145	5.93E-01	0.990
Strain-specific	H3K9ac	CRE	0	0	1.00E+00	0.000
Not expressed	H3K9ac	CRE	570	762	5.46E-01	0.995
Stable	H3K4me3	CRE	4237	3024	6.09E-02	1.049
Variable	H3K4me3	CRE	1098	899	9.67E-01	0.915
Strain-specific	H3K4me3	CRE	0	0	1.00E+00	0.000
Not expressed	H3K4me3	CRE	720	612	9.86E-01	0.881
Stable	H3K4ac	CRE	4802	2459	6.51E-01	0.988
Variable	H3K4ac	CRE	1357	640	9.29E-02	1.073
Strain-specific	H3K4ac	CRE	0	0	1.00E+00	0.000
Not expressed	H3K4ac	CRE	873	459	7.45E-01	0.962
Stable	H3K36me3	CRE	943	6318	1.20E-01	1.056
Variable	H3K36me3	CRE	223	1774	9.41E-01	0.890
Strain-specific	H3K36me3	CRE	0	0	1.00E+00	0.000
Not expressed	H3K36me3	CRE	145	1187	9.48E-01	0.865
Stable	H3K16ac	CRE	2486	4775	2.93E-01	1.018
Variable	H3K16ac	CRE	677	1320	4.86E-01	1.003
Strain-specific	H3K16ac	CRE	0	0	1.00E+00	0.000

Not expressed	H3K16ac	CRE	420	912	9.57E-01	0.901
Stable	H3K27me3	CRE	2396	4865	3.65E-02	1.061
Variable	H3K27me3	CRE	619	1378	7.43E-01	0.967
Strain-specific	H3K27me3	CRE	0	0	1.00E+00	0.000
Not expressed	H3K27me3	CRE	343	989	1.00E+00	0.747

* H3K9me3 is not calculated at CREs since it is absent

CRE stands for *cis*-regulatory region

Supplementary Table 3. Sparse multiple regression analysis based on seven histone modifications and expression conservation and levels in *Verticillium dahliae*.

Variable	Predictor	Estimate in full model	*p-value* full model	Variance Inflation	Genetic element	Strain
Conservation	H3K9ac	0.006	6.43E-07	1.8708	Coding seq.	CQ2
Conservation	H3K4me3	0.007	9.95E-09	1.9	Coding seq.	CQ2
Conservation	H3K4ac	0.002	5.19E-02	1.8178	Coding seq.	CQ2
Conservation	H3K36me3	0.002	3.79E-01	3.7571	Coding seq.	CQ2
Conservation	H4K16ac	0.010	3.61E-10	2.196	Coding seq.	CQ2
Conservation	H3K27me3	0.006	1.27E-03	2.3035	Coding seq.	CQ2
Conservation	H3K9me3	-0.002	1.08E-05	1.8899	Coding seq.	CQ2
Conservation	Expression	0.030	2.00E-16	1.3343	Coding seq.	CQ2
Conservation	Stability	0.043	2.60E-10	1.0953	Coding seq.	CQ2
Expression	H3K9ac	0.142	4.44E-16	1.8638	Coding seq.	CQ2
Expression	H3K4me3	0.284	2.00E-16	1.8581	Coding seq.	CQ2
Expression	H3K4ac	0.141	2.00E-16	1.807	Coding seq.	CQ2
Expression	H3K36me3	0.280	4.87E-13	3.7396	Coding seq.	CQ2
Expression	H4K16ac	0.049	1.98E-02	2.2028	Coding seq.	CQ2
Expression	H3K27me3	-0.434	2.00E-16	2.2347	Coding seq.	CQ2
Expression	H3K9me3	0.028	2.61E-04	1.8909	Coding seq.	CQ2
Expression	Conservation	5.664	2.00E-16	1.1906	Coding seq.	CQ2
Expression	Stability	1.409	2.00E-16	1.0765	Coding seq.	CQ2
Exp. Stability	H3K9ac	-0.010	4.24E-09	1.8707	Coding seq.	CQ2
Exp. Stability	H3K4me3	-0.010	2.82E-08	1.9008	Coding seq.	CQ2
Exp. Stability	H3K4ac	-0.006	1.88E-04	1.8167	Coding seq.	CQ2
Exp. Stability	H3K36me3	-0.010	1.24E-02	3.7551	Coding seq.	CQ2
Exp. Stability	H4K16ac	-0.001	5.00E-01	2.2038	Coding seq.	CQ2
Exp. Stability	H3K27me3	0.032	2.00E-16	2.2689	Coding seq.	CQ2
Exp. Stability	H3K9me3	0.001	4.35E-01	1.8932	Coding seq.	CQ2
Exp. Stability	Conservation	-0.084	1.68E-10	1.4291	Coding seq.	CQ2

Exp. Stability	Expression	-0.016	2.00E-16	1.574	Coding seq.	CQ2
Conservation	H3K9ac	0.001	7.02E-01	3.6959	Coding seq.	JR2
Conservation	H3K4me3	0.007	5.21E-08	1.8181	Coding seq.	JR2
Conservation	H3K4ac	0.004	5.25E-08	1.3609	Coding seq.	JR2
Conservation	H3K36me3	0.038	2.00E-16	5.6689	Coding seq.	JR2
Conservation	H4K16ac	-0.015	2.51E-08	3.4948	Coding seq.	JR2
Conservation	H3K27me3	-0.008	4.33E-05	2.3461	Coding seq.	JR2
Conservation	H3K9me3	0.000	9.64E-01	1.3837	Coding seq.	JR2
Conservation	Expression	0.031	2.00E-16	1.3558	Coding seq.	JR2
Conservation	Stability	0.054	4.44E-15	1.0773	Coding seq.	JR2
Expression	H3K9ac	0.465	2.00E-16	3.6327	Coding seq.	JR2
Expression	H3K4me3	0.040	1.13E-02	1.8219	Coding seq.	JR2
Expression	H3K4ac	0.013	1.21E-01	1.3642	Coding seq.	JR2
Expression	H3K36me3	0.613	2.00E-16	5.6309	Coding seq.	JR2
Expression	H4K16ac	0.090	4.35E-03	3.502	Coding seq.	JR2
Expression	H3K27me3	-0.522	2.00E-16	2.2469	Coding seq.	JR2
Expression	H3K9me3	0.003	5.30E-01	1.3836	Coding seq.	JR2
Expression	Conservation	4.488	2.00E-16	1.1766	Coding seq.	JR2
Expression	Stability	1.003	2.00E-16	1.0692	Coding seq.	JR2
Exp. Stability	H3K9ac	-0.027	1.22E-12	3.6825	Coding seq.	JR2
Exp. Stability	H3K4me3	0.001	6.55E-01	1.8229	Coding seq.	JR2
Exp. Stability	H3K4ac	-0.001	5.57E-01	1.3645	Coding seq.	JR2
Exp. Stability	H3K36me3	-0.010	6.15E-02	5.715	Coding seq.	JR2
Exp. Stability	H4K16ac	-0.002	5.04E-01	3.5046	Coding seq.	JR2
Exp. Stability	H3K27me3	0.040	2.00E-16	2.3101	Coding seq.	JR2
Exp. Stability	H3K9me3	-0.001	2.47E-01	1.3835	Coding seq.	JR2
Exp. Stability	Conservation	-0.105	4.44E-16	1.3577	Coding seq.	JR2
Exp. Stability	Expression	-0.014	2.00E-16	1.5527	Coding seq.	JR2

Supplementary Table 3. Continued

Variable	Predictor	Estimate in full model	*p-value* full model	Variance Inflation	Type	Strain
Conservation	H3K9ac	0.003	1.85E-04	1.1736	CRE	CQ2
Conservation	H3K4me3	0.001	3.44E-01	1.1563	CRE	CQ2
Conservation	H3K4ac	0.003	3.75E-03	1.2088	CRE	CQ2
Conservation	H3K36me3	-0.005	2.53E-03	1.7596	CRE	CQ2
Conservation	H4K16ac	0.000	8.36E-01	1.2728	CRE	CQ2
Conservation	H3K27me3	0.006	2.73E-09	1.4661	CRE	CQ2
Conservation	H3K9me3	-0.005	1.46E-12	1.1823	CRE	CQ2
Conservation	Expression	0.019	2.00E-16	1.113	CRE	CQ2
Conservation	Stability	-0.010	1.93E-01	1.0785	CRE	CQ2
Expression	H3K9ac	0.000	7.65E-01	1.1728	CRE	CQ2
Expression	H3K4me3	0.001	3.53E-01	1.1517	CRE	CQ2
Expression	H3K4ac	-0.001	2.61E-01	1.2084	CRE	CQ2
Expression	H3K36me3	0.002	2.10E-01	1.7609	CRE	CQ2
Expression	H4K16ac	0.000	6.48E-01	1.2727	CRE	CQ2
Expression	H3K27me3	0.006	8.30E-07	1.4537	CRE	CQ2
Expression	H3K9me3	0.004	4.18E-06	1.174	CRE	CQ2
Expression	Conservation	0.012	3.18E-01	1.0207	CRE	CQ2
Expression	Stability	-0.022	2.00E-16	1.0199	CRE	CQ2
Exp. Stability	H3K9ac	0.047	3.56E-06	1.1752	CRE	CQ2
Exp. Stability	H3K4me3	0.070	4.41E-11	1.1563	CRE	CQ2
Exp. Stability	H3K4ac	0.045	6.39E-04	1.2096	CRE	CQ2
Exp. Stability	H3K36me3	0.028	2.25E-01	1.761	CRE	CQ2
Exp. Stability	H4K16ac	-0.011	3.62E-01	1.2728	CRE	CQ2
Exp. Stability	H3K27me3	-0.166	2.00E-16	1.4678	CRE	CQ2
Exp. Stability	H3K9me3	-0.117	2.00E-16	1.1861	CRE	CQ2
Exp. Stability	Conservation	3.799	2.00E-16	1.1007	CRE	CQ2
Exp. Stability	Expression	2.641	2.00E-16	1.1349	CRE	CQ2

Conservation	H3K9ac	0.003	4.17E-02	1.6085	CRE	JR2
Conservation	H3K4me3	0.005	3.76E-07	1.1322	CRE	JR2
Conservation	H3K4ac	0.001	6.32E-02	1.0445	CRE	JR2
Conservation	H3K36me3	-0.005	5.57E-02	2.3803	CRE	JR2
Conservation	H4K16ac	-0.006	4.87E-04	1.8651	CRE	JR2
Conservation	H3K27me3	-0.002	1.32E-01	1.5177	CRE	JR2
Conservation	H3K9me3	-0.001	1.69E-02	1.035	CRE	JR2
Conservation	Expression	0.019	2.00E-16	1.0844	CRE	JR2
Conservation	Stability	0.002	8.45E-01	1.0684	CRE	JR2
Expression	H3K9ac	-0.003	9.24E-02	1.6053	CRE	JR2
Expression	H3K4me3	0.002	6.90E-02	1.1346	CRE	JR2
Expression	H3K4ac	0.000	9.56E-01	1.0448	CRE	JR2
Expression	H3K36me3	0.005	7.62E-02	2.3811	CRE	JR2
Expression	H4K16ac	0.002	4.04E-01	1.867	CRE	JR2
Expression	H3K27me3	0.012	2.00E-16	1.4951	CRE	JR2
Expression	H3K9me3	0.000	7.51E-01	1.0355	CRE	JR2
Expression	Conservation	-0.007	5.37E-01	1.0146	CRE	JR2
Expression	Stability	-0.021	2.00E-16	1.0249	CRE	JR2
Exp. Stability	H3K9ac	0.096	4.21E-07	1.6088	CRE	JR2
Exp. Stability	H3K4me3	-0.022	5.93E-02	1.1346	CRE	JR2
Exp. Stability	H3K4ac	-0.002	7.99E-01	1.0448	CRE	JR2
Exp. Stability	H3K36me3	-0.010	7.43E-01	2.3806	CRE	JR2
Exp. Stability	H4K16ac	-0.025	2.61E-01	1.8671	CRE	JR2
Exp. Stability	H3K27me3	-0.186	2.00E-16	1.5101	CRE	JR2
Exp. Stability	H3K9me3	-0.006	4.06E-01	1.0355	CRE	JR2
Exp. Stability	Conservation	2.943	2.00E-16	1.0748	CRE	JR2
Exp. Stability	Expression	2.075	2.00E-16	1.1019	CRE	JR2

*CRE stands for *cis*-regulatory region

Supplementary Table 4. Differential gene expression (log2FC) is linked to H3K27me3-associated chromatin states

Genetic element	State differentially expressed[1]	Comparison state	*p-value*	*p-value adjusted*	Significance [1]	Genes upregulated[3]	Genes downregulated[3]	Strain
Coding seq.	13	20	0.82000	1.00000	ns	119	330	JR2
Coding seq.	13	19	0.00072	0.00580	***	134	781	JR2
Coding seq.	13	18	0.00003	0.00031	****	127	915	JR2
Coding seq.	13	17	0.00002	0.00020	****	149	1142	JR2
Coding seq.	13	16	0.00001	0.00008	****	81	616	JR2
Coding seq.	13	15	0.00001	0.00015	****	526	2865	JR2
Coding seq.	13	14	0.66700	1.00000	ns	1	2	JR2
Coding seq.	13	13	NA	NA	ns	208	621	JR2
Coding seq.	13	12	0.05010	0.30000	ns	31	49	JR2
Coding seq.	13	11	0.61100	1.00000	ns	139	423	JR2
Coding seq.	13	10	0.60800	1.00000	ns	154	535	JR2
Coding seq.	13	9	0.06250	0.31000	ns	143	486	JR2
Coding seq.	13	8	0.00073	0.00580	***	180	981	JR2
Coding seq.	13	7	0.00001	0.00015	****	176	1005	JR2
Coding seq.	13	6	0.00000	0.00001	****	180	1194	JR2
Coding seq.	13	5	0.00000	0.00001	****	77	619	JR2
Coding seq.	13	4	0.00008	0.00076	****	108	649	JR2
Coding seq.	13	3	0.00000	0.00001	****	215	1373	JR2
Coding seq.	13	2	0.00000	0.00001	****	171	1225	JR2
Coding seq.	13	1	0.00000	0.00001	****	368	2338	JR2
CREs	13	20	0.70500	1.00000	ns	44	83	JR2
CREs	13	19	0.31700	1.00000	ns	48	107	JR2
CREs	13	18	0.00046	0.00600	***	48	201	JR2

CREs	13	17	0.05860	0.35000	ns	54	142	JR2
CREs	13	16	0.00123	0.01500	**	11	60	JR2
CREs	13	15	0.00002	0.00037	****	275	960	JR2
CREs	13	14	0.84500	1.00000	ns	4	5	JR2
CREs	13	13	NA	NA	ns	98	148	JR2
CREs	13	12	0.88300	1.00000	ns	13	17	JR2
CREs	13	11	0.41600	1.00000	ns	37	67	JR2
CREs	13	10	0.00869	0.07800	**	44	136	JR2
CREs	13	9	0.04150	0.29000	*	52	114	JR2
CREs	13	8	0.00444	0.04900	**	74	247	JR2
CREs	13	7	0.00007	0.00095	****	63	237	JR2
CREs	13	6	0.02660	0.21000	*	60	159	JR2
CREs	13	5	0.00731	0.07300	**	17	74	JR2
CREs	13	4	0.00000	0.00000	****	35	194	JR2
CREs	13	3	0.00001	0.00012	****	81	355	JR2
CREs	13	2	0.00004	0.00067	****	71	283	JR2
CREs	13	1	0.00000	0.00000	****	138	686	JR2
Coding seq.	13	20	0.00000	0.00000	****	202	238	CQ2
Coding seq.	13	19	0.64200	1.00000	ns	385	269	CQ2
Coding seq.	13	18	0.00000	0.00000	****	343	332	CQ2
Coding seq.	13	17	0.00000	0.00000	****	454	788	CQ2
Coding seq.	13	16	0.00000	0.00000	****	233	292	CQ2
Coding seq.	13	15	0.00000	0.00000	****	1055	1333	CQ2
Coding seq.	13	14	0.00810	0.03000	**	2	8	CQ2
Coding seq.	13	13	NA	NA	ns	696	456	CQ2
Coding seq.	13	12	0.00000	0.00000	****	261	270	CQ2

Supplementary Table 4. Continued

Type	State differentially expressed[1]	Comparison state	*p-value*	*p-value adjusted*	Significance [1]	Genes upregulated[3]	Genes downregulated[3]	Strain
Coding seq.	13	11	0.64100	1.00000	ns	280	204	CQ2
Coding seq.	13	10	0.00000	0.00000	****	519	578	CQ2
Coding seq.	13	9	0.00758	0.03000	**	249	201	CQ2
Coding seq.	13	8	0.00000	0.00000	****	172	183	CQ2
Coding seq.	13	7	0.00000	0.00000	****	145	222	CQ2
Coding seq.	13	6	0.00000	0.00000	****	371	438	CQ2
Coding seq.	13	5	0.00000	0.00000	****	293	533	CQ2
Coding seq.	13	4	0.00000	0.00000	****	476	539	CQ2
Coding seq.	13	3	0.00000	0.00000	****	721	1317	CQ2
Coding seq.	13	2	0.00000	0.00000	****	293	601	CQ2
Coding seq.	13	1	0.00000	0.00000	****	391	520	CQ2
CREs	13	20	0.00000	0.00002	****	66	72	CQ2
CREs	13	19	0.04190	0.13000	*	139	86	CQ2
CREs	13	18	0.00000	0.00000	****	88	96	CQ2
CREs	13	17	0.00000	0.00000	****	148	245	CQ2
CREs	13	16	0.00039	0.00230	***	65	51	CQ2
CREs	13	15	0.00000	0.00000	****	473	480	CQ2
CREs	13	14	0.73800	1.00000	ns	2	1	CQ2
CREs	13	13	NA	NA	ns	261	114	CQ2
CREs	13	12	0.01210	0.04800	*	51	33	CQ2
CREs	13	11	0.69900	1.00000	ns	83	37	CQ2
CREs	13	10	0.00027	0.00190	***	127	111	CQ2
CREs	13	9	0.00468	0.02300	**	61	51	CQ2

CREs	13	8	0.00016	0.00230	***	66	61	CQ2
CREs	13	7	0.00000	0.00001	****	30	38	CQ2
CREs	13	6	0.00000	0.00000	****	111	114	CQ2
CREs	13	5	0.00000	0.00000	****	68	126	CQ2
CREs	13	4	0.00000	0.00000	****	197	205	CQ2
CREs	13	3	0.00000	0.00000	****	233	358	CQ2
CREs	13	2	0.00000	0.00000	****	66	148	CQ2
CREs	13	1	0.00000	0.00000	****	136	164	CQ2

(1) *p-value* after one-sided Wilcoxon rank sum test, compared to state 13 (H3K27me3-only)

*p<=0.05, **p<= 0.01, ***p<= 0.001, ****p <=0.0001, NA no significative

(3) number of a given chromatin state over up/down regulated genes

Supplementary Table 5. Structural variants are depleted at specific chromatin states in *Verticillium dahliae* strain JR2

States	SV type	Synteny breaks overlap	Coverage(bp) overlap	log10(score for breaks)[1]	log10(score for coverage)[1]
1	Deletion	43	203958	-0.335116284	-1.258515313
1	Duplication	3	391190	-0.375092121	-0.984049374
1	Inversion	6	537975	-0.366333622	-1.089100806
1	PolTEs	8	832	-0.667777308	-1.556799943
1	Translocation	64	30	-0.185870797	-0.514440721
2	Deletion	6	95083	-0.841343439	-1.240863131
2	Duplication	0	161600	0	-1.01824893
2	Inversion	0	218200	0	-1.130941905
2	PolTEs	0	400	0	-1.526190383
2	Translocation	11	9	-0.600884162	-0.687545543
3	Deletion	22	136868	-0.364080619	-1.169672696
3	Duplication	1	229923	-0.590543408	-0.953184734
3	Inversion	3	354636	-0.412607872	-1.015324335
3	PolTEs	1	1279	-1.307996633	-1.107182062
3	Translocation	36	24	-0.172717125	-0.348319589
4	Deletion	13	74106	-0.238481772	-1.08204307
4	Duplication	1	104251	-0.236183772	-0.942327286
4	Inversion	0	142600	0	-1.054368739
4	PolTEs	8	2607	-0.051072302	-0.444077232
4	Translocation	18	12	-0.118874845	-0.294477308
5	Deletion	7	41167	-0.391471729	-1.221491867
5	Duplication	1	84985	-0.121480812	-0.916362277
5	Inversion	0	113400	0	-1.039292342
5	PolTEs	0	0	0	0

5	Translocation	2	2	-0.958026042	-0.957537247
6	Deletion	25	125937	-0.164936799	-1.062194781
6	Duplication	0	138800	0	-1.027054396
6	Inversion	0	258200	0	-1.011267795
6	PolTEs	4	441	-0.562740348	-1.426417723
6	Translocation	11	5	-0.543726061	-0.885659946
7	Deletion	31	203883	-0.27093282	-1.052384816
7	Duplication	3	273984	-0.169603898	-0.933223695
7	Inversion	2	389753	-0.644318404	-1.029936962
7	PolTEs	13	1356	-0.250634058	-1.138373695
7	Translocation	70	26	0.059589279	-0.370046618
8	Deletion	25	129096	-0.194038154	-1.080536694
8	Duplication	1	164600	-0.474760115	-0.982554028
8	Inversion	4	243668	-0.168267632	-1.058907153
8	PolTEs	9	379	-0.239630877	-1.52129015
8	Translocation	40	24	-0.012681221	-0.234041176
9	Deletion	17	99498	-0.070643662	-0.90274955
9	Duplication	0	74400	0	-1.032573907
9	Inversion	6	174254	0.286587097	-0.925759425
9	PolTEs	3	564	-0.423705242	-1.055603366
9	Translocation	11	9	-0.279826948	-0.366488328
10	Deletion	40	232197	0.059959725	-0.775714979
10	Duplication	3	144064	0.057176104	-0.985613425
10	Inversion	18	371466	0.509614649	-0.851116815
10	PolTEs	15	1521	0.036461194	-0.863556825
10	Translocation	134	38	0.564554842	0.017722436

Supplementary Table 5. Continued

States	SV type	Synteny breaks overlap	Coverage(bp) overlap	log10(score for breaks)[1]	log10(score for coverage)[1]
11	Deletion	36	236091	0.158461985	-0.624232402
11	Duplication	3	117873	0.205214881	-0.92471578
11	Inversion	6	392473	0.14985558	-0.709863913
11	PolTEs	11	1119	0.048149872	-0.8504687
11	Translocation	23	15	-0.052159558	-0.23730734
12	Deletion	282	1210083	0.927652269	-0.039252997
12	Duplication	4	191777	0.276968207	-0.766519008
12	Inversion	5	780212	-0.041168755	-0.523304162
12	PolTEs	128	243115	1.059527	1.432072812
12	Translocation	47	25	0.202686546	-0.070982509
13	Deletion	59	344631	0.162894488	-0.670074501
13	Duplication	10	199714	0.515705773	-0.90810946
13	Inversion	18	570761	0.431173963	-0.74302228
13	PolTEs	44	4220	0.439254968	-0.484941231
13	Translocation	72	44	0.231023063	0.017632038
14	Deletion	136	757105	0.524483546	-0.329370565
14	Duplication	35	759084	1.038997752	-0.349004198
14	Inversion	90	1221033	1.090980568	-0.451912569
14	PolTEs	105	120921	0.84622667	1.001483127
14	Translocation	1204	260	1.123448446	0.458284102
15	Deletion	246	1029893	0.170025584	-0.807588734
15	Duplication	13	921468	0.014769535	-0.858917302
15	Inversion	51	1563431	0.30451262	-0.884356865
15	PolTEs	80	46670	0.091535025	-0.048573135
15	Translocation	600	292	0.527482701	0.215203098

16	Deletion	3	44933	-0.748501334	-1.172528491
16	Duplication	0	66800	0	-1.009619542
16	Inversion	0	103200	0	-1.068760739
16	PolTEs	1	14	-0.828711714	-2.588639652
16	Translocation	9	3	-0.294936012	-0.771568472
17	Deletion	11	89406	-0.466923169	-1.156420506
17	Duplication	0	122400	0	-1.027554843
17	Inversion	1	169800	-0.683128099	-1.128568441
17	PolTEs	17	532	0.119358306	-1.291234956
17	Translocation	8	6	-0.627922789	-0.75237273
18	Deletion	14	74898	-0.362753101	-1.233882234
18	Duplication	1	119800	-0.391424892	-1.037191818
18	Inversion	0	145400	0	-1.193461586
18	PolTEs	3	491	-0.634723453	-1.326819189
18	Translocation	26	12	-0.117295704	-0.452599011
19	Deletion	69	332028	0.243565936	-0.673579762
19	Duplication	14	180086	0.6728654	-0.942006427
19	Inversion	14	487471	0.337021948	-0.796535332
19	PolTEs	14	1956	-0.046235303	-0.807050462
19	Translocation	30	18	-0.135956789	-0.357316743
20	Deletion	28	175544	0.040650598	-0.761592761
20	Duplication	6	122404	0.491688868	-0.922890516
20	Inversion	4	316710	-0.022317206	-0.799093449
20	PolTEs	14	913	0.140602643	-0.951110589
20	Translocation	24	18	-0.045700038	-0.170149979

(1) log10 score calculated based in the formula of Fudenberg and Pollard, 2019

Supplementary Table 6. Genes with conserved chromatin states in each expression category

Strain	Gene category	Expression	Classification*	Genes	% Total genes
JR2	Other genes	Not expressed	Conserved	587	5.58%
JR2	Other genes	Variable	Conserved	1253	11.90%
JR2	Other genes	Stable	Conserved	5259	49.97%
JR2	Other genes	Not expressed	Not Conserved	477	4.53%
JR2	Other genes	Variable	Not Conserved	530	5.04%
JR2	Other genes	Stable	Not Conserved	1795	17.05%
JR2	Other genes	Not expressed	NA	93	0.88%
JR2	Other genes	Variable	NA	52	0.49%
JR2	Other genes	Stable	NA	139	1.32%
JR2	Effectors	Not expressed	Conserved	51	15.94%
JR2	Effectors	Variable	Conserved	55	17.19%
JR2	Effectors	Stable	Conserved	64	20.00%
JR2	Effectors	Not expressed	Not Conserved	47	14.69%
JR2	Effectors	Variable	Not Conserved	37	11.56%
JR2	Effectors	Stable	Not Conserved	37	11.56%
JR2	Effectors	Not expressed	NA	10	3.13%
JR2	Effectors	Variable	NA	5	1.56%
JR2	Effectors	Stable	NA	4	1.25%
JR2	CAZy	Not expressed	Conserved	42	18.34%
JR2	CAZy	Variable	Conserved	43	18.78%
JR2	CAZy	Stable	Conserved	75	32.75%
JR2	CAZy	Not expressed	Not Conserved	22	9.61%
JR2	CAZy	Variable	Not Conserved	11	4.80%
JR2	CAZy	Stable	Not Conserved	29	12.66%
JR2	CAZy	Not expressed	NA	2	0.87%
JR2	CAZy	Variable	NA	0	0.00%
JR2	CAZy	Stable	NA	3	1.31%
JR2	Other secreted	Not expressed	Conserved	38	10.44%
JR2	Other secreted	Variable	Conserved	80	21.98%
JR2	Other secreted	Stable	Conserved	126	34.62%
JR2	Other secreted	Not expressed	Not Conserved	24	6.59%
JR2	Other secreted	Variable	Not Conserved	30	8.24%
JR2	Other secreted	Stable	Not Conserved	48	13.19%
JR2	Other secreted	Not expressed	NA	1	0.27%
JR2	Other secreted	Variable	NA	5	1.37%
JR2	Other secreted	Stable	NA	2	0.55%
CQ2	Other genes	Not expressed	Conserved	638	6.06%

Strain	Gene category	Expression	Classification*	Genes	% Total genes
CQ2	Other genes	Variable	Conserved	1293	12.29%
CQ2	Other genes	Stable	Conserved	5194	49.35%
CQ2	Other genes	Not expressed	Not Conserved	521	4.95%
CQ2	Other genes	Variable	Not Conserved	545	5.18%
CQ2	Other genes	Stable	Not Conserved	1763	16.75%
CQ2	Other genes	Not expressed	NA	94	0.89%
CQ2	Other genes	Variable	NA	55	0.52%
CQ2	Other genes	Stable	NA	134	1.27%
CQ2	Effectors	Not expressed	Conserved	33	9.32%
CQ2	Effectors	Variable	Conserved	55	15.54%
CQ2	Effectors	Stable	Conserved	140	39.55%
CQ2	Effectors	Not expressed	Not Conserved	20	5.65%
CQ2	Effectors	Variable	Not Conserved	28	7.91%
CQ2	Effectors	Stable	Not Conserved	51	14.41%
CQ2	Effectors	Not expressed	NA	5	1.41%
CQ2	Effectors	Variable	NA	4	1.13%
CQ2	Effectors	Stable	NA	5	1.41%
CQ2	CAZy	Not expressed	Conserved	20	9.52%
CQ2	CAZy	Variable	Conserved	30	14.29%
CQ2	CAZy	Stable	Conserved	82	39.05%
CQ2	CAZy	Not expressed	Not Conserved	9	4.29%
CQ2	CAZy	Variable	Not Conserved	16	7.62%
CQ2	CAZy	Stable	Not Conserved	37	17.62%
CQ2	CAZy	Not expressed	NA	4	1.90%
CQ2	CAZy	Variable	NA	1	0.48%
CQ2	CAZy	Stable	NA	4	1.90%
CQ2	Other secreted	Not expressed	Conserved	27	8.06%
CQ2	Other secreted	Variable	Conserved	53	15.82%
CQ2	Other secreted	Stable	Conserved	138	41.19%
CQ2	Other secreted	Not expressed	Not Conserved	20	5.97%
CQ2	Other secreted	Variable	Not Conserved	19	5.67%
CQ2	Other secreted	Stable	Not Conserved	58	17.31%
CQ2	Other secreted	Not expressed	NA	3	0.90%
CQ2	Other secreted	Variable	NA	2	0.60%
CQ2	Other secreted	Stable	NA	5	1.49%

*NA depicts number of genes with chromatin states 16-20. Classification based on the CQV (Figure 2).
**Genes with conserved chromatin states are defined as =>1 chromatin state conserved between homologous genes in JR2-CQ2

Supplementary Table 7. Chromatin states conserved[(1)] in homologous genes in each expression category

Chrom. States	Not expressed	Differentially expressed	Conserved expressed	Not expressed (%)	Differentially expressed (%)	Conserved expressed (%)	log10 (odds ratio)	X^2	*p-value*	Type
1	54	250	1205	3.58%	16.57%	79.85%	-0.071	6.20	0.01277	Other genes
2	20	130	726	2.28%	14.84%	82.88%	0.135	11.86	0.00058	Other genes
3	39	285	1452	2.20%	16.05%	81.76%	0.100	14.14	0.00017	Other genes
4	22	99	453	3.83%	17.25%	78.92%	0.042	1.28	0.25852	Other genes
5	6	67	360	1.39%	15.47%	83.14%	0.113	5.23	0.02216	Other genes
6	26	110	578	3.64%	15.41%	80.95%	0.106	5.89	0.01523	Other genes
7	12	45	298	3.38%	12.68%	83.94%	0.206	9.26	0.00234	Other genes
8	14	58	245	4.42%	18.30%	77.29%	0.005	0.21	0.64613	Other genes
9	35	76	162	12.82%	27.84%	59.34%	-0.299	23.74	0.00000	Other genes
10	89	150	438	13.15%	22.16%	64.70%	-0.164	16.31	0.00005	Other genes
11	110	109	198	26.38%	26.14%	47.48%	-0.373	57.77	0.00000	Other genes
12	21	20	35	27.63%	26.32%	46.05%	-0.380	11.47	0.00071	Other genes
13	215	244	500	22.42%	25.44%	52.14%	-0.334	96.99	0.00000	Other genes
14	0	0	0	0.00%	0.00%	0.00%	0.000	0.00	0.00000	Other genes
15	220	673	3016	5.63%	17.22%	77.16%	0.044	3.47	0.06265	Other genes
1	2	12	27	4.88%	29.27%	65.85%	0.081	0.29	0.59261	secreted*
2	1	2	20	4.35%	8.70%	86.96%	0.599	3.88	0.04895	secreted*
3	1	10	32	2.33%	23.26%	74.42%	0.090	0.44	0.50775	secreted*
4	1	3	11	6.67%	20.00%	73.33%	0.145	0.25	0.61434	secreted*
5	0	4	16	0.00%	20.00%	80.00%	0.186	0.73	0.39127	secreted*
6	0	5	25	0.00%	16.67%	83.33%	0.294	2.09	0.14829	secreted*
7	1	3	7	9.09%	27.27%	63.64%	-0.058	0.04	0.84893	secreted*
8	1	2	8	9.09%	18.18%	72.73%	0.182	0.27	0.60360	secreted*

9	3	6	6	20.00%	40.00%	40.00%	-0.440	2.93	0.08692	secreted*
10	5	8	18	16.13%	25.81%	58.06%	-0.078	0.36	0.55085	secreted*
11	3	10	9	13.64%	45.45%	40.91%	-0.499	5.82	0.01582	secreted*
12	0	2	2	0.00%	50.00%	50.00%	-0.430	1.15	0.28268	secreted*
13	11	16	21	22.92%	33.33%	43.75%	-0.343	5.09	0.02402	secreted*
14	0	0	0	0.00%	0.00%	0.00%	0.000	0.00	0.00000	secreted*
15	6	25	85	5.17%	21.55%	73.28%	0.145	1.33	0.24911	secreted*
1	2	3	20	8.00%	12.00%	80.00%	-0.331	1.36	0.24313	CAZy
2	0	1	7	0.00%	12.50%	87.50%	0.332	0.67	0.41319	CAZy
3	0	3	17	0.00%	15.00%	85.00%	0.251	1.09	0.29738	CAZy
4	0	3	7	0.00%	30.00%	70.00%	-0.165	0.44	0.50686	CAZy
5	0	0	4	0.00%	0.00%	100.00%	0.000	1.40	0.23711	CAZy
6	0	2	16	0.00%	11.11%	88.89%	0.407	1.86	0.17237	CAZy
7	0	0	5	0.00%	0.00%	100.00%	0.000	1.75	0.18624	CAZy
8	1	3	3	14.29%	42.86%	42.86%	-0.544	2.38	0.12325	CAZy
9	0	2	6	0.00%	25.00%	75.00%	-0.049	0.13	0.72209	CAZy
10	9	5	11	36.00%	20.00%	44.00%	-0.199	2.20	0.13777	CAZy
11	3	2	8	23.08%	15.38%	61.54%	0.082	0.26	0.60960	CAZy
12	0	2	0	0.00%	100.00%	0.00%	0.000	7.53	0.00608	CAZy
13	8	6	14	28.57%	21.43%	50.00%	-0.175	1.42	0.23376	CAZy
14	0	0	0	0.00%	0.00%	0.00%	0.000	0.00	0.00000	CAZy
15	2	15	39	3.57%	26.79%	69.64%	-0.152	1.08	0.29883	CAZy
1	4	6	35	8.89%	13.33%	77.78%	-0.417	4.24	0.03946	Effector
2	3	7	18	10.71%	25.00%	64.29%	0.018	0.02	0.89411	Effector
3	5	3	33	12.20%	7.32%	80.49%	0.703	7.38	0.00661	Effector
4	1	3	5	11.11%	33.33%	55.56%	-0.176	0.30	0.58267	Effector

Supplementary Table 7. Continued

Chrom. States	Not expressed	Differentially expressed	Conserved expressed	Not expressed (%)	Differentially expressed (%)	Conserved expressed (%)	log10 (odds ratio)	X^2	*p-value*	Type
5	0	3	8	0.00%	27.27%	72.73%	0.034	0.21	0.64715	Effector
6	0	6	10	0.00%	37.50%	62.50%	-0.181	0.96	0.32740	Effector
7	0	1	5	0.00%	16.67%	83.33%	0.311	0.60	0.43918	Effector
8	1	1	8	10.00%	10.00%	80.00%	0.521	1.42	0.23363	Effector
9	3	5	4	25.00%	41.67%	33.33%	-0.507	2.91	0.08814	Effector
10	4	16	10	13.33%	53.33%	33.33%	-0.661	13.46	0.00024	Effector
11	7	8	9	29.17%	33.33%	37.50%	-0.364	3.00	0.08316	Effector
12	1	1	3	20.00%	20.00%	60.00%	0.085	0.06	0.80946	Effector
13	11	12	12	31.43%	34.29%	34.29%	-0.431	5.56	0.01838	Effector
14	0	0	0	0.00%	0.00%	0.00%	0.000	0.00	0.00000	Effector
15	5	19	65	5.62%	21.35%	73.03%	0.187	2.16	0.14138	Effector

(1)Total number of chromatin states considered over homologous genes with >=1 chromatin state shared between JR2 and CQ2 strain. Genes can depict >1 chromatin state
Chi-square statistic (X^2) and associated *p-value* after comparison of differentially expressed genes *versus* conserved expressed genes.
*'Other secreted' coding genes.

Supplementary Table 8. Differentially expressed genes between strains are enriched for H3K27me3-associated chromatin states.

Strain	State	Other secreted[(1)]	Effectors[(1)]	CAZy[(1)]	Other genes[(1)]	Other secreted X^2	Other secreted *p-value*	FDR	Effectors X^2	Effectors *p-value*	FDR	CAZy X^2	CAZy *p-value*	FDR	Other genes X^2	Other genes *p-value*	FDR
CQ2	1	1.042	0.766	0.833	0.907	0.223	6.37E-01	NA	0.371	5.42E-01	NA	0.164	6.86E-01	NA	0.065	7.99E-01	NA
CQ2	2	0.786	0.747	1.157	0.757	0.114	7.35E-01	NA	0.015	9.02E-01	NA	0.214	6.44E-01	NA	0.153	6.96E-01	NA
CQ2	3	0.509	0.283	0.802	0.736	2.894	8.89E-02	NA	3.910	4.80E-02	NA	3.752	5.27E-02	NA	2.004	1.57E-01	NA
CQ2	4	1.189	1.448	0.689	0.949	0.323	5.70E-01	NA	0.761	3.83E-01	NA	0.965	3.26E-01	NA	0.238	6.26E-01	NA
CQ2	5	0.488	0.549	0.784	0.717	0.178	6.73E-01	NA	0.083	7.74E-01	NA	0.321	5.71E-01	NA	0.157	6.92E-01	NA
CQ2	6	0.635	0.949	0.854	0.898	0.216	6.42E-01	NA	0.292	5.89E-01	NA	0.533	4.65E-01	NA	0.014	9.06E-01	NA
CQ2	7	0.444	0.429	0.543	0.619	2.608	1.06E-01	NA	1.191	2.75E-01	NA	2.703	1.00E-01	NA	3.973	4.62E-02	NA
CQ2	8	0.880	1.287	1.618	1.054	4.161	4.14E-02	NA	2.280	1.31E-01	NA	3.888	4.86E-02	NA	3.698	5.45E-02	NA
CQ2	9	1.814	1.596	1.467	1.703	13.984	1.84E-04	****	26.624	2.47E-07	****	14.762	1.22E-04	****	5.888	1.52E-02	****
CQ2	10	1.175	1.589	1.126	1.120	0.646	4.22E-01	NA	0.421	3.31E-03	****	0.240	6.24E-01	NA	0.110	7.40E-01	NA
CQ2	11	1.332	1.585	0.873	1.577	5.368	2.05E-02	****	11.151	8.40E-04	****	8.627	3.31E-03	****	3.751	5.28E-02	NA
CQ2	12	0.605	0.909	1.233	1.275	5.508	1.89E-02	****	8.627	5.17E-01	NA	5.554	1.84E-02	****	4.292	3.83E-02	NA
CQ2	13	2.329	1.155	1.370	1.725	0.032	8.58E-01	NA	0.163	6.86E-01	NA	0.180	6.71E-01	NA	0.082	7.75E-01	NA
CQ2	14	0.000	0.000	0.000	4.796	0.000	NA	NA	NA	NA	NA	0.000	NA	NA	0.000	NA	NA
CQ2	15	0.650	1.133	1.294	1.004	9.219	2.39E-03	****	9.335	2.25E-03	****	9.809	1.74E-03	****	7.491	6.20E-03	****
CQ2	16	1.512	1.221	0.991	1.088	19.447	1.03E-05	****	5.610	1.79E-02	****	9.273	2.33E-03	****	5.422	1.99E-02	****
CQ2	17	1.231	0.773	0.609	0.727	0.893	3.45E-01	NA	1.082	2.98E-01	NA	1.046	3.06E-01	NA	0.459	4.98E-01	NA
CQ2	18	1.817	1.100	0.942	1.055	0.651	4.20E-01	NA	0.193	6.60E-01	NA	2.201	1.38E-01	NA	0.198	6.57E-01	NA
CQ2	19	1.801	1.562	1.360	1.605	1.375	2.41E-01	NA	1.106	2.93E-01	NA	0.256	6.13E-01	NA	0.955	3.29E-01	NA
CQ2	20	0.411	0.836	1.040	1.077	0.365	5.46E-01	NA	4.536	3.32E-02	****	0.949	3.30E-01	NA	1.940	1.64E-01	NA
JR2	1	0.788	0.894	0.930	0.854	3.075	7.95E-02	NA	2.510	1.13E-01	NA	3.847	1.19E-01	NA	2.435	1.19E-01	NA
JR2	2	1.024	0.626	1.107	0.860	0.151	6.97E-01	NA	0.042	8.37E-01	NA	0.206	8.36E-01	NA	0.043	8.36E-01	NA
JR2	3	0.783	0.693	0.864	0.916	0.836	3.61E-01	NA	0.761	3.83E-01	NA	0.660	6.88E-01	NA	0.161	6.88E-01	NA

Supplementary Table 8. Continued

Strain	State	Other secreted[(1)]	Effectors[(1)]	CAZy[(1)]	Other genes[(1)]	Other secreted X^2	Other secreted *p-value*	FDR	Effectors X^2	Effectors *p-value*	FDR	CAZy X^2	CAZy *p-value*	FDR	Other genes X^2	Other genes *p-value*	FDR
JR2	4	1.115	0.535	1.782	0.868	0.211	6.46E-01	NA	0.776	3.78E-01	NA	1.462	3.37E-01	NA	0.923	3.37E-01	NA
JR2	5	1.183	1.019	1.555	1.028	3.544	5.98E-02	NA	2.495	1.14E-01	NA	8.025	4.49E-02	NA	4.024	4.49E-02	NA
JR2	6	1.179	0.859	0.621	0.908	0.473	4.92E-01	NA	0.010	9.21E-01	NA	0.187	7.05E-01	NA	0.143	7.05E-01	NA
JR2	7	1.144	1.236	0.485	0.975	0.052	8.20E-01	NA	0.084	7.72E-01	NA	0.681	6.89E-01	NA	0.160	6.89E-01	NA
JR2	8	0.528	1.179	1.207	0.935	0.251	6.17E-01	NA	0.191	6.62E-01	NA	0.037	9.43E-01	NA	0.005	9.43E-01	NA
JR2	9	0.988	1.245	1.673	1.339	2.425	1.19E-01	NA	6.371	1.16E-02	****	8.764	6.09E-02	NA	3.514	6.09E-02	NA
JR2	10	1.656	0.788	1.482	1.675	6.174	1.30E-02	****	5.968	1.46E-02	****	3.379	2.70E-01	NA	1.218	2.70E-01	NA
JR2	11	2.142	1.820	0.777	1.603	11.765	6.03E-04	****	6.586	1.03E-02	****	13.153	5.48E-02	NA	3.688	5.48E-02	NA
JR2	12	1.932	2.005	0.963	3.227	0.748	3.87E-01	NA	0.915	3.39E-01	NA	0.757	6.15E-01	NA	0.253	6.15E-01	NA
JR2	13	1.222	2.073	1.075	1.725	0.551	4.58E-01	NA	0.314	5.75E-01	NA	1.688	5.83E-01	NA	0.301	5.83E-01	NA
JR2	14	0.000	0.000	0.000	3.930	0.000	1.10E-01	NA	0.000	1.12E-01	NA	0.000	1.18E-01	NA	0.000	1.18E-01	NA
JR2	15	0.727	1.207	1.149	0.872	8.861	2.91E-03	****	8.418	3.71E-03	****	9.552	4.87E-03	****	7.928	4.87E-03	****
JR2	16	1.213	0.696	0.497	0.828	5.815	1.59E-02	****	9.857	1.69E-03	****	5.228	7.12E-02	NA	3.255	7.12E-02	NA
JR2	17	0.963	0.651	0.741	0.882	0.315	5.75E-01	NA	0.718	3.97E-01	NA	0.388	5.36E-01	NA	0.383	5.36E-01	NA
JR2	18	0.451	0.809	1.185	0.856	0.182	6.69E-01	NA	0.056	8.12E-01	NA	0.259	6.17E-01	NA	0.250	6.17E-01	NA
JR2	19	1.305	1.181	0.773	1.026	0.042	8.37E-01	NA	0.024	8.77E-01	NA	0.088	6.10E-01	NA	0.260	6.10E-01	NA
JR2	20	0.965	1.196	1.652	1.711	9.517	2.04E-03	****	3.467	6.26E-02	NA	5.621	5.15E-02	NA	3.792	5.15E-02	NA

[(1)]Differential fold enrichment ratio for each chromatin state and each pathogenicity-related gene categories [variable expressed genes/ conserved expressed genes]; based in the formula of Ernst et al., 2017

Chi-square statistic (X^2) and associated *p-value* after comparison of a given chromatin state enrichment for variable expressed genes *versus* conserved expressed genes with the same chromatin state.

FDR (false discovery rate) **** $p<0.05$; NA depicts $p>0.05$.

```
# core genes
core.25ka<- read.table("~/Dropbox/rebuttal_raw/materi-
al/core.25k.a.bed",sep="\t",fill=T,
                  col.names=c("Chr_bnd","start_bnd","end_bnd","BND","Ins","X-
","Chr","start","end","strand","genes","ID",
                              "baseMean","log2FoldChange","lfcSE","stat","pvalue","pad-
j","O","P","Q","R","S","W","Z"))%>%
  dplyr::select(-X,-O,-P,-Q,-R,-S,-W,-Z)%>%dplyr::mutate(region=rep("Core"))
ov <- core.25ka%>%group_by(Chr_bnd)%>%mutate(a=start-end_b-
nd,b=end-start_bnd)%>%mutate(ov=ifelse(a<0&b>0,"ov","no"))%>%filter(ov=="ov")
%>%
  dplyr::select(-a,-b)%>%mutate(dist=rep(0),cat=rep("0k"))
ov2 <- core.25ka%>%group_by(Chr_bnd)%>%mutate(a=start-end_b-
```

CHAPTER 5

Three-dimensional chromatin organization determines the evolution of adaptive genomic regions in the plant pathogen *Verticillium dahliae*

David E. Torres[1,2], H. Martin Kramer[1], Vittorio Tracanna[3], Gabriel L. Fiorin[1], David E. Cook[1], Michael F. Seidl[1,2], and Bart P.H.J. Thomma[1,3]

[1]Wageningen University and Research, NL
[2]Utrecht University, NL
[3]University of Cologne, DE
[4]Kansas State University, USA

Modified version submitted for publication *Biorxiv* (2023): 2023-04

Abstract

The spatial organization of the eukaryotic nuclear genome is intimately linked to its biological functions. Beyond the linear organization of genetic elements in the DNA, chromatin folding and organization play an important role in the regulation of gene expression and genome evolution. In fungi, the three-dimensional (3D) organization of the genome is still under-investigated. Therefore, the role of the 3D organization in transcriptional regulation, genome organization, and evolution remains unclear. The 3D genome organization could be relevant in the genomes of fung'l plant pathogens that typically display a two-tier genome organization. The genome of the plant pathogen *Verticillium dahliae* contains designated plastic regions, known as adaptive genomic regions (AGRs), that are enriched in transposable elements and in *in planta*-induced genes that mediate virulence during host infection. Here, we utilize chromatin conformation capture (Hi-C) to uncover the 3D organization of the plant pathogenic fungus *Verticillium dahliae* and its influence on transcription and genome evolution. Our analysis reveals the presence of topologically associating domains (TADs) in *V. dahliae.* Interestingly, we observe that TAD boundaries are gene-rich, display lower gene expression and are depleted in genomic variation when compared with the genome-wide average. We observe that TAD boundaries in AGRs are weaker than in the core genome, which is associated with the enrichment of facultative heterochromatin. Interestingly, for most of the *Verticillium* species, we observe that TADs in AGRs physically associate. Moreover, we show that the physical interaction of AGRs is associated with abundant segmental duplications in *V. dahliae* and we confirmed the same physical association in some other *Verticillium* species. Thus, our analysis demonstrates that the 3D organization is conserved within the *Verticillium* genus, and indicates that an increased genomic variation over TAD boundaries and the colocalization of segmental duplications potentially contribute to the evolution of AGRs.

Introduction

The spatial organization of the eukaryotic nuclear genome is directly linked to its biological functions (Lieberman-Aiden et al. 2009; Hoencamp et al. 2021). Besides the organization of genetic elements on the linear DNA strands, the association of DNA and nuclear proteins, such as histones, leads to an organized nucleoprotein complex known as chromatin (Campos and Reinberg 2009). Chromatin states can be broadly divided in the relatively weakly compacted euchromatin, in which genetic features can be transcriptionally active, and the strongly compacted heterochromatin that is generally inaccessible to the DNA-binding components that act in the transcriptional machinery, and therefore transcriptionally repressive (Bannister and Kouzarides 2011). The difference in chromatin organization is mainly mediated by chemical modifications to nucleosomes, the building blocks of chromatin that consist of an octamer of four histone proteins, wrapped by 146 bp of DNA, with unstructured tails sticking out of the complex (Luger et al. 1997). Typically, tri-methylation of lysines 9 and 27 on the tails of histone 3 (H3K9me3 and H3K27me3) are hallmarks of heterochromatin, whereas di-methylation of lysine 4 on the tails of histone 3 (H3K4me2) is a hallmark of euchromatin (Xhemalce et al. 2012; Bannister and Kouzarides 2011). Besides these histone modifications, many more have been described, yet their role in genome organization often remains unclear (Zhao and Garcia 2015).

On a global scale, the genome displays a spatial three-dimensional (3D) structure that brings in close proximity genomic sites that are physically separated on the linear DNA strand, or lie on different chromosomes, and conversely, separates proximal genomic sites through specific folding barriers (Lieberman-Aiden et al. 2009; Jerkovic and Cavalli 2021). Such 3D chromosome structure has revealed multiple levels of organization across genomic scales. These levels of organization range from small-scale chromatin loops of a few kilobases that could contribute to transcriptional regulation based on DNA-DNA contacts, to large-scale subdomains composed of hundreds of kilobases that arrange large chromosomal regions into active or silent chromatin regions in 'A' and 'B' subdomains, respectively (Rao et al. 2014; Holwerda and De Laat 2012; Rowley and Corces 2018; Eagen et al. 2017).

Local 3D chromosome interactions shape chromosome structure into discrete genomic regions, commonly known as topologically associating domains (TADs). TADs are physically self-interacting genomic regions that are delineated by TAD boundaries that display less interaction to adjacent genomic regions (Dixon et al. 2012; Szabo et al. 2019). Although the function of TADs is controversial and still under debate (Ghavi-Helm et al. 2019; Kaushal et al. 2021; Cavalheiro et al. 2021; Arzate-Mejía et al. 2020), several studies have associated the organization of TADs with transcriptional regulation by limiting the interaction of regulatory sequences with their gene targets (Dixon et al. 2012; Rao et al. 2014; Cavalheiro et al. 2021). Other studies implicate TADs in genome replication by keeping origins of replication synchronized and active within TADs (Eser et al. 2017; Yang et al. 2019). Studies into various groups of related organisms uncovered evolutionary conservation of TAD

organization (Harmston et al. 2017; Fudenberg and Pollard 2019; Golicz et al. 2020; Krefting et al. 2018; Liu et al. 2017; McArthur and Capra 2021; Rowley et al. 2017; Rao et al. 2014; Yang et al. 2019; Huynh and Hormozdiari 2019). In metazoans, genes and TADs co-localize by their evolutionary age, suggesting a high degree of conservation of TAD organization (James et al., 2021). Moreover, conserved TADs between different species display similar transcription patterns of genes residing within TADs (Yang et al. 2019; Krefting et al. 2018; Rao et al. 2014; Liao et al. 2021ª; Torosin et al. 2020). Furthermore, the genomic regions comprising particular TADs appear to have reshuffled integrally during metazoan evolution (Liao et al. 2021ª; Krefting et al. 2018; Yang et al. 2019; Szabo et al. 2019). Unexpected changes in TAD organization potentially lead to variations in gene expression patterns, and therefore to important phenotypic modifications (Szabo et al. 2019; Niu et al. 2021). For example, changes in TAD organization are associated with developmental alterations and particular human diseases (Anania and Lupiáñez 2020; Lupiáñez et al. 2016). Recently transient modifications of TAD organization were also associated with quick transcriptional changes upon environmental cues, suggesting an active role of TADs in transcriptional regulation (Li et al. 2015; Liang et al. 2021; Kainth et al. 2021). Thus, it seems clear that changes in 3D genome organization are potentially relevant for phenotypic variations and quick transcriptional responses in changing environments.

In the fungal model organisms *Saccharomyces cerevisiae, Schizosaccharomyces pombe*, and *Neurospora crassa*, the 3D chromosomal organization is linked to heterochromatin distribution (Mizuguchi et al. 2014; Schalbetter et al. 2019; Muller et al. 2018; Rodriguez et al. 2022; Galazka et al. 2016; Klocko et al. 2016). Structurally, heterochromatic regions in TAD boundaries are associated with permissive cohesin and condensin I binding (Mizuguchi et al. 2014; Eser et al. 2017; Schalbetter et al. 2019; Hoencamp et al. 2021). Cohesin and condensin I are architectural DNA-binding proteins widely known to cooperate in chromosome folding, and assist in chromosome organization during meiosis and mitosis (Hirano 2016; Kalitsis et al. 2017). Additionally, heterochromatin is typically associated with repeat-rich regions (Lippman et al. 2004). In the endophytic fungus *Epichlöe festucae*, repeat-rich regions often colocalize with TAD boundaries and are therefore associated with genome folding (Winter et al. 2018). Moreover, genes that are highly expressed *in planta* are enriched near those repeat-regions (Winter et al. 2018). In several other fungi it has been observed that heterochromatic regions are associated with the regulation of environmentally responsive genes, such as *in planta* induced genes or genes expressed upon heat-shock (Kainth et al. 2021; Meile et al. 2020; Kramer et al. 2022; Zhang et al. 2021b; Connolly et al. 2013). Collectively, these findings suggest a structural correlation between the 3D chromosome organization, heterochromatin, and repeat-rich regions, putatively linked with transcriptional regulation in response to environmental challenges in fungal organisms.

The genomes of many plant pathogenic fungi display a two-tier organization in which particular gene-poor and transposable element (TE)-rich genomic regions are evolutionary

more dynamic than the relatively stable core genome (Raffaele and Kamoun 2012; Dong et al. 2015; Torres et al. 2020). Such dynamic regions often contain environmentally responsive genes, including genes that encode *in planta* secreted proteins, and display increased frequencies of nucleotide substitutions, genomic rearrangements, presence/absence polymorphism and are typically associated with facultative heterochromatin (Rouxel et al. 2011; Raffaele and Kamoun 2012; de Jonge et al. 2013; Dong et al. 2015; Dutheil et al. 2016; Peng et al. 2019; Tsushima et al. 2019; Cook et al. 2020; Frantzeskakis et al. 2019; Kramer et al. 2022; Goodwin et al. 2011; Ma et al. 2010). Conversely, TE-poor regions are gene-dense, harbor primary metabolic genes, and are often associated with euchromatin (Seidl and Thomma 2014; Dong et al. 2015; Frantzeskakis et al. 2019; Cook et al. 2020; Schotanus et al. 2015; Kramer et al. 2022). Collectively, this two-tier organization is typically referred to as a 'two-speed genome' (Raffaele and Kamoun 2012; Raffaele et al. 2010; Dong et al. 2015). TE-rich genomic compartments play important roles in the coevolutionary 'arms-race' between pathogens and their hosts, as the increased frequency of genomic variation can more rapidly delete or diversify genes encoding proteins recognized by plants, or it could generate genes encoding proteins with novel functions in pathogenicity (Seidl and Thomma 2017; Croll and McDonald 2012; Frantzeskakis et al. 2019; Torres et al. 2020). Presently, it remains unclear what role the 3D genome organization plays in this genomic compartmentalization.

The asexual soil-borne fungal plant pathogen *Verticillium dahliae* is a notorious vascular wilt pathogen that can infect hundreds of plant species (Fradin and Thomma 2006). Comparative genomics among *V. dahliae* strains has revealed the presence of extensive large-scale genomic rearrangements associated with discrete TE-rich regions (de Jonge et al. 2013; Faino et al. 2015, 2016; Torres et al. 2021). These rearrangements are associated with the occurrence of extensive segmental duplications that underwent substantial reciprocal gene losses leading to a high degree of presence/absence polymorphism that, collectively, contributed to the formation of the genomic compartments formerly known as lineage-specific regions (de Jonge et al. 2013; Faino et al. 2015, 2016; Klosterman et al. 2011). Recent work on the chromatin landscape in *V. dahliae* revealed that these regions display unique chromatin characteristics that are shared by additional regions in the genome that were not previously recognized as lineage-specific based on comparative genomics alone (Cook et al. 2020). Collectively, these regions are now referred to as adaptive genomic regions (AGRs) (Cook et al. 2020). Importantly, AGRs are enriched in *in planta*-induced genes including those that encode effector proteins, but also for genes that are differentially expressed between *in vitro* growth media, suggesting that these regions contain conditionally responsive genes that contribute to host colonization and environmental adaptation (de Jonge et al. 2013; Cook et al. 2020; Torres et al. 2021; Kramer et al. 2022). While a key role of TEs in driving the formation and maintenance of centromeres and AGRs in *V. dahliae* has been revealed (Faino et al. 2016; Cook et al. 2020; Seidl et al. 2020; Torres et al. 2021), it presently remains unclear how chromosome folding affects the organization and evolution of the core genome and AGRs. Here, we explore the chromatin conformation of *V. dahliae* with DNA proximity

ligation followed by sequencing (Hi-C) to uncover the spatial organization of the core genome and the AGRs in detail. In our analysis, we include the whole *Verticillium* genus. Our analysis reveals a unique chromatin conformation of AGRs, and further unveils ancestral conservation of chromosome organization in the core genome in the *Verticillium* genus.

Results

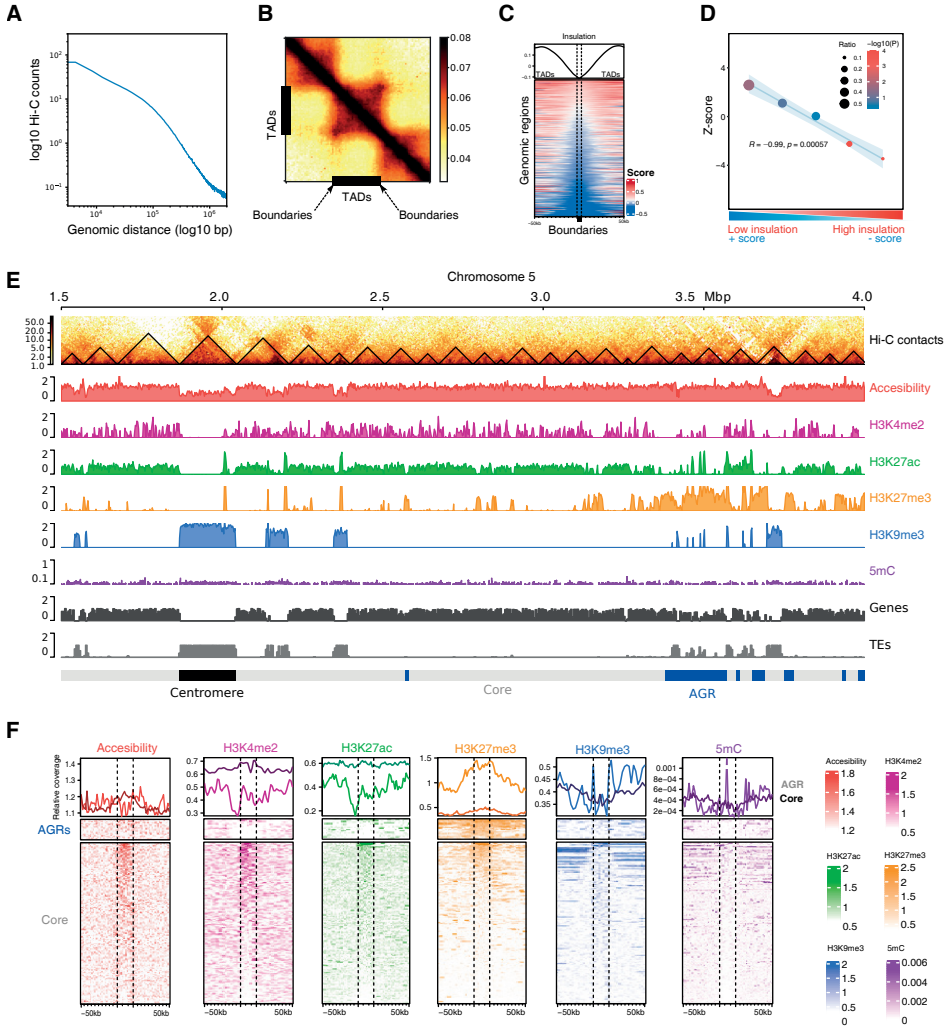
The *Verticillium dahliae* genome is locally organized in topologically associating domains (TADs)

We sought to determine if the differential association of particular histone marks to the core genome and to AGRs in *V. dahliae* (Cook et al. 2020) may be correlated with a differential spatial organization of DNA in the nucleus (Rowley et al. 2017; Depotter et al. 2019). To investigate the chromatin organization in *V. dahliae*, we performed chromatin conformation capture sequencing (Hi-C) in two biological replicates of *V. dahliae* strain JR2 cultivated for 6 days in potato dextrose broth (PDB). As the Hi-C data between the replicates displayed a high correlation (Supplementary Fig.1A-D), we combined these in a single interaction matrix. As expected, we observed a negative correlation between interaction strength and genomic distance (Fig. 1A) that arises from genomic regions interacting strongly with neighboring genomic regions (Dekker et al. 2013). To further investigate the occurrence of discrete clusters of strong DNA interactions, we predicted TAD organization based on the insulation score method (Ramírez et al. 2018). Building on the notion that a TAD is a self-interacting genomic region with sequences that physically interact more with each other than with sequences outside the TAD, we calculated the insulation score of each bin on the Hi-C interaction matrix (average bin size ~4 kb) by determining the interaction strength with the adjacent bins. Bins that display a low insulation score weakly interact with neighboring bins and consequently were assigned as a TAD boundary region. The bins between two TAD boundaries were therefore assigned to a single TAD. Using this approach, we identified in total 353 TADs (mean size 102,394 bp) separated by 347 TAD boundaries (mean size 4,747 bp) that are distributed along the eight chromosomes of *V. dahliae* strain JR2 (Fig. 1B,C,E, Supplementary Fig.1E,F). Taken together, the high reproducibility between two independent samples and the identification of clear TAD boundaries in the combined interaction matrix, suggests a confident prediction of TADs as units of chromosomal organization in *V. dahliae*.

In various eukaryotes, chromatin loops and TAD boundary regions are enriched in sequence motifs for DNA-binding insulator proteins, such as the zinc-finger CTCF and the cohesin ringed complexes in vertebrates, or condensin I in yeasts (de Wit et al. 2015; Dong et al. 2017; Ramírez et al. 2018; Anderson et al. 2019; Hoencamp et al. 2021). To investigate the presence of similar protein-binding motifs in the boundaries of *V. dahliae* TADs, we queried the 347 boundary sequences for potential motifs *de novo* using MEME (Bailey et al. 2015). This analysis revealed two significantly enriched motifs; an 11 bp GAAG

motif that is present in 68.6% of the TAD boundaries (p=6.5×10^{-10}; Supplementary Fig.1G), and a 24 bp TATA motif in 18.2% of the boundaries (p =1.2×10^{-66}; Supplementary Fig.1G). To further analyze these two motifs, we queried these in the JASPAR 2018 and YEASTRACT 2018 databases using TomTom (Fornes et al. 2020; Monteiro et al. 2020; Bailey et al. 2015). The GAAG motif showed a significant match to the Azflp transcription factor of *S. cerevisiae* (p =8.74×10^{-5}), a Zinc-finger protein known to regulate the expression of genes under different carbon sources (Mirzaei et al. 2013; Newcomb et al. 2002; Slattery et al. 2006). Similarly, we found a significant match for the TATA motif (p =2.9×10^{-8}) to an HMG (High Mobility Group) nucleosome remodeler, know to slide and eject nucleosomes to regulate gene transcription (Schlichter et al. 2020; Hepp and Gutierrez 2014; Ragab and Travers 2003). Presence of the motifs coincides with a decline in chromatin accessibility, as determined by the assay for transposase-accessible chromatin (ATAC; Supplementary Fig.1G,H), suggesting DNA-protein interactions at these sites. Even though we find clear enrichment of these motifs in TAD boundaries, the putative function of their protein binding partners does not appear to be that of the DNA-binding insulator proteins CTCF, cohesin and condensin. Although we cannot rule out that these motifs in *V. dahliae* are recognized by proteins that function in TAD separation, it Is more likely that they function in other boundary-associated processes.

We reasoned that the previously reported genome compartmentalization of *V. dahliae* into core genome and AGRs could be associated with a differential TAD organization between these two compartments. Therefore, we investigated the TADs and TAD boundaries within the context of both genomic compartments. Globally, of the 353 TADs and 347 TAD boundaries that we identified in the *V. dahliae* genome, 277 TADs (78.47%) and 308 TAD boundaries (88.76%) could be assigned to the core genome, and 76 TADs (21.53%) and 39 TAD boundaries (11.24%) to an AGR. Interestingly, we observed that AGRs are enriched for relative weakly insulated TAD boundaries (Fig. 1D), indicating that the insulation of TADs in AGRs is 'weaker' than in the core genome. Moreover, we observed a correlation of weak boundaries with smaller TADs in AGRs (Supplementary Fig.1I), suggesting a correlation between insulation and length of TADs.


A
log10 Hi-C counts
Genomic distance (log10 bp)
B
TADs
Boundaries
TADs
Boundaries
C
Insulation
TADs
TADs
Genomic regions
Score
Boundaries
D
Ratio
−log10(P)
Z-score
R = −0.99, p = 0.00057
Low insulation
+ score
High insulation
- score
E
Chromosome 5
1.5
2.0
2.5
3.0
3.5 Mbp
4.0
Hi-C contacts
Accesibility
H3K4me2
H3K27ac
H3K27me3
H3K9me3
5mC
Genes
TEs
Centromere
Core
AGR
F
Accesibility
H3K4me2
H3K27ac
H3K27me3
H3K9me3
5mC
Relative coverage
AGRs
Core
AGR
Core
−50kb
50kb

<<

Figure 1. The *Verticillium dahliae* genome is organized in topological associating domains (TADs).
(A) Distance-dependent interaction frequency (Hi-C contacts) in JR2 shows that most physical interactions occur between regions in close proximity on the linear space. **(B)** Hi-C contact matrix showing local interaction frequency, aggregated over all TADs (black bars) with 50 kb up- and down-stream sequence. The drop in intensity at boundaries at either side of the TADs indicates stronger interaction within TADs than with neighboring genomic regions. **(C)** Heatmap showing insulation scores centered over boundaries with 50 kb up- and down-stream sequence as rows, ordered on insulation score with weakest insulated boundaries (i.e. having the highest insulation score) on top. The top plot displays the average insulation score as shown in the heatmap below. **(D)** TADs in AGRs are less well insulated when compared with TADs in the core genome. The X-axis indicates quintiles of boundaries, separated based on insulation scores. The Y-axis indicates Z-score and the -log10(p-value) color-scale after a permutation test (10,000 iterations). The plot displays a linear regression (blue line) and confidence interval (light blue) as well as the R and p-value after linear regression. **(E)** TAD distribution in V. dahliae strain JR2, with a section of chromosome 5 as an example. From top to bottom: Hi-C contact matrix depicting TADs as black triangles, open chromatin regions as determined with ATAC-seq, H3K4me2, H3K27ac, H3K27me3, and H3K9me3 normalized over a micrococcal nuclease digestion control, GC methylation, as well as gene and transposable element (TE) densities in 10 kb windows. Adaptive genomic regions (AGRs) (Cook et al. 2020) and the centromeric region (Seidl et al. 2020) are indicated in blue and black, respectively. **(F)** Chromatin characteristics are differentially associated with TAD boundaries in the core genome and in AGRs. On top, distribution of each chromatin feature in **(E)** centered for boundaries with 50 kb up- and down-stream sequence, for the core genome (dark color) and AGRs (light color). On the bottom the corresponding heatmaps are shown for the core genome and AGRs.

Chromosome organization is typically associated with chromatin characteristics, such as DNA methylation and histone modifications (Rowley and Corces 2018). Therefore, we analyzed the distribution of a set of histone marks, and of DNA methylation, over the TADs and boundaries. In line with previous results (Cook et al. 2020; Kramer et al. 2021, 2022; Seidl et al. 2020; Torres et al. 2021), we observed that the gene-rich core genome is enriched in H3K27ac and H3K4me2, while the centromeres and TE-rich core regions are enriched in H3K9me3 and DNA methylation, and AGRs are enriched in H3K27me3 (Fig. 1E,F, Supplementary Fig.2). We observed that such broad chromatin associations are maintained similarly on TADs and boundaries, suggesting that chromatin characteristics associate with the overall separation into core genome and AGRs, rather than with TAD organization (Supplementary Fig.3). Interestingly, for the core genome, our analysis revealed that boundary regions show higher chromatin accessibility, and reduced presence of heterochromatin-associated marks (H3K9me3, DNA methylation) than TADs (Fig. 1F, Supplementary Fig.3). In contrast, we did not observe major differences in chromatin accessibility at TADs and boundaries in AGRs. However, boundary regions in AGRs are depleted in H3K9me3 and in activation marks (H3K4me2, H3K27ac), but enriched in H3K27me3 and DNA methylation when compared with TADs in AGRs (Fig. 1F, Supplementary Fig.3). These differences in chromatin state between TADs and boundaries for both the core genome and AGRs suggests that TADs and TAD boundaries may differ in functionality, not only between each other, but also between the two genomic compartments.

Local TAD organization impacts transcriptional regulation

Considering the differential enrichment of particular chromatin marks between TADs and boundaries in AGRs, and the enrichment of AGRs in conditionally responsive genes that contribute to host colonization and environmental adaptation (de Jonge et al. 2013; Cook et al. 2020; Kramer et al. 2022), we hypothesized that genes located in the core genome and in AGRs differ in their transcriptional profile between TADs and boundaries. To investigate this, we first determined the occurrence of genes over TADs and boundaries for both genomic compartments. Previous studies uncovered that the core genome contains more genes and fewer TEs compared with AGRs (de Jonge et al. 2013; Faino et al. 2016; Torres et al. 2021). Interestingly, within the core genome we observed a significant enrichment of genes, and a corresponding depletion of TEs, in the TAD boundaries when compared with TADs (Fig. 2A). However, within AGRs we do not observe enrichment or depletion of genes and TEs in boundaries (Fig. 2A). Genes localized at TAD-like boundaries are enriched in functions associated with carbohydrate metabolism, RNA processes, and translation and mostly depleted in functions associated with pathogenicity (Fig. 2B). The genetic differences between TADs and boundaries in the core genome, and the epigenetic differences between TADs and boundaries in AGRs, suggests that TAD organization may impact transcriptional regulation.

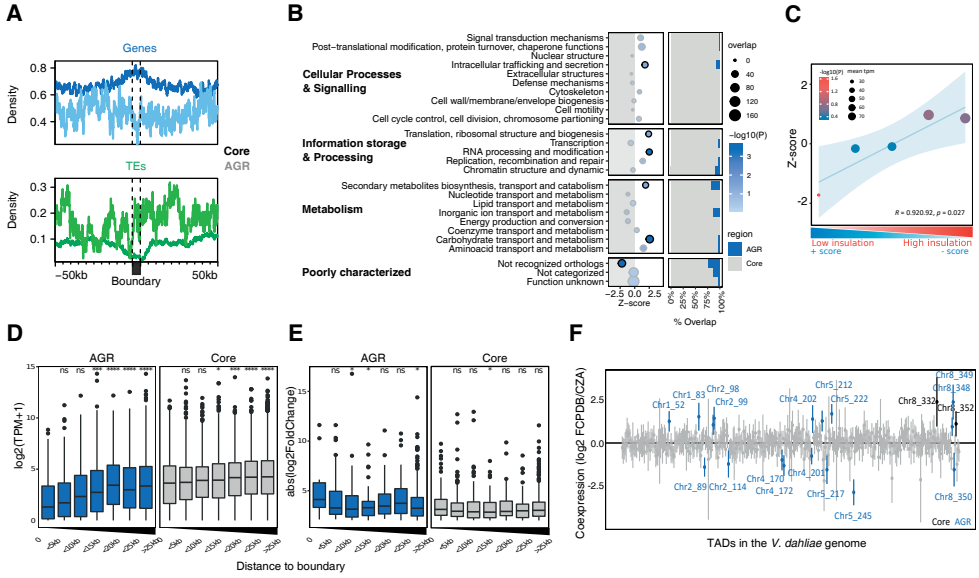


Figure 2. Topological associating domain (TAD) organization affects transcription in *Verticillium dahliae*.
(A) Average density of genes and transposable elements (TEs) per 1 kb window centered over boundaries with 50 kb up- and down-stream sequence in the core genome and in adaptive genomic regions (AGRs) of *V. dahliae* strain JR2. The core genome (dark color) and in AGRs (light color). **(B)** Cluster for orthologs (COGs) gene enrichment in TAD-like boundaries. The y-axis depicts the COG categories and the x-axis the z-score after a permutation test (10,000 iterations); negative z-score indicates depletion, while positive z-score shows enrichment. The significance is shown as –log10(FDR adjusted *p-value*) and is color-coded; circle size is relative to the number of genes overlapping in boundaries per each category. The bar charts on the right indicates the proportion of genes in boundaries that occur in the core genome and AGRs. **(C)** Genes in lowly insulated boundaries are lower expressed than those in more highly insulated boundaries. Association between TAD boundary quintiles, separated on insulation score, and transcription of genes located in TAD boundaries. The Y-axis depicts the Z-score, colour of the datapoints indicates the -log10(*p-value*) after a permutation test (10,000 iterations) and size of datapoints indicates mean transcription value (TPM) of represented genes. A linear regression (blue line), with 95% confidence interval (light blue), between boundary quintiles is displayed. **(D)** Transcription values for *V. dahliae* cultivated for 6 days in potato dextrose broth (PDB) and **(E)** absolute log2-fold change in expression between cultivation in PDB or in Czapec-Dox medium (CZA), for all genes grouped based on their distance to the closest boundary in the core genome (grey) or in AGRs (blue). Statistically significant differences in average transcription level for the distance groups was compared to the group of genes located in boundaries (distance 0) and determined by the Wilcoxon Rank-Sum test (* $p < 0.05$). **(F)** Linear regression effect size of each TAD on differential gene expression between cultivation for 6 days in PDB or in CZA. Mean effect size of each TAD is shown as a point, with 95% confidence interval, and TADs with a significant effect (95% confidence interval is significantly different from 0) are shown in color and labelled by corresponding chromosome and TAD number, for TADs in the core genome (black labels) and in AGRs (blue labels).

To further study the impact of TAD organization on gene expression, we queried previously generated expression data of *V. dahliae* cultivated for 6 days in PDB (Kramer et al. 2022), which is the same cultivation condition as used for our Hi-C data. To this end, we performed a Uniform Manifold Approximation and Projection for Dimensional Reduction (UMAP) on all *V. dahliae* genes, based on DNA methylation, CRI (Composite Repeat Index), H3K27ac, H3K27me3, H3K4me2, and H3K9me3. In line with previous observations, genes are mainly separated in clusters representing the core genome and AGRs (Supplementary Fig.4) (Cook et al. 2020). Additionally, when considering only boundary genes, we observe that the UMAP cluster containing AGR genes is enriched for H3K27me3, while the cluster containing core genes is enriched for H3K27ac. Furthermore, we observe a clear separation based on transcriptional activity, with genes in boundaries of the core genome generally displaying higher transcription levels, associated with increased H3K4me2 levels, than genes in AGR boundaries (Supplementary Fig.4). As we reported above that AGRs are enriched for weak boundaries (Fig. 1D), we sought to investigate a potential relationship between gene expression and insulation strength of boundaries. Interestingly, we observed a positive correlation between boundary insulation and expression level (Fig. 2C, Supplementary Fig.4), indicating that genes in weakly insulated TAD boundaries are generally lower expressed than genes in strongly insulated boundaries. To investigate if transcriptional activation is dependent on distance to boundaries, we grouped *V. dahliae* genes based on their distance to the closest boundary and quantified their expression. We observed that genes located within TAD boundaries in the core genome and in the AGRs, are lower expressed than those located further in the TADs (Fig. 2D). However, the expression of genes within TAD boundaries in the core genome is notably higher than the expression of genes within boundaries in AGRs, whereas genes further away from boundaries in the core genome and AGRs are similarly expressed (Fig. 2D). Taken together, these findings indicate that genes localized in proximity to weak AGR boundaries are strongly silenced *in vitro*, yet strong boundaries in the core genome are equally transcriptionally active as the adjacent TADs.

As TADs are thought to function as regulatory units for differential gene expression (Cavalheiro et al. 2021; Gonzalez-Sandoval and Gasser 2016), we hypothesized that differentially expressed genes (DEGs) are enriched within TADs, and thus are depleted in boundaries. To test this hypothesis, we first examined the occurrence of differentially expressed genes (DEGs) between cultivation in potato dextrose broth (PDB) and in Czapec-Dox medium (CZA) *in vitro.* Previously, we have shown that DEGs are enriched in AGRs as well as in H3K27me3 domains in the core genome (Cook et al. 2020; de Jonge et al. 2013; Kramer et al. 2022), indicating that TADs in both the core genome and in AGRs may function as regulatory units for differential gene expression. We identified a total of 1,844 DEGs (1,005 higher expressed in PDB and 839 higher expressed in CZA), yet we did not observe differences in DEG presence between TADs and boundaries in the core genome nor in AGRs (Fig. 2E). Additionally, when inspecting the absolute log2 fold-change for the expression of

genes in relation to their distance to the closest boundary, we observed that only genes at a distance of 5-15 kb and >25 kb from TAD boundaries in AGRs are significantly stronger differentially expressed than genes located in AGR boundaries (Fig. 2E), suggesting that differential gene expression *in vitro* does not depend on the local physical TAD organization, and thus rather is a general feature of genes In AGRs.

To investigate whether genes localizing within the same TAD in *V. dahliae* display transcriptional co-regulation, we fitted a linear model in which differential expression between cultivation in PDB and CZA of each gene is predicted by TAD association. Co-expressed genes in TADs will result in a positive or negative score for that TAD, depending on the prevalent direction of differential gene expression, while opposite direction of differential gene expression within a TAD results in a mean effect of zero. We identified 19 TADs with confidence intervals that are not zero, and therefore have a significant effect on transcription (Fig. 2F). Of these TADs, 17 are associated with AGRs and two with the core genome (Fig. 2F), suggesting that transcriptional co-regulation of expression mainly occurs in AGRs. To corroborate these findings, we checked whether TADs in the core genome and those in AGRs contain genes that display a common transcriptional pattern of higher expression in PDB than in CZA, or vice versa. First, we selected only TADs with more than five DEGs, as we did not consider TADs with fewer DEGs to be co-regulated. In total, 68 out of 258 (26.4%) in the core genome, and 64 out of 96 (6.7%) TADs in AGRs, contain more than five DEGs (Supplementary Fig.4B). Of these TADs, 13 out of 68 (19.1%) and 19 out 64 (29.7%) in the core genome and AGRs, respectively, contain more than twice the number of genes that are higher expressed in one of the growth media than in the other one, and thus display co-regulation of differential transcription (Supplementary Fig.4B). Taken together, our results suggest that TAD organization affects transcription *in vitro*, and that although some TADs display transcriptional co-regulation of gene expression, this occurs only for a subset of TADs that predominantly locate in AGRs.

TAD boundaries are depleted of genomic variation

TADs are often considered to be conserved between closely related organisms (Spielmann et al. 2018; McArthur and Capra 2021). To study whether TADs are conserved between strains of *V. dahliae*, we performed chromatin conformation capture sequencing (Hi-C) in two biological replicates of *V. dahliae* strain VdLs17 that is 98% syntenic to strain JR2 (Fig. 3A) (Klosterman et al. 2011; de Jonge et al. 2013; Faino et al. 2015). Following the same methodology as for strain JR2, we combined the biological replicates (between replicate correlation >0.89; Supplementary Fig.5A,B) in a single interaction matrix and predicted 365 TADs (mean length=98,558 bp) and 357 boundaries (mean length=4,506 bp) (Supplementary Fig.5C). Remarkably, the TAD organization in VdLs17 displays similar patterns of insulation scores, gene-enrichment, TE-depletion, and DNA motif enrichment in boundaries as in strain JR2, suggesting that these TAD characteristics are conserved in *V. dahliae* (Supplementary Fig.5C-H). To investigate if also TAD localizations are conserved

between the two strains, we normalized the boundary distribution over syntenic regions and tested their overlap between VdLs17 and JR2 (*n*=342 and *n*=330 TADs in syntenic portion, respectively). Interestingly, we observed a significant overlap between the boundaries of the two strains (*n*=225, *p*=9.6×10-4, one-way Fisher exact test; Fig. 3A), and an overall overlap in TAD positions (Fig. 3B; Supplementary Fig.5G). Originally, AGRs in strain JR2 were defined based on absence of synteny with genomic regions in other *V. dahliae* strains (de Jonge et al. 2013; Klosterman et al. 2011; Faino et al. 2016), whereas more recently AGRs have been defined based on their epigenetic profile (Cook et al. 2020). As we do not presently have the required complex epigenetic data to determine AGRs in strain VdLs17, we divide the genome of VdLs17 into syntenic and non-syntenic compartments based on the genomic comparison to JR2. Non-syntenic compartments of the VdLs17 genome are enriched for weak TAD boundaries (z-score=2.3858, *p*=0.00001, permutation test after 10,000 iterations; Supplementary Fig.5H), which is similar to what we observed for AGRs in the JR2 strain.

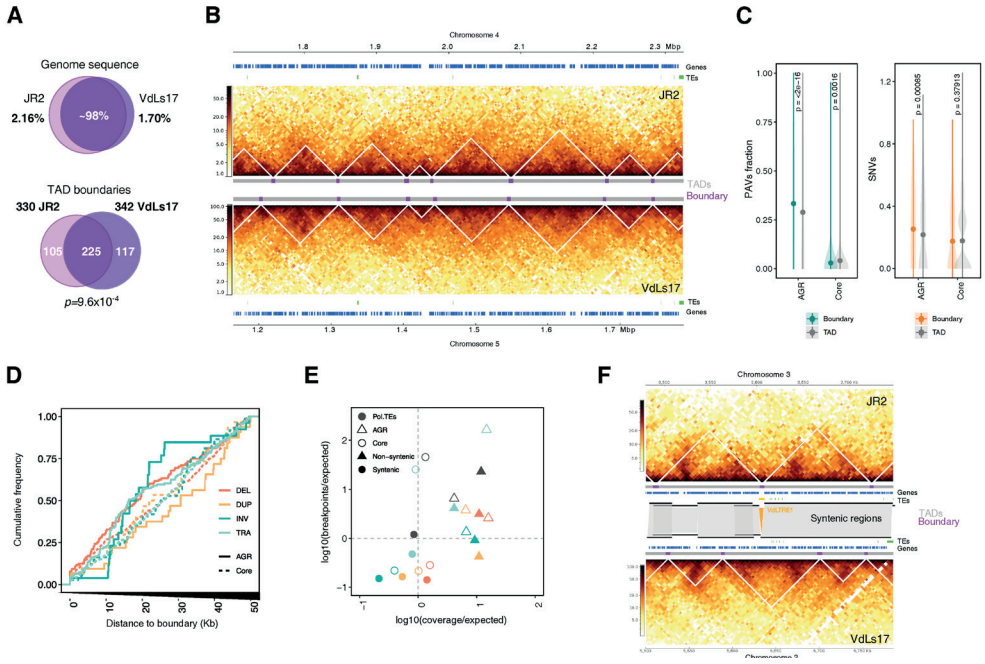


Figure 3. TAD organization is conserved between two *Verticillium dahliae* strains.
(A) Top: V. dahliae strains JR2 and VdLs17 are highly similar as 97.84% and 98.30% of their respective genomes are syntenic (Faino et al. 2016). Bottom: Most of the TAD boundaries overlap significantly between JR2 and VdLs17. **(B)** Syntenic block between JR2 chromosome and VdLs17 chromosome 5 shows conserved distribution of TADs and boundaries. Heatmaps represent contact matrixes of JR2 (top) and VdLs17 (bottom) with TADs (white triangles). TADs are also displayed as grey bars between heatmaps. Genes and transposable elements (TEs) are displayed above. **(C)** Boundaries are not enriched for genomic variation in a set of 42 V. dahliae strains. presence/absence variation (PAVs) (Cook et al. 2020). Right: single nucleotide variants (SNVs) (Torres et al. 2021). **(D)** Cumulative frequency plot of structural variant (SV) breakpoints over distance from boundaries in the core genome (dashed line) and in AGRs (solid line), overlaps with boundaries (distance = 0) are included. SVs are separated in deletions (DEL, orange), duplications (DUP, yellow), inversions (INV, green) and translocations (TRA, blue) (Torres et al. 2021). **(E)** TAD boundaries in AGRs contain more SVs than expected by chance, whereas TAD boundaries in the core genome contain fewer SVs than expected by chance. Color code of SVs is similar as in **(D)** for SVs in boundaries in the core genome (open circles) and in AGRs (open triangles), as well as for boundaries in syntenic (solid circles) and non-syntenic (solid triangles) genomic regions and for polymorphic TEs (grey circles (Torres et al. 2021)). **(F)** Synteny breaks caused by the insertion of transposable elements may give rise to new TAD boundaries in V. dahliae strain JR2. Heatmaps represent contact matrixes of JR2 (top) and VdLs17 (bottom) with TADs (white triangles), and TADs, genes and TEs are displayed in between. Synteny between JR2 and VdLs17 indicated as grey blocks. A VdLTRE1 insertion in strain JR2 is indicated in yellow.

As genomic rearrangements directly impact genome organization, while we observe that TAD boundaries are conserved between *V. dahliae* strains JR2 and VdLs17, we hypothesized that TAD boundaries may be depleted of such genomic variation. Previously, genomic comparisons between *V. dahliae* strains have revealed extensive genomic rearrangements and structural variations (SVs) (de Jonge et al. 2013; Faino et al. 2015; Torres et al. 2021; Depotter et al. 2019; Shi-Kunne et al. 2018). To study the association between SVs and TAD boundaries, we used previously generated data for a set of 42 *V. dahliae* strains (Cook et al. 2020; Torres et al. 2021) to query the distribution of single nucleotide variants (SNVs) and presence/absence polymorphisms (PAVs) over TAD boundaries in *V. dahliae* strain

JR2. The PAV data was generated by summarizing genomic segments of *V. dahliae* strain JR2 that were absent in the other strains and therefore concerns absence counts relative to the genome of strain JR2 (Cook et al. 2020). Interestingly, we observed a depletion of SNVs and of PAVs in TAD boundaries in the core genome (Fig. 3C,D). However, TAD boundaries in AGRs showed more PAVs (Supplementary Fig.6C), together with a lower nucleotide diversity (Supplementary Fig.6A,B), indicating that TAD boundaries in the core genome are strongly depleted in genomic variation, while boundaries in AGRs are evolutionary less stable. As boundaries in the core genome and in AGRs differ with respect to genomic variation, we next questioned if presence of genomic variation in boundaries correlated with insulation. However, we observed no significant correlation between insulation on the one hand and either SNVs (R = -0.56, p = 0.16, Supplementary Fig.6B) or PAVs (R = -0.76, p = 0.068, Supplementary Fig.6C) on the other hand.

To continue our investigations into genomic variation in relation to TAD organization, we next studied different categories of SVs. We observed that deletions, duplications, inversions, and translocations occur more commonly in TADs than in boundaries, indicating depletion of genomic variation from TAD boundaries (Fig. 3D). To assess whether this depletion may be due to purifying selection, we calculated the expected amount of SV breakpoints and SV coverage occurring in boundaries, based on their genome wide occurrences, and compared this to the measured number of breakpoints and their coverage. SV categories displaying fewer breakpoints and lower coverage than expected in boundaries are considered to be under purifying selection, as SVs displaying negative selection are expected to display lower frequency of occurrence as well as lower total coverage (Fudenberg and Pollard 2019). As expected, the different SV categories appear to be under purifying selection in boundaries in syntenic regions (Fig. 3E). Conversely, SVs occur more commonly over TAD boundaries in non-syntenic regions and in AGRs, which is in line with previous observations that suggested that SVs in non-syntenic regions are tolerated (Fig. 3E) (Torres et al. 2021). Thus, our results show that SVs in boundaries in the syntenic core genome undergo purifying selection, suggesting that genomic stability at TAD boundaries is important for species adaptation.

Prior studies have indicated that SVs in *V. dahliae* often colocalize with polymorphic TEs that display PAV between strains and are characterized as evolutionary young, scarcely methylated and highly expressed (Torres et al. 2021). As TE activity may have been involved in the generation of SVs (Faino et al. 2016; Torres et al. 2021), we investigated whether polymorphic TEs occur more frequently in TADs than in boundaries. We identified 36 polymorphic TEs (21.8% of the total) that display PAV between *V. dahliae* strains JR2 and VdLs17, mainly concerning TEs present in *V. dahliae* strain JR2 that are absent from VdLs17. However, we observed no overrepresentation nor depletion of polymorphic TEs in boundary regions (Fig. 3E). Nevertheless, interestingly, some TE insertions in *V. dahliae* strain JR2 occur at a 'new' TAD boundary or at a site of boundary rearrangement (Fig. 3F, Supplementary Fig.6D,E), suggesting that polymorphic TEs may drive changes in the TAD organization.

TAD organization in the core genome is evolutionary conserved in the *Verticillium* genus

Genomic regions that are syntenic between species, often also display conservation of TAD organization (Dixon et al. 2012; Yang et al. 2019). As we observed that TADs and boundaries are conserved between strains of *V. dahliae* (Fig. 3A,B), we hypothesized that TAD organization is similarly conserved in other within the *Verticillium* species. Therefore, we aligned the genomes of all ten *Verticillium* species and retained the genomic regions that aligned to *V. dahliae* strain JR2 (Supplementary Fig.6F). As expected, relatively distantly related species such as *V. albo-atrum* shared less syntenic content (78.30%) with *V. dahliae* JR2 than closely related species such as *V. alfalfae* (94.56%). The allodiploid species *V. longisporum*, which is composed by three lineages that each arose from a different hybridization event (strain VLB2=A1/D1, PDB589=A1/D3) between two *V. dahliae* strains or between *V. dahliae* and an unknown species (Depotter et al. 2021), shared 50.57% and 44.70% for the strain VLB2 and PD589, respectively (Supplementary Fig.6F). To investigate conservation of TAD boundaries between the *Verticillium* species, we used the generated alignments to calculate conservation score (Siepel et al. 2005). As expected, we observed a higher conservation score in the core genome than in AGRs (Fig. 4A). Interestingly, we observed a higher conservation score in TAD boundaries in the core but not in the AGRs (Fig. 4A), which is similar to our comparison between *V. dahliae* strains JR2 and VdLs17 (Fig. 3C-E; Supplementary Fig.6G), indicating that TAD organization is conserved within the *Verticillium* genus as well.

To further assess the conservation of TAD organization in the *Verticillium* genus, we used previously generated Hi-C data (Seidl et al. 2020; Depotter et al. 2021). Using the same approach as for *V. dahliae* strains JR2 and VdLs17, we performed matrix normalization, correction, and insulation score calculation for each species. As TAD boundaries of *V. dahliae* strain JR2 display increased sequence conservation when compared with TADs, we hypothesize that conserved TADs can be recovered as lowly insulated regions in the Hi-C data of other *Verticillium* species. As we did not generate replicates of Hi-C data for these species, we were unable to identify their TAD organization as reliably as for *V. dahliae* strains JR2 and VdLs17 using the insulation method. Therefore, we conceived a sequence-based method to predict boundaries in the other *Verticillium* species (Supplementary Fig.7). To this end, we compared the sequence of all TAD boundaries in *V. dahliae* strain JR2 to the genome sequences of the other *Verticillium* species. We first removed the boundaries containing TE insertions, as TEs often occur abundantly in the genome, which may lead to improper identification of TAD boundaries. Additionally, as we do not expect TAD sizes to differ fundamentally among *Verticillium* spp., we removed putative boundaries that were separated from contiguous boundaries by a distance shorter than the smallest TAD in *V. dahliae* strain JR2 (Supplementary Fig.7A). Finally, we used the Hi-C data of each species to assess whether the insulation score of the predicted boundaries is lower than for adjacent genomic regions (Supplementary Fig.7A). To verify the validity of our sequence-based

method for boundary identification, we first used it on the genome of *V. dahliae* VdLs17, of which the boundaries were also annotated using the insulation method (Fig. 3). Using this method, we recovered 269 boundaries in VdLs17 (Supplementary Fig.7B-D) that display significant positional overlap with the boundaries as determined by the insulation method (z-score=27.1264, p=9.99×10-5; Fig. 4B).

As our sequence-based method for boundary identification proved reliable for *V. dahliae* strain VdLs17, we next used this method to predict boundaries in the other *Verticillium* spp. As expected, we observed that boundaries of *V. dahliae* strain JR2 are more likely to be shared with phylogenetically close species (Fig. 4C). For instance, only 80 boundaries were recovered in the more distantly related *V. albo-atrum*, whereas 254 boundaries were predicted in *V. alfalfae* and 283 boundaries in *V. longisporum* strain PD589, that are both relatively closely related to *V. dahliae* (Fig. 4C). In general, the predicted boundaries in the *Verticillium* genus depict a drop in insulation score with the adjacent genomic regions (Supplementary Fig.7E), indicating that we confidently assign TAD boundaries. Additionally, we could recover full length TAD structures in the syntenic regions of the other species when compared with JR2, possibly indicating a high conservation of TADs in the core genome (Fig. 4D). Collectively, our results suggest that boundary sequences display a high degree of conservation between species, especially those that are phylogenetically close, and that not only the sequence, but also the depletion of physical interactions in the boundaries is conserved.

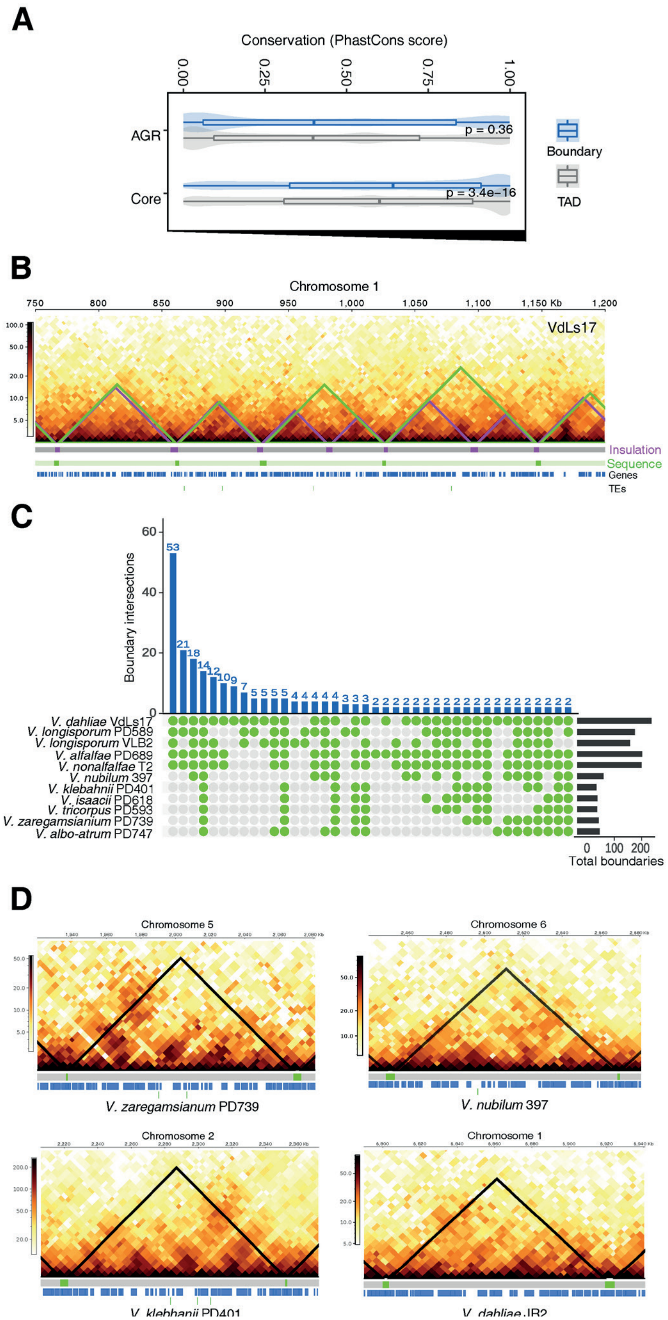


Figure 4. TAD boundaries show signs of conservation in the *Verticillium* genus.
(A) TAD boundaries are more conserved than TADs. Boxplots display the conservation score of each TAD (grey) and boundary (blue) in the core genome and in AGRs. P-values shown represent the result of a one-way Wilcoxon rank sum test. **(B)** Overlap in TAD prediction for strain VdLs17 using the insulation method (purple triangles) and the JR2-sequence-based method (green triangles), with partial chromosome 1 of the strain VdLs17 (755,315-1167941 bp) as example. Tracks depict in grey and purple, TADs and boundaries by the insulation method respectively; in light green and green, TADs and boundaries by the sequence-based method respectively; Genes in blue and TEs in green. **(C)** TAD boundaries predicted in the ten Verticillium species ordered according to JR2-sequence-based method, on the bottom-right the total number of boundaries predicted in each species. On the bottom, a combination matrix depicting combinations of species (in green) in which boundaries are shared. On the top, the total amount of boundaries shared for each species combination. **(D)** (D) TADs are conserved in syntenic regions in the Verticillium genus. One syntenic TAD is shown in four species of the Verticillium genus. Black tringle indicates the TAD. Grey and green tracks depict TADs and boundaries respectively. Genes in blue and TEs in green.

Verticillium dahliae adaptive genomic regions colocalize within the nucleus

Next to investigating local genome architecture, we sought to study the overall organization of the DNA in the nucleus by assessing physical associations between regions that are distant in the linear genome, here defined as >20 kb apart on the same chromosome or those which are located on separate chromosomes (Lieberman-Aiden et al. 2009). We have previously demonstrated that *Verticillium* centromeres colocalize within the nucleus (Seidl et al. 2020). Thus, making use of these interactions as references, we identified additional genomic regions that consistently colocalize based on a measurement of expected-observed interaction counts. This measure considers contact frequencies and the relative distance between regions by identifying those that interact at a higher frequency than the expected average centromere to centromere interaction rate. For *V. dahliae* strain JR2, this resulted in 1,842 long-range colocalization events, among which are 953 centromeric interactions (Fig 5A, Supplementary Table 1). Interestingly, 475 of the 889 remaining non-centromeric interactions (53.4%) involve AGRs (Fig 5A,B, Supplementary Table 1), which is a strong overrepresentation (chi-squared test; $p<0.05$, Supplementary Table 2) given that ARGs represent only 3.33 Mb (9.6%) of the 36.15 Mb total genome size of *V. dahliae* strain JR2 (Cook et al. 2020). Moreover, colocalization events that associate AGR regions with other AGR regions comprise nearly one-third of the non-centromeric interactions (28.1%), whereas AGR-core and core-core co-localization events make up 25.3% and 46.6%, respectively.

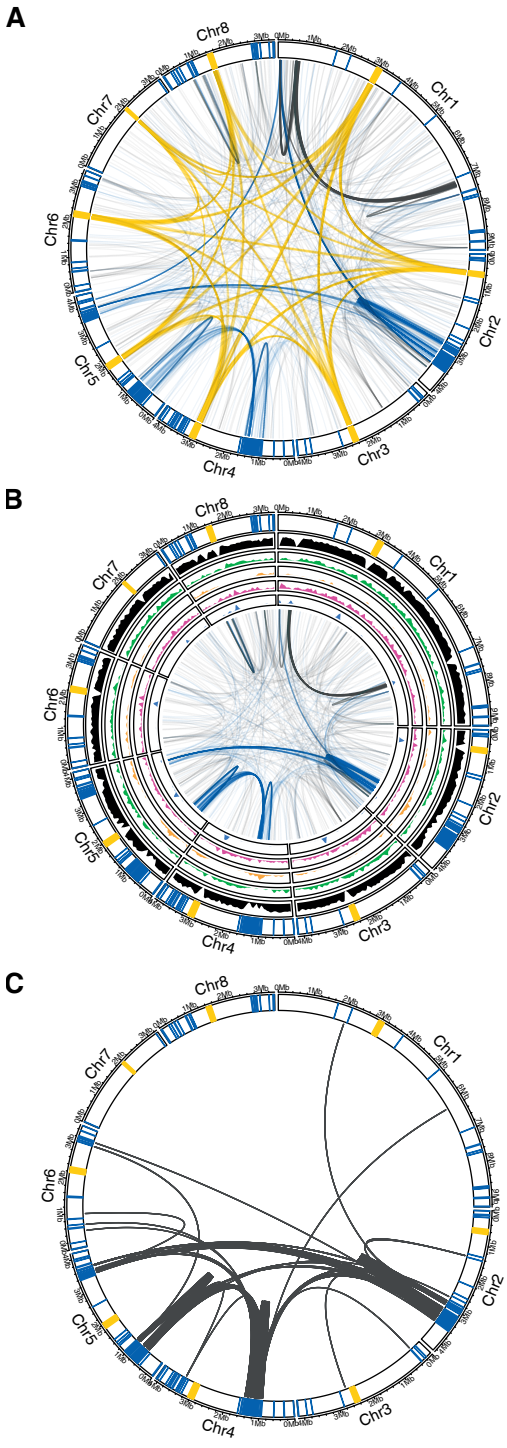


Figure 5. Adaptive genomic regions physically colocalize within the nucleus of *Verticillium dahliae* strain JR2. All circular plots display in the outer track the eight chromosomes of *V. dahliae* strain JR2 with centromeres highlighted in yellow, adaptive genomic regions (AGRs) in blue, and core regions in white. **(A)** Long-range interactions that exceed the average interaction strength of centromeres are shown as edges. Edges for centromeric interactions are shown in yellow, AGR interactions in blue, and core interactions in grey. **(B)** Non-centromeric long range interactions are shown as edges, with inner tracks depicting histone modification densities. From outside to inside: gene density (10kb; black), H3K27ac, green; H3K27me3, yellow; H3K4me2, purple; and H3K9m3, blue. **(C)** Edges represent segmental duplications as identified previously in Faino *et al.* (2016).

Intriguingly, whereas all centromeres interact with each other (Fig. 5A) (Seidl et al. 2020), AGRs interactions concern bipartite colocalization events (Fig 5B). For example, AGR regions from chromosome 2 (located at 2,719,271–3,070,641 bp and 3,232,606–3,474,600 bp) colocalize with AGR regions on chromosome 5 (3,672,718–3,699,257 bp), whereas other AGR regions on chromosome 5 (556,510–987,441 bp) colocalize with AGR regions on chromosome 4 (1,361,005–1,401,989 bp). The function and colocalization of centromeres in *V. dahliae* strain JR2 correlates to the presence of CENH3 nucleosomes, presence of the AT-rich transposable element *VdLTRE9*, and the heterochromatin-associated H3K9me3 (Seidl et al. 2020). To identify epigenetic drivers of long-range interactions, we assessed the presence of the histone marks H3K9me3, H3K27me3, H3K4me2, and H3K27ac as well as chromatin accessibility (Cook et al. 2020; Seidl et al. 2020; Kramer et al. 2021, 2022) with the identified long-range interactions. Importantly, we found no correlation between colocalizing AGR regions and any of the assessed histone modifications (Fig 5B). In contrast, colocalizing regions in the core genome displayed a significant increase of H3K9me3 (t-test, adjusted $p = 6.68\times10^{-127}$), and a decrease of H3K27me3 ($p = 0.007$), H3K4me2 ($p = 4.39\times10^{-6}$), H3K27ac ($p = 0.001$), and chromatin accessibility ($p = 1.07\times10^{-28}$), but given the low number of core-core interactions, this pattern is strongly determined by the two colocalizing regions on chromosome 1 (Seidl et al. 2020). Collectively, our results suggest that different mechanisms govern the colocalization events of centromeric regions and of AGRs in *V. dahliae*.

The evolution of *V. dahliae* involved large-scale genomic rearrangements including segmental duplications and reciprocal gene loss (Faino et al. 2016; Torres et al. 2021; Shi-Kunne et al. 2018). Given that segmental duplications involving distal regions likely require proximity of these regions, colocalization of AGR regions may be a driver for such events (Ebert et al. 2014). Therefore, we hypothesize that the colocalization of AGR regions may be associated with duplication events. Intriguingly, we observed that colocalizing AGR regions are associated with duplication events (Fig, 5C); of the 475 colocalization events that involve AGR regions, 260 involve segmental duplications (Fig. 5C, Supplementary Fig.8A, Supplementary Table 3), which is a significant enrichment not only genome-wide (z-test, $p<0.05$), but even within AGRs (z-test, $p<0.05$). Moreover, whereas genome-wide 264 interactions were recorded that involve segmental duplications, 260 (98.5%) concern AGRs.

To assess whether similar colocalization of AGRs occurs in other strains than JR2 in the *V. dahliae* population, we analyzed the HiC data of the strain VdLs17. Following the same methods as in JR2, we identified 1,978 colocalization events, indicating that all centromeres interact as expected in VdLs17 (Fig. 6A) (Seidl et al. 2020). Remarkably, similar as in JR2, the VdLs17 non-centromeric colocalization events are enriched for AGR interactions that involve 452 out of 1,451 (31.2%) non-centromeric colocalization events (chi-squared test; $p = 8.5\times10^{-56}$). Moreover, we also observed that AGRs that interact are strongly associated with genomic duplications (Fig. 6B, Supplementary Fig.8B, Supplementary Table 1), indicating that long-range AGR organization within the nucleus is conserved among *V. dahliae* strains.

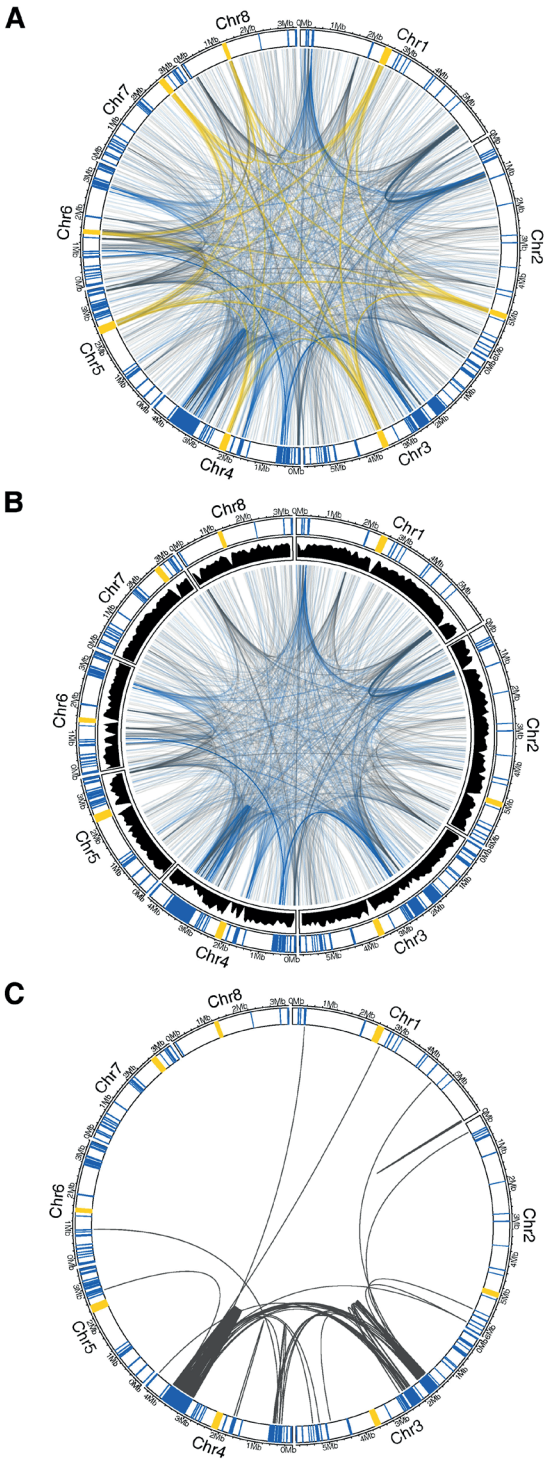


Figure 6. Adaptive genomic regions physical interaction is conserved in the *Verticillium dahliae* strain VdLs17. All circular plots display in the outer track the eight chromosomes of *V. dahliae* strain VdLs17 with centromeres highlighted in yellow, adaptive genomic regions (AGRs) in blue, and core regions in white. **(A)** Long-range interactions that exceed the average interaction strength of centromeres are shown as edges similar colored as in Figure 5. Centromeric interactions are shown in yellow, AGR interactions in blue and core interactions in grey. **(B)** Non-centromeric long range interactions are shown as edges. The inner track depicts coding region density (black). **(C)** Edges represent segmental duplications as in Faino *et al.,* 2016

The colocalization of adaptive genomic regions is evolutionary conserved in the Verticillium genus

To assess whether the long-range colocalization patterns observed in *V. dahliae* similarly occur in other *Verticillium* species (Fig 7, Supplementary Table 3,4), we also assessed the Hi-C for other species. To consistently define AGRs throughout the *Verticillium* genus, we queried the genome assembly of each species for the JR2-AGRs, and in addition included the non-syntenic regions. We observed a similar overrepresentation of AGRs in long-range interactions as observed for *V. dahliae* for several other *Verticillium* species (Supplementary Table 5), namely *V. albo-atrum*, *V. klebahnii*, *V. nubilum*, and *V. nonalfalfae*. Interestingly, also in these species, we found that AGR interactions are dominated by interactions between segmental duplications. Moreover, although we were not able to demonstrate enrichment for interactions between AGRs in *V. tricorpus*, *V. alfalfae*, *V. isaacii*, and *V. longisporum*, long-range interactions also occur predominantly among segmental duplications in all of these except for *V. alfalfae* (Supplementary Fig.9). This implies that the colocalization of AGRs and the association with segmental duplications could contribution to the formation of AGRs (Faino et al. 2016; Depotter et al. 2019), and thus our results indicate that the local and long-range 3D organization of the genome impact the evolution of *V. dahliae* and the *Verticillium* genus.

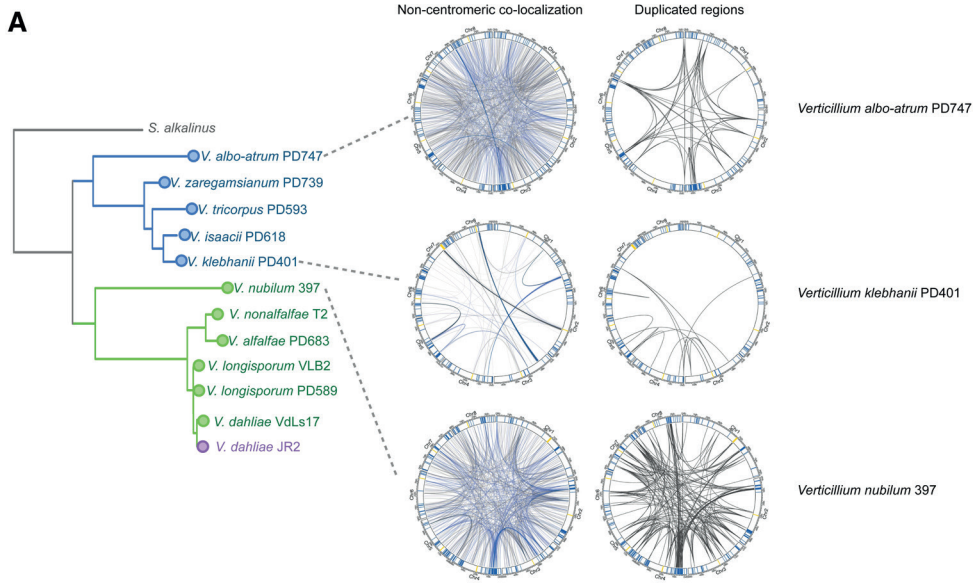


Figure 7. Adaptive genomic regions physically colocalize within the nucleus across the *Verticillium* genus.
(A) Simplified *Verticillium* phylogeny showing the relationships between *Verticillium* species. Non-centromeric interactions occur preferentially between adaptive genomic regions (AGRs) in different *Verticillium* species. For all circos plot, outer track depicts centromeres in yellow, AGR regions in blue, and core regions in white. For every genome, non-centromeric interactions exceeding the average interaction strength of centromeres are showed. On the left: simplified *Verticillium* species phylogeny, Flavexudans and Flavnonexudans clades in blue and green, respectively. Purple dot depicts *V. dahliae* position as reference (Supplementary Fig.6F). On the right: segmental duplications within each genome for every species.

Discussion

In eukaryotes, the DNA within the nucleus typically is subject to three-dimensional (3D) folding, resulting in the local organization of chromosomes into so-called topologically associating domains (TADs), and in the global organization of chromosomes, as for example in the bundle of heterochromatic and euchromatic regions within the nucleus. Here, we used chromatin conformation capture and high-throughput sequencing (Hi-C) to identify the local and global 3D genome organization in the fungal plant pathogen *Verticillium dahliae*. Our analyses reveal the presence of TADs in *V. dahliae*. Remarkably, we find that TADs in the evolutionary dynamic AGRs of *V. dahliae* are less well insulated than TADs in the core genome, indicating that TADs in AGRs are not as well established as those in the core genome. Moreover, TADs in AGRs display significantly more co-regulation of gene expression than TADs in the core genome. Furthermore, genes located in TAD boundaries are generally lower expressed in AGRs *in vitro*, while stronger differentially expressed between *in vitro* conditions, than genes located in TADs in AGRs. We find that TAD boundaries are depleted for structural variation between *Verticillium* species, and that TADs are generally conserved in the *Verticillium* genus. Interestingly, we show that the physical interaction of AGRs is associated with abundant segmental duplications in *V. dahliae*, as we similarly confirmed the same physical association in other *Verticillium* species. Collectively, our results provide evidence for the local organization of the genome in TAD units and show that 3D organization of AGRs is potentially associated with the genome evolution of the *Verticillium* genus.

The existence of TADs has been demonstrated in numerous eukaryotes (Wang et al. 2018; Dixon et al. 2012; Liao et al. 2021a; Dong et al. 2017), and in fungi, similar TAD structures have been observed (Mizuguchi et al. 2014; Eser et al. 2017; Tsochatzidou et al. 2017; Schalbetter et al. 2019; Rodriguez et al. 2022; Winter et al. 2018; Galazka et al. 2016). Based on local interactions, we identified genomic regions with an average size of 100 kb that display stronger interactions within these regions than with neighboring genomic regions, reminiscent of metazoans TADs. The size of the regions is smaller than typical mammalian TADs (200 kb - 2.5 Mb) (Dixon et al. 2012), yet similar to TADs in *Drosophila*, and reminiscent TAD structures in the filamentous fungi *N. crassa* and *E. festucae*, and the yeast *S. cerevisiae* (100 - 150 kb) (Eser et al. 2017; Liao et al. 2021a; Sexton et al. 2012; Winter et al. 2018; Galazka et al. 2016). Our TAD prediction was performed for two strains of *V. dahliae*, and for each of the strains we used independent biological replicates. Importantly, among strains as well as among replicates we obtained a high degree of reproducibility, suggesting that we have confidently predicted TADs as units of chromosomal organization in *V. dahliae*.

As TADs are genomic regions that interact more strongly within the region than with neighboring regions, they separate genomes into discrete units (Szabo et al. 2018; Chang et al. 2020). In this manner, TADs function in physically separating clusters of replicating DNA (Pope et al. 2014). Replication generally occurs with coordinated timing within genomic domains, called replication domains, and in several animal species as well

as in *S. cerevisiae*, these domains largely correlate with a TAD structure (Eser et al. 2017; Kolesnikova et al. 2018; Pope et al. 2014), suggesting that TADs can be seen as a unit of replication. In addition to TADs functioning in DNA replication, for some TADs it has been shown that transcriptional co-regulation of genes within the TADs occurs (Le Dily and Beato 2015; Le Dily et al. 2014; Jin et al. 2013; Nora et al. 2012; Gonzalez-Sandoval and Gasser 2016; Kim and Dean 2021; Yildirir et al. 2021; Winter et al. 2018). However, even for TADs that display transcriptional co-regulation it currently remains unclear whether TAD organization facilitates such transcriptional co-regulation, or whether shared transcriptional profiles and the epigenetic status of the genes within a particular genomic region predispose a region to become organized into a TAD (Rocha et al. 2015; Cavalheiro et al. 2021). Thus, for TADs in *V. dahliae* it can presently only be concluded that a correlation between TAD organization and gene expression exists. Only few TADs correlate with transcriptional co-regulation, as observed in the arbuscular mycorrhizal fungus *Rhizophagus irregularis* and in *E. festucae* (Yildirir et al. 2021; Winter et al. 2018). Similarly, in *V. dahliae* we find that transcriptional co-regulation occurs in only 19 out of 353 TADs. It is not known what makes that these particular TADs display transcriptional co-regulation, while most TADs do not show signs of transcription co-regulation. Transcriptional co-regulation within TADs could be determined by epigenetic characteristics of TADs. Remarkably, we find that 17 of the 19 TADs displaying transcriptional co-regulation localize in AGRs. Previously we have shown that AGRs are epigenetically distinct from the core genome (Cook et al. 2020). This difference between AGRs and the core genome involves a lack of DNA methylation over TEs, as well as a general enrichment in the histone modification H3K27me3 combined with accessible DNA in AGRs (Cook et al. 2020). The enrichment in H3K27me3 at AGRs plays a role in transcriptional regulation, as loss of H3K27me3 in a *V. dahliae* mutant strain induces the expression of particular genes (Kramer et al. 2021). Additionally, we showed that changes in H3K27me3 levels between *in vitro* growth media only partially explains differential gene expression (Kramer et al. 2021). An association between H3K27me3 and transcriptional regulation has similarly been observed in various plant associated fungi (Studt et al. 2016; Möller et al. 2019; Zhang et al. 2021b; Connolly et al. 2013; Soyer et al. 2014; Carlier et al. 2021) and it is therefore conceivable that the presence of H3K27me3 over TADs in *V. dahliae* is associated with transcriptional co-regulation of the genes within these. Nevertheless, we observed that also within the AGRs only a fraction of TADs displays transcriptional co-regulation, suggesting that the presence of H3K27me3 alone is not sufficient to regulate transcription, and consequently other factors involved in regulating transcription are required.

TAD boundaries are the regions that mediate the separation of TADs (Dixon et al. 2012; Van Bortle et al. 2014). Such boundaries have frequently been shown to display a relatively high degree of conservation between closely related organisms (Rao et al. 2014; Fudenberg and Pollard 2019; McArthur and Capra 2021; Krefting et al. 2018), yet their conservation has not been addressed for fungal species. Here, we show that the boundaries between TADs display a high degree of conservation in the *Verticillium* genus.

The conservation of TAD boundaries suggests that these are important for evolutionary stability, or that boundaries serve additional roles, for instance in regulation of the global 3D genome organization. Interestingly, whereas we find that TAD boundaries in general are conserved in *Verticillium*, TAD boundaries located in AGRs are not particularly conserved between *V. dahliae* strains. This could mean that TAD organization in the evolutionary young AGRs has not yet been firmly established, or that a clear TAD organization is important in the core genome yet can be more relaxed in evolutionary dynamic AGRs. However, these possibilities are not mutually exclusive. Remarkably, we found clear signs of TE insertions coinciding with putatively newly generated TAD boundaries as well as with extensively rearranged TAD boundaries. Interestingly, a recent study in cotton uncovered that cultivar-specific TAD boundaries generally harbor more TEs than conserved TAD boundaries that are shared between cultivars (Wang et al. 2021b). Moreover, it has been shown that *de novo* insertions of HERV-H TEs in humans can introduce new TAD boundaries (Zhang et al. 2019). As TE activity in *V. dahliae* is largely confined to AGRs (Faino et al. 2016; Torres et al. 2021), we speculate that such TE activity may be involved in generating new TAD boundaries. In addition to the low conservation of TAD boundaries in AGRs, we also observe that these boundaries are generally weaker insulated than those in the core genome, suggesting that TADs in AGRs are not as well established as those in the core genome. However, single-cell research in human and *Drosophila* cells show that TADs are dynamic structures that are continuously broken and formed, and that boundary positions can vary between cells, especially in a heterogeneous population containing distinct cell types (Finn et al. 2019; Luppino et al. 2020; Bintu et al. 2018; Szabo et al. 2018). Therefore, it cannot be excluded presently that the weaker insulation of TAD boundaries in AGRs than in the core genome is a consequence of TAD boundaries in AGRs being more dynamic among the cells used as input for our Hi-C experiments. In metazoans, TAD boundaries are bound by CTCF proteins, which are thought to be involved in maintaining TAD organization (Kentepozidou et al. 2020), and the amount of CTCF proteins bound to the TAD boundary correlates with the insulation score (Gong et al. 2018; Barutcu et al. 2018). Therefore, the difference in TAD boundary dynamics in AGRs and in the core genome may be caused by differential binding of proteins with CTCF-like function. However, such proteins have not been identified in any filamentous fungi so far, including in *V. dahliae*. Additionally, the functional implications of the here observed differential dynamics of TAD boundaries in AGRs and the core genome remain unclear. Potentially, the TAD boundary dynamics observed in AGRs confers a more flexible transcriptional response of the environmental responsive genes that predominantly locate in AGRs (de Jonge et al. 2013; Cook et al. 2020; Kramer et al. 2021).

Besides local interactions in the context of TADs, the 3D genome is known to also display long-distance interactions, in which distant regions on the same chromosome, or regions on distinct chromosomes, interact (Dekker and Misteli 2015). For example, in some eukaryotes, long-distance interactions occur between centromeres that, despite localizing on different chromosomes, colocalize within the nucleus (Smith et al. 2012; Marie-Nelly et

al. 2014; Muller et al. 2019a; Hoencamp et al. 2021). The colocalization of centromeres in the nucleus in proximity to the nuclear membrane is a commonly observed pattern in fungi (Tanizawa et al. 2017; Smith et al. 2012; Hoencamp et al. 2021; Duan et al. 2010; Mizuguchi et al. 2014; Zaccaron et al. 2022; Liang et al. 2022; Yildirir et al. 2021; Edwards et al. 2022; Sperschneider et al. 2021; Henningsen et al. 2022; Li et al. 2022; Winter et al. 2018; Galazka et al. 2016), and we have previously observed centromere clustering in *V. dahliae* as well (Seidl et al. 2020). Additionally, long-distance interactions among chromosomes of *N. crassa* occur between heterochromatic regions, for instance, those that are marked with H3K27me3 (Galazka et al. 2016; Klocko et al. 2016). Therefore, we hypothesized that H3K27me3-marked AGRs in *V. dahliae* similarly display long-distance interactions. Remarkably, we observed nuclear clustering of facultative heterochromatic AGRs that is distinct from the constitutive heterochromatic centromeres, suggesting for two different mechanisms are responsible for the physical clustering of H3K27me3-rich AGRs and centromeres within the nucleus.

The physical clustering of AGRs in the nucleus may lead to the formation of nuclear bodies through Liquid-Liquid Phase Separation (Sawyer and Dundr 2016; Penagos-Puig and Furlan-Magaril 2020; Larson et al. 2017). Nuclear bodies are membrane-less sub-compartments in the nucleus, in which a micro-environment exists that allows spatial segregation of nuclear activities, such as transcription and DNA-repair (Mao et al. 2011; Morimoto and Boerkoel 2013; Sawyer and Dundr 2016). Additionally, the spatial segregation of AGRs could be caused by their association with the nuclear membrane. In mammalian nuclei, heterochromatic regions are associated with Lamin proteins and additional anchor proteins to form LADs (Lamina Associated Domains) at the nuclear membrane (Harr et al. 2015, 2016; Briand and Collas 2020; van Steensel and Belmont 2017). However, to date, the genes encoding for the LADs-responsible proteins have not been found in fungi (Wagner and Krohne 2007; Koreny and Field 2016). Nevertheless, fungi have alternative processes that anchor facultative heterochromatin to the nuclear membrane, and these are mostly associated with the presence of H3K27me3 (Jamieson et al. 2018; Möller et al. 2019; Schotanus et al. 2021; Dumesic et al. 2015; Galazka et al. 2016; Klocko et al. 2016). Thus, the physical clustering of H3K27me3-rich AGRs in the nucleus may be caused by the association of AGRs at the nuclear membrane. Furthermore, chromosome rearrangements could potentially explain why inter-AGR interactions preferentially occur within AGRs (de Jonge et al. 2013; Faino et al. 2016), as we observed a colocalization of these interactions with segmental duplications. Homolog sequences mediate the physical inter-chromosomal interactions in polyploid wheat (Jia et al. 2021; Yuan et al. 2022), similar to some repetitive elements that colocalize in some metazoan genomes (Cournac et al. 2016; Choudhary et al. 2020). Intriguingly, the physical clustering of AGRs is observed across the *Verticillium* genus, indicating that the potential mechanism driving this physical association could be conserved in the *Verticillium* species.

The 3D organization of chromosomes influences the replication timing of the genome (Comoglio et al. 2015; Ryba et al. 2010; Dileep et al. 2015; Pope et al. 2014; Gindin et al.

2014; Rhind and Gilbert 2013; Solovei et al. 2016). Therefore, the physical colocalization of AGRs can result in a differential replication timing, compared to the core genome (Wolfe et al. 1989; Heun et al. 2001; Nair et al. 2017; Jørgensen et al. 2007; Pichugina et al. 2016; Janssen et al. 2018). Differences in replication timing have been observed in H3K27me3-rich regions, and it has been associated with chromosome instability in *Z. tritici* (Möller et al. 2019; Pentzold et al. 2021; Admire et al. 2006; Habig et al. 2021; Schotanus et al. 2015). Chromosome instability and the physical co-localization of highly similar sequences may decrease the efficiency of separation of the DNA content during mitosis resulting in an increase in DNA double-strand breaks (Patel et al. 2019; Zuo et al. 2021). The DNA double-strand breaks repair in H3K27me3-rich regions is associated with the Microhomology-Mediated End Joining repair pathway (MMEJ) (Her and Bunting 2018; Schep et al. 2021). Typically, the MMEJ repair pathway introduces abundant mutations (Boulton and Jackson 1996; Ma et al. 2003) and could potentially explain the increase in genomic rearrangements in AGRs (Huang and Cook 2022; de Jonge et al. 2013; Faino et al. 2016; Seidl and Thomma 2017; Cook et al. 2020; Torres et al. 2021; Depotter et al. 2019). Therefore, we speculate that the physical colocalization of homologous sequences in AGRs could increase the probability of DNA double-strand breaks that, unfaithfully repaired, could contribute to the emergence of the plastic genome fraction.

AGRs were originally discovered by their high degree of PAV between *V. dahliae* strains (de Jonge et al. 2013; Faino et al. 2016), and subsequently further characterized and refined by their distinct epigenomic profile (Cook et al. 2020). Such plastic genome compartments are epigenetically distinct from core genomic regions, which is increasingly recognized as a common pattern in various plant pathogens (Cook et al. 2020; Fokkens et al. 2018; Torres et al. 2020; Rojas-Rojas and Vega-Arreguín 2021; Schotanus et al. 2015; Wang et al. 2017; Soyer et al. 2021). Here, we have added yet another layer to the divergence between plastic regions and the core genome, by showing that also the 3D organization differs between the core genome and AGRs in *V. dahliae*. The exact ramifications of this divergent 3D organization on evolution and transcriptional regulation are still unclear. Future research into global 3D genome organization within the nucleus should show the mechanisms driving AGRs interactions and their role in the emergence of the plastic genome. This holistic view with genetic, epigenetic, and spatial characterization of plastic genome compartments will aid further understanding of the genome function and evolution in fungi.

Material and Methods

Hi-C analysis and TAD prediction

Hi-C library preparation was performed with *V. dahliae* strains JR2 and VdLs17 as previously described (Seidl et al. 2020), and paired-end (2 × 150 bp) sequenced on the NextSeq500 platform at USEQ (Utrecht, the Netherlands). Additional Hi-C datasets of *V. dahliae* strains JR2 and VdLs17, *V. albo-atrum* strain PD747, *V. alfalfa* strain PD683, *V. isaacii* strain PD618, *V. klebahnii* strain PD401, *V. longisporum* strains PD589 and VLB2, *V. nonalfalfae* strain T2, *V. nubilum* strain 397, *V. tricorpus* strain PD593, and *V. zaregamsianum* strain PD739 were previously generated (Seidl et al. 2020; Depotter et al. 2021).

Sequenced read-pairs were quality-filtered and trimmed using Trimmomatic (v 0.36) in paired-end mode with default settings (Bolger et al. 2014). Filtered and trimmed reads were mapped to the corresponding genomes (Faino et al. 2015; Seidl et al. 2020) using Burrows-Wheeler aligner (BWA mem, settings: -A1 -B4 -E50 -L0) (Li and Durbin 2010). Hi-C interaction matrices were built and analysed using HiCExplorer tools (Wolff et al. 2020). First, we used hicBuildMatrix to generate the interaction matrix based on the *in silico Dpn*II restriction digested corresponding genome. Matrix resolution was reduced by merging five adjacent bins using hicMergeMatrixBins. For *V. dahliae* strains JR2 and VdLs17, replicates were corrected separately according to the iterative correction and eigenvector decomposition (ICE) method (Imakaev et al. 2012) using hicCorrectMatrix, and TADs were predicted using hicFindTADs (settings: --delta 0.01). Correlation between replicates was determined by using a reproducibility score based on a stratified cross-correlation using the HiCRep package (Yardımcı et al. 2019).

To combine replicate matrices, resolutions of raw matrices were reduced by merging 5 adjacent bins using hicMergeMatrixBins, normalized between replicates using hicNormalize (settings: --setToZeroThreshold 1), corrected separately according to the ICE method using hicCorrectMatrix, and finally combined using hicSumMatrices (Wolff et al. 2020). For the other *Verticillium* species, matrix resolution reduction and correction was performed as above, and hicFindTADs was used to generate a table with per bin insulation scores.

Characterization of epigenetic profiles

Chromatin immunoprecipitation followed by sequencing (ChIP-seq) for H3K4me2, H3K9me3, H3K27me3, and H3K27ac, and the assay for transposase-accessible chromatin followed by sequencing (ATAC-seq) were performed for *V. dahliae* strain JR2 as described previously (Seidl et al. 2020; Cook et al. 2020; Kramer et al. 2021). ChIP datasets were normalized over MNase control samples.

We used the umap-learn implementation through the R/umap package. This implementation make use of the python UMAP algorithm (McInnes et al. 2018). For the gene analysis, the following variables were used: GC content, ATAC-seq, 5mC, H3K27ac,

H3K27me3, H3K9me3 and log2(PDB *in vitro* expression +1), with the following parameters random_state=42, n_neighbors=50, n_components=2, min_dist=0.01, metric=cosine. The resulting two-dimensional values from UMAP fit.transform were used for plotting and further statistical analysis using Matplolib, Numbpy and Seaborn V0.8.1 (Hunter 2007; Waskom et al. 2017; van der Walt et al. 2011).

Characterization of transcriptional regulation

RNA sequencing of *V. dahliae* strain JR2 cultivated for six days in potato dextrose broth (PDB) and Czapec-Dox medium (CZA) was previously performed (Cook et al. 2020; Kramer et al. 2022). Analyses of gene (Faino et al. 2015) and TE presence (Seidl et al. 2020; Torres et al. 2021) over TADs and TAD boundaries were performed using the EnrichedHeatmap package in R (Gu et al. 2018; Team 2013). To assess co-regulation of genes within TADs, we used R to fit a linear model with log2 fold-change in expression of target genes between PDB and CZA as the response variable and TAD membership as a predictor, similarly as previously described (Winter et al. 2018).

Characterization of *Verticillium dahliae* genomic variation

Structural variants (duplications, deletions, inversions and translocations), single nucleotide variants and polymorphic transposable elements were previously identified using paired-end sequencing reads of each 42 previously sequenced *V. dahliae* strains (Torres et al. 2021). Briefly, structural variants were predicted using the 'sv-callers' workflow with few modifications that enabled parallel execution of multiple SV callers (Kuzniar et al. 2020), an approach that is considered optimal as it exploits complementary information to predict SVs (Goerner-Potvin and Bourque 2018; Cameron et al. 2019). Single nucleotide variants were identified using the -HaplotypeCaller of the Genome Analysis Toolkit (GATK) v.4.0 (Poplin et al. 2018) and transposable element PAV was analyzed using TEPID v.2.0 (Stuart et al. 2016). To investigate if SVs and polymorphic TEs co-localize with TAD boundaries, we summarized the overlap of each set of variants by their breakpoint frequency (start or ends ±1 bp of the feature) and coverage (number of bases covered) across the genome of *V. dahliae* strain JR2 (Fudenberg and Pollard 2019). Similar to Fudenberg and Pollard (2019), we calculated the log10(observed/expected) of each feature representing the deviation from a uniform distribution across the genome, therefore accounting for the proportion of the genome covered by a specific genomic feature. Finally, we considered two scenarios: core genome vs AGRs, and syntenic regions between JR2 and VdLs17 versus non-syntenic regions. Syntenic regions between *V. dahliae* strains JR2 and VdLs17 were previously determined (Faino et al. 2016). Briefly, whole-genome alignments between the eight chromosomes was performed using MUMmer 3.0 and GEvo (Kurtz et al. 2004; Lyons et al. 2008), where only gene-coding regions were used as anchors between syntenic chromosomal regions.

To further expand our analysis of *V. dahliae* to the full genus, we used the recently available Hi-C-corrected genomes of all *Verticillium* species (Seidl et al. 2020; Depotter et

al. 2021). The phylogenetic tree was generated using Realphy v. 1.12 using a maximum likelihood inferred by RAxML (Langmead and Salzberg 2012; Bertels et al. 2014). We aligned the *Verticillium* genomes using ProgressiveCactus (Armstrong et al. 2020). This approach allowed us to reduce the reference-bias and consider more accurate further analysis. We obtained the specific MAF alignments on JR2 and syntenic regions using the HAL package (Hickey et al. 2013). To analyze the nucleotide conservation throughout the genus, we used PhastCons, a hidden Markov model-based method that estimates the probability that each nucleotide belongs to a conserved element based on a multiple sequence alignment (Siepel et al. 2005). Briefly, for each independent JR2 chromosome we assumed a neutral evolution model and correction for indels. For further analyses, we summarized the PhastCons score over TADs and TAD boundaries in the core genome and in AGRs.

TAD boundary prediction throughout the *Verticillium* genus

The Hi-C datasets of the *Verticillium* species (excluding *V. dahliae*) were available without replicates (Faino et al. 2015; Seidl et al. 2020; Depotter et al. 2021). Therefore, we decided to predict TAD boundaries based on sequence homology to boundaries in *V. dahliae* strain JR2. We first filtered the boundary sequences that do not have a TE insertion and queried them to the *Verticillium* genomes using Blastn (Chen et al. 2015), retaining those with >50% coverage that were contiguous in the same syntenic block. Finally, we cross-referenced those putative TAD boundary regions with the previously calculated insulation score for each independent species.

Statistical analysis and visualization

Hi-C matrix and TAD visualizations were performed using HiCExplorer and FAN-C (Kruse et al. 2020). Heatmap and enrichment visualization of insulation scores over boundaries, normalized chromatin marks, structural and nucleotide variants, as well as the PhastCons score, were performed using the R/EnrichedHeatmap v1.2 package (Gu et al. 2018). Permutation tests were computed using R/Bioconductor regionR v1.18.1 package (Gel et al. 2016) and performed with 10,000 iterations, using overlaps between TAD boundaries divided by the insulation score quantiles and the predefined AGRs, and circular randomization to maintain the order and distance of the regions in the chromosomes. All statistical analyses and comparison tests were performed in R v.3.6.3 (Team 2013), and visualized with ggplot2 (Wickham 2016).

Identification of significant colocalization events from Hi-C data

Expected-Observed counts were obtained from HiCExplorer (version 3.7) using the export function with expected-observed nonzero from the corrected and summed matrices. Expected-observed count matrixes containing the interactions between all the merged bins were filtered to keep only interaction between genetically distant regions. Two regions were considered genetically distant if the bins belonged to different chromosomes or if two

bins were more than 20kb distant from each other. Next, the bins were annotated based on which genomic compartment they belong to (centromere or AGR) and the remaining bins were annotated as core genome. The mean expected-observed count for centromere-centromere interactions is calculated and only bin to bin colocalization events with expected-observed counts above this threshold are kept. Genomic bins which still have at least one interaction are considered as genomic region involved in long range interaction. It needs to be noted that our analysis is limited by the fact that we only assess those interactions that are stronger than the average interaction strength that occurs between centromeres in each species but, given that the constitution of centromeres differs between *Verticillium* spp. (Seidl et al., 2020), this interaction strength may consequently not be uniform between species. Visualization of the interaction and associated genetic and epigenetic features was done in R with the circos package.

Identification of duplicated regions and association with long-range interactions

Self-whole-genome alignments of the genome assemblies of *Verticillium* species were performed using MUMmer (Kurtz et al., 2004), regions that mapped elsewhere on the genome with a nucleotide identity above 80% and above 1kb in size are considered duplicated. For each colocalization event that associate two genetically distant regions, we verify if there was a duplicated region that borders with or overlaps with both genetically distant regions. We consider the colocalization event to be neighboring a duplicated region if the average distance between the colocalizing region and the duplicated region is below 50kb.

We calculated the enrichment of colocalization events neighboring duplicated regions through a permutation test. We simulated random interactions genome wide and calculated the number of observed interactions neighboring a duplicated region. We repeated the process 100 times to generate a distribution of expected interactions neighboring duplicated regions and performed a t-test versus the real observed value. This permutation test was performed using the genome wide-duplicated regions distribution and AGR compartment specific.

de novo motif prediction of AGR genomic regions involved in colocalization events

We identify de novo motifs enriched in genomic bins involved in long range interactions using the MEME suite with the following parameters: -dna -revcomp -objfun de -nmotifs 5 -mod anr using seed 42. This searches forward and reverse complement for up to 5 motifs which are enriched in the positive set (genomic regions involved in colocalization events) versus the negative set (genomic regions not involved in colocalization events). The search is limited to the genomic compartment. eg. AGR involved in colocalization events version AGR regions not involved in colocalization events. We searched for related annotated

motifs using the Jaspar fungal and general database (version 2016) and manually inspected significant results.

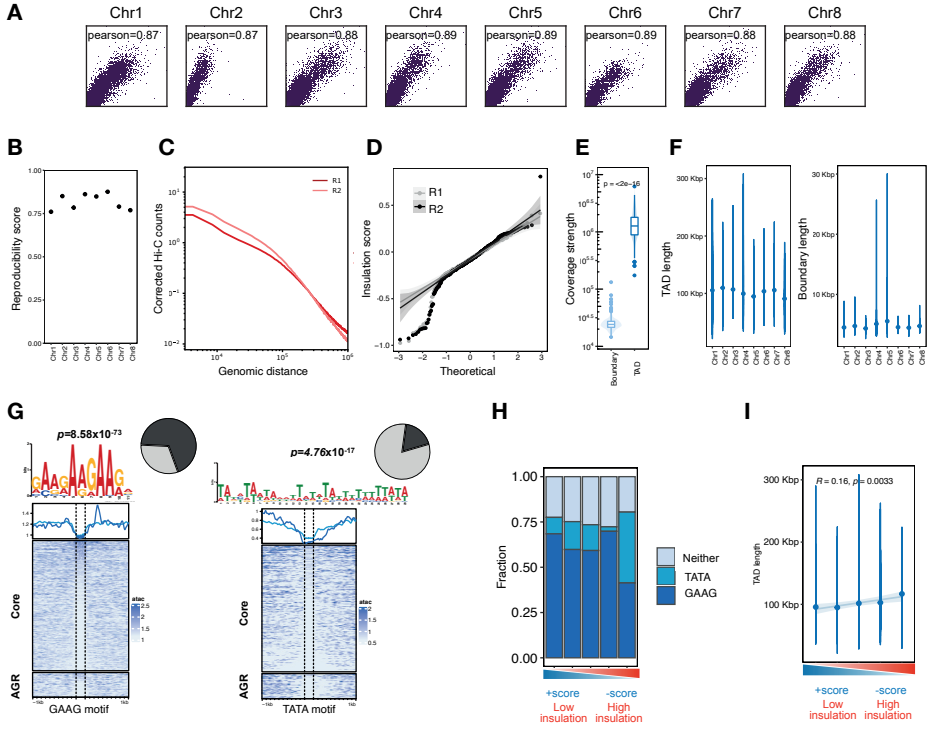
Adaptive genomic regions in non-*V. dahliae* strain JR2 genomes

Nucleotide sequences from the *V. dahliae* strain JR2 AGR compartment were aligned versus each of the *Verticillium* genomes included in this work using ProgressiveCactus (Armstrong et al. 2020). Syntenic regions in the respective genomes are considered AGR regions in the respective organism. Additionally, each *Verticillium* strain genome was aligned using ProgressiveCactus (Armstrong et al. 2020) with the other Verticillium strain genomes included in this work. Genomic regions which were unique and not found in any other genome are also considered AGR in that strain.

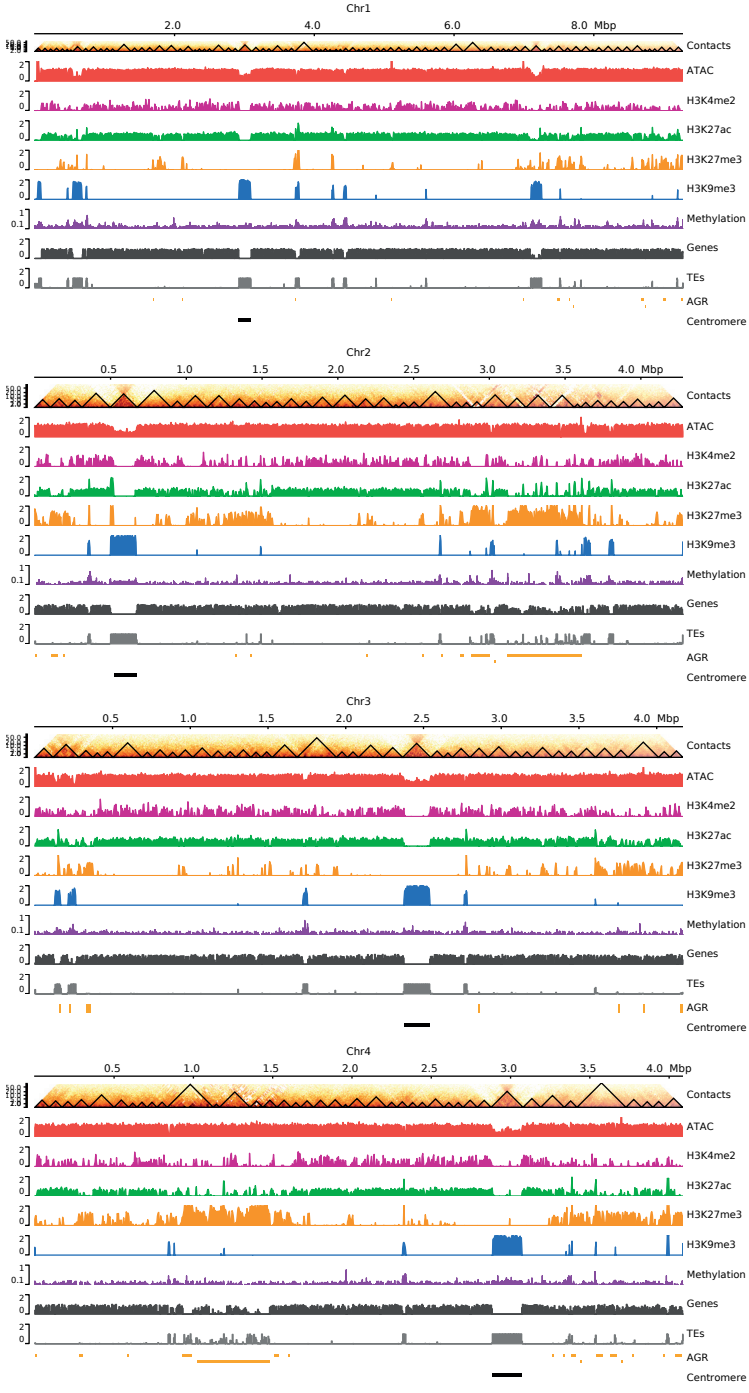
Acknowledgments

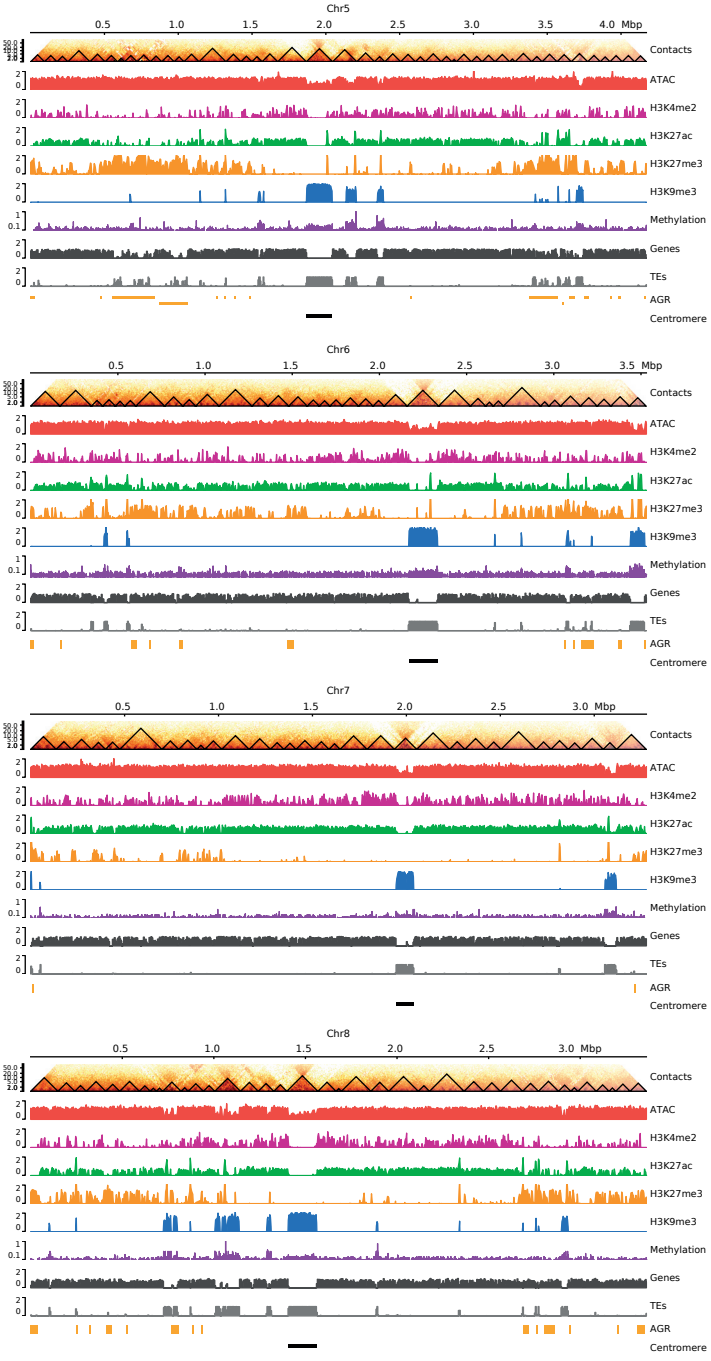
This work was supported by the Consejo Nacional de Ciencia y Tecnología, México to DET, by a PhD grant of the Research Council Earth and Life Sciences (project 831.15.002) to HMK, and by Human Frontier Science Program Postdoctoral Fellowship (HFSP, LT000627/2014-L), by USDA's National Institute of Food and Agriculture (award no. 2018-67013-28492) through the Plant Biotic Interactions Program, and by the National Science Foundation (award no. 1936800) Division of Molecular and Cellular Biosciences to DEC. Work in the laboratories of M.F.S and B.P.H.J.T. is supported by the Research Council Earth and Life Sciences (ALW) of the Netherlands Organization of Scientific Research (NWO). B.P.H.J.T acknowledges funding by the Alexander von Humboldt Foundation in the framework of an Alexander von Humboldt Professorship endowed by the German Federal Ministry of Education and Research is furthermore supported by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) under Germany´s Excellence Strategy – EXC 2048/1 – Project ID: 390686111. The authors declare no conflict of interest.

Supplementary data



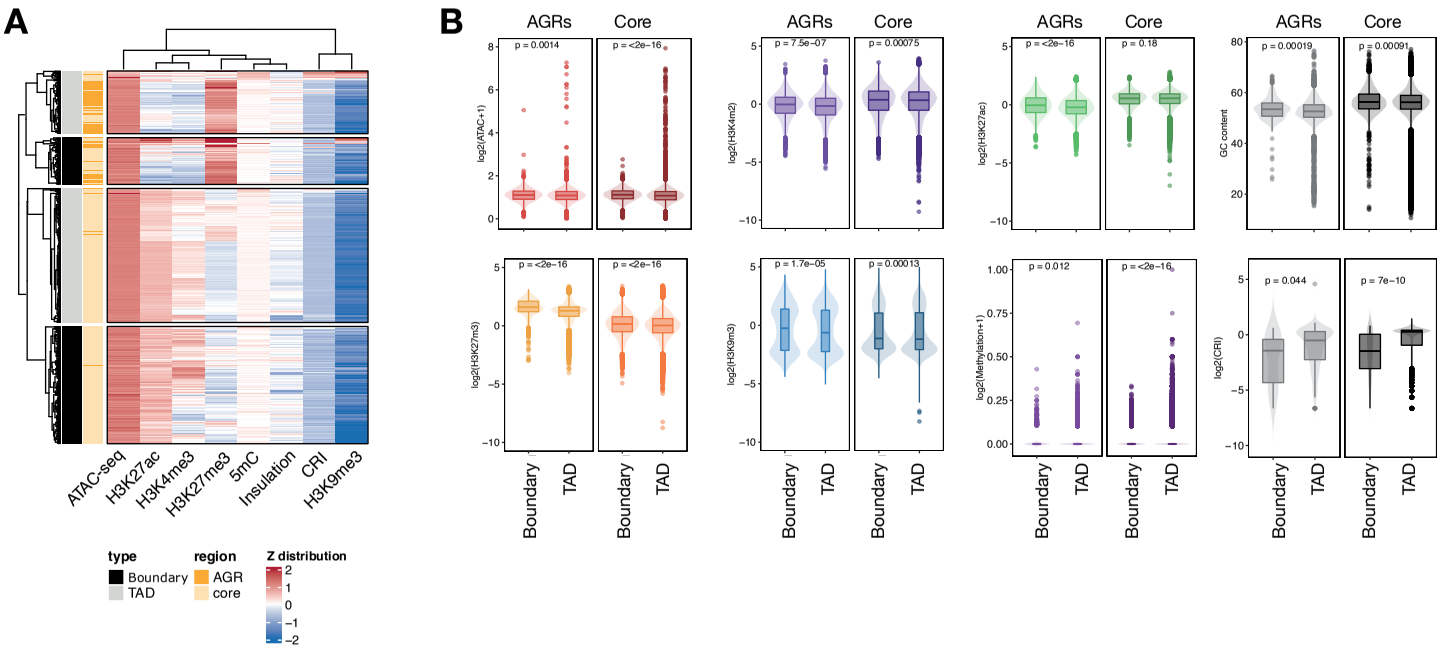
Supplementary Figure 1. Characteristics of TAD and boundaries in *Verticillium dahliae* strain JR2.
(A) Pearson correlation between replicates of Hi-C interactions for *V. dahliae* strain JR2 divided per chromosome. **(B)** Reproducibility score based on a stratified cross-correlation of interactions between replicates of Hi-C interactions for *V. dahliae* strain JR2 divided per chromosome. **(C)** Distance-dependent interaction frequency in replicates of Hi-C data for *V. dahliae* strain JR2. **(D)** QQ plot of insulation scores calculated for each Hi-C replicate, confirming the non-significant differences between replicates. **(E)** Coverage strength (Hi-C counts) over boundaries and bodies of TADs in merged replicates, showing that boundaries are depleted of interactions. P-value after Wilcoxon rank sum test. **(F)** Length of TADs bodies and boundaries predicted on the merged replicates, divided per chromosome. **(G)** A GAAG-motif (left) and TATA-motif (right) is enriched in TAD boundaries of *V. dahliae* strain JR2 and correlate with a decrease in chromatin accessibility as determined by the assay for transposase accessible chromatin (ATAC). The top plot displays the average ATAC signal over TATA motifs with 1 kb up- and down-stream sequence in the core genome (light blue) and in AGRs (dark blue). Heatmap display ATAC signal around the GAAG and TATA motifs as rows in the core genome and in AGRs, respectively. **(H)** Occurrence of the GAAG-motif and the TATA-motif in boundaries separated on their insulation score. **(I)** Correlation between boundary insulation score and length of adjacent TADs for boundary quintiles separated based on their insulation score.


Chr1
2.0
4.0
6.0
8.0 Mbp
Contacts
ATAC
H3K4me2
H3K27ac
H3K27me3
H3K9me3
Methylation
Genes
TEs
AGR
Centromere
Chr2
0.5
1.0
1.5
2.0
2.5
3.0
3.5
4.0 Mbp
Contacts
ATAC
H3K4me2
H3K27ac
H3K27me3
H3K9me3
Methylation
Genes
TEs
AGR
Centromere
Chr3
0.5
1.0
1.5
2.0
2.5
3.0
3.5
4.0 Mbp
Contacts
ATAC
H3K4me2
H3K27ac
H3K27me3
H3K9me3
Methylation
Genes
TEs
AGR
Centromere
Chr4
0.5
1.0
1.5
2.0
2.5
3.0
3.5
4.0 Mbp
Contacts
ATAC
H3K4me2
H3K27ac
H3K27me3
H3K9me3
Methylation
Genes
TEs
AGR
Centromere

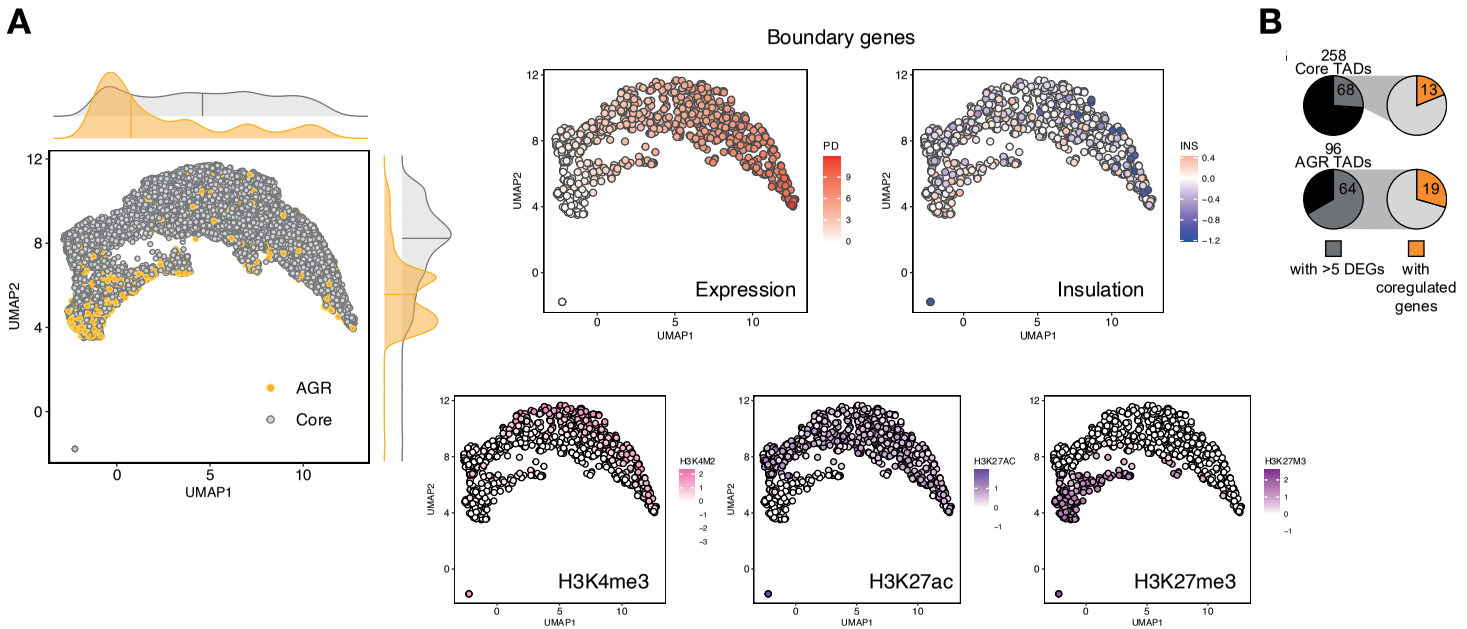


Supplementary Figure 2. Distribution of TADs in *Verticillium dahliae* strain JR2 for all chromosomes.

From top to bottom: Hi-C contact matrix depicting TADs, chromatin structures as black triangles, open chromatin regions as determined with ATAC-seq as well as histone modifications H3K4me2, H3K27ac, H3K27me3, and H3K9me3 normalized over a micrococcal nuclease digestion control, DNA methylation, gene and transposable element (TE) densities in 10 kb non-sliding windows, and adaptive genomic regions (AGRs) (Cook et al. 2020) and centromeric region (Seidl et al. 2020). Chromosome sizes are displayed above the matrix.

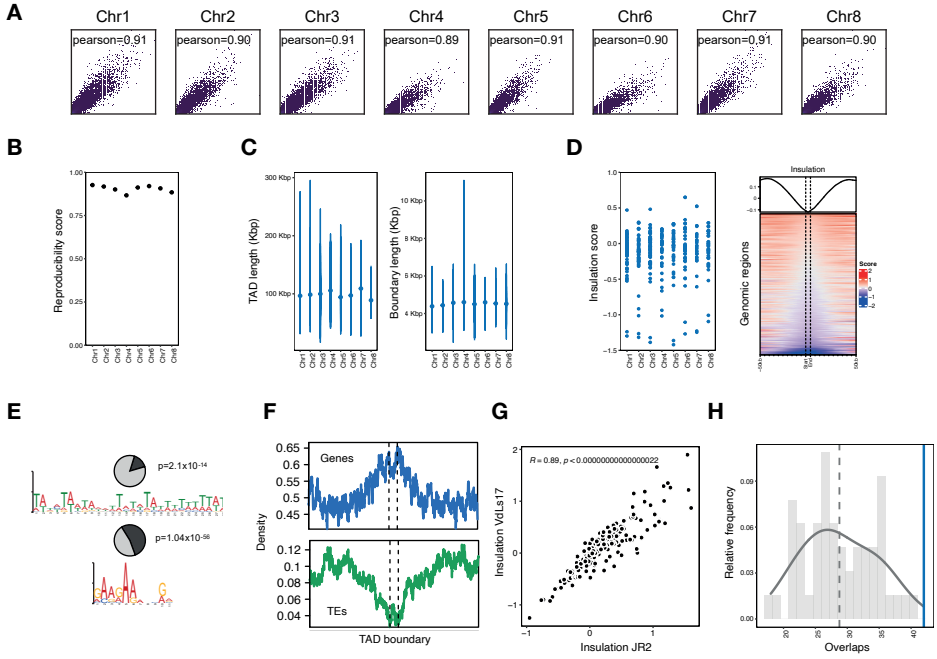


Supplementary Figure 3. Chromatin characteristics are differentially associated with TADs and boundaries.
(A) k-means clustering (*k*=2) of TADs and boundaries using data presented in Fig. 1E combined with insulation scores and composite repeat-induced-point mutation (RIP) index (CRI) values averaged for 100 bp non-overlapping windows and summarized per TAD and boundary region. **(B)** TADs and TAD boundaries differ in histone modifications, CRI, and RIP. *p-value* after one-way Wilcoxon rank sum test.

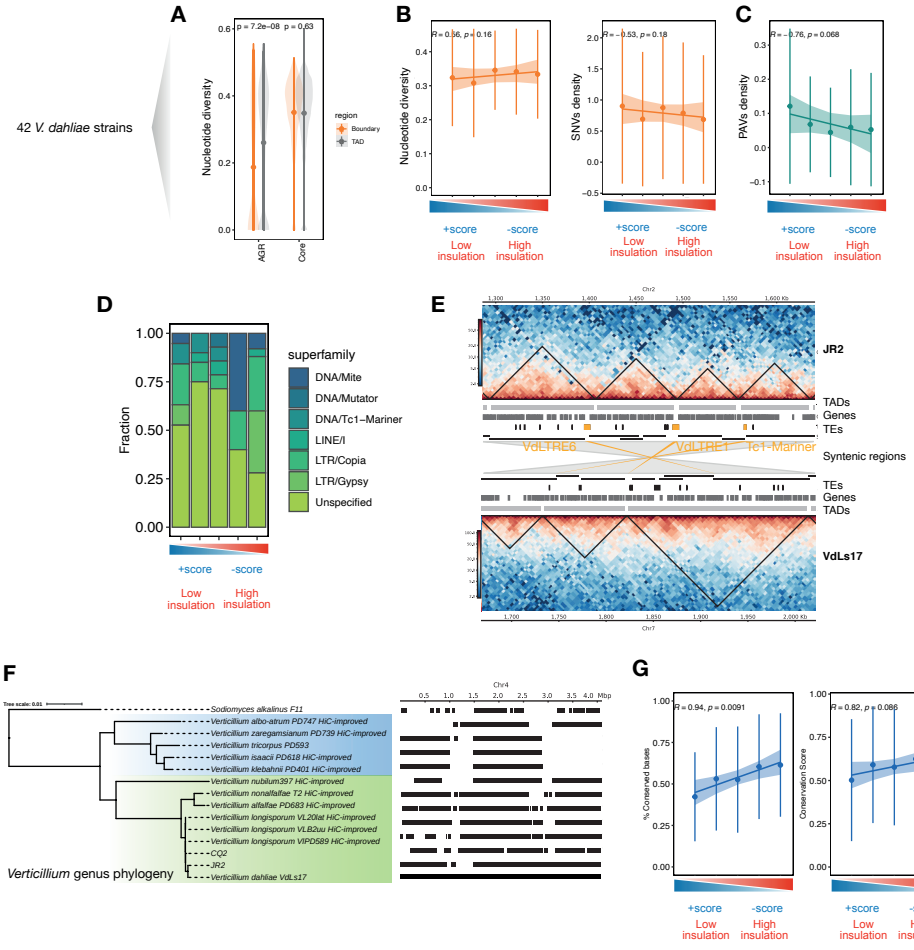


Supplementary Figure 4. Uniform Manifold Approximation and Projection for Dimensional Reduction (UMAP) to separate genes of *Verticillium dahliae* strain JR2 according to their epigenetic profile.

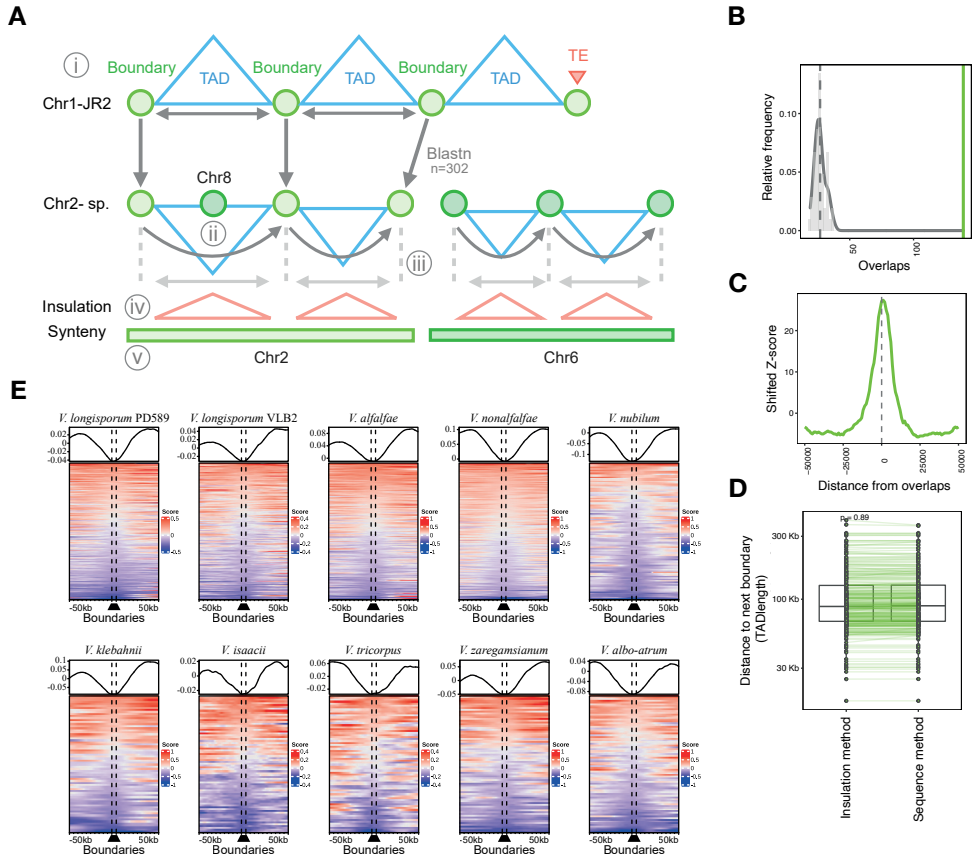
The left top plot displays all genes in the core genome (grey) and in AGRs (yellow). Distributions are shown on top and right of the plot. The remaining plots display only genes located at TAD boundaries and are coloured according to their transcription in PDB (TPM PDB), the insulation score of the boundary they locate in, their H3K4me2 ChIP coverage signal, their H3K4ac ChIP coverage signal, and their H3K27me3 ChIP coverage signal. **(B)** Pie chart showing the proportion of TAD in the core genome and AGRs containing more than five differentially expressed genes. Between cultivation for 6 days in PDB or in CZA (Dark grey), and for which DEGs display directionality of differential expression (orange)



Supplementary Figure 5. Characteristics of TADs and boundary in *Verticillium dahliae* strain VdLs17.
(A) Pearson correlation between replicates of Hi-C interactions for *V. dahliae* strain VdLs17, divided per chromosome. **(B)** Reproducibility score based on a stratified cross-correlation of interactions between replicates of Hi-C interactions for *V. dahliae* strain VdLs17 divided per chromosome. **(C)** Length of TADs and boundaries predicted on the merged replicates, divided per chromosome. **(D)** On the left: Insulation score of boundaries predicted in *V. dahliae* strain VdLs17. On the right: insulation score over boundaries with 50kb up- and down-stream sequence. **(E)** Presence of the TATA- and GAAG-motifs in boundaries of *V. dahliae* strain VdLs17. **(F)** Enrichment of genes and depletion of transposable elements (TEs) in and near boundaries in *V. dahliae* strain VdLs17. **(G)** Correlation of insulation score for boundaries predicted in genomic regions that are syntenic between strains JR2 and VdLs17. **(H)** Lowly insulated boundaries of *V. dahliae* strain VdLs17 are enriched in genomic regions that are non-syntenic between JR2 and VdLs17.

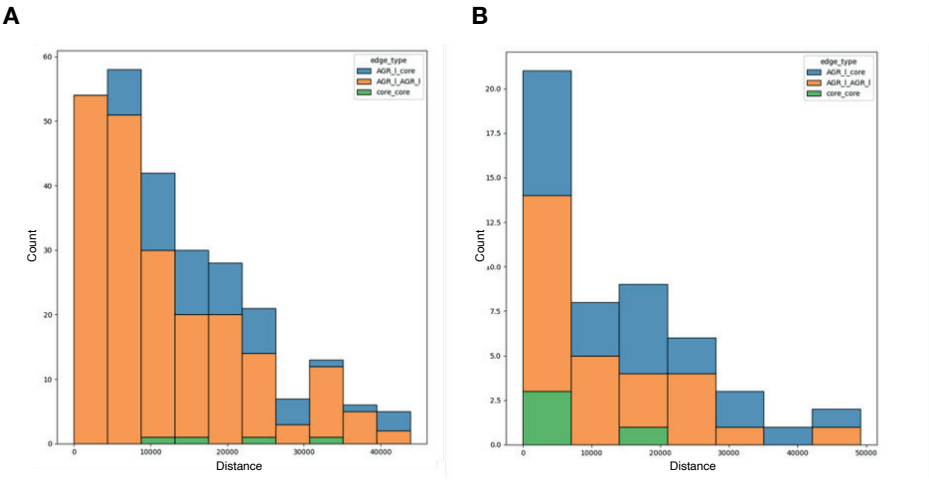


Supplementary Figure 6. Conservation of TADs and boundaries in a collection of 42 *Verticillium dahliae* strains and in the *Verticillium* genus.
(A) Nucleotide diversity in AGRs and in the core genome. **(B)** Correlation between insulation score and nucleotide diversity (left) and SNVs density (right) for boundary quintiles separated on their insulation strength. **(C)** Correlation between insulation score and presence/absence variation (PAV) for boundary quintiles separated on their insulation strength. **(D)** Abundance of polymorphic transposable elements (*n*=86) for boundary quintiles separated by their insulation strength in the *V. dahliae* strain JR2. **(E)** Multiple TE insertions in *Verticillium dahliae* strain JR2 coincide with boundary rearrangements between *V. dahliae* strains. Heatmaps represent contact matrixes with TADs (black triangles) over a section of chromosome 2 of JR2 (top) and chromosome 7 of VdLs17 (bottom) that are syntenic. The region occurs inverted in VdLs17 relative to JR2. TADs, genes and TEs are displayed between heatmaps. Synteny between JR2 and VdLs17 indicated as grey blocks. TE insertions in strain JR2 are indicated in yellow. **(F)** Phylogeny of the ten species in the *Verticillium* genus using *Sodiomyces alkalinus* to root the phylogeny; Flavexudans and Flavnonexudans clades are depicted in blue and green, respectively. On the left: syntenic regions (black blocks) between *Verticillium* genomes when compared with *V. dahliae* strain JR2, based on chromosome 4 of *V. dahliae* JR2 strain as example. **(G)** Correlation between insulation score and fraction of conserved bases (left) and conservation score (right) for boundary quintiles separated on their insulation strength.

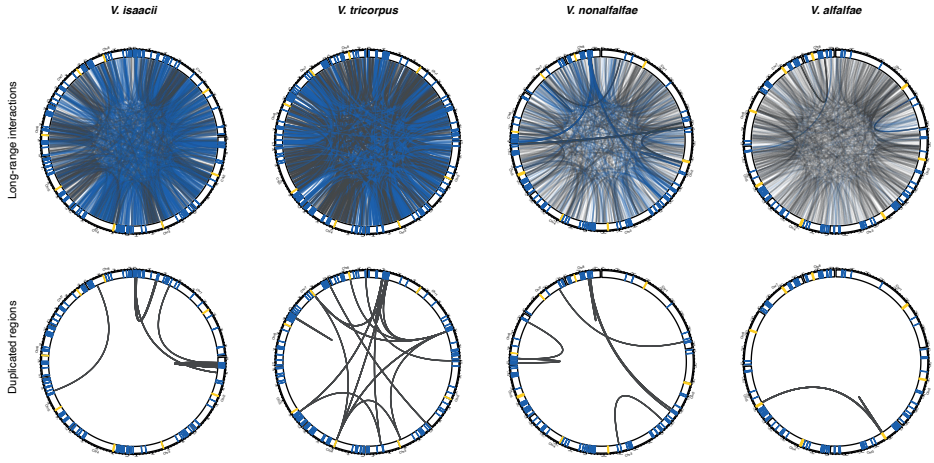


Supplementary Figure 7. Predicted TAD boundaries in the *Verticillium* genus display low insulation compared to adjacent genomic regions.

(A) Overview of our approach to predict the TAD organization in other *Verticillium* species with a schematic representation of the *V. dahliae* JR2 reference TAD organization on top, boundaries as green circles and TAD as blue triangles. i) 302 query DNA sequences from boundaries (excluding those with TE insertions) were obtained and compared to the genomes of *Verticillium* spp. ii) Contiguous sequences belonging to same chromosome were retained and single interrupting matches from other chromosomes were removed. iii) The distances between contiguous boundaries were calculated in reference and subject. iv) Predicted boundaries were compared to the calculated insulation score from the Hi-C data. v) Predicted boundary locations were compared to the progressive Cactus syntenic regions to crosscheck boundary distribution. **(B)** Distribution of 10,000 iterations of the permutation test for overlaps between the boundaries as predicted for *V. dahliae* strain VdLs17 by the sequence-based method in **(A)** and by the insulation method (grey distribution), green line indicates significant overlaps between predicted and 'reference' boundaries from *V. dahliae* VdLs17 (Fig. 3), Z-score=27.1264, p=9.9×10-5. **(C)** Z-score shifts from boundaries indicate a high enrichment of overlapping TAD boundaries for the two methods. **(D)** TAD length prediction distribution by the insulation method and the sequence method in **(A)** is not significantly different. P-value after one-way Wilcoxon rank sum test. **€** Insulation score over TAD boundaries predicted by the sequence-based method **(A)**, with 50 kb up- and down-stream sequence, for each *Verticillium* species. Line plots display average signal over boundaries and up/down-stream sequence, bottom plots display predicted boundaries in rows, ordered on insulation score.



Supplementary Figure 8. Distance between colocalizing regions and neighboring duplicated regions in *Verticillium dahliae*.
(A) Duplicated regions occur in close proximity to physical co-located regions in *V. dahliae* JR2. Y-axis depict distance from co-located regions, and colors show the type of physical co-location event (AGR-to-Core, AGR-to-AGR, Core-to-Core). **(B)** Duplicated regions occur near physically co-located regions in *V. dahliae* VdLs17. Y-axis depict distance from co-located regions, and colors show the type of physical co-location event (AGR-to-Core, AGR-to-AGR, Core-to-Core). Colocalization events involving AGR regions with other AGR regions are shown in orange, AGR region with core regions in blue and core regions with core regions in green.



Supplementary Figure 9. Non-centromeric long-range interactions co-locate with duplicated regions in species of the *Verticillium* genus.
All circular plots display the chromosomes with centromeres in yellow, AGR regions in blue, and core regions in white. For every genome, the upper plot shows as edges non-centromeric long-range interactions that exceed the average interaction strength of centromeres and are color-coded according to the respective genomic annotation of the regions they connect, with AGR interactions in blue and core interactions in grey. In the lower plot, the edges represent segmental duplications.

Supplementary Table 1. Aggregated counts of long-range interactions and their overlap with segmental duplications in Verticillium dahliae strains

	V. dahliae strain JR2	*V. dahliae* strain VdLs17
Genome-wide	1842	1978
Centromere	953	527
Non-centromere	889	1451
AGR	475	452
Core-AGR	225	407
AGR-AGR	250	45
Core-core	414	999
Segmental duplication	264	50
AGR-AGR	207	21
Core-AGR	53	20

Supplementary Table 2. Expected and observed non-centromeric colocalization events between genetically distant regions in *Verticillium* genus genomes.

Species	Non-centromeric colocalization events	Expected core	Observed core	Expected AGR	Observed AGR	*p-value*
V. dahliae JR2	889	880	639	162	475	6.8695E-148
V. dahliae VdLs17	1451	1442	999	217	452	6.1638E-87
V. albo-atrum PD747	897	893	880	111	143	0.002152792
V. klebahnii PD401	573	571	571	65	178	1.24565E-44
V. nubilum 397	582	578	572	81	173	1.52804E-24
V. nonalfalfae T2	891	887	873	105	157	3.46163E-07
V. isaacii PD618	1645	1639	1632	191	132	1.93215E-05
V. alfalfae PD683	1455	1450	1453	163	71	5.74392E-13
V. tricorpus PD593	1053	1048	1051	126	73	2.32935E-06
V. longisporum VlPD589	2426	2420	2407	231	234	0.741519345
V. longisporum VL20	1201	1198	1195	110	91	0.069731441

p-value is based on chi-square test between observed and expected colocalization events in the different genomic compartments.

Supplementary Table 3. Observed and expected association between genomic colocalization events among genetically distant regions and duplicated genomic regions.

Species	Observed duplication-colocalizations	Expected duplication-colocalizations	Duplication-colocalizations core	Duplication-colocalizations AGR	*p-value*
V. dahliae JR2	264	0.38	57	260	1.06E-267
V. dahliae VdLs17	50	0.54	29	41	4.21E-184
V. albo-atrum PD747	11	0.23	8	6	2.93E-133
V. klebahnii PD401	24	0.06	24	0	3.60E-200
V. nubilum 397	16	0.69	15	13	6.92E-132
V. nonalfalfae T2	6	0.07	3	5	4.01E-137
V. isaacii PD618	13	0.11	7	6	1.03E-157
V. alfalfae PD683	0	0	0	0	1.00E+00
V. tricorpus PD593	4	0.09	4	2	2.35E-114
V. longisporum VD589	2400	3.63	2381	232	0.00E+00
V. longisporum VL20	296	1.82	295	9	3.19E-235

Expected overlaps values represent the average after 100 permutations. *p-value* is based on t-test of the observed duplicated regions neighboring colocalization events versus the expected distribution (100 permutations).

Supplementary Table 4. Association between genomic colocalization events among genetically distant regions and duplicated genomic regions.

Species	AGR-AGR	AGR-core	core-core	Total
V. dahliae JR2	207	53	4	264
V. dahliae VdLs17	21	20	9	50
V. albo-atrum PD747	3	3	5	11
V. klebahnii PD401	0	0	24	24
V. nubilum 397	1	12	3	16
V. nonalfalfae T2	3	2	1	6
V. isaacii PD618	6	0	7	13
V. alfalfae PD683	0	0	0	0
V. tricorpus PD593	0	2	2	4
V.longisporum VlPD589	19	213	2168	2400
V. longisporum VL20	1	8	287	296

Association events are separated based on the genomic compartments of the genetically distant regions.

```
# core genes
core.25ka<- read.table("~/Dropbox/reb
al/core.25k.a.bed",sep="\t",fill=T,
,"O","P","Q","R","S","W","Z"))%>%
dplyr::select(-X,-O,-P,-Q,-R,-S,-W,-Z)%>%dplyr::mutate(region=rep("Core"))
ov <- core.25ka%>%group_by(Chr_bnd)%>%mutate(a=start-end_b-
nd,b=end-start_bnd)%>%mutate(ov=ifelse(a<0&b>0,"ov","no"))%>%filter(ov=
dplyr::select(-a,-b)%>%mutate(dist=rep(0),cat=rep("0k"))
ov2 <- core.25ka%>%group_by(Chr_bnd)%>%mutate(a=start-end_bnd,b=end-sta plyr::select(-a,-b)%>%
mutate(-
at=ifelse(dist>=0&dist<5000,"<5k",ifelse(dist>5000&dist<10000,"<10k",ifelse(dist>=10000&dist<15000,"<15k",ifelse(dist>
15000&dist<20000,"<20k",ifelse(dist>20000&dist<250
00,"<25k","Ot"))))))
core.25ka <- rbind(ov,ov2)%>%dplyr::select(-start_b-
```

CHAPTER 6

Nuclear genome organization in Fungi: From Gene folding to Rabl Chromosomes

David E. Torres[1,2], Andrew T. Reckard[3], Andrew D. Klocko[3], and Michael F. Seidl[1]

[1]Utrecht University, NL
[2] Wageningen University and Research, NL
[3]University of Colorado Colorado Spings, USA

Published in *FEMS Microbiology Reviews, 47 (3)*, fuad021.

Abstract

Comparative genomics has recently provided unprecedented insights into the biology and evolution of the fungal lineage. In the post-genomics era, a major research interest focuses now on detailing the functions of fungal genomes, i.e., how genomic information manifests into complex phenotypes. Emerging evidence across diverse eukaryotes has revealed that the organization of DNA within the nucleus is critically important. Here, we discuss the current knowledge on the fungal genome organization, from the association of chromosomes within the nucleus to topological structures at individual genes and the genetic factors required for this hierarchical organization. Chromosome conformation capture followed by high-throughput sequencing (Hi-C) has elucidated how fungal genomes are globally organized in Rabl configuration, in which centromere or telomere bundles are associated with opposite faces of the nuclear envelope. Further, fungal genomes are regionally organized into Topologically Associated Domain-like (TAD-like) chromatin structures. We discuss how chromatin organization impacts the proper function of DNA-templated processes across the fungal genome. Nevertheless, this view is limited to a few fungal taxa given the paucity of fungal Hi-C experiments. We advocate for exploring genome organization across diverse fungal lineages to ensure the future understanding of the impact of nuclear organization on fungal genome function.

Introduction

In eukaryotes, the DNA within the nucleus is organized as chromatin, a dynamic complex formed by DNA in association with proteins (Campos and Reinberg 2009). Approximately 146 bp of DNA is wrapped around an octamer of histone proteins, two copies of histones H2A, H2B, H3, and H4, that together form the nucleosome core particle (Luger et al. 1997), which represents the smallest organizational unit of chromatin (Campos and Reinberg 2009). DNA and histone proteins are subjected to a plethora of chemical modifications such as, but not limited to, methylation or acetylation of lysine residues in histone proteins, and specific combinations of these modifications - the histone code - have been associated with the formation and maintenance of, as well as dynamic transitions between, different local chromatin organization (Strahl and Allis 2000; Jenuwein and Allis 2001). Depending on its local compaction state, chromatin can be broadly subdivided into two types where the DNA is more (euchromatin) or less (heterochromatin) accessible to the transcriptional machinery, thereby directly influencing gene expression (Bannister and Kouzarides 2011; Allshire and Madhani 2018). Euchromatin is considered to be transcriptionally active, while heterochromatin is typically transcriptionally silent (Jenuwein and Allis 2001). Heterochromatin can be further subdivided into constitutive heterochromatin that is largely devoid of genes, and facultative heterochromatin that typically contains genes that are transcriptionally repressed during specific developmental or environmental conditions (Wang et al. 2016; Trojer and Reinberg 2007; Allshire and Madhani 2018; Saksouk et al. 2015).

Local chromatin can be folded in the three-dimensional (3D) space of the nucleus to structurally organize and package DNA, while also enabling precise gene expression (Hoencamp et al. 2021; Sexton and Cavalli 2015; Bonev and Cavalli 2016; Lieberman-Aiden et al. 2009). For instance, higher-order chromatin structures allow the spatial association of genomic elements that are physically separated on the linear DNA strand or occur on different chromosomes, and conversely such structures can physically segregate nearby genomic sites through specific folding barriers (Tolhuis et al. 2002; Lieberman-Aiden et al. 2009; West and Fraser 2005). Consequently, the disruption of DNA folding can result in ectopic interactions of genes and their regulatory regions (Krumm and Duan 2019; Sexton and Cavalli 2015; Bonev and Cavalli 2016), for instance, by hijacking enhancers as observed for proto-oncogenes (Northcott et al. 2014; Flavahan et al. 2016). Consequently, it is important to understand how the 3D genome organization in eukaryotes is established, how it affects genome function, and how this chromatin organization and the processes that establish it are conserved between eukaryotes. While much about genome organization and its function is known in a few animal and plant model systems, there is a dearth of knowledge about these processes in most eukaryotes, including fungi.

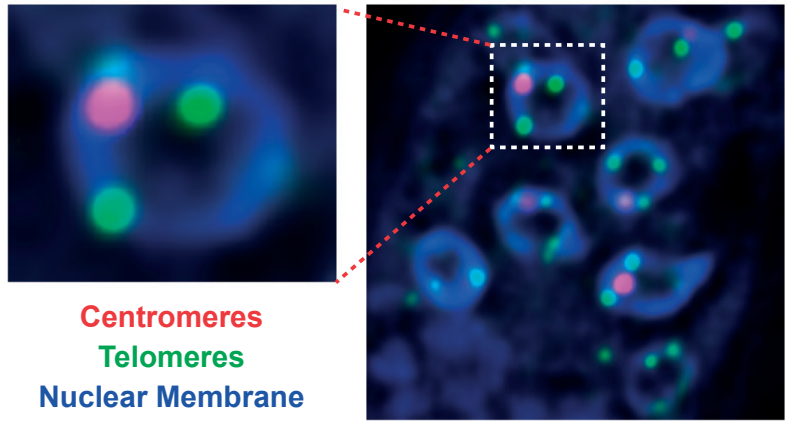
The fungal kingdom is estimated to contain millions of different species Fungi are important components of worldwide ecosystems as decomposers of organic material derived

from plants and animals, and many have been exploited for decades in food production and biotechnology (Stajich et al. 2009). Many fungi are also symbionts that can engage in mutualistic to parasitic interactions with other organisms, which can have significant economic and ecological impact (Fisher et al. 2020, 2012). For instance, in agriculture, plant pathogenic fungi can cause devastating epidemics in staple and commodity crops necessary for the survival of billions of humans (Fisher et al. 2020). Importantly, many fungi are also outstanding model systems for higher eukaryotes, including humans; essential eukaryotic functions are often conserved in fungi, yet fungi are typically more simplistic, genetically tractable, and research is cost-efficient. In the last decade, research into chromatin and the organization of fungal genomes has largely focused on the bakers' yeast *Saccharomyces cerevisiae* and fission yeast *Schizosaccharomyces pombe*, given their ease of study and the plethora of available mutant strains and genetic tools. Due to the ecological and economic relevance of fungi, it is important to further advance research into the function of fungal genomes with the goal of elucidating the regulation of gene expression and discovering how this regulatory control is influenced by nuclear genome organization. Here, we review the current knowledge of fungal genome organization, organized hierarchically from topological structures at the level of individual genes through the association of chromosomes in the nucleus to form the Rabl chromosome conformation, in which centromeres and telomeres cluster distinctly at the nuclear periphery (Box 1). Recent work in fungal model systems and increasingly in more diverse fungi has started to elucidate pertinent sub-nuclear chromatin structures that will be pivotal to our understanding of the function and conservation of genome organization and will provide the framework to detail how nuclear processes impact genome functions in fungi. To date, only a few high-resolution fungal Hi-C datasets are available, but an increasing number of datasets with lower resolution (>5 kb) from a wider diversity of fungi now enable researchers to draw general conclusions about hierarchical 3D structures observed that organize the fungal genome.

Box 1. The history of deciphering genome organization

Historically, studies on genome organization date to the late 1800s when Carl Rabl made his seminal observations on chromosome organization in eukaryotic nuclei (Cremer and Cremer 2010, 2006). Using light microscopy, he observed that centromeres are located on one side of the nucleus and the telomeres are found at the opposing side, an observation that was consistently maintained during the cell cycle. Further advances in the 1950s, made by examining cells with electron microscopy, elucidated the partitioning of the active and silent chromatin: the densely compacted heterochromatin predominantly localizes at the nuclear periphery while the more-open euchromatin is mostly found in nucleus center (Cremer and Cremer 2006; Shatkin and Tatum 1959). The advent of fluorescent microscopy allowed individual genomic features to be examined (Renz 2013), either with dyes that generally stained the DNA (e.g., DAPI), fluorescent probes to examine individual loci (e.g., Fluorescent In Situ Hybridization [FISH] experiments), fluorescently-tagged proteins to examine the nuclear location and dynamics of proteins *in vivo* in live cells, or fluorescently-labeled antibodies that highlight the localization of individual proteins in the nucleus in fixed cells. These advances elucidated the alternating banding pattern of euchromatin and heterochromatin in *Drosophila* salivary gland polytene chromosomes (Bridges 1938; Zhimulev et al. 2014), how some genes co-localize with RNA Polymerase II in possible co-transcriptional hubs (Schoenfelder et al. 2010), and the Rabl chromosome conformation that organizes fungal genomes into weak chromosome territories; in a Rabl conformation, the centromeres from each chromosome cluster into a single focus at the nuclear periphery, while at a different inner nuclear membrane location, the telomeres of each chromosome associate into several foci (Guacci et al. 1994; Klocko et al. 2016; Funabiki et al. 1993; Goto et al. 2001; Gasser 2002) (Box 1 figure). However, these methods are limited by the number of fluorescent dyes with unique emission wavelengths and the resolution of microscopic images (Carlton 2008; Lichtman and Conchello 2005), thereby restricting the number of loci and/or proteins that could be examined at the same time and the resolution by which we can study the interactions of individual proteins and/or co-localization of genes.

Recently, these challenges have been addressed by the introduction of **c**hromosome **c**onformation **c**apture (3C)-based experiments, which revolutionized the field of nuclear genome organization (Dekker et al. 2002). In 3C, chromatin is crosslinked with non-specific crosslinking agents (e.g., formaldehyde), underlying DNA is digested with restriction enzymes, and interacting genomic loci are ligated into a single DNA molecule. Traditional 3C uses a PCR reaction with target-specific primers to show interacting genomic loci. Coupling 3C to high-throughput sequencing (Hi-C) allows researchers to examine on a genome-wide scale how genomic regions interact (Lieberman-Aiden et al. 2009). Typically, Hi-C results are displayed as a two dimensional (2D) heatmap, with the x- and y-axes plotting the genomic location on the chromosome(s), and the intensity of color showing the strength of interactions as measured by the contact probability (Fig. 4), although the 3D chromosome folding can also be modeled from Hi-C data (Oluwadare et al. 2019; Galazka et al. 2016), thereby allowing researchers to infer the folding of chromosomes in exquisite detail and derive functional and structural mechanisms for how this organization occurs.



Box 1 Figure. Fluorescent microscopy can assess specific chromosomal features in fungi.
The clustering of centromeres, which is independent of the co-localization of telomeres, and the association of these features with the nuclear membrane is instrumental to the formation of a Rabl chromosome conformation in fungal nuclei. Fluorescent micrograph of conidia from a *N. crassa* strain expressing CenH3::iRFP (red) to illuminate centromeres, TZA1::GFP (previously reported as TRF1::GFP; green) to highlight telomeres, and ISH1::BFP (blue) demarcate the nuclear membrane (Klocko et al. 2016). The image on the left is an enhanced image of a single nucleus from the conidia image on the right. Methods for imaging are detailed in Supplementary Methods.

The composition of fungal genomes

Over the last few decades, advances in genome sequencing technologies have provided a wealth of novel insights about the genome composition of fungi, which is a prerequisite to study nuclear genome organization in detail. Yeasts are arguably the most well-studied fungal model organisms (Liti 2015; Botstein and Fink 2011), in part due to their relatively small and well-characterized genomes. For example, the genome of the budding yeast *Saccharomyces cerevisiae* is ~12 Megabases (Mb) in size divided into sixteen chromosomes, while the fission yeast *Schizosaccharomyces pombe* has a 13 Mb genome divided over only three chromosomes (Goffeau et al. 1996; Engel et al. 2014; Wood et al. 2002). Relative to the more simplistic yeasts, the genomes of filamentous fungi are typically larger and more complex, but vastly smaller than those of higher metazoans whose genomes often comprise billions of base pairs. Specifically, the genomes of filamentous fungi are typically 30 to 50 Mb in size, divided into variable chromosome numbers (Mohanta and Bae 2015), with several species' genomes being much larger (Kiran et al. 2016; Porto et al. 2019). Several examples with nearly-complete genome assemblies of more well-studied fungal organisms include the saprophyte and model organism *Neurospora crassa* (41 Mb divided into seven Linkage Groups [LG] or chromosomes), the soil-borne plant pathogen *Verticillium dahliae* (34 Mb divided into eight chromosomes), the human pathogens *Aspergillus fumigatus* (28 Mb across

eight chromosomes) and *Cryptococcus neoformans* (19 Mb across five chromosomes), the wheat head blight fungus *Fusarium graminearum* (36 Mb across four chromosomes), and the grass endophyte *Epichloë festucae* (35 Mb across seven chromosomes) (Faino et al. 2015; Loftus et al. 2005; King et al. 2015; Nierman et al. 2005; Galagan et al. 2003; Bowyer et al. 2022; Winter et al. 2018).

Fungal genomes are well-known to be highly dynamic with discrete regions enriched with polymorphisms and chromosomal rearrangements (Dong et al. 2015; Raffaele and Kamoun 2012; Torres et al. 2020; Möller and Stukenbrock 2017). These variable regions are often embedded in different chromosomes or even comprise complete chromosomes. For instance, regions in proximity to telomeres can be highly variable in budding yeast, *Aspergillus, Neurospora*, or *Magnaporthe* species where they display frequent presence/absence of variation and chromosomal rearrangements (Brown et al. 2010; Yue et al. 2017; Farman and Kim 2005; Farman 2007; Starnes et al. 2012; McDonagh et al. 2008; Chang and Ehrlich 2010; Jamieson et al. 2013, 2018). Other species like *Verticillium dahliae* contain similar Adaptive Genomic Regions (AGRs) that are embedded within different chromosomes and contain *in planta* expressed genes (de Jonge et al. 2013, 2012; Faino et al. 2016; Cook et al. 2020). Entire chromosomes are variable within strains of *Fusarium oxysporum* or *Zymoseptoria tritici*, as they contain genes important for the biology of these fungi (Möller et al. 2019; van Dam et al. 2017). It is becoming increasingly evident that dynamic regions in fungal genomes are often composed of highly repetitive relicts of transposable elements (Möller and Stukenbrock 2017; Frantzeskakis et al. 2019). Consequently, these regions are known to have a high number of adenine and thymine (AT) base pairs (bp), which lower the overall guanine and cytosine (GC; can include G:C base pairs or GpC/CpG dinucleotides in a single strand) content of fungal genomes from an approximately 50% GC bp composition. These AT-rich regions, termed AT-isochores to specifically delineate small genomic regions depleted of GC bp (Testa et al. 2016), can be interspersed in the core chromosomes or localized on accessory chromosomes. For example, one third of the genome of the fungus *Leptosphaeria maculans,* a Dothideomycete pathogen of the canola plant crops for producing rapeseed oil, contains blocks of dense AT-rich sequences derived from transposable elements that often house virulence genes (Fudal et al. 2007; Rouxel et al. 2011). Similarly, dense AT- and repeat-rich regions are distributed throughout the *Epichlöe festucae* genome, comprising roughly 25% of its genome (Testa et al. 2016; Winter et al. 2018), and ~16% of the *Neurospora crassa* genome is interspersed with AT-rich, yet gene poor isochores (Fig. 1) (Selker et al. 2003; Lewis et al. 2009; Testa et al. 2016). Many AT-rich sequences in fungal genomes are derived from the action of Repeat Induced Point-Mutation (RIP) (Selker and Stevens 1987; Selker 1990; Freitag et al. 2002), where the GC nucleotides in duplicated or repeated sequences are heavily mutated to contain numerous AT transition mutations, thereby inactivating the underlying sequence and increasing the local AT-content. RIP has been shown to be active in a plethora of fungal species (Gladyshev 2017; Clutterbuck 2011; Hane et al. 2015; Cambareri et al. 1991), and specifically in *Neurospora*,

RIP'd transposable element relicts comprise essentially most of the AT-rich sequences across the genome (Lewis et al. 2009) (Fig. 1).

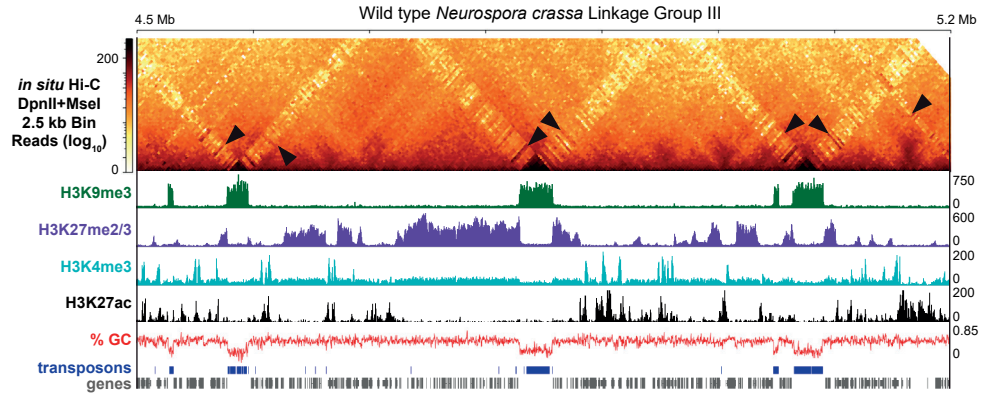


Figure 1. Chromatin profile and evidence for cross-compartment (heterochromatic-euchromatic) contacts in *Neurospora crassa*. Histone post-translational modifications (both activating [H3K4me3 and H3K27ac] and repressive [H3K9me3 and H3K27me2/3]) demarcate the chromatin of *Neurospora crassa*. Regions of the *N. crassa* genome that contain numerous transposon relicts are often rich in adenine/thymine base pairs (AT-rich) and devoid of genes. The occurrence of Hi-C interactions between euchromatin (or facultative heterochromatin) and H3K9me3-marked constitutive heterochromatin may be important for gene regulation. (**Top**) Corrected *in situ* Hi-C heatmap at 2.5 kb resolution showing interactions across a ~700 kb section of the right arm on Linkage Group III. Arrowheads highlight interactions between H3K9me3-marked heterochromatic regions and euchromatin. (**Bottom**) The different tracks show enrichment of histone post-translational modifications based on Chromatin Immunoprecipitation sequencing (ChIP-seq) experiments, the calculated percent guanine/cytosine bases in the genome, predicted transposable elements, and genes. Methods for Hi-C and ChIP-seq data processing and image generation are detailed in Supplementary Methods

Conserved heterochromatic region features in fungal genomes

AT-rich repetitive isochores direct the formation of the silent constitutive heterochromatin (Miao et al. 2000; Lewis et al. 2009). As initially elucidated in *Neurospora*, AT-rich isochores, by an as of yet unknown mechanism, recruit the histone methyltransferase complex DCDC (***D***IM-5/-7/-9 ***C***UL4 ***D***DB1$^{dim\text{-}8}$ ***C***omplex) (Miao et al. 2000; Lewis et al. 2009; Rountree and Selker 2010; Courtney et al. 2020; Lewis et al. 2010a). DCDC catalyzes the tri-methylation of lysine 9 on histone 3 (H3K9me3) of nucleosomes within AT isochores (Fig. 1), with KMT1$^{DIM\text{-}5}$ (Lysine [***K***] ***M***ethyl ***T***ransferase-***1*** / ***D***efective ***I***n ***M***ethylation-***5***) being the SET domain-containing subunit with histone methyltransferase catalytic activity (Lewis et al. 2010a, 2010b; Freitag 2017). H3K9me3 is bound by HP1 (***H***eterochromatin ***P***rotein-***1***) to directly recruit the DNA methyltransferase DIM-2 for methylation on the 5th intracyclical carbon of cytosine bases (5^{m}C) (Freitag et al. 2004; Honda and Selker 2008; Nielsen et al. 2002). Consequently, AT-rich isochores are readily observable across fungal chromosome sequences as they appear as peaks of enrichment of H3K9me3 (Figure 1) and cytosine methylation interspersed among the gene-rich euchromatin (Lewis et al. 2009). One important subclass

of AT-rich isochores in fungal genomes are centromeres, which are critical for chromosomal function as they facilitate homologous chromosome pairing during mitosis/meiosis (Yadav et al. 2018a; Previato de Almeida et al. 2019; Kurdzo et al. 2017; Krassovsky et al. 2012; Wells et al. 2006; Stewart and Dawson 2004). Chromosomes of the budding yeast *S. cerevisiae* contain a 'point' centromere that comprises ~200 bp in a single nucleosome, while most other fungi have 'regional' centromeres spanning thousands of base pairs (Lefrançois et al. 2013; Yadav et al. 2018a; Galazka et al. 2016; Malik and Henikoff 2009; Smith et al. 2012; Seidl et al. 2020). Similar to AT-rich isochores, centromeres are densely enriched for H3K9me3 and $5^{m}C$ (Lewis et al. 2009; Galazka et al. 2016; Smith et al. 2012; Cook et al. 2020; Seidl et al. 2020), yet are differentiated from other AT-rich transposon relicts by the deposition of the centromere-specific histone variant CenH3 (Galazka et al. 2016; Yadav et al. 2018a; Meluh et al. 1998; Malik and Henikoff 2009; Smith et al. 2012; Seidl et al. 2020).

Facultative heterochromatin can dynamically transition between being densely compacted and thereby silencing the expression of the underlying DNA to looser compaction resulting in chromatin being more open for activating transcription (Trojer and Reinberg 2007). Facultative heterochromatin in fungi is typically delineated by either the di- or tri-methylation of lysine 27 on histone H3 (H3K27me2/3) (Figure 1) catalyzed by the PRC2 (***P***olycomb ***R***epressive ***C***omplex-***2***; the SET domain protein $KMT6^{SET-7}$ is the subunit that specifically methylates histones) (Wiles and Selker 2017; Freitag 2017). This facultative heterochromatin mark is densely enriched over predominantly gene-rich regions of the genome (Fig. 1) to repress gene expression (Jamieson et al. 2013; Basenko et al. 2015; Dumesic et al. 2015). In *N. crassa*, H3K27me2/3 is enriched over sub-telomeric chromosomal regions that are positionally dependent on the presence of telomere repeats, although chromosome-internal, position independent peaks of H3K27me2/3 are readily observed (Jamieson et al. 2013, 2018; Klocko et al. 2016; Basenko et al. 2015). In other fungi, large stretches of H3K27me3 enrichment can be observed along core chromosome arms (Zhang et al. 2021b; Schotanus et al. 2015; Kramer et al. 2022; Connolly et al. 2013; Carlier et al. 2021; Cook et al. 2020; Soyer et al. 2021). For example, the fusion of multiple subtelomeric domains, which form the four chromosomes in *F. graminearum*, result in large chromosome-internal domains of H3K27me3 (Connolly et al. 2013), while H3K27me3 represses genes in the dynamic AGRs regions in *V. dahliae* (Cook et al. 2020; Kramer et al. 2022), and across accessory chromosomes in *Z. tritici* or *F. oxysporum* (Schotanus et al. 2015; Fokkens et al. 2018). Interestingly, the deposition of the H3K27me2/3 mark varies with the protein subunits associated with PRC2, as loss of PRC2-associated proteins NPF (***N****eurospora* **p5**5 ortholog) or PAS (***P***RC2 ***A***ccessory ***S***ubunit) abolishes subtelomeric H3K27me2/3 in *N. crassa*, while the conserved protein EPR-1 is known to 'read' the H3K27me2/3 mark for gene repression; in support, deletion of the EPR1 homolog in *F. graminearum*, BP1, phenocopies loss of KMT6 and BP1 directly binds DNA to stabilize nucleosomes for enhancing transcriptional repression (Wiles et al. 2020; McNaught et al. 2020; Jamieson et al. 2013; Klocko et al. 2016). Genomic regions that display properties of both constitutive

and facultative heterochromatin are typically not observed (Galazka et al. 2016; Klocko et al. 2016), yet some species-specific exceptions occur. For instance, the genome of *Z. tritici* contains large domains enriched for both H3K27me2/3 and H3K9me3, while only a few larger interspersed heterochromatic regions in *V. dahliae* or only the telomeres of *N. crassa* are enriched for both marks (Schotanus et al. 2015; Klocko et al. 2016; Cook et al. 2020). In addition, a newly emerging yet understudied facultative heterochromatic mark required for repressing gene expression in *N. crassa* and *Fusarium fujikuroi* is di- or tri-methylation of lysine 36 on histone H3 (H3K36me2/3), which is catalyzed by the SET-domain containing histone methyltransferase ASH1 (Bicocca et al. 2018; Janevska et al. 2018). In *N. crassa*, ASH1-specific H3K36me2 is enriched over lowly-expressed genes that are also enriched with H3K27me2/3, and loss of ASH1 causes altered enrichment of H3K27me2/3 (Bicocca et al. 2018). Similarly, in *F. fujikuroi*, ASH1 methylates H3K36 to establish facultative heterochromatin at sub-telomeres; ASH1 deletion increases sub-telomeric H3K27me3 enrichment, yet genes at the sub-telomeres become unstable (Janevska et al. 2018). These data argue that ASH1-specific H3K36me2/3 is epistatic to the action of the PRC2 complex and highlighting the need for additional study on the dynamics of facultative heterochromatin marks in multiple fungal species.

Another subclass of AT-rich isochores with distinct functions in fungal genomes are found at the telomeres, which, similar to other eukaryotes, delineate and protect chromosome ends (Rahnama et al. 2021). Like those in higher metazoans, fungal telomeres require telomerase and a shelterin complex (Jamieson et al. 2018; Yadav et al. 2021; Nandakumar and Cech 2013; Smogorzewska and de Lange 2004). Telomerase uses a ribonucleic acid template to synthesize short DNA repeats ($[5'\ TTAGGG]_n$) onto the chromosome ends (Jain and Cooper 2010; Qi et al. 2013; Greider and Blackburn 1985, 1989; Szostak and Blackburn 1982), while the shelterin complex protects the single stranded telomeric repeats from being degraded and/or recognized as a double strand DNA (dsDNA) break (Jain and Cooper 2010; Zinder et al. 2022).

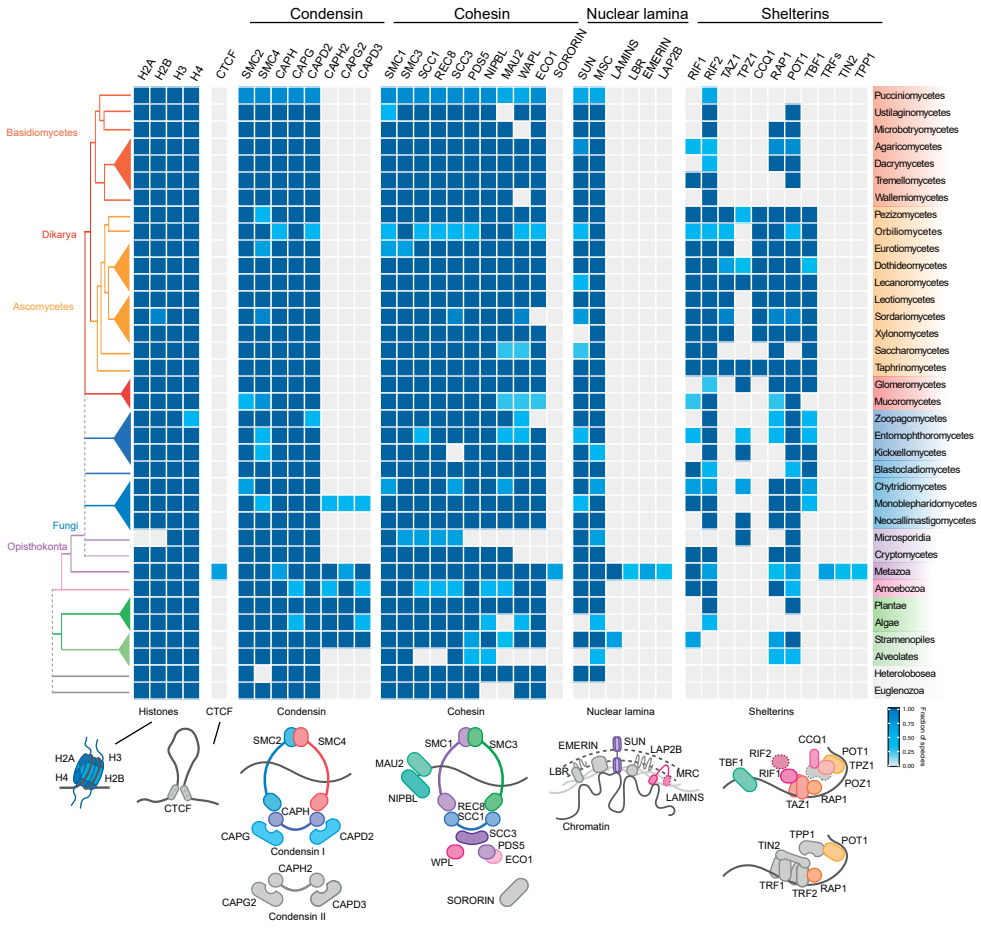


Figure 2. The protein landscape of chromatin organization in fungi.
Phylogenetic profile of 41 well-studied proteins known to impact chromatin organization encoded in 88 eukaryotes. These proteins are divided into the core histone proteins, CTCF, condensin subunits (types I & II), cohesin subunits, nuclear envelope-associated proteins, and shelterin-associated proteins (see Supplementary Table 2 for specific gene information). Gene orthologs encoding most of the proteins involved in genome organization are conserved in fungi, with the exceptions of CTCF proteins associated with loop formation, condensin II for chromosome territories, Lamins at the nuclear periphery, and multiple components of the shelterin complex for telomere protection/organization. The heatmap color code reflects the fraction of species of a specific taxonomic lineage in which orthologs were found, and gray boxes indicate the absences of that specific protein. Known phylogenetic relationships between the different fungal lineages and other taxonomic groups (**right**) are shown on the left. Schematic representations that summarize the composition of the analyzed protein complexes are shown on the bottom; colored subunits depict presence in fungi, while gray colored subunits depict absence. Methods for ortholog searches are detailed in Supplementary Methods.

In the mammalian shelterin complex, the proteins TRF1 (***T***elomeric ***R***epeat-binding ***F***actor ***1***), TRF2 (***T***elomeric ***R***epeat-binding ***F***actor ***2***), and POT1 (***P***rotection ***O***f ***T***elomeres protein ***1***) bind the telomeric repeats while TIN2 (***T***ERF1-***I***nteracting ***N***uclear factor ***2***), TPP1 (***T***INT1, ***P***TOP, ***P***IP***1*** complex), and RAP1 (***R***epressor ***A***ctivator ***P***rotein ***1***) mediate protein intra-complex interactions (Palm and de Lange 2008; Myler et al. 2021; de Lange

2005). However, only RAP1 and POT1 have clear fungal homologs (Fig. 2). In fact, fungal shelterin complexes are quite diverse, as highlighted by the divergence of shelterin complex members in two closely related yeast species **(**Fig. 2) (Xue et al. 2017; Steinberg-Neifach and Lue 2015; Erlendson et al. 2017). In *S. cerevisiae*, RAP1 directly recognizes telomeric dsDNA and recruits RIF1/2 (***R***AP1 ***I***nteracting ***F***actor ***1/2***) to regulate telomere length, while TBF1 (***T***TAGGG ***B***inding ***F***actor ***1***), a yeast ortholog of human TRF1, binds to single stranded telomeric DNA to antagonize improper telomere lengthening (Li et al. 2000; Kaizer et al. 2015; Irie et al. 2019; Kabir et al. 2010; Pitt et al. 2008; Bilaud et al. 1996; Hediger et al. 2006). In contrast, the *S. pombe* shelterin complex contains the proteins TAZ1 (***T***elomere-***A***ssociated in *Schi***Z***osaccharomyces pombe **1***; a homolog of mammalian TRF1/2), RAP1, RIF1, POT1, TPZ1 (**T***elomeres* **P***rotection protein in Schi***Z***osaccharomyces pombe **1***), and CCQ1 (***C***oiled-***C***oil **Q**uantitatively-enriched protein ***1***) for protecting telomeric DNA and silencing sub-telomeric chromatin (Irie et al. 2019; Baumann and Cech 2001; Fujita et al. 2012; Deng et al. 2015; Zofall et al. 2016; Cooper et al. 1997; Kanoh and Ishikawa 2001; Miyoshi et al. 2008). In filamentous fungi, the shelterin complexes of *N. crassa* and other ascomycetes appear to be an amalgam of budding and fission yeast shelterins, based on protein conservation (Fig. 2). *Neurospora* encodes a TAZ1 homolog to bind to dsDNA and a TBF1 homolog to bind to single stranded [5' TTAGGG]$_n$ telomeric DNA; TAZ1 (labeled TRF1 in (Galazka et al. 2016; Klocko et al. 2016)) is a functional homolog of mammalian TRFs that is restricted to ascomycetes (Figure 2). Other fungal species may employ unique shelterin complex proteins, including the basidiomycete corn smut pathogen *Ustilago maydis* that uses the telomere-repeat binding proteins UmTay1, UmPot1, and UmTrf2 (Yu et al. 2020; Zahid et al. 2022; Yu et al. 2018). Variability at yeast telomere repeats might partially explain this shelterin complex diversity in fungal species, given that different proteins would be required to recognize alternative telomere repeat sequences or differing telomere lengths (Červenák et al. 2021, 2017; Malyavko et al. 2014, 2019; Sepsiova et al. 2016; Lue 2021). For example, *S. cerevisiae* telomeres have the sequence [T(G)$_{2\text{-}3}$(TG)$_{1\text{-}6}$], while *S. pombe* telomeres have [TTACAG$_{2\text{-}5}$] (Červenák et al. 2021). In contrast, *N. crassa* and other filamentous fungi use the canonical [5' TTAGGG$_n$] repeat found in metazoans to recruit the shelterin complex (Casas-Vila et al. 2015; Erlendson et al. 2017), arguing filamentous fungi could be more appropriate models for studies on telomere protection in higher organisms.

Conserved chromosomal features in euchromatic regions in fungal genomes

Euchromatic genes have the potential to be transcriptionally active, but presumably control of gene expression can be driven by transcription factor recruitment, deposition of specific histone post-translational modifications, or even the 3D genome organization. Many fungi share similar epigenetic marks as higher eukaryotes over actively transcribed genes, including the extensive acetylation of histone proteins and the di- or tri-methylation of lysine 4 on histone H3 (H3K4me2/3) (Lewis et al. 2009; Bicocca et al. 2018; Courtney et

al. 2020; Xiong et al. 2010; Anderson et al. 2010; Smith et al. 2010; Zhu et al. 2019; Pokholok et al. 2005). As in metazoans, H3K4me3 is primarily enriched at promoter regions (Fig. 1), while H3K4me2 is predominantly deposited across gene bodies (Pokholok et al. 2005; Liu et al. 2005; Zhu et al. 2019; Cemel et al. 2017). H3K4me2/3 is catalyzed by the conserved COMPASS complex, which contains the SET-domain protein KMT2^{SET-1} as its catalytic subunit, but the composition of associated subunits can vary (Miller et al. 2001; Roguev et al. 2001; Freitag 2017). All four histone proteins within euchromatic nucleosomes are also substantially acetylated, with multiple lysine residues being subject for acetylation, including lysine 27 on histone H3 (H3K27ac; Fig. 1) (Bicocca et al. 2018). H3K27ac is also observed in *M. oryzae*, where loss of H3K27me3 and gain in H3K27ac increases transcription of genes necessary for plant infection (Zhang et al. 2021b). In general, acetylation appears dense over gene bodies and intergenic spacing, but little acetylation is found in AT-rich isochores, as deacetylase complexes, such as the HCHC in *N. crassa*, act upon heterochromatic regions to remove acetyl groups to induce chromatin condensation (Honda et al. 2012, 2016).

The distribution of these 'activating' histone modifications can influence the formation of nucleosomes in euchromatin by altering the association of histone proteins with DNA, thereby changing the accessibility of the underlying DNA (Bannister and Kouzarides 2011; Smolle and Workman 2013). Euchromatin accessibility has been assessed by ATAC-seq (**A**ssay for **T**ransposase **A**ccessible **C**hromatin-sequencing). For example, *Neurospora* displays highly accessible chromatin regions (ACRs) that are often found upstream of genes: most are small, but a subset of these intergenic ACRs are large (>2,000 bp), presumably for multiple transcription factors to control of expression of ACR-proximal genes; these open chromatin regions are enriched with acetylation of H3K27 to possibly 'open' the chromatin (Ferraro et al. 2021). Euchromatin is characterized by a more open nucleosome conformation that would also allow the underlying DNA to be more accessible. Specifically, the promoter DNA of expressed genes is typically devoid of histone proteins, forming a nucleosome free region that allows transcription factor binding to *cis*-regulatory elements for the recruitment of the RNA Polymerase II machinery for transcription initiation (Struhl and Segal 2013; Bai and Morozov 2010; Radman-Livaja and Rando 2010). Nucleosome-free regions (NFR) at promoters with more accessible chromatin have been widely observed in fungi, including *S. cerevisiae* (Oberbeckmann et al. 2019), *N. crassa* (Ferraro et al. 2021; Klocko et al. 2019), *C. albicans* (Jenull et al. 2020), *Aspergillus niger* (Huang et al. 2021), *V. dahliae* (Cook et al. 2020), and *Ustilaginoidea virens* (Chen et al. 2021b). Histone variants that alter nucleosome composition are also known to impact chromatin accessibility across eukaryotic genomes (Talbert and Henikoff 2010; Henikoff and Smith 2015; Talbert and Henikoff 2017). For example, the placement of the histone variant H2A.Z, which replaces H2A in the histone octamer, is enriched at the first nucleosome immediately downstream of the transcription start site of actively transcribed genes, possibly functioning to impact transcription by altering the chromatin structure at fungal promoters (Dong et al. 2018; Chen and Ponts 2020; Martire and Banaszynski 2020). Additionally, modifying the position of nucleosomes

in the open reading frames of genes or within intergenic regions can impact the accessibility or composition of chromatin. Nucleosome positions are altered by chromatin remodelers, and in fungi, well-studied chromatin remodelers include the SWI-SNF complex and the AAA-ATP member DIM-1/CATP, the latter of which impacts the enrichment of H3K9me3 and $5^{m}C$ in intergenic regions in euchromatin (Wiles et al. 2022; Kamei et al. 2021; Cha et al. 2013; Klocko et al. 2019).

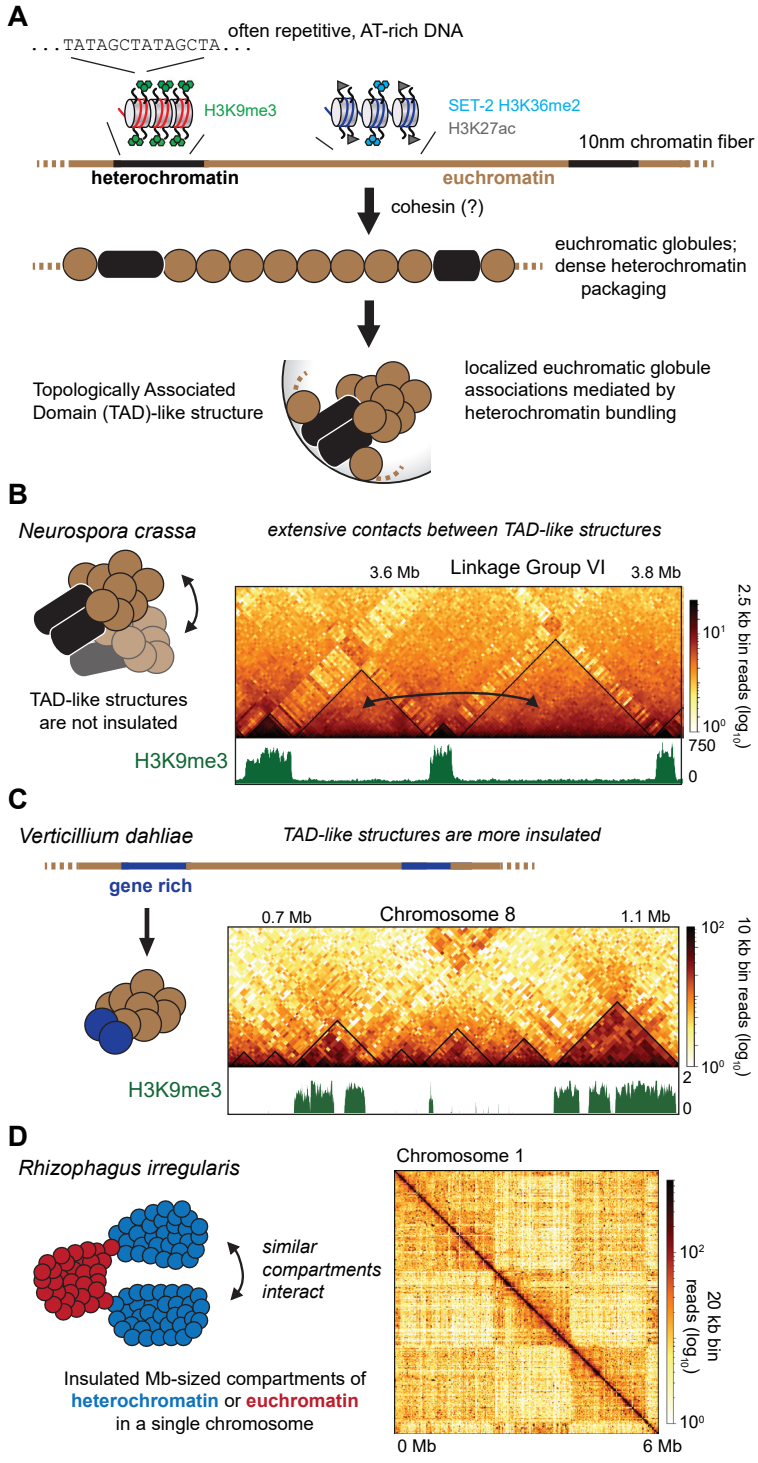
The spatial organization of fungal chromatin

During interphase, the nucleosomes package DNA into 10 nanometer (nm) chromatin fibers, a beads on a string conformation that in turn may aggregate further into 30 nm chromatin fibers (Hansen et al. 2018b; Bernardi 2015; Wako et al. 2020; Maeshima et al. 2016). Presumably, these chromatin fibers would be spatially organized in the nucleus to facilitate proper genome function, including the initiation of transcription to enable timely gene expression (Misteli 2007; Cook 1999). To facilitate this nuclear organization, the chromatin fibers in eukaryotic genomes form a series of subnuclear structures of hierarchically increasing sizes – in fungi, these range from chromatin loops to the Rabl chromosome conformation (Box 1) – that both allow critical genome features to aggregate into nuclear euchromatic/heterochromatic compartments and prevent the formation of knots (or other non-viable DNA strand folding). This hierarchical organization of fungal chromatin folding allows euchromatic genes to have the potential to dynamically associate in spatially-close proximity, thereby forming non-random interactions or higher-order 3D structures critical for gene expression. However, one can speculate that the large-scale folding of chromatin fibers is also necessary to compact the fungal genome in the nucleus to ensure the proper functioning of the DNA templated processes necessary for viability in an identical manner to the role of genome organization in metazoan genome function (Misteli 2007; Cavalli and Misteli 2013; Sexton and Cavalli 2015). It should be noted, however, that most genome organization experiments in fungi are derived from Ascomycetes, and consequently additional work in Basidiomycetes as well as in earlier diverging fungi will be needed to determine which aspects of the hierarchical genome organization detailed below are conserved across fungi.

Chromatin loops/globules

Chromatin loops (often called ‘globules’) are among the most prevalent 3D structures in metazoan genomes (Heger et al. 2012; Lieberman-Aiden et al. 2009; Kadauke and Blobel 2009). Here, chromatin loops occur when two loci physically separated on the linear chromosome associate, and the intervening chromatin condenses. In metazoans, these ~300 kilobases (kb) chromatin loops may functionally influence gene expression by facilitating interactions between distant regulatory sequences, such as enhancers and silencers,

although the possibility exists that chromatin loops only form to structurally organize metazoan genomes (Cavalli and Misteli 2013; Rao et al. 2014, 2017; Dekker and Heard 2015). Two critical components in the formation of chromatin loops in humans are cohesin and condensin I, which are architectural DNA-binding protein complexes important for DNA replication and chromosome folding, as well as for condensing chromosomes for meiosis and mitosis (Rao et al. 2017; Davidson and Peters 2021; Hirano 2012; Green et al. 2012; Haarhuis et al. 2017; Hoencamp et al. 2021; Jeppsson et al. 2022). In the cohesin complex, the ***S***tructural ***M***aintenance of ***C***hromosomes (SMC)-***1***, SMC-3, Scc1, and Scc3 proteins form the core machinery for chromatin looping (Peric-Hupkes and van Steensel 2008; Skibbens 2019; Dorsett 2007). In addition, five to six accessory proteins are responsible for dynamically loading or unloading the core cohesin complex onto chromatin and activating the cohesin ATPase activity (Haarhuis et al. 2017; Yoshida et al. 2022; Davidson and Peters 2021). The loading and activation of the SMC complex onto chromatin in budding yeast is facilitated by the yeast homologs Scc2 and Scc4, as signaled by the acetyltransferase Eco1, while cohesin dissociation from chromatin is mediated by WAPL and PDS5 (Ciosk et al. 2000; Murayama and Uhlmann 2014; Çamdere et al. 2015; Petela et al. 2018; Rolef Ben-Shahar et al. 2008; Chan et al. 2012; Haarhuis et al. 2017). Together, SMC proteins comprise a ring-shape structure in cohesin through which chromatin is actively extruded, thereby forming chromatin loops (Ganji et al. 2018; Fudenberg et al. 2016; Bauer et al. 2021). Chromatin extrusion continues until a boundary element is encountered, which both signals for cessation of extrusion mechanism and the anchoring of the resulting chromatin loop (Ganji et al. 2018). Cohesin remains associated with chromatin through interphase, potentially to stabilize loops, until mitosis dependent Scc1 cleavage occurs (Murayama and Uhlmann 2014; Uhlmann et al. 1999; Nasmyth 2001). Borders of chromatin loops are typically demarcated in metazoan Hi-C datasets by the visualization of a 'focus' or point of enriched contacts off diagonal (Rao et al. 2014). These strong interactions at loop bases are often centered over binding sites for the CTCF (**C**CC**TC** binding **f**actor) (Fudenberg et al. 2016; Brackley et al. 2018; Banigan and Mirny 2020; Davidson and Peters 2021). CTCF-mediated loops are known to insulate nearby active and repressive chromatin regions by blocking further chromatin extrusion (de Wit et al. 2015; Sanborn et al. 2015). Here, two CTCF proteins bind to convergently-oriented asymmetric 14 bp sequences and dimerize to form a loop of chromatin in the genomes of higher eukaryotes (Rao et al. 2014; Ong and Corces 2014; Nora et al. 2017; de Wit et al. 2015). Consequently, depletion of CTCF disrupts loop formation in many cases (Xu et al. 2021; Zuin et al. 2014; Nora et al. 2017).


A
often repetitive, AT-rich DNA
...TATAGCTATAGCTA...
H3K9me3
SET-2 H3K36me2
H3K27ac
10nm chromatin fiber
heterochromatin
euchromatin
cohesin (?)
euchromatic globules;
dense heterochromatin
packaging
Topologically Associated
Domain (TAD)-like structure
localized euchromatic globule
associations mediated by
heterochromatin bundling
B
Neurospora crassa
extensive contacts between TAD-like structures
3.6 Mb
Linkage Group VI
3.8 Mb
2.5 kb bin reads (log10)
TAD-like structures
are not insulated
H3K9me3
750
0
C
Verticillium dahliae
TAD-like structures are more insulated
gene rich
0.7 Mb
Chromosome 8
1.1 Mb
10 kb bin reads (log10)
H3K9me3
2
0
D
Rhizophagus irregularis
similar
compartments
interact
Chromosome 1
20 kb bin
reads (log10)
Insulated Mb-sized compartments of
heterochromatin or euchromatin
in a single chromosome
0 Mb
6 Mb

<<

Figure 3. The formation of regional structures in several fungal organisms.
In filamentous fungi, TAD-like structures generally form through hierarchical clustering of local chromatin structures. (**A**) A generalized schematic of the folding of chromatin from the 'beads-on-a-string' model of the 10 nm fiber in which DNA is bound to histone proteins to form nucleosome, to the formation of loops/globules with the action of cohesin, to the folding of those globules into TAD-like/Regional Globule Clusters in the fungal genome. The folding of TAD-like structures inferred from Hi-C contact data in (**B**) *Neurospora crassa* (Rodriguez *et al.* 2022) and (**C**) *Verticillium dahliae* (Torres *et al.* 2023). (**D**) The compartmentalization of heterochromatic (stronger interactions) and euchromatic regions can be observed in the genome of *Rhizophagus irregularis* (Yildirir *et al.* 2022). Methods for Hi-C data processing and image generation are detailed in Supplementary Methods.

The cohesin core complex and most accessory proteins are highly conserved throughout eukaryotes (Fig. 2), so it is reasonable to expect that chromatin loop formation occurs by a loop extrusion mechanism in fungi as well (Fig. 3A). In fungi, the role of cohesin has been only examined in the fission yeast *S. pombe*, where in a temperature-sensitive cohesin mutant strain the ~40 kb sized globules are no longer visible across the genome (Mizuguchi et al. 2014; Tanizawa et al. 2017). Chromatin loops/globules have been also observed in filamentous fungi. In *N. crassa* slightly larger (~60 to 80 kb) globules are readily visible in Hi-C datasets (Rodriguez et al. 2022). Given the conservation of cohesin subunits in these filamentous fungi (Fig. 2), cohesin should also act across these genomes. However, it is possible that cohesin and its activity are not fully conserved throughout all fungal lineages. In Microsporidia, almost all cohesin complex components are absent (Fig. 2), suggesting that these spore-forming unicellular fungal parasites evolved distinct mechanisms to topologically organize their genome. Further, WAPL and Eco1 seem to be absent in fungi belonging to the Zoopagomycetes (Fig. 2), suggesting novel mechanisms for signaling loop dissociation might exist in these obligate soil nematode parasites. Interestingly, and in contrast to cohesin subunits, the gene encoding CTCF is restricted to higher metazoan species, which typically have larger genomes. One can speculate that the CTCF bound to the base of a chromatin loop anchors, and possibly stabilizes, the larger-sized globules/loops in metazoans (Pugacheva et al. 2020; Li et al. 2020). Other boundary elements apart from CTCF have been found in plants (Dong et al. 2017; Xie et al. 2019), *D. melanogaster* (Ramírez et al. 2018), *C. elegans* (Anderson et al. 2019), and humans (Valton et al. 2022; Anania et al. 2022). However, the specific boundary elements restricting chromatin loop size by repressing cohesin loop extrusion in fungi have not been fully characterized (Mizuguchi et al. 2014; Rodriguez et al. 2022; Schalbetter et al. 2019). Fungi may employ diverse mechanisms to negatively regulate loop formation, including the enrichment of convergent genes that delineate loop 'boundaries' in *S. pombe* (Mizuguchi et al. 2014), or the incorporation of AT-rich, repetitive heterochromatic isochores in the genome, as observed in *N. crassa* (Rodriguez et al. 2022) and *E. festucae* (Winter et al. 2018). While future experiments are needed to assess the requirement of cohesin for forming chromatin loops in fungi (Schalbetter et al. 2019), the hypothesis that constitutive heterochromatic regions possibly act as loop anchors would render CTCF or similar boundaries unnecessary in fungi.

Topologically Associated Domains

Topologically Associated Domains (TADs) are 3D chromosomal structures that regionally organize the genome by subdividing chromatin compartments (Dixon et al. 2012; Acemel and Lupiáñez 2023) suggesting that a single TAD may contain several chromatin loops or become visible when multiple loops are averaged over a cell population (Dekker and Heard 2015; Hansen et al. 2018a) (Fig. 3). TADs in metazoans are typically megabase-sized genomic regions in which chromatin is more apt to contact yet is insulated from chromatin outside of the TAD (Lieberman-Aiden et al. 2009; Dixon et al. 2012). In metazoan Hi-C datasets, TADs appear as triangles of increased contact probability immediately off-diagonal, with few distant contacts with chromatin beyond the TAD border (Fig. 3). Interestingly, TAD borders in metazoans are often enriched for housekeeping genes, tRNA genes, and retrotransposons (Dixon et al. 2012). However, it is currently unclear if TADs are primarily structural, i.e., TAD formation organizes the genome in the nucleus, or are essential for genome function (Beagan and Phillips-Cremins 2020), possibly required to regulate gene expression by increasing the contact probability of distant enhancers/silencers with cognate promoters (Flavahan et al. 2016; Dixon et al. 2018). While altered TAD borders may cause misregulation of gene expression in some human cancers (Valton and Dekker 2016; Taberlay et al. 2016; Akdemir et al. 2020; Flavahan et al. 2016; Dixon et al. 2018), multiple datasets report only minimal changes in gene expression when individual TADs are altered (Ghavi-Helm et al. 2019; Rao et al. 2017). Additionally, TADs change little through interphase of the cell-cycle, but TADs are consistently lost upon entry into mitosis when chromosomes are locally condensed, and are reformed before entry to G1, suggesting TADs are primarily structural in nature (Naumova et al. 2013; Abramo et al. 2019).

TAD formation in higher metazoans appears to require the action of condensin complexes. Specifically, the SMC complexes condensin type I and II are known to shape individual chromosomes by forming higher order looping structures (Hagstrom and Meyer 2003; Hirano 2012). In both complexes, the yeast homologs SMC2 and SMC4 form a highly conserved core ring structure (Schleiffer et al. 2003), but individual condensin types can be distinguished by other subunits: condensin I contains CAPH, CAPG, and CAPD2, while condensin II has CAPH2, CAPG2, and CAPD3 (Ono et al. 2003; Schleiffer et al. 2003). Condensin I has the greatest impact during the mitotic phase of the cell cycle, where it laterally compacts sister chromatids by forming smaller supercoiled chromatin loops (Kimura and Hirano 1997; Green et al. 2012; Hagstrom and Meyer 2003; Golfier et al. 2020; Kong et al. 2020). Condensin II affects high order chromosome organization during all cell cycle stages, as it is found in the nucleus during S phase; during mitosis, condensin II may involve the formation of large chromatin loops, in conjunction with condensin I activity, for establishing a 'nested loop architecture' critical for sister chromatid/homologous chromosome pairing (Kong et al. 2020; Yu and Koshland 2005; Ono et al. 2013).

Emerging evidence from multiple species suggests that filamentous fungi have TAD-like structures analogous to metazoan TADs (Rodriguez et al. 2022; Mizuguchi et

al. 2014; Tsochatzidou et al. 2017; Eser et al. 2017; Winter et al. 2018; Schalbetter et al. 2019). Fungal TAD-like structures are several hundred kilobases in size and the euchromatin internal to these TAD-like structures is more apt to contact (Galazka et al. 2016; Rodriguez et al. 2022; Dixon et al. 2012). In *N. crassa*, TAD-like structures were originally termed Regional Globule Clusters (RGCs) named for the aggregation of several ~40 kilobase euchromatic globules into larger, compact structures ~250 kb in size that could be interpreted as a large chromatin aggregates analogous to metazoan TADs (Fig. 3B). RGCs display extensive yet random internal euchromatic contacts that are not restricted from outside chromatin, as strong inter-RGC contacts readily occur, arguing proteins insulating internal euchromatin are not encoded in *Neurospora crassa* (Rodriguez et al. 2022). RGCs are flanked by constitutive heterochromatic regions to delineate RGC borders. The clustering of heterochromatin regions, possibly through Liquid Liquid Phase Separation (LLPS) condensates (Larson et al. 2017), may act as RGC anchor (Fig. 3B) in an analogous manner to CTCF at chromatin loops (Rodriguez et al. 2022); consequently, cohesin would act specifically to form the smaller globules internal to and comprising this TAD-like structure. Similar patterns of TAD-like structures have been observed in *S. cerevisiae*, where globule structures are delimited by transcriptionally active genes that are often in a convergent orientation (Tsochatzidou et al. 2017; Schalbetter et al. 2019). Additionally, TAD-like structures can be seen in *E. festucae*, where RIP'd AT-rich heterochromatic regions strongly interact to form large structures to compact chromatin (Winter et al. 2018), and in the fungal pathogen *Puccinia striiformis*, which may form un-insulated RGCs across each chromosome arm (Xia et al. 2022). Another example of TAD-like structures is found in *V. dahliae* and related *Verticillium* species where Hi-C datasets display TAD-like structures with increased internal chromatin contact probabilities and few intrachromosomal contacts beyond the TAD-like structure boundaries (Fig. 3C) (Torres et al. 2023). In contrast to the situation in *N. crassa*, TAD-like boundaries in *V. dahliae* and *E. festucae* are enriched for expressed genes, suggesting that TADs in these fungi are necessary for proper gene expression. Additional evidence from yeasts suggests that TAD-like structures are critical for other genome functions apart from gene expression, such as repressing recombination or promoting genome evolution (Tsochatzidou et al. 2017; Mizuguchi et al. 2014; Gu et al. 2022). For example, TADs may be essential for fungal chromosome replication during S-phase of the cell cycle, as ~200-kb TAD-like structures across the *S. cerevisiae* genome separate clusters of early or late timed origins of replication across its 16 chromosomes (Eser et al. 2017).

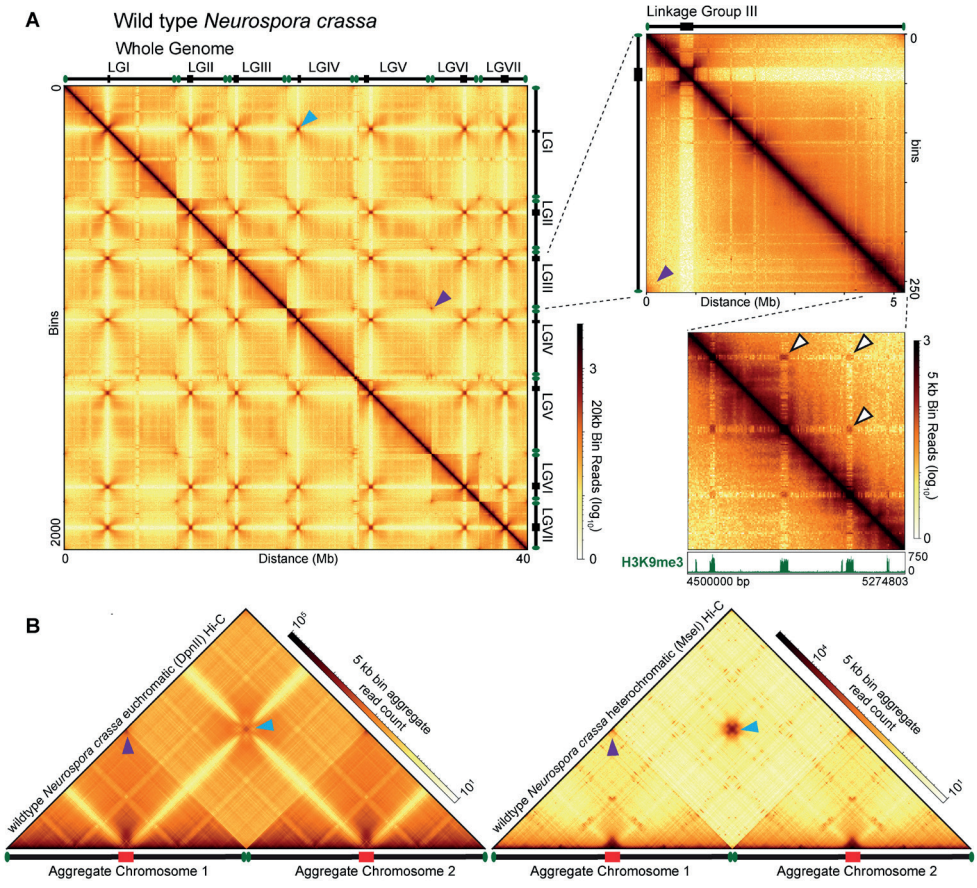


Figure 4. Genome organization of the filamentous fungus *Neurospora crassa*.
The genome organization of the seven chromosomes comprising the entire *Neurospora crassa* genome is characterized by inter-chromosomal centromeric contacts, as well as inter-chromosomal contacts between telomeres in a Rabl chromosome conformation. Individual chromosomes have strong telomeric interactions, while centromeric chromatin strongly self-interacts yet is isolated from other genomic foci. Interestingly, the strongest long-range interactions occur between H3K9me3-enriched constitutive heterochromatic regions. (**A**) Corrected *in situ* Hi-C data of combined DpnII (euchromatin-specific) and MseI (heterochromatin-specific) Hi-C data (Rodriguez *et al.* 2022) at 20 kb resolution, showing the interactions across the entire genome (all seven chromosomes indicated on the top and right of the Hi-C contact map), one chromosome (Linkage Group III), and a zoomed image of the right arm of LG III. Blue arrowheads indicate inter-chromosomal centromere interactions, purple arrowheads indicate inter- or intra-chromosomal telomeric interactions, and white arrowheads indicate strong intra-chromosomal heterochromatic interactions. ChIP-seq track of H3K9me3 enrichment shows regions of constitutive heterochromatin. (**B**) Aggregate chromosome plots, at 5 kb resolution, of DpnII (euchromatin-specific, left) or MseI (heterochromatin-specific, right) *in situ* Hi-C data (Rodriguez *et al.* 2022). Two aggregate chromosome schematics are below the Hi-C contact map, with centromeres denoted by red boxes and telomeres shown by green ovals. Blue arrowheads indicate inter-chromosomal centromere interactions and purple arrowheads indicate intra- or inter-chromosomal telomeric interactions. Methods for Hi-C data processing and image generation are detailed in Supplementary Methods.

Organization of interspersed constitutive heterochromatic regions

AT-rich isochores comprising constitutive heterochromatin can be found embedded throughout fungal chromosomes as well as at centromeres and telomeres (Lewis et al. 2009; Seidl et al. 2020; Winter et al. 2018). *N. crassa* has approximately 300 AT- and H3K9me3-enriched isochores interspersed throughout the genome, which range in size from <1 kb to ~400 kb (Galazka et al. 2016; Klocko et al. 2016). These regions readily associate in the nucleus, as exceptionally strong contacts between silent chromatin regions both within a chromosome and across chromosomes are frequently observed in fungal Hi-C datasets **(**Fig. 4), arguing that the clustering of constitutive heterochromatic regions in the fungal nucleus may be particularly important for the proper chromosome conformation in fungi. However, it is readily apparent in fungal Hi-C datasets that any interspersed constitutive heterochromatic region has the potential to interact, as uniformly strong contacts between all H3K9me3-marked silent regions are readily observed within Hi-C data derived from a population of fungal nuclei (Galazka et al. 2016; Rodriguez et al. 2022; Winter et al. 2018). Moreover, the chromatin inside every constitutive heterochromatic region strongly and consistently interacts across the entire length of that silent region, with the strongest interactions occurring on the level of individual nucleosomes, suggesting that silent chromatin forms dense globule-like structures consisting of a stochastic nucleosome aggregation (Galazka et al. 2016; Rodriguez et al. 2022; Winter et al. 2018). At the highest resolutions, dense globules are visible at the boundaries between heterochromatic and euchromatic regions, implying the formation of 3D chromatin structures to prevent heterochromatin spread (Rodriguez et al. 2022). However, the loss of the known constitutive heterochromatin machinery has little impact on the folding of individual heterochromatic regions. In *Neurospora*, deletion of the gene encoding the $KMT1^{DIM-5}$ histone methyltransferase or its cognate binding partner HP1 reduces the dense internal compaction of heterochromatic regions and leads to reduced contacts between the euchromatin bordering these silent regions (Galazka et al. 2016). This suggests that folding of constitutive heterochromatic regions is dependent on proper deposition of different chromatin modifications and that the primary role of H3K9me3 and HP1 is to compact individual silent regions, thereby restricting contacts even between distant heterochromatic regions (Galazka et al. 2016; Zenk et al. 2021). Notably, AT-rich DNA forms few contacts with the surrounding euchromatin despite active and silent chromatin being in close proximity on the linear chromosome (Galazka et al. 2016; Rodriguez et al. 2022; Klocko et al. 2016), suggesting the heterochromatin-internal nucleosomes are isolated from active chromatin in fungal nuclei. Similarly, H3K9me3-enriched AT-rich sequences in *V. dahliae* and AT-rich isochores in *E. festucae* appear to be insulated from euchromatic contacts (Seidl et al. 2020; Winter et al. 2018).

Despite the overall segregation of heterochromatic and euchromatic DNA, recent work has shown that contacts that form between active and silent chromatin may regulate fungal gene expression (Rodriguez et al. 2022). Specifically, in *N. crassa*, small 'bands' of strong interactions between H3K9me3-marked constitutive heterochromatic regions and

select genes in more GC-rich genomic regions, which are possibly marked with a unique combination of histone post-translational modifications, are readily observed at the highest resolution Hi-C datasets (Fig. 1) (Rodriguez et al. 2022). Many of these genes display drastic changes in gene expression when constitutive heterochromatin is compromised (e.g., in a Δ*dim-5*/Δ*kmt1* mutant strain) (Rodriguez et al. 2022). The possibility of constitutive heterochromatin regulating gene expression has been observed previously (Yang et al. 2014) and may not be limited to *Neurospora* as several hundred genes that significantly change expression have been observed in *V. dahliae* upon loss of the lysine 9-specific histone methyltransferase *Dim-5* (Kramer et al. 2021).

To segregate heterochromatic genomic loci from those that are actively transcribed, and possibly to prevent aberrant RNA synthesis of silent chromatin, heterochromatin clusters at the nuclear membrane in virtually all eukaryotic nuclei (Falk et al. 2019; Solovei et al. 2016; Gonzalez-Sandoval and Gasser 2016). In mammals, most heterochromatin associates with lamin filaments and additional anchor proteins to form LADs (***L***amina ***A***ssociated ***D***omains) at the nuclear envelope (Guelen et al. 2008; Briand and Collas 2020). Mammalian proteins required for LAD formation include the lamin B receptor, Emerin, and ***L***amina-***A***ssociated-***P***olypeptide ***2***-b (LAP-2), all of which contain a LEM (***L***AP-2, ***E***merin, ***M***AN1) domain capable of binding chromatin at the nuclear envelope (Wagner and Krohne 2007; Buchwalter et al. 2019; Lin et al. 2000). However, fungi do not encode these proteins (Fig. 2), nor has the LEM domain been observed in any fungal proteins (Wagner and Krohne 2007; Koreny and Field 2016). However, fungi encode two general classes of integral nuclear membrane proteins that facilitate heterochromatin interactions with the nuclear membrane: MSC and SUN proteins (Koreny and Field 2016). Members of the MSC family of proteins include the integral membrane proteins Src1/Heh1, Heh2, Man1, and Lem2 (Grund et al. 2008; Mekhail and Moazed 2010; Taddei and Gasser 2012; Wagner and Krohne 2007; Brachner et al. 2005; King et al. 2006). All MSC proteins contain N-terminal LEM-like and MSC (***M***an1-***S***rc1p ***C***-terminal) domains in their primary structures. In *S. pombe*, Lem2 facilitates the anchoring of heterochromatic regions to the nuclear envelope, where the chromatin silencing machinery, including histone deacetylase complexes targeted to the telomeres, are recruited for chromatin repression (Hirano et al. 2020; Barrales et al. 2016; Sugiyama et al. 2007). Further, individual heterochromatic regions require specific proteins for nuclear envelope association: centromeres require Csi1 and telomeres use Dsh1 and Bqt3 (or Bqt4) (Harr et al. 2016; Ebrahimi et al. 2018; Barrales et al. 2016). The SUN (**S**ad-1, **Un**c-84) protein family similarly mediates heterochromatin interactions with the nuclear membrane (Tzur et al. 2006). Specifically, the *S. cerevisiae* SUN protein Mps3 associates with Sir silencing proteins, including the Sir4-Sir3 complex that binds deacetylated histone H4 in silent chromatin (Mekhail and Moazed 2010; Taddei and Gasser 2012; Harr et al. 2016; Bupp et al. 2007). These observations highlight a direct nuclear membrane-heterochromatin contact being important for genome organization. The MSC and SUN proteins are also

widely conserved across filamentous fungi (Fig. 2), implying similar mechanisms might also be employed in these species to tether heterochromatic regions to the nuclear membrane.

Chromatin compartmentalization in fungi

Euchromatin and heterochromatin in metazoans typically segregates into two distinct nuclear compartments: the euchromatic (active) 'A' and heterochromatic (silent) 'B' compartments. This compartmentalization of chromatin is readily observed as a 'checkerboard' pattern in Hi-C contact maps (Fig. 3D). The functional interpretation of this pattern is that genomic regions that have a similar transcriptional activity (e.g., heterochromatic regions that are silent) are spatially interacting within the nucleus (Lieberman-Aiden et al. 2009; Rao et al. 2014; Dixon et al. 2012; Dong et al. 2017; Rowley et al. 2017; Nichols and Corces 2021). Further studies have demonstrated that chromatin in each compartment physically associates: in the A-compartment euchromatin interacts in the central nucleus region, while in the B-compartment heterochromatin associates at the nuclear periphery (Beagan and Phillips-Cremins 2020; Rao et al. 2017; Buchwalter et al. 2019; Rao et al. 2014; Lieberman-Aiden et al. 2009). Mechanistically, this segregation may occur due to the aggregation of heterochromatic regions, possibly through Liquid-Liquid Phase Separation, at the nuclear membrane causing euchromatin to associate in the nucleus center (Falk et al. 2019; Larson et al. 2017), or due to the forces emerging from the activity of DNA-templated proteins in euchromatin forcing the segregation of silent chromatin to the nuclear periphery (Mahajan et al. 2022). In contrast to the extensive compartmentalization seen in metazoan Hi-C datasets, few fungi have evidence of prominent compartmentalization. To date, only the arbuscular mycorrhizal fungus *Rhizophagus irregularis*, a member of the Glomeromycetes clade, displays clear A/B compartments (Fig. 3D) (Xia et al. 2022; Yildirir et al. 2022). In contrast, other fungi display minimal chromatin compartmentalization, possibly reflecting the presence of smaller heterochromatic regions integrated among larger euchromatin domains (Xia et al. 2022). However, segregation of fungal chromatin in a manner analogous to A/B compartments, where heterochromatic regions aggregate yet are separated from euchromatic TAD-like structures, has been observed in the Hi-C datasets of multiple fungal species, including *S. cerevisiae* (Duan et al. 2010), *S. pombe* (Mizuguchi et al. 2014), *N. crassa* (Galazka et al. 2016; Rodriguez et al. 2022), *E. festucae* (Winter et al. 2018), *A. bisporus* (Hoencamp et al. 2021), and *V. dahliae* (Seidl et al. 2020). Presumably, interactions between heterochromatic regions, even when the small AT-rich isochores across fungal genomes are formed into heterochromatin, could be crucial for phase separation into the active 'A' and silent 'B' compartments (Falk et al. 2019). Chromatin compartmentalization is also supported by historical electron microscopy data, which shows clusters of densely-stained heterochromatin, which are often at the nuclear periphery but can be in the nucleus center, interspersed with lightly-stained euchromatin (Shatkin and Tatum 1959), thus arguing the chromatin composition organizes the nuclear genome in fungi.

Organization of fungal chromosomes into a Rabl conformation

The compartmentalization of the heterochromatic centromeres and telomeres of fungal chromosomes onto the nuclear membrane would facilitate the formation of Rabl chromosome conformations within the fungal nucleus (Box 1) (Hoencamp et al. 2021; Duan et al. 2010; Kim et al. 2017; Schalbetter et al. 2019). Rabl chromosome conformation is typically characterized by the clustering of centromeres on one side of the nuclear envelope and chromosomal arms extending outwards towards the opposing nuclear periphery on which the (sub)telomeres associate (Fig. 5A; e.g., (Jin et al. 2000)). Microscopic (e.g., (Guacci et al. 1994; Jin et al. 2000; Gasser 2002; Laroche et al. 1998; Schober et al. 2008; Goto et al. 2001)) as well as Hi-C experiments (Hoencamp et al. 2021; Duan et al. 2010; Kim et al. 2017; Schalbetter et al. 2019) also confirmed that *S. cerevisiae* organizes its 16 chromosomes in Rabl conformation during interphase, with a centromere cluster in proximity to the spindle pole body while the 32 (sub)telomeres associate non-randomly in four to eight foci at the nuclear membrane opposite the centromere bundle (Bystricky et al. 2004; Therizols et al. 2010; Duan et al. 2010; Kim et al. 2017; Schalbetter et al. 2019). Similarly, *S. pombe* organizes its three chromosomes into a Rabl structure (Mizuguchi et al. 2014). In filamentous fungi, the Rabl conformation was initially observed in *N. crassa* Hi-C experiments (Klocko et al. 2016; Galazka et al. 2016; Rodriguez et al. 2022). These data were corroborated by fluorescent microscopy of *N. crassa* nuclei in which a single centromere focus and three to four telomeric foci associate with the nuclear membrane (Fig. 5, Box 1). Additional Hi-C data from a plethora of filamentous fungi show that the centromeres contact independent of - and distinct from - telomere clustering and thus confirmed the existence of Rabl chromosomes, including in ascomycetes [*Penicillium oxalicum* (Li et al. 2022), *E. festucae* (Winter et al. 2018), *V. dahliae* (Torres et al. 2023; Seidl et al. 2020), *Candida albicans* (Burrack et al. 2016), *Fusarium verticillioides* (Yao et al. 2023) and [*Cladosporium fulvum* (Zaccaron et al. 2022)] and basidiomycetes [*Puccinia polysora* (Liang et al. 2022), *Puccinia graminis* (Sperschneider et al. 2021; Henningsen et al. 2022), *Autropuccinia psiddi* (Edwards et al. 2022), and *A. bisporus* (Hoencamp et al. 2021)] (Supplementary Table 1**)**. Interestingly, *R. irregularis* does not show clear centromere bundling for organizing its chromosomes, but it does seem to exhibit telomere bundling (Yildirir et al. 2022). Thus, these observations collectively suggest that the vast majority of fungi exhibit at least some of the features associated with Rabl-like chromosomal conformation.

One of the most prominent features observed in Hi-C contact maps of species employing a Rabl chromosome conformation is the strong inter-chromosomal interactions indicative of centromere bundling (Fig. 4). Centromere bundles manifest as dark 'spots' between chromosomes in whole genome contact maps, as observed in yeasts (*S. cerevisiae* and *S. pombe*), *N. crassa* (Fig. 4), *V. dahliae*, and *E. festucae*, among others (Rodriguez et al. 2022; Mizuguchi et al. 2014; Kim et al. 2017; Belton et al. 2015; Tanizawa et al. 2017; Seidl et al. 2020; Winter et al. 2018). The size of the inter-chromosomal centromere contacts is dependent on whether the centromeres in that species are point (small inter-centromere

interactions) or regional (large inter-centromere interactions) centromeres (Tanizawa et al. 2017; Yadav et al. 2018a; Galazka et al. 2016; Rodriguez et al. 2022; Belton et al. 2015; Seidl et al. 2020). Work in budding yeast has shown that centromeres bundle at the spindle pole body during interphase, an interaction which may require the centromeres to associate with microtubules as the application of nocodazole, a toxin that disrupts microtubule assembly, leads to reduced centromere clustering (Jin et al. 2000; Goto et al. 2001). The yeast CBF3 complex, which deposits the centromere specific histone variant Cse4 in centromeric DNA, also contains the critical kinetochore protein Ndc10p, highlighting the connection between centromere clustering and microtubule binding (Yan et al. 2018; Lechner and Carbon 1991). Surprisingly, few interactions between the heterochromatic centromeres and other interspersed heterochromatic regions are observed in *N. crassa*, even when heterochromatic features are in close proximity on the linear chromosome (Galazka et al. 2016; Rodriguez et al. 2022). This argues that the centromeric bundle is refractory to interacting with other heterochromatic regions, implying that the centromere bundle forms a dense, compact structure that isolates centromeric DNA (Fig. 4,5). Specificity for centromere bundling may be derived from deposition of CenH3 into centromeric nucleosomes, as CenH3 enrichment is only observed at the centromeres in *N. crassa* and other fungi (Smith et al. 2012; Galazka et al. 2016; Seidl et al. 2020). Presumably, kinetochore proteins that specifically associate with these centromeric histone variants might play a role in establishing and maintaining the centromere bundles at the nuclear membrane across filamentous fungi, in a manner similar to that of yeast (Westermann et al. 2007; Biggins 2013; Pidoux and Allshire 2004). Centromere clustering may also require some general property of heterochromatin independent of H3K9me3 deposition, as yeasts lacking H3K9me3 and SET-domain histone methyltransferases still have extensive inter-chromosomal centromere bundling, suggesting centromere clusters could rely on the silent chromatin in these regions forming LLPS condensates.

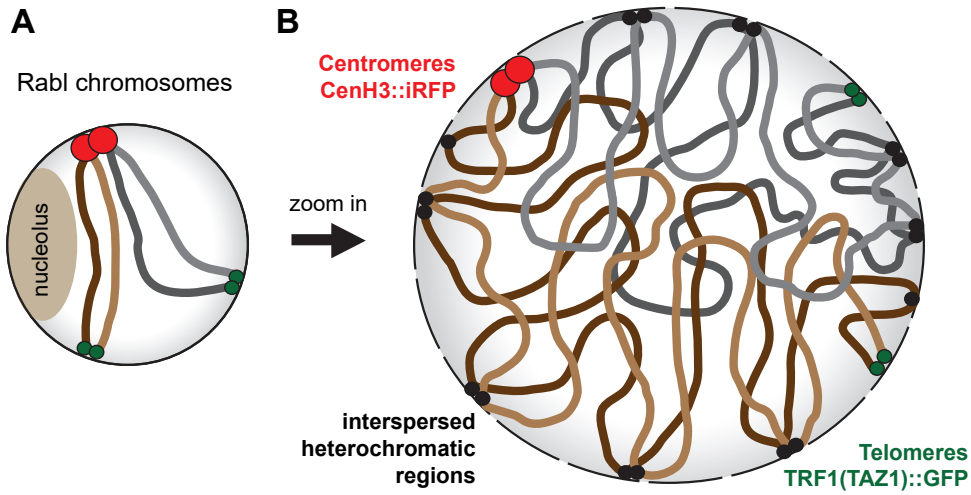


Figure 5. The formation of the Rabl chromosome conformation in fungal nuclei. (**A**) A schematic representation of two chromosomes (colored brown and gray for distinction) in the Rabl chromosome conformation (Mizuguchi, Barrowman and Grewal 2015), where centromeres (red circles) cluster on one side of the nucleus and the telomeres (green circles) cluster on the opposite side; the nucleolus is a distinct structure apart from these interphase chromosomes. (**B**) Detailed schematic of how interspersed heterochromatic regions (black circles) facilitate chromatin associating with the nuclear membrane and the compartmentalization of active and silent chromatin. Each chromosome forms a weak territory, with more intra-chromosomal contacts, but some inter-chromosomal interactions can occur.

Increased contacts between telomeres that indicate clustering of chromosome ends are another prevalent feature of most fungal Hi-C interaction maps (Fig. 4), and strong interactions originating at chromosome ends in Hi-C datasets can extend ~200 kb internal to the chromosome into the sub-telomeres (Fig. 4) (Galazka et al. 2016; Klocko et al. 2016; Mizuguchi et al. 2014; Rodriguez et al. 2022; Kim et al. 2017; Tanizawa et al. 2017). The relative position and distance between telomere bundles in the nucleus are governed by the chromosomal arm length, the position of the centromere, and the nuclear volume (Bystricky et al. 2004; Therizols et al. 2010). Individual telomeres may also have a unique landscape of chromatin modifications. For instance, in *Neurospora* these chromosome ends are the only loci in the genome enriched for both H3K9me3 and H3K27me2/3 (Klocko et al. 2016; Jamieson et al. 2018). However, the terminal H3K9me3-enriched telomere repeats minimally interact with nearby interspersed facultative or constitutive heterochromatic regions (Rodriguez et al. 2022). Importantly, no experimental data thus far has observed the co-localization of both centromeres and telomeres in the same nuclear region (Fig. 4), suggesting that mechanisms exist that independently organize these chromosomal structures. One possibility is that fungi may co-opt facultative heterochromatin to restrict (sub)telomeric contacts, thereby ensuring proper genome organization. In support, the Δ*set-7* (Δ*kmt6*) *N. crassa* strain devoid of H3K27me2/3 can have microscopic co-localization of centromeres and telomeres when facultative heterochromatin is compromised (Klocko et al. 2016). In Hi-C datasets, this loss of H3K27me2/3 reduces sub-telomeric interactions and

causes general genome organization disorder consistent with reduced interactions between the sub-telomeres and the nuclear membrane (Klocko et al. 2016). However, only sub-telomeric H3K27me2/3 is critical for nuclear membrane interactions, as a Δ*npf* (*Neurospora* p55 homolog) strain that loses sub-telomeric H3K27me2/3 exhibits genome disorder consistent with compromised sub-telomere interactions despite internal H3K27me2/3 enrichment being maintained (Klocko et al. 2016). Thus, current data seems to suggest that the unique chromatin landscape at the centromeres and telomeres ensure the isolation of these chromosomal features in a Rabl conformation despite both exhibiting properties of constitutive heterochromatin.

Loss of shelterin complex members eliminates telomere clustering and impacts the localization of telomeres to the nuclear periphery, as increased numbers of telomere foci are observed in loss of function shelterin mutants (Chikashige and Hiraoka 2001; Kanoh and Ishikawa 2001; Palladino et al. 1993). The heterodimeric Yku complex (Yku70/80) may be required for telomere clustering, as Yku70/80 anchors telomeres to the nuclear envelope (Hediger et al. 2002; Laroche et al. 1998), but also plays a role in shielding telomeric ends from shortening or fusion, and in telomere silencing (Ponnusamy et al. 2008; Polotnianka et al. 1998; Boulton and Jackson 1998). In filamentous fungi, recent work elucidated the role of an unexpected protein contributing to telomere tethering. In the Neurospora *dim-3* strain, which encodes a mutant allele of Importin α/Karyopherin α, some telomere foci no longer associate with the nuclear membrane, and a *dim-3* strain exhibits a drastically altered genome organization in a manner consistent with compromised telomere anchoring (Galazka et al. 2016). This phenotype may be indirectly caused by an increase in nuclear volume, which under normal situations would physically constrain telomere anchoring: *dim-3* nuclei consistently have a larger nuclear membrane diameter than the nuclei in wild-type strains (Galazka et al. 2016). Consistently, previous work in metazoans suggest that Importin α/Karyopherin α is necessary to constrict nuclear volumes (Levy and Heald 2010). Further experiments aimed at elucidating the proteins necessary for telomere-nuclear membrane anchoring, based on the groundwork established in yeast systems, should prove fruitful to uncover if filamentous fungi use similar mechanisms for telomere anchoring.

Quantitative computational models suggest that the specific tethering of heterochromatic chromosomal regions, including centromeres and telomeres to distinct nuclear landmarks, is sufficient to explain higher order organization of fungal genomes (Tjong et al. 2012; Wong et al. 2012). For example, in budding yeast, the complex Rabl nuclear organization emerges in computational models in which chromosomes are allowed to form random contacts, yet chromosomes are constrained by the tethering of chromosomal features to the nuclear envelope, by the distances between telomeres, and by the co-localization of functionally related loci (Tjong et al. 2012; Wong et al. 2012). Therefore, these geometrical constraints alone are sufficient to explain highly organized nuclear genome organization, including the Rabl-like chromosomal conformation, in *S. cerevisiae* (Tjong et al. 2012). Despite the genomes of filamentous fungi being larger

and often containing more chromosomes, the bundling of centromeres, telomeres, and interspersed heterochromatic regions at the nuclear periphery would be expected to drive the Rabl chromosome organization in a similar manner.

One corollary effect of a Rabl conformation in fungal nuclei is that chromosomes cannot form distinct territories, a property of higher metazoan genomes in which each chromosome occupies a defined space in the nucleus (Cremer and Cremer 2010; Tanabe et al. 2002; Parada et al. 2004; Manuelidis 1985; Cremer and Cremer 2001). In humans, chromosomal territories are evident by an enhanced intra-chromosome contact frequency, and minimal inter-chromosomal interactions (Lieberman-Aiden et al. 2009; Imakaev et al. 2012; Hoencamp et al. 2021; Falk et al. 2019). Eukaryotes encoding a complete condensin II complex form these territories, arguing that the presence of condensin II either strengthens chromosomal territories or suppresses Rabl conformation formation (Hoencamp et al. 2021). Specifically, deletion of the condensin II subunit CAPH2 in human cells promotes the formation of a Rabl-like chromosome conformation by increasing inter-chromosomal and trans-centromeric contacts while lowering lengthwise compaction of chromosomes (Hoencamp et al. 2021; Yoshida et al. 2022). However, nuclear architecture is more variable when longer evolutionary time scales are considered (Hoencamp et al. 2021). Specifically, fungal genomes do not encode condensin II accessory subunits (Fig. 2), as previously noted in *S. cerevisiae*, *S. pombe*, and *N. crassa* (Hudson et al. 2009; Hirano 2012; Hoencamp et al. 2021; Rodriguez et al. 2022), and consequently strong chromosome territories rarely form in fungal nuclei. Thus, the formation of a Rabl chromosome conformation (Fig. 5) in fungi is near ubiquitous. In this model, each chromosome exhibits extensive inter-chromosomal contacts (Fig. 4) that could be necessary for proper genome function, including the regulation of fungal gene expression. However, several intriguing observations directly contrast the Rabl chromosome model being applicable to all fungi. First, in *R. irregularis*, no clear Rabl chromosomes can be observed in Hi-C experiments (Fig. 3D) (Yildirir et al. 2022) yet condensin II orthologs are absent (Fig. 2), suggesting novel proteins facilitate chromatin compartmentalization into novel sub-nuclear structures. Further, orthologs of the condensin II accessory subunits CAPH2, CAPG2, and CAPD3 are present in species of the Monoblepharidomycetes clade (Fig. 2), suggesting that chromosome territories may exist in these taxa, but no Hi-C data currently exists to test this hypothesis.

Concluding remarks and future research

The spatial organization of the eukaryotic nuclear genome is closely linked to its biological functions (Lieberman-Aiden et al. 2009; Hoencamp et al. 2021; Bonev and Cavalli 2016; Sexton and Cavalli 2015). Here, we addressed the occurrence, formation, and functional implications on the spatial organization of the nuclear genome in fungi. Yeasts have been important model systems to study nuclear genome organization in the last decades, but

only recently data on the composition and organization of chromatin on more diverse fungi became available. Based on these, we sought to summarize and discuss structures homologous to those found in model eukaryotes and examined the protein complexes that are implicit in establishing these structures. We argued that the folding of chromatin fibers in fungi is similarly hierarchical as in other eukaryotes, ranging from small-scale chromatin loops of a few kilobases to large-scale subdomains comprising hundreds of kilobases that segregate chromatin into A or B compartments. The self-interacting domains similar to metazoan TADs have been observed in yeasts (Mizuguchi et al. 2014; Duan et al. 2010; Tsochatzidou et al. 2017; Schalbetter et al. 2019), and TAD-like structures that regionally organize the genome are prevalent in most studied filamentous fungi. However, the precise nature of the boundary or insulator regions, that allow the loading of cohesin or restriction of the extrusion remains to be examined in detail. Furthermore, based on experimental data from several diverse fungi, we argued that Rabl chromosome conformation is the hallmark of fungal genome organization, and the independent bundling of centromeres and telomeres drives the overall nuclear organization of the fungal genome. Implicit in this hierarchical organization is that different 'levels' are interconnected and that changes in local chromatin organization have significant impact on the global nuclear organization, e.g., in formation of heterochromatic structure at centromeres and telomeres and *vice versa*. In addition to the critical link between gene expression and chromatin folding, several conserved DNA-templated processes may be directly tied to genome organization, including DNA replication and repair that can alter genome and nuclear organization. DNA repair events can contribute to fungal evolution, as the occurrence of a double stranded DNA break and its concomitant repair onto DNA strands spatially close in 3D may be the first step in structurally varying a species' genome (Huang and Cook 2022; Hanson et al. 2021; Zhang et al. 2012). Thus, DNA-templated processes in the nucleus are often influenced by the organization of the genome, yet improper functioning of these genomic functions can feedback and alter genome organization, highlighting the intricate interconnection between chromosome conformations and genome function. While we here strived to paint a complete picture on the genome organization in fungi, it is important to note that detailed, high-resolution datasets on chromosomal conformation as well as chromatin composition and organization are only available for very few fungal model species. Even though the hierarchical organization of nuclear genome organization is largely conserved between the fungal species examined to date, we also highlighted intriguing differences between species, including the species-specific characteristics defining TAD-like structures. Recent work comparing the position of TAD-like structures and conservation of TAD boundaries between fungal strains and species of the same genus suggests that there is limited TAD variation (Torres et al. 2023), which is reminiscent of observations in other eukaryotes (Rao et al. 2014; Rowley et al. 2017; Harmston et al. 2017; Krefting et al. 2018; Fudenberg and Pollard 2019; McArthur and Capra 2021; Liao et al. 2021b). These data favor a hypothesis where TAD-like structures are generally conserved in related fungi but may

be less conserved when greater fungal diversity is considered. Nevertheless, in *Drosophila* species, rearrangements occur predominantly at TAD boundaries and not in TAD bodies (Wright and Schaeffer 2022; Liao et al. 2021b), suggesting that TAD-like structures play important roles for the evolution of genome organization. Therefore, comparative studies that systematically probe genome organization throughout the fungal lineage will help deepen understanding the establishment, conservations, and the functional implications of genome organization. We therefore advocate that exploring genome organization in the context of the extensive fungal biodiversity will be essential to uncover in the future how nuclear organization impacts fungal genome function and evolution.

Acknowledgments

The authors wish to thank Dr. Jolien J. E. van Hooff for providing access to profile hidden Markov models as well as the members of the Seidl and Klocko labs for discussions and support during the crafting of this review.

Funding

This work was supported by a PhD grant from the Consejo Nacional de Ciencia y Tecnología, México to D.E.T. Work in the laboratory of M.F.S. is supported by the Research Council Earth and Life Sciences (ALW) of the Netherlands Organization of Scientific Research (NWO). Work in the laboratory of A.D.K. is supported by an Academic Research Enhancement Award (AREA) grant from the National Institutes of Health (NIH) (1R15GM140396-01), start-up funds from the University of Colorado Colorado Springs (UCCS) College of Letters, Arts, and Sciences, and a UCCS Committee on Research and Creative Works (CRCW) internal seed grant. The authors declare no conflict of interest.

Supplementary data

Methods

Imaging analysis

Imaging of the *Neurospora crassa* strain N6224 (*ish-1::foTagBFP2::hph::LoxP; hpo-TagRFP-T::hph::loxP; CenH3::iRFP670::hph::LoxP; Trf-1::GFP::hph::LoxP;* Δ*sad-2::nat1*) was performed under standard imaging conditions (conidia from N6224 were resuspended in a 50% glycerol solution, and 3 µL was spotted on glass microscope slides and covered with a glass coverslip) (Klocko et al. 2016) on a Leica Thunder Imaging System; composite and deconvoluted images of GFP, BFP, and iRFP were cropped in Adobe Photoshop.

Hi-C data analysis

Paired-end sequence fastq files of Hi-C data were downloaded from the NCBI GEO (*N.crassa*) or the Sequence Read Archive (SRA; *R. irregularis*) using the program fastq-dump, mapped with bowtie2 (Langmead and Salzberg 2012) or BWA mem (*V. dahliae,* (Li and Durbin 2009)), and processed with hicExplorer (Ramírez et al. 2018), using the subprograms hicBuildMatrix, hicCorrect using Knight Ruiz correction for *N. crassa* and *R. irregulatis* (Knight and Ruiz 2013) or Iterative Correction (ICE) correction for *V. dahliae* (Imakaev et al. 2012), hicPlotMatrix, and hicPlotTADs. Aggregate read count matrices for *N. crassa* Hi-C data were built with Juicer (Hoencamp et al. 2021; Rao et al. 2014) using fastq files of DpnII (euchromatin-specific) or MseI (heterochromatin-specific) Hi-C datasets (Rodriguez et al. 2022), and the resultant .mcool matrix files were displayed with hicExplorer. The contact matrix for *V. dahliae* Hi-C data was built with hicBuildMatrix (Ramírez et al. 2018) from DpnII (euchromatin-specific) proximity ligation Hi-C data and displayed using hicPlotMatrix.

Chromatin Immunoprecipitation data analysis

ChIP-seq data was mapped to version 14 of the *Neurospora crassa* genome using bowtie2 (Langmead and Salzberg 2012), converted to a bam file and sorted with samtools (Li et al. 2009), and the sequence enrichment was calculated by the bamCoverage program in DeepTools (Ramírez et al. 2016) and displayed as a bigwig file on Integrative Genomics Viewer (IGV) (Robinson et al. 2011). Similar methods were used to display H3K9me3 ChIP-seq data for *V. dahliae* using the reference strain JR2 (Faino et al. 2015). Percent guanine-cytosine bases (%GC) was calculated for the nc12 fixed genome and displayed as a .wig file (Klocko et al. 2019); given the differences between versions 12 and 14 of the N. crassa genome only affect chromosomes (Linkage Groups) I and V (Rodriguez et al. 2022), the same wig file can be used for Linkage Group III images. Transposon location bed files were displayed with IGV, with the locations modified from (Nguyen et al. 2022).

Phylogenetic profile of chromatin topology associated proteins in fungi

Database

To identify proteins involved in chromatin topology in fungi, we used 89 eukaryotic (predicted) proteomes (Supplementary Table 2). Proteomes of non-fungal taxa (15) were obtained from an in-house dataset described previously (Deutekom et al. 2019). The vast majority of fungal proteomes were obtained from the Joint Genome Institute (JGI) Mycocosm fungal reference database (https://mycocosm.jgi.doe.gov/mycocosm/home). Additionally, the predicted proteomes of *Austropuccinia psidii* (Edwards et al. 2022) and *Hemileia vastratix* (Tobias et al. 2022) were manually added after prediction of protein-coding genes by Augustus with default settings (Stanke et al. 2006).

HMM model assignment

Hmmersearch (HMMER v.3.1 (Eddy 2011)) was used with in-house profile hidden Markov models (HMMs), Pfam 31.0 (Finn et al. 2016), and histone profiles (https://www.ncbi.nlm.nih.gov/research/HistoneDB2.0/) to search the in-house database for homologous sequences. Briefly, in house profile HMMs were generated by obtaining diverse eukaryotic sequences for lamins (Kollmar 2015), shelterin subunits (Myler et al. 2021; Finn et al. 2016), SUN and SMC (Koreny and Field 2016). The condensin (I, II), cohesin, and CTCF profile HMMs were kindly provided by Dr. Jolien J. E. van Hooff, Université Paris-Saclay, CNRS, AgroParisTech, Ecologie Systématique Evolution, Orsay, France. Homologous sequences in our fungal database for each protein family were aligned using MAFFT v7.453 --localpair --maxiterate 1000 (Katoh and Standley 2013), trimmed to remove non-informative sites using ClipKit v.1.3.0 –smartgap (Steenwyk et al. 2020) The final profile HMMs was generated using hmmbuild with default settings (Eddy 2011). To identify the final set of homologs, we used hmmersearch with -E 1e-10, --domE 1e-10, --incE 1-e-10, --incdomE 1-e-10 as settings. Matching proteins for each profile HMM in our database were extracted, aligned using MAFFT v7.453 --auto --maxiterate 1000 (Katoh and Standley 2013) and multiple sequence alignments were trimmed using trimal v.1.4.1 --gt 0.1 (Capella-Gutiérrez et al.). We built a phylogenetic tree for each alignment using IQ-TREE v.2.2.0.3 (Nguyen et al. 2015) with LG model (--mset LG (Le et al. 2012)) and 1,000 ultrafast bootstraps (--B 1000, (Minh et al. 2020)). The presence of similar Pfam domains for each protein were verified in a multiple sequence alignment using hmmerscan and the Pfam database with default settings (Finn et al. 2016; Eddy 2011). Finally, we only considered sequences as an ortholog if after the profile HMMs search the tree topology and the Pfam domain supports a monophyletic relationship with well-characterized eukaryotic proteins. Protein sequences for the orthologous families are available at Zenodo doi:10.5281/zenodo.7635847.

Data Availability

Hi-C data is publicly available at either the Gene Expression Omnibus (GEO; for *N. crassa* [accession number GSE173593]) (Rodriguez et al. 2022) or the Sequence Read Archive (SRA; for *R. irregularis* [accession number PRJNA748024] and *V. dahliae* [accession PRJNA641329]) (Yildirir et al. 2022; Seidl et al. 2020) at the National Center of Biotechnology Information (NCBI) at the National Institutes of Health, USA. Publicly available Chromatin Immunoprecipitation-sequencing (ChIP-seq) from *N. crassa* wild type (WT) datasets of H3K9me3 (merged from GSE68897 and GSE98911), H3K27me2/3 (merged from GSE82222 and GSE100770), H3K27ac (GSE118495), and H3K4me3 (GSE121356) are available at the GEO from published manuscripts (Galazka et al. 2016; Klocko et al. 2016, 2019, 2020; Jamieson et al. 2016; Bicocca et al. 2018; Zhu et al. 2019). Publicly available ChIP-seq from *V. dahliae* wild type (WT) datasets of H3K9me3 (Cook et al. 2020) is accessible from the NCBI/SRA under BioProject PRJNA592220.

Supplementary Table 1. Published Hi-C datasets of Fungi.

Fungal Species	Genotype	Public Repository	Reference
Neurospora crassa	WT, Δdim-5, Δhpo, dim-3	NCBI GEO; GSE71024	(Galazka et al. 2016)
Neurospora crassa	Δset-7, Δnpf, Δset-7;Δdim-5	NCBI GEO; GSE82222	(Klocko et al. 2016)
Neurospora crassa	WT	NCBI GEO; GSE173593	(Rodriguez et al. 2022)
Verticillium dahliae	WT	NCBI SRA; PRJNA641329	(Seidl et al. 2020)
Verticillium longisporum	WT	NCBI SRA; PRJNA473305	(Depotter et al. 2021)
Verticillium spp.	WT	NCBI SRA; PRJNA641329	(Seidl et al. 2020)
Epichlöe festucae	WT	NCBI SRA; PRJNA431450	(Winter et al. 2018)
Penicillium oxalicum	WT	NCBI SRA; PRJNA772803	(Li et al. 2022)
Fusarium oxysporum	WT	NCBI SRA; SRR11742970	(Fokkens et al. 2020)
Puccinia triticina	WT	NCBI SRA; PRJNA725323	(Duan et al. 2022)
Puccinia graminis f.sp tritici	WT	NCBI SRA; PRJNA516922	(Sperschneider et al. 2021)
Lentinula edodes	WT	NCBI SRA; SRP256131	(Yu et al. 2022)
Rhizophagus irregularis	WT	NCBI SRA; PRJNA748024	(Yildirir et al. 2022)
Sparasis latifolia	WT	NCBI SRA; PRJNA686158	(Yang et al. 2021)
Agaricus bisporus	WT	NCBI SRA; PRJNA512907	(Hoencamp et al. 2021)
Austropuccinia psidii	WT	NCBI SRA; PRJNA810573/ PRJNA81057	(Edwards et al. 2022)
Hemileia vastatrix	WT	NCBI SRA; PRJNA837996	(Tobias et al. 2022)
Passalora fulva	WT	NCBI SRA; SRR16292147	(Zaccaron et al. 2022)
Puccinia polysora	WT	NCBI SRA; PRJNA881038	(Liang et al. 2022)
Schizosaccharomyces pombe	WT, Δcut14SO	NCBI GEO;GSE94478	(Kakui et al. 2017)
Schizosaccharomyces pombe	WT, Δcut14+, Δcut14SO	NCBI GEO;GSE143338	(Kakui et al. 2020)
Schizosaccharomyces pombe	WT, Δrad21, Δclr4	NCBI GEO;GSE56849	(Mizuguchi et al. 2014)
Schizosaccharomyces pombe	WT, Δsap1	NCBI GEO;GSE96883	(Mizuguchi et al. 2017)
Schizosaccharomyces pombe	WT at 20, 30, 40, 50, 60, 70, 80, 120min and Δcut3, Δcut14, Δrad21, Δcut14 & Δcut3	NCBI GEO;GSE93198	(Tanizawa et al. 2017)

Saccharomyces cerevisiae	WT, SCRaMbLE	NCBI SRA; SRP070421	(Mercy et al. 2017)
Saccharomyces cerevisiae	WT, ΔFkh1,ΔFkh2	NCBI SRA; SRP101770	(Eser et al. 2017)
Saccharomyces cerevisiae	WT and several stages arrested of cell cycle	NCBI GEO; GSE90902	(Lazar-Stefanita et al. 2017)
Saccharomyces cerevisiae	WT (0h, 2h, 3h, 4h, 5h, 6h, 8h)	NCBI GEO; GSE127940	(Schalbetter et al. 2019)
Saccharomyces cerevisiae	WT, metaphase and G1 arrested, Δcdc45-degron	NCBI GEO; GSE87311	(Schalbetter et al. 2017)
Saccharomyces cerevisiae	WT, SMC3-AID,smc3-RR, ECO1-AID	NCBI GEO; GSE186987	(Bastié et al. 2022)
Saccharomyces cerevisiae	WT and several stages arrested of cell cycle	NCBI SRA; PRJNA427106	(Ohno et al. 2019)
Saccharomyces cerevisiae	WT and several stages arrested of cell cycle	NCBI SRA; PRJNA486513	(Garcia-Luis et al. 2019)
Saccharomyces cerevisiae	WT and several stages arrested of cell cycle	NCBI SRA; PRJNA528616	(Dauban et al. 2020)
Saccharomyces cerevisiae	WT, Δssu72-2	NCBI GEO; GSE68016	(Hsieh et al. 2015)

Supplementary Table 2. Phylogenetic profile of chromatin topology-associated proteins in fungi. Matrix displays the presence (1) and absences (0) for each analyzed species/taxa.

Species	Division	Class	H3	H4	H2A	H2B	CTCF	SMC2	SMC4	CAPH	CAPG	CAPD2	CAPH2	CAPG2	CAPD3	SMC1	SMC3	SCC1	REC8	SCC3	PDS5	NIPBL	MAU2	WAPL	SORORIN	ECO1	LAMIN	LAMIN_A	LAMIN_B	LAMIN_C	LAMIN_D	LBR	EMERIN	LAP2B	SUN	MSC	TAZ1	RIF1	RIF2	TPZ1	CCQ1	RAP1	POT1	TBF1	TRF1	TRF2	TIN2	TPP1
Aspergillus nidulans	Ascomycota	Eurotiomycetes	1	1	1	1	0	1	1	1	1	1	0	0	0	1	1	1	1	1	1	1	1	1	0	1	0	0	0	0	0	0	0	0	1	1	1	1	1	1	1	1	1	0	0	0	0
Acarospora strigata	Ascomycota	Lecanoromycetes	1	1	1	1	0	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	0	1	0	0	0	0	0	0	0	0	1	1	0	0	1	0	0	0	1	0	0	0	0	0
Botrytis cinerea	Ascomycota	Leotiomycetes	1	1	1	1	0	0	0	0	0	0	0	0	0	0	0	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Blumeria graminis f. sp. hordei	Ascomycota	Leotiomycetes	1	1	1	1	0	0	1	1	1	1	0	0	0	1	1	1	1	1	1	1	1	1	0	1	0	0	0	0	0	0	0	0	1	1	0	1	1	0	0	1	1	1	0	0	0	0
Candida albicans	Ascomycota	Saccharomycetes	1	1	1	1	0	1	1	1	1	1	0	0	0	1	1	1	1	1	1	1	1	1	0	1	0	0	0	0	0	0	0	0	1	1	1	1	1	0	1	1	1	1	0	0	0	0
Cladonia grayi	Ascomycota	Lecanoromycetes	1	1	1	1	0	1	1	1	1	1	0	0	0	1	1	1	1	1	1	1	1	1	0	1	0	0	0	0	0	0	0	0	1	1	1	1	1	0	1	1	1	1	0	0	0	0
Colletotrichum higginsianum	Ascomycota	Sordariomycetes	1	1	1	1	0	1	1	1	1	1	0	0	0	1	1	1	1	1	1	1	1	1	0	1	0	0	0	0	0	0	0	0	1	1	0	1	0	0	0	0	1	1	0	0	0	0
Drechslerella stenobrocha	Ascomycota	Orbillomycetes	1	1	1	1	0	1	0	1	1	1	0	0	0	1	1	1	1	1	1	1	1	1	0	1	0	0	0	0	0	0	0	0	1	1	0	1	1	1	0	1	1	1	0	0	0	0
Epichloe festucae	Ascomycota	Sordariomycetes	1	1	1	1	0	1	1	1	1	1	0	0	0	1	1	1	1	1	1	1	1	1	0	1	0	0	0	0	0	0	0	0	1	1	0	0	0	0	0	0	1	1	0	0	0	0
Fusarium graminearum	Ascomycota	Sordariomycetes	1	1	1	1	0	1	1	1	1	1	0	0	0	1	1	1	1	1	1	1	0	0	0	1	0	0	0	0	0	0	0	0	0	1	0	1	1	0	0	1	0	1	0	0	0	0
Fusarium oxysporum f. sp. lycopersici	Ascomycota	Sordariomycetes	1	1	1	1	0	1	1	1	1	1	0	0	0	1	1	1	1	1	1	1	1	1	0	1	0	0	0	0	0	0	0	0	1	1	0	0	0	0	0	1	1	0	0	0	0	0
Kluyveromyces lactis	Ascomycota	Saccharomycetes	1	1	1	1	0	1	1	1	1	1	0	0	0	1	1	0	1	1	1	1	0	0	0	1	0	0	0	0	0	0	0	0	0	1	0	0	1	0	0	0	1	0	0	0	0	0
Leptosphaeria maculans	Ascomycota	Dothideomycetes	1	1	1	1	0	1	1	0	1	1	1	1	1	1	1	1	1	1	1	1	1	1	0	1	1	1	1	1	1	0	1	0	1	1	0	1	0	0	0	0	1	0	0	0	0	0
Mycosphaerella graminicola	Ascomycota	Dothideomycetes	1	1	1	1	0	1	1	1	1	1	0	0	0	1	1	1	1	1	1	1	1	1	0	1	0	0	0	0	0	0	0	0	0	1	1	1	1	1	1	1	1	1	0	0	0	0
Monacrosporium haptotylum	Ascomycota	Orbillomycetes	1	1	1	1	0	1	1	1	1	1	0	0	0	1	1	1	1	1	1	1	1	1	0	0	0	0	0	0	0	0	0	0	0	1	1	1	1	1	0	1	1	1	1	0	0	0	0
Morchella importuna	Ascomycota	Pezizomycetes	1	1	1	1	0	1	1	1	1	1	0	0	0	0	1	1	1	0	1	1	0	1	0	1	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	1	0	1	1	0	0	0	0
Neurospora crassa	Ascomycota	Sordariomycetes	1	1	1	1	0	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	0	0	1	0	1	0	0	0	0	0	0	0	0	1	0	0	1	0	0	0	0	0	0	0	0	0
Penicillium chrysogenum	Ascomycota	Eurotiomycetes	1	1	1	1	0	1	1	1	1	1	0	0	0	1	1	1	1	1	1	1	1	1	0	1	0	0	0	0	0	0	0	0	1	1	0	1	1	0	0	0	1	0	0	0	0	0
Penicillium oxalicum	Ascomycota	Eurotiomycetes	1	1	1	1	0	1	1	1	0	1	1	1	0	1	1	1	1	1	1	1	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Passalora fulva	Ascomycota	Dothideomycetes	1	1	1	1	0	1	1	1	1	1	0	0	0	1	1	1	1	0	1	1	1	1	0	1	0	0	0	0	0	0	0	0	1	1	0	0	1	1	0	0	1	1	0	0	0	0
Pyricularia oryzae	Ascomycota	Sordariomycetes	1	1	1	1	0	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	0	1	0	0	0	0	0	0	0	0	1	1	0	1	1	0	0	1	0	0	0	0	0	0

Pichia pastoris	Ascomycota	Saccharomycetes	1 1 1 1 1 1 1 1 1 1 1 1 0 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 0 0 0 1 1 0 1 1 0 0 0 0 0 0 0 0 0
Saccharomyces cerevisiae	Ascomycota	Saccharomycetes	1 1 1 1 0 1 1 1 1 1 0 0 0 1 1 1 1 1 1 1 1 1 0 1 0 0 0 0 0 0 0 0 0 1 1 0 0 1 0 0 1 1 0 0 0 0 0
Saitoella complicata	Ascomycota	Taphrinomycetes	1 1 1 1 0 1 1 0 1 0 0 0 0 0 1 0 0 0 0 1 1 0 0 0 0 0 0 0 0 0 0 0 0 1 0 0 0 0 0 1 1 0 1 0 0 0 0
Schizosaccharomyces pombe	Ascomycota	Taphrinomycetes	1 1 0 0 0 1 1 1 1 1 0 0 0 1 1 1 1 1 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 1 1 0 0 0 1 0 0 1 0 0 0 0 0
Tuber melanosporum	Ascomycota	Pezizomycetes	1 1 0 0 0 1 1 1 1 1 0 0 0 1 1 1 1 1 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 1 1 0 0 0 1 0 0 1 0 0 0 0 0
Verticillium dahliae	Ascomycota	Sordariomycetes	1 1 1 0 0 1 0 1 1 1 0 0 0 1 1 1 1 0 1 1 1 0 0 0 0 0 0 0 0 0 0 0 0 1 1 1 1 1 1 0 1 1 0 1 0 0 0 0
Xylona heveae	Ascomycota	Xylonomycetes	1 1 1 1 0 1 1 1 1 1 0 0 0 1 1 1 1 1 1 1 1 1 0 0 0 0 0 0 0 0 0 0 0 1 1 0 1 1 0 1 0 1 1 0 0 0 0
Agaricus bisporus	Basidiomycota	Agaricomycetes	1 1 1 1 0 1 1 1 1 1 0 0 0 1 1 1 1 1 1 1 1 1 0 0 0 0 0 0 0 0 0 0 0 1 1 1 1 1 1 0 1 1 1 1 0 0 0 0
Austropuccinia psidii	Basidiomycota	Pucciniomycetes	1 1 1 1 0 1 0 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 0 1 0 0 0 0 0 0 0 0 0 1 0 1 1 0 0 1 1 1 0 0 0 0
Calocera cornea	Basidiomycota	Dacrymycetes	1 1 1 1 0 1 1 1 1 1 0 0 0 1 1 1 1 1 1 1 1 1 0 1 0 0 0 0 0 0 0 0 0 1 1 0 0 1 1 0 1 1 1 0 0 0 0
Coprinopsis cinerea	Basidiomycota	Agaricomycetes	1 1 1 1 0 1 1 1 1 1 0 0 0 1 1 0 0 1 1 1 0 0 0 1 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
Cryptococcus neoformans	Basidiomycota	Tremellomycetes	1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 0 1 1 0 0 1 1 0 1 1 1 1
Dacryopinax primogenitus	Basidiomycota	Dacrymycetes	1 1 1 1 0 1 1 1 1 1 0 0 0 1 1 1 1 1 1 1 1 1 0 1 0 0 0 0 0 0 0 0 0 1 1 0 1 1 0 0 1 1 0 0 0 0 0
Hemileia vastatrix	Basidiomycota	Pucciniomycetes	1 1 1 1 0 1 0 1 1 1 0 0 0 1 1 1 1 0 1 1 1 1 0 1 0 0 0 0 0 0 0 0 0 1 0 0 0 1 1 0 0 1 1 0 0 0 0
Laccaria bicolor	Basidiomycota	Agaricomycetes	1 1 1 1 0 1 1 1 1 1 0 0 0 1 1 1 1 1 1 1 0 0 0 1 0 0 0 0 0 0 0 0 0 1 1 0 1 1 0 0 1 0 1 0 0 0 0
Melampsora larici-populina	Basidiomycota	Pucciniomycetes	1 1 1 1 0 1 1 1 1 1 0 0 0 1 1 1 1 1 1 1 1 1 0 1 0 0 0 0 0 0 0 0 0 1 1 0 1 1 0 0 1 1 0 0 0 0 0
Naganishia vishniacii	Basidiomycota	Tremellomycetes	1 1 1 1 0 1 1 1 1 1 0 0 0 1 1 1 1 1 1 1 1 1 0 1 0 0 0 0 0 0 0 0 0 1 1 0 1 1 0 1 1 1 0 0 0 0 0
Puccinia graminis f. sp. tritici	Basidiomycota	Pucciniomycetes	1 1 1 1 0 1 1 1 1 1 0 0 0 1 1 1 1 1 1 1 0 0 0 0 0 0 0 0 0 0 0 0 0 1 1 0 0 1 0 0 0 1 0 0 0 0 0
Puccinia polysora	Basidiomycota	Pucciniomycetes	1 1 1 1 0 1 1 1 1 1 0 0 0 1 1 1 1 1 1 1 1 1 0 1 0 0 0 0 0 0 0 0 0 1 1 0 1 1 0 0 1 1 0 0 0 0 0
Rhizoctonia solani	Basidiomycota	Agaricomycetes	1 1 1 1 0 1 1 1 1 1 0 0 0 1 1 1 1 1 1 1 1 1 0 1 0 0 0 0 0 0 0 0 0 1 1 1 1 1 1 1 1 1 0 0 0 0 0
Rhodotorula sp.	Basidiomycota	Microbotryomycetes	1 1 1 1 0 1 1 1 1 1 0 0 0 1 1 1 1 1 1 1 1 1 0 1 0 0 0 0 0 0 0 0 0 1 1 1 1 1 1 0 1 1 1 1 0 0 0 0
Schizophyllum commune	Basidiomycota	Agaricomycetes	1 1 1 1 0 1 1 1 1 1 0 0 0 1 1 1 1 1 1 1 1 1 0 1 0 0 0 0 0 0 0 0 0 1 1 1 1 1 1 0 1 1 1 1 0 0 0 0
Ustilago maydis	Basidiomycota	Ustilaginomycetes	1 1 1 1 0 1 1 1 1 1 0 0 0 1 1 1 1 1 1 1 1 1 0 1 0 0 0 0 0 0 0 0 0 1 1 0 0 1 0 0 0 0 0 0 0 0 0
Violaceomyces palustris	Basidiomycota	Ustilaginomycetes	1 1 1 1 0 1 1 1 1 1 0 0 0 1 1 1 1 1 1 1 1 1 0 0 0 0 0 0 0 0 0 0 0 1 1 1 1 1 1 0 1 1 1 1 0 0 0 0

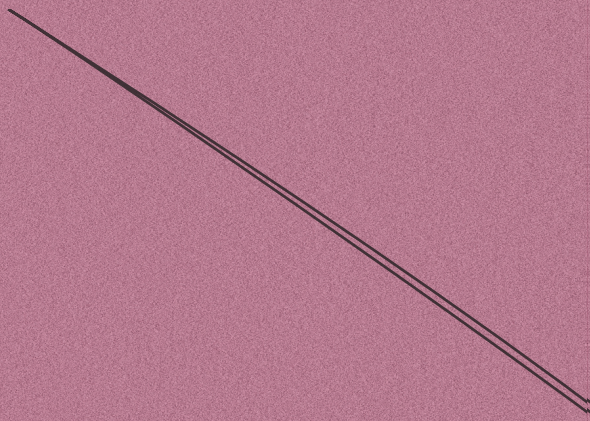
Supplementary Table 2. Continued

Species	Division	Class	H3	H4	H2A	H2B	CTCF	SMC2	SMC4	CAPH	CAPG	CAPD2	CAPH2	CAPG2	CAPD3	SMC1	SMC3	SCC1	REC8	SCC3	PDS5	NIPBL	MAU2	WAPL	SORORIN	ECO1	LAMIN	LAMIN_A	LAMIN_B	LAMIN_C	LAMIN_D	LBR	EMERIN	LAP2B	SUN	MSC	TAZ1	RIF1	RIF2	TPZ1	CCQ1	RAP1	POT1	TBF1	TRF1	TRF2	TIN2	TPP1
Wallemia mellicola	Basidiomycota	Wallemiomycetes	1	1	1	1	0	1	0	1	1	1	0	0	0	1	1	1	1	1	1	1	1	1	0	1	0	0	0	0	0	0	0	0	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0
Allomyces macrogynus	Blastocladiomycota	Blastocladiomycetes	1	1	1	1	0	1	1	1	1	1	0	0	0	1	1	1	1	1	1	1	1	1	0	1	0	0	0	0	0	0	0	0	1	1	0	0	1	1	0	1	1	0	0	0	0	0	0
Blastocladiella britannica	Blastocladiomycota	Blastocladiomycetes	1	1	1	1	0	1	1	1	1	1	0	0	0	1	1	1	1	1	1	1	1	1	0	1	0	0	0	0	0	0	0	0	1	1	0	1	1	0	0	0	1	0	0	0	0	0	0
Batrachochytrium dendrobatidis	Chytridiomycota	Chytridiomycetes	1	1	1	1	0	1	0	1	1	1	0	0	0	0	0	1	1	1	1	1	1	1	0	1	0	0	0	0	0	0	0	0	1	1	1	1	1	1	0	1	1	1	1	0	0	0	0
Chytridium lagenaria	Chytridiomycota	Chytridiomycetes	1	1	1	1	0	1	1	1	1	1	0	0	0	1	1	1	1	1	1	1	1	1	0	1	0	0	0	0	0	0	0	0	1	1	1	1	1	1	0	1	1	1	1	0	0	0	0
Spizellomyces punctatus	Chytridiomycota	Chytridiomycetes	1	1	1	1	0	1	1	1	1	1	0	0	0	1	1	1	1	1	1	1	1	1	0	1	0	0	0	0	0	0	0	0	1	1	1	1	1	1	0	1	1	1	1	0	0	0	0
Rozella allomycis	Cryptomycota	Cryptomycetes	1	1	1	1	0	1	1	1	1	1	0	0	0	1	1	1	1	1	1	1	1	1	0	1	0	0	0	0	0	0	0	0	1	1	0	0	1	0	0	0	0	0	0	0	0	0	0
Basidiobolus meristosporus	Entomopthoromycota	Entomophthoromycetes	1	1	1	1	0	1	1	1	1	1	1	1	1	1	1	1	1	1	1	0	1	0	1	0	1	0	0	0	0	0	0	0	1	1	0	0	0	0	0	0	1	0	0	0	0	0	0
Conidiobolus coronatus	Entomopthoromycota	Entomophthoromycetes	1	1	1	1	0	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	0	1	0	1	1	1	1	1	1	0	0	0	1	1	0	1	0	0	0	1	1	0	0	0	0	0	0
Gigaspora rosea	Glomeromycota	Glomeromycetes	1	1	1	1	0	1	1	1	1	1	0	0	0	1	1	1	1	1	1	1	1	1	0	0	0	0	0	0	0	0	0	0	0	1	1	1	1	0	1	1	1	1	0	0	0	0	0
Rhizopaghus irregularis A1	Glomeromycota	Glomeromycetes	1	1	1	1	0	1	1	1	1	1	0	0	0	1	1	1	1	1	1	1	0	0	0	1	0	0	0	0	0	0	0	0	0	1	0	1	1	0	0	1	0	1	0	0	0	0	0
Rhizopaghus irregularis C2	Glomeromycota	Glomeromycetes	1	1	1	1	0	1	1	1	1	1	0	0	0	1	1	1	1	1	1	1	1	1	0	1	0	0	0	0	0	0	0	0	1	1	0	0	1	1	0	1	1	0	0	0	0	0	0
Rhizopaghus irregularis	Glomeromycota	Glomeromycetes	1	1	1	1	0	1	1	1	1	1	0	0	0	1	1	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0
Coemansia reversa	Kickzellomycota	Kickxellomycetes	1	1	1	1	0	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	0	1	1	1	1	0	1	0	0	0	1	1	0	1	0	0	0	1	1	0	0	0	0	0	0
Kickxella alabastrina	Kickzellomycota	Kickxellomycetes	1	1	1	1	0	1	1	1	1	1	0	0	0	1	1	1	1	1	1	1	1	1	0	0	0	0	0	0	0	0	0	0	1	1	0	1	1	0	0	1	1	0	0	0	0	0	0
Encephalitozoon cuniculi	Microsporidia	Microsporidia	1	1	1	1	0	1	1	1	1	1	0	0	0	1	1	1	1	1	1	1	1	1	0	1	0	0	0	0	0	0	0	0	1	1	0	0	0	1	0	1	1	1	1	0	0	0	0
Encephalitozoon intenstinalis	Microsporidia	Microsporidia	1	1	1	1	0	1	1	1	1	1	0	0	0	1	1	1	1	1	1	1	1	1	0	1	0	0	0	0	0	0	0	0	1	1	0	0	0	1	0	1	1	1	1	0	0	0	0
Vavraia culicis	Microsporidia	Microsporidia	1	1	1	1	0	1	1	1	1	1	0	0	0	1	1	1	1	1	1	1	1	1	0	1	0	0	0	0	0	0	0	0	1	1	0	0	0	1	0	1	1	1	1	0	0	0	0
Gonapodya prolifera	Monophleradyiomycota	Monoblepharidomycetes	1	1	1	1	0	0	1	1	1	1	0	0	0	1	1	1	1	1	1	1	0	0	0	0	0	0	0	0	0	0	0	0	1	1	0	0	1	0	0	0	1	0	0	0	0	0	0
Hyaloraphidium curvatum	Monophleradyiomycota	Monoblepharidomycetes	1	1	1	1	0	1	1	1	1	1	0	0	0	1	1	1	1	1	1	1	1	1	0	1	0	0	0	0	0	0	0	0	0	1	0	1	0	0	0	1	1	0	0	0	0	0	0
Mucor circinelloides	Mucoromycota	Mucoromycetes	1	1	1	1	0	1	1	1	1	1	0	0	0	1	1	1	1	1	1	1	1	1	0	0	1	0	0	0	0	0	0	0	0	1	1	0	0	1	0	0	1	1	0	0	0	0	0

Mortierella elongata	Mucoromycota	Mucoromycetes	1	1	1	1	0	1	1	1	1	1	0	0	0	1	1	1	1	1	1	1	1	1	0	1	0	0	0	0	0	0	0	0	1	1	0	0	0	0	0	0	1	0	0	0	0	0
Phycomyces blakesleeanus	Mucoromycota	Mucoromycetes	1	1	1	1	0	1	1	1	1	1	0	0	0	1	1	1	1	1	1	1	1	1	0	1	0	0	0	0	0	0	0	0	0	1	0	1	1	0	0	1	0	1	0	0	0	0
Rhizopus microsporus	Mucoromycota	Mucoromycetes	1	1	1	1	0	1	1	1	1	1	0	0	0	1	1	1	1	1	1	1	1	1	0	1	0	0	0	0	0	0	0	0	1	1	1	1	1	1	1	1	1	1	0	0	0	0
Neocallimastix lanati	Neocallimastigomycota	Neocallimastigomycetes	1	1	1	1	0	1	1	1	1	1	0	0	0	1	1	1	1	1	1	1	1	1	0	1	0	0	0	0	0	0	0	0	1	1	1	1	1	1	1	1	1	1	0	0	0	0
Piromyces sp. finn	Neocallimastigomycota	Neocallimastigomycetes	1	1	1	1	0	1	1	1	1	1	0	0	0	1	1	1	1	1	1	1	1	1	0	1	0	0	0	0	0	0	0	0	1	1	0	1	1	1	0	1	1	1	0	0	0	0
Syncephalis truncata	Zoopagomycota	Zoopagomycetes	1	1	1	1	0	1	1	1	1	1	0	0	0	1	1	1	1	1	1	1	1	0	0	1	0	0	0	0	0	0	0	0	1	1	0	0	1	0	0	0	0	0	0	0	0	0
Piptocephalis tieghemiana	Zoopagomycotina	Zoopagomycetes	1	1	1	1	0	1	1	1	1	1	0	0	0	1	1	1	1	1	1	1	1	1	0	1	0	0	0	0	0	0	0	0	0	1	1	1	1	0	1	1	1	1	0	0	0	0
Caenorhabditis elegans	Opisthokonta	Metazoa	1	0	1	1	0	1	1	1	1	1	0	0	0	1	1	1	1	1	1	1	1	1	0	0	0	0	0	0	0	0	0	0	1	1	0	0	1	0	0	1	1	1	0	0	0	0
Drosophila melanogaster	Opisthokonta	Metazoa	1	1	1	1	0	1	1	1	1	1	0	0	0	1	1	1	1	1	1	0	0	1	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Homo sapiens	Opisthokonta	Metazoa	1	1	1	1	0	1	0	1	1	1	0	0	0	1	1	1	1	1	1	1	1	1	0	1	0	0	0	0	0	0	0	0	1	1	1	1	1	1	1	1	1	1	0	0	0	0
Naegleria gruberi	Heterolobosea	Heterolobosea	1	1	1	1	0	1	1	1	1	1	0	0	0	1	1	1	1	1	1	1	1	1	0	1	0	0	0	0	0	0	0	0	1	1	0	0	1	0	0	0	0	0	0	0	0	0
Trypanosoma brucei	Euglenozoa	Euglenozoa	1	1	1	1	0	1	1	1	1	0	0	0	0	1	1	1	1	1	1	1	1	0	0	0	0	0	0	0	0	0	0	0	1	1	0	0	1	0	0	0	1	0	0	0	0	0
Paramecium tetraurelia	Alveolates	Alveolates	1	1	1	1	0	1	1	1	1	1	0	0	0	1	1	1	1	1	1	1	1	1	0	1	0	0	0	0	0	0	0	0	1	1	0	0	1	0	0	1	0	0	0	0	0	0
Tetrahymena thermophila	Alveolates	Alveolates	1	1	1	1	0	1	1	1	0	1	0	1	0	1	1	0	0	0	1	0	0	1	0	1	0	0	0	0	0	0	0	0	1	1	0	1	1	0	0	1	1	0	0	0	0	0
Acanthamoeba castellanii	Amoebozoa	Amoebozoa	1	1	1	1	0	1	1	1	1	1	0	0	0	1	1	1	1	1	1	1	1	1	0	1	0	0	0	0	0	0	0	0	1	1	0	1	1	0	0	0	0	1	0	0	0	0
Dictyostelium discoideum	Amoebozoa	Amoebozoa	1	1	1	1	0	1	1	1	1	1	0	0	0	1	1	1	1	1	1	1	1	1	0	1	0	0	0	0	0	0	0	0	1	1	1	1	1	0	1	1	1	1	0	0	0	0
Phaeodactylum tricornutum	Stramenopiles	Stramenopiles	1	1	1	1	0	1	1	1	1	1	0	0	0	1	1	1	1	1	1	1	0	1	0	1	0	0	0	0	0	0	0	0	1	1	0	0	1	0	0	0	1	0	0	0	0	0
Phytophthora infestans	Stramenopiles	Stramenopiles	1	1	0	0	0	1	1	1	1	1	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	1	0	0	1	0	0	0	0	0
Pythium ultimum	Stramenopiles	Stramenopiles	1	1	1	1	0	1	1	1	1	1	0	0	0	1	1	1	1	1	1	1	1	1	0	0	0	0	0	0	0	0	0	0	0	1	1	1	1	0	1	1	1	1	0	0	0	0
Arabidopsis thaliana	Tracheophyta	Plantae	1	1	1	1	0	1	1	1	1	1	0	0	0	0	1	1	1	1	1	1	0	1	0	1	0	0	0	0	0	0	0	0	1	1	0	0	1	0	0	0	1	0	0	0	0	0
Cyanidioschyzon merolae	Rhodophyta	Algae	1	1	1	1	0	1	1	1	1	1	0	0	0	1	1	0	0	1	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	1	0	0	0	0	0	0
Chlamydomonas reinhardtii	Chlorophyta	Algae	1	1	1	1	0	0	0	1	1	1	0	0	0	1	1	1	1	1	1	1	0	0	0	0	0	0	0	0	0	0	0	0	1	1	0	0	1	0	0	0	1	0	0	0	0	0

Detailed metadata can be found at Zenodo doi:10.5281/zenodo.7635847 and FEMS Microbiology Reviews, fuad021; https://doi.org/10.1093/femsre/fuad021

6



```
,"O","P","Q","R","S","W","Z"))%>%
dplyr::select(-X,-O,-P,-Q,-R,-S,-W,-Z)%>%dplyr::mutate(region=rep("Core"))
ov <- core.25ka%>%group_by(Chr_bnd)%>%mutate(a=start-end_b-
nd,b=end-start_bnd)%>%mutate(ov=ifelse(a<0&b>0,"ov","no"))%>%filter(ov=
="ov")%>%
dplyr::select(-a,-b)%>%mutate(dist=rep(0),cat=rep("0k"))
ov2 <- core.25ka%>%group_by(Chr_bnd)%>%mutate(a=start-end_b-
nd,b=end-sta plyr::select(-a,-b)%>%
mutate(-
cat=ifelse(dist>=0&dist<5000,"<5k",ifelse(dist>5000&dist<10000,"<10k",ifelse
(dist>=10000&dist<15000,"<15k",ifelse(dist>15000&dist<20000,"<20k",ifelse(
dist>20000&dist<25000,"<25k","Ot"))))))

dplyr::select(-a,-b)%>%mutate(dist=rep(0),cat=rep("0k"))
ov2 <- core.25ka%>%group_by(Chr_bnd)%>%mutate(a=start-end_bnd,b=end-sta plyr::select(-a,-b)%>%
mutate(-
(dist>=0&dist<5000,"<5k",ifelse(dist>5000&dist<10000,"<10k",ifelse(dist>=10000&dist<15000,"<15k",ifelse(dist>15

15000&dist<20000,"<20k",ifelse(dist>20000&dist<250
00,"<25k","Ot"))))))
core.25ka <- rbind(ov,ov2)%>%dplyr::select(-start_b-
```

CHAPTER 7

General Discussion

```
#
##
# nucleosome distance between them
pdf("nucleosomes.distance.3.strains.pdf",width=6, height=5)
ggplot(distribution,aes(x=distance,y=..scaled..,color=strain)) +
 geom_density(size=1) +
 #scale_y_sqrt(labels=percent) +
 ylab('Frequency (%)') + xlab('Distance (bp)') +
 scale_x_sqrt(breaks=c(0,1000,5000,10000,20000)) +
 #scale_x_continuous(breaks=c(1000,5000,10000,15000,20000)) +
 scale_color_manual(values = c("#228EE3", "#9A5BD4", "#62dc1c")) +
 theme_bw() +
 theme(axis.text = element_text(size=12,color='black'),
       panel.border = element_rect(colour = "black", fill=NA, size=1),panel.grid = element_blank())
####
##
#
#
#
```

```
  #scale_y_sqrt(labels=percent) +
  ylab('Frequency (%)') + xlab('Size (bp)') +
  scale_x_continuous(breaks=c(0,150,300,450,600,750,900),expand =
  scale_color_manual(values = c("#228EE3", "#9A5BD4", "#62dc1c"))
  theme_bw() +
  theme(axis.text = element_text(size=12,color='black'),
        panel.border = element_rect(colour = "black", fill=NA, size=1),par
element_blank())
dev.off()
```

Genomes are highly organized inside eukaryotic nucleus (Lynch and Walsh 2007). On the linear DNA strand, genes are not randomly organized but usually occur in clusters that are separated from repetitive elements (Lynch 2006). Histone proteins fold the DNA into chromatin, which can be accessible to the transcription machinery as in euchromatin, or inaccessible as in heterochromatin (Olins and Olins 2003). Additionally, the chromatin is organized within the three-dimensional (3D) space of the nucleus (Lieberman-Aiden et al. 2009). Thus, the mechanisms that control the genome organization can be summarized in three genomic dimensions. The genome in three dimensions comprises the architecture of the linear DNA, the chromatin, and the 3D organization of the genome. These three dimensions and the interactions between them compose the epigenome. The work described in this thesis combines complementary analyses across these three genomic dimensions focusing on the genome evolution of *Verticillium dahliae*, a plant-pathogenic fungus capable of colonizing hundreds of plant species (Fradin and Thomma 2006).

Genome evolution of plant pathogens

The genomes of many plant pathogens are thought to be organized in two compartments, one compartment is gene-rich and repeat-poor while the other one is gene-poor and repeat-rich (Dong et al. 2015; Möller and Stukenbrock 2017; Torres et al. 2020; Seidl and Thomma 2017). The repeat-rich genomic compartments are considered to be evolutionary dynamic as they usually contain abundant genomic variation (Erlendson et al. 2017; Seidl et al. 2016), such as in the accessory chromosomes of *Zymoseptoria tritici* (Schotanus et al. 2015) or the adaptive genomic regions (AGRs) in *V. dahliae* (Cook et al. 2020). Interestingly, many genes that are thought to contribute to host and environmental adaptation reside in these dynamic genomic compartments (Dong et al. 2015; Upson et al. 2018; Torres et al. 2020; Frantzeskakis et al. 2019; Erlendson et al. 2017). Furthermore, it is thought that genes localized in dynamic genomic compartments evolve faster because of the typically observed high abundance of genomic variation, in contrast to genes localized in repeat-poor regions that are depleted of genomic variation (Dong et al. 2015; Möller and Stukenbrock 2017; Upson et al. 2018; Torres et al. 2020; Seidl and Thomma 2017). These observations gave rise to the 'two-speed' genome model, as a model to understand the genome organization and evolution of filamentous plant pathogens (Raffaele and Kamoun 2012; Raffaele et al. 2010; Dong et al. 2015). However, to date, it remains challenging to test the assumption of the increased evolutionary speed in dynamic genomic compartments, mostly due to the lack of knowledge of mutation rates and the evolutionary forces shaping genome organization (**Chapter 2**).

Single-base mutations are those mutations in which a single nucleotide base is substituted, inserted, or deleted. The rate of single-base mutations has been explored in only a few fungal organisms (Hiltunen et al. 2019; Ene et al. 2018; Long et al. 2016; Álvarez-Escribano et al. 2019; Habig et al. 2021; de la Peña et al. 2023). When compared with the core genome, accessory chromosomes and repetitive regions show increased rates of

single-base mutation in *Candida albicans* and *Z. tritici* (Habig et al. 2021; Ene et al. 2018). Thus, it has been suggested that increased mutation rates play a role in the emergence of dynamic genomic compartments. However, the single-base mutation rate in *Verticillium dahliae* is currently unknown. Moreover, genes located in AGRs are depleted in nucleotide polymorphisms between species (Depotter et al. 2019). Our data support this notion, as we similarly observed a reduction of nucleotide polymorphisms measured within a population of *V. dahliae* (**Chapters 3-5**). The Darwinian view assumes that after mutations occur, natural selection will maintain beneficial mutations by positive selection or purge harmful mutations by negative selection. Actually, many environmentally responsive genes in dynamic genomic compartments have been found to be directly subject to strong positive natural selection in some plant pathogens (Aguileta et al. 2009; Stukenbrock et al. 2010; Sperschneider et al. 2015; Schweizer et al. 2018; Grandaubert et al. 2019). Thus, it is thought that a differential effect of natural selection on genes across the genome could promote the emergence of dynamic genomic compartments. However, genes located in AGRs do not show signs of positive selection in *V. dahliae* (Depotter et al. 2019). In line with this observation, our data shows sequence conservation in AGRs when *V. dahliae* is compared to other species of the *Verticillium* genus (**Chapters 3-5**). Instead of abundant nucleotide mutations and detectable differences in selection regimes, large-scale genomic rearrangements and abundant duplications followed by reciprocal gene losses have been shown to play a role in the emergence of AGRs (de Jonge et al. 2013; Faino et al. 2016).

Genome compartmentalization analogous to the two-speed genome organization in plant pathogens have been observed all across the tree of life, and in many cases dynamic genomic compartments are similarly enriched for environmentally responsive genes (**Chapter 2**). It is therefore conceivable that the two-speed genome organization directly or indirectly contributes to adaptation by the capacity of dynamic genomic compartments to generate beneficial genetic variation.

Plant pathogen evolution in three dimensions

The capacity of a species to generate genetic variation is its capacity to evolve (Gregory 2009). Genetic variation can be achieved by different types of mutations, like single-base mutations or genomic rearrangements (Shastry 2009; Escaramís et al. 2015; Ho et al. 2020). Single-base mutations may change amino acids by changes to codons or by introducing stop codons, thus known as nonsynonymous mutations, or may not alter protein sequences, thus known as synonymous mutations (Ohta and Ina 1995; Nei and Kumar 2000). Typically, nonsynonymous mutations are under strong negative selection because of their potential to change the phenotype (Ohta and Ina 1995; Nei and Kumar 2000; Conrad and Hurles 2007; Feulner and De-Kayne 2017). Conversely, synonymous mutations are considered mostly neutral and irrelevant in terms of natural selection because of their insufficient phenotypic effect (Ohta and Ina 1995; Kimura 1968; Nei and Kumar 2000), or at least, nearly neutral as they potentially could have limited effects on the phenotype (Lynch and Conery 2003; Lynch

and Walsh 2007; Chamary et al. 2006; Ohno 2013; Bergthorsson et al. 2007). The rate that mutations actually occurs is higher than expected if all mutations in a species population would contribute to adaptation by natural selection (Kimura 1968, 1983; Nei and Kumar 2000; Lynch 2010; Lynch et al. 2016). Therefore, the vast majority of random mutations that occur in a population are evolutionary neutral (Kimura 1983). The accumulation of neutral mutations is usually referred to as genetic robustness against deleterious mutations (Wagner 2013; Masel and Trotter 2010). Thus, the more neutral mutations, the higher the genetic robustness. It is expected then that populations will favor genetic robustness when mutations are already adaptive for a given environment (Meyers and Bull 2002). However, if the environment changes, and thus the selective pressure, it is unlikely that a beneficial mutation arises in a heterogenous environment if most mutations are random and neutral.

The capacity to generate beneficial or adaptive genetic variation is known as evolvability (Wagner 2013; Pigliucci 2008). For evolvability to occur, the current pool of mutations of an organism is important, but the additional mutations that may be generated in the future might even be more important (Wagner and Altenberg 1996; Payne and Wagner 2018; Draghi and Wagner 2008; Riederer et al. 2022). Evolvability can be achieved by shaping mechanisms that generate and maintain variation (e.g. mutation accumulation rate bias), by shaping the effect of variation (e.g. phenotypic developmental bias), or by shaping the process of selection (e.g. shortening the generation time to allow for faster adaptation) (Riederer et al. 2022). Thus, it is thought that evolvability can shape the mutational outcome of a species.

At first sight, evolvability contradicts the common assumption that mutations are random and that any evolutionary force acts blindly using the random genetic variation already present in a population (Wagner 2008). However, adaptive evolution integrates information from the environment into the organism by selecting variation that is relevant to that specific environment and by further acting on such genetic variation (Draghi and Wagner 2008). For example, fluorescent proteins under a strong selection for a yellow fluorescence phenotype evolve faster to a 'new' green fluorescence phenotype when the environment changes the green fluorescence would be selected for (Zheng et al. 2020). Strong selection in the yellow fluorescence phenotype vastly removed deleterious mutations but accumulate foldability-improving mutations (Zheng et al. 2020). In doing so, these mutations promote the formation of correctly folded proteins after the additional mutations that are necessary to evolve the new green phenotype. Thus, yellow fluorescent proteins evolved a mutation bias for further a beneficial 'new' phenotype. Thus, a mutation bias can shorten the path, i.e. the number of mutations needed, to generate beneficial and adaptive mutations thereby enabling swift adaptation to a heterogenous environment compared with completely random mutations.

The 'exploration' for beneficial mutations could be driven by abundant neutral mutations. Paradoxically, as random neutral mutations arise and the genetic robustness increase, there is fewer mutations for natural selection to act upon because simply these

mutations are neither beneficial nor deleterious to be selected on. Higher genetic robustness is typically reconciled with natural selection if we consider genetic robustness as a transient property in which abundant neutral mutations can occur, but at some point become visible to natural selection (Wagner 2005, 2008; Masel and Trotter 2010; Payne and Wagner 2018). Therefore, the 'exploration' for beneficial mutations by high transient genetic robustness allows for evolvability to arise. In this manner, abundant neutral mutations could shape the 'adaptive valleys' for a swift rise of beneficial mutations (Fig. 2). High transient robustness could be achieved by increasing the effective population size. Thus, the more reproductive individuals, the more neutral mutations will be in a population. However, this population increase could be virtually absent in asexual organisms that supposedly have low effective population sizes, such as for example *V. dahliae* (Gurung et al. 2014; Short et al. 2014; Milgroom et al. 2014). Instead, the accumulation of mutations, for a high transient genetic robustness, can be facilitated by variations of mutation rates (Sniegowski et al. 2000; Hodgkinson and Eyre-Walker 2011; Martincorena et al. 2012; Martincorena and Luscombe 2013), as well as by the aggregation of genes with similar effect on the phenotype along the genome (Pepper 2003; Gerhart and Kirschner 2007; Yona et al. 2012). Thus, one can speculate that high transient robustness to facilitate evolvability can be achieved by genome compartmentalization.

Theoretical models have shown that genomic compartments can emerge from a genome that is randomly populated by genes and transposable elements (TEs) through TE-driven genomic rearrangements (Crombach and Hogeweg 2007, 2008, 2011). In these models, TEs can copy themselves through the genome and create double-strand breaks that can lead to rearrangements, and hence via those rearrangements, a highly organized genome emerges in which essential genes co-occur and are separated from TE-rich regions that contain genes for environmental adaptation, resembling the two-speed genome organization (Crombach and Hogeweg 2007, 2011). Genes that cluster with TEs have a higher probability of mutations, which enables adaptation to changing environments (Crombach and Hogeweg 2007; Cuypers et al. 2017). Indeed, abundant TEs and enhanced TE activity have been observed in dynamic genomic compartments in many fungal pathogens (Croll and McDonald 2012; Sánchez-Vallet et al. 2018). TEs have also been shown to play a key role in the formation and maintenance of the AGRs in *V. dahliae* (Amyotte et al. 2012; Faino et al. 2016; Depotter et al. 2019; Cook et al. 2020; Seidl et al. 2020). We observed that a small fraction of the total TE content remains active and is associated with abundant structural variations (SVs) in AGRs (**Chapter 3**; Fig. 1) (Faino et al. 2016). Active TEs are expressed during *in vitro* conditions (Amyotte et al. 2012; Faino et al. 2016; Cook et al. 2020) and are also highly expressed *in planta* (**Chapter 3**). Additionally, some *V. dahliae* TEs are induced by heat stress (Amyotte et al. 2012). Structural variations can emerge rapidly through the action of TEs and thus TEs likely drive the formation and maintenance of AGRs. Nevertheless, the mechanisms by how TE insertions cause genomic variation that is restricted to AGRs remains largely unexplored. TE insertion and deletion polymorphisms, associated with TE activity, tend to accumulate in TE-rich regions, but not in the core genome (Fouché et al. 2022; Lorrain et al. 2021; Seidl and Thomma 2017). Indeed, as

previously shown (de Jonge et al. 2013; Faino et al. 2016), we also observed that TE insertions are associated with AGRs (**Chapter 3**), and consequently the mutational bias of the TE activity to AGRs could increase the evolvability of *V. dahliae*.

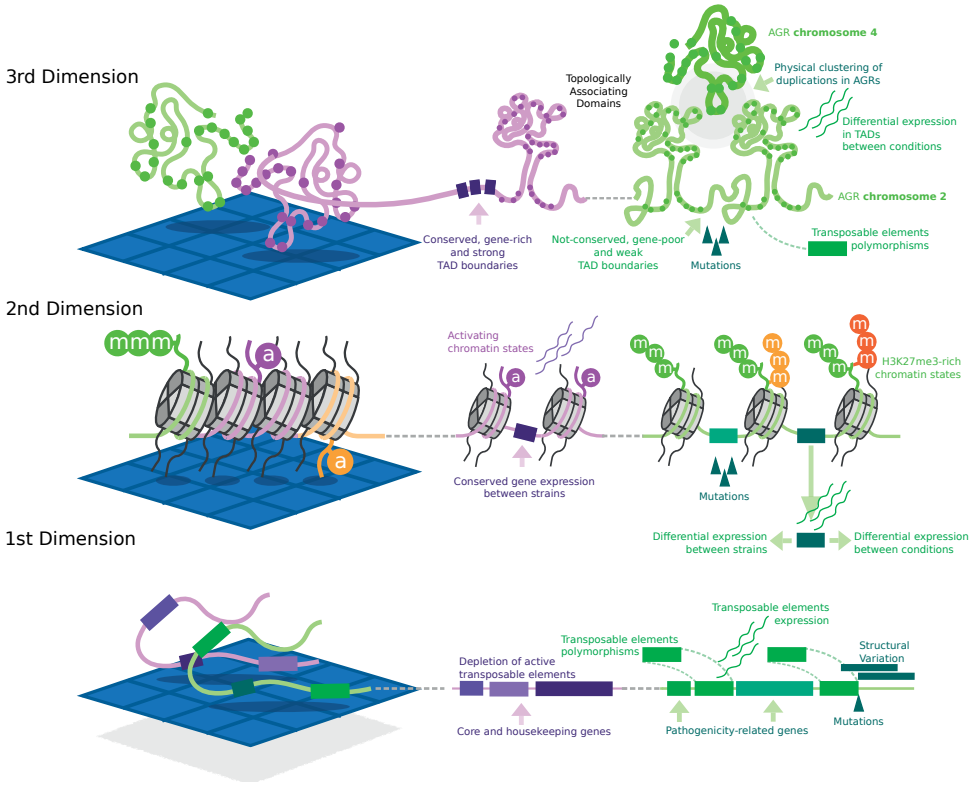


Figure 1. The genome of the plant pathogen *Verticillium dahliae* in three dimensions.
Schematic representation of the major findings in this thesis summarized in the three dimensions of genome organization. **In the first dimension** (bottom), active and polymorphic transposable elements (TEs) are depleted from the core genome and are mostly restricted to adaptive genomic regions (AGRs). Polymorphic TEs co-locate with structural variants. Polymorphic TEs are active, highly expressed, and locate close to pathogenicity-related genes. **In the second dimension** (middle), the core genome is enriched in activating chromatin states, while AGRs are characterized by the co-occurrence of the heterochromatic histone modification H3K27me3 (green methylations) with activating histone modifications (i.e. H3K4me3 in yellow or H3K36me3 in red). Genes enriched in H3K27me3-associated chromatin states show low sequence conservation and are differentially expressed between *V. dahliae* strains and conditions. Most of the genes enriched in those chromatin states have pathogenicity-related functions. **In the third dimension** (upper), chromatin is organized locally in topologically associating domains (TADs). In the core genome, TADs are separated by boundary regions that lack chromatin interactions, are enriched in genes, and are highly conserved across the *Verticillium* genus. In the AGRs, TADs are separated by weak boundary regions, enriched in polymorphic TEs, and have variable sequence conservation. Moreover, duplications in AGRs physically interact.

Chromatin can influence the likelihood that mutations occur and are tolerated at different regions across the genome (Martincorena and Luscombe 2013). Typically, euchromatic regions are associated with the depletion of genetic variation (Makova and Hardison 2015; Hazarika et al. 2022; Quiroz et al. 2022). In contrast, heterochromatin is

associated with abundant genetic variation (Lewis et al. 2009; Makova and Hardison 2015; Janssen et al. 2018; de la Peña et al. 2023; Liu et al. 2020a; Yasuhara et al. 2005). Despite these observations, in the Darwinian view, mutations occur regardless of their fitness effect. Recently, it has been observed that the chromatin landscape reduces the mutation rate and the occurrence of deleterious mutations in housekeeping genes in *Arabidopsis thaliana* (Monroe et al. 2022b). Highly debated due to its implications, the observations in *A. thaliana* suggest that chromatin can bias the direction of natural selection (Monroe et al. 2022a, 2022b; Liu and Zhang 2022; Charlesworth and Jensen 2023). Thus, chromatin could shape the mutational bias at different regions along the genome and contribute to evolvability.

In many plant pathogens, dynamic genomic compartments are enriched for the tri-methylation of the lysine 27 on histone 3 (H3K27me3) (Möller and Stukenbrock 2017; Seidl et al. 2016), like in *V. dahliae* AGRs (Cook et al. 2020; Kramer et al. 2022). H3K27me3 has been associated with high mutation rates (Habig et al. 2021; de la Peña et al. 2023), double-strand breaks (Her and Bunting 2018; Schep et al. 2021; Boulton and Jackson 1996), and reduced sequence conservation (Jamieson et al. 2013; Tralamazza et al. 2022). In line with those observations, we find that genes in AGRs that are enriched in H3K27me3 show increased presence/absence mutations across the *Verticillium* genus when compared with genes depleted in H3K27me3 (**Chapter 4**). Hence, H3K27me3 correlates with high genetic variation. Additionally, genes enriched for H3K27me3 are differentially expressed between strains (**Chapter 4**), suggesting that AGRs show high levels of expression variation together with abundant presence/absence variation (Cook et al. 2020; de Jonge et al. 2013). Incremental evidence suggests that H3K27me3 is not the only factor that could drive the functions associated with dynamic genomic compartments. For example, H3K27me3 is not essential for differential expression of the genes localized in AGRs in *V. dahliae* (Cook et al. 2020; Kramer et al. 2022). Furthermore, the removal of H3K27me3 does not have an effect on the accumulation of genomic rearrangements in *Z. tritici* (Habig et al. 2021). Thus, we explored the combinations of seven histone modifications into the so-called chromatin states (**Chapter 4**). We observed that H3K27me3 in AGRs occurs in discrete combinations with other histone modifications typically associated with active gene expression such as H3K9ac, H3K4me3, or H3K36me3 (**Chapter 4**; Fig. 1). Those H3K27me3-associated chromatin states are conserved between strains and correlate with gene expression and sequence variation in AGRs (**Chapter 4**). Importantly, H3K27me3-associated chromatin states are mostly present at conditionally responsive genes, which are thought to play a role in the host- and environmental adaptation (**Chapter 4**). Hence, it is conceivable that the combinations of H3K27me3 could be associated with a bias for the accumulation of mutations in AGRs. Nevertheless, experimental evidence that H3K27me3-associated chromatin states increase mutation rates is still lacking for *V. dahliae*. Thus, admittedly, we cannot discern if H3K27me3 directly prevents the repair of mutations, or the accumulation of mutations is a byproduct of the repression of gene expression (Habig et al. 2021; Monroe et al. 2022b). Alternatively, the occurrence of H3K27me3 might be an indirect response

from the silencing of recently evolved genes or recent TE insertions (Jamieson et al. 2013; Sentmanat and Elgin 2012).

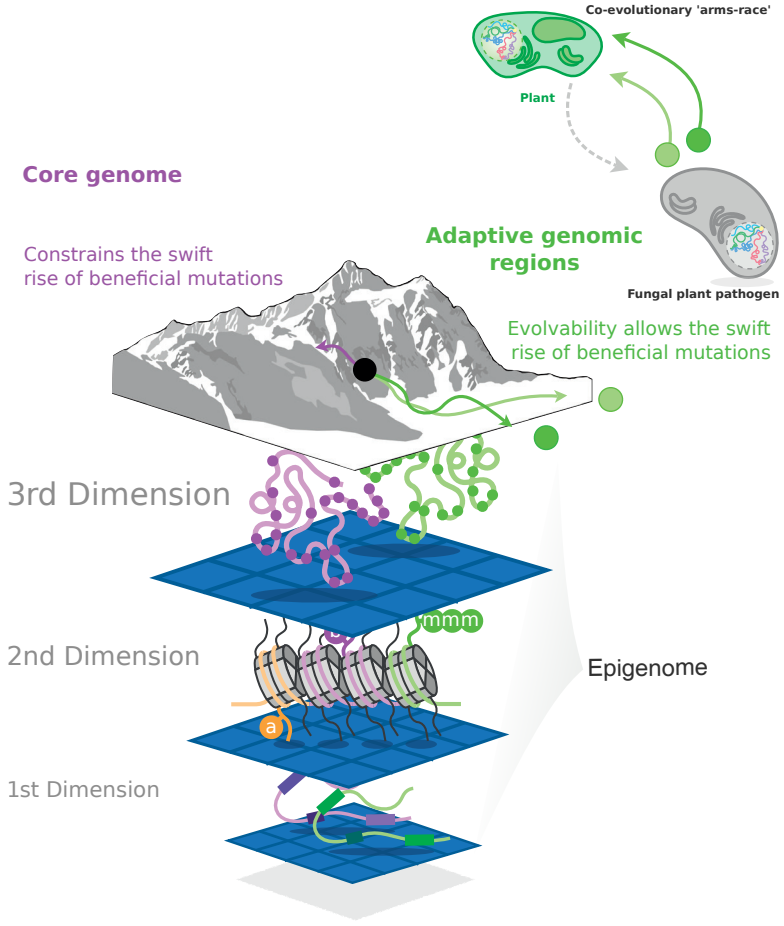


Figure 2. The epigenome in three dimensions shapes the evolvability of adaptive genomic regions.
Collectively, three genomic dimensions and their interactions form the epigenome. The epigenome shapes the mutational landscape in *V. dahliae*, as shown on top. The mountains are resembling the functional constraints of the core genome for mutations to occur. Thus, the rise of a beneficial mutation would be difficult in the core genome (purple arrow). The valleys resemble the already abundant mutations that 'pave' the way for the swift rise of beneficial mutations in the adaptive genomic regions (AGRs, green arrows). Changes in the environment allow the generation of quick beneficial mutations (light and dark green arrows). Therefore, the swift rise of beneficial mutations in the AGRs could allow the fungal plant pathogen *V. dahliae* to adapt in the coevolutionary 'arms race' with plant hosts.

Besides H3K27me3 roles in gene expression and functioning as the direct or indirect regulator of mutation accumulation in diverse fungi (Habig et al. 2021; de la Peña et al. 2023), it has been hypothesized that the 3D folding of the H3K27me3-rich regions

inside the nucleus could increase the likelihood of mutations (Möller et al. 2019; Rodriguez et al. 2022). H3K27me3 has been shown to affect the local and global 3D organization of the genome in fungi, including a few studied plant pathogens (**Chapter 6**). For example, sub-telomeric H3K27me3-rich regions bundle close together within the nucleus of *Neurospora crassa* (Rodriguez et al. 2022; Klocko et al. 2016; Galazka et al. 2016). Additionally, it has been suggested that the H3K27me3-rich accessory chromosomes may be physically separated from core chromosomes in *Z. tritici* (Schotanus et al. 2015; Möller et al. 2019). We observed the physical clustering of AGRs across the *Verticillium* genus, indicating that the mechanism driving this physical association is conserved (**Chapter 5**; Fig. 1). The physical association of H3K27me3-rich regions has been associated with abundant accessory chromosome losses, and thus chromosome instability in *Z. tritici* (Habig et al. 2021; Möller et al. 2019). Late replication timing of the H3K27me3-rich regions could be the cause of chromosome instability (Habig et al. 2021). In *V. dahliae*, we observed that the physical association corresponds with segmental duplications in AGRs (**Chapter 5**). The physical co-localization of highly similar sequences may decrease the efficiency of replication, resulting in an increase in DNA double-strand breaks (Patel et al. 2019; Zuo et al. 2021). Hence, we speculate that the physical colocalization of homologous sequences in AGRs could increase the probability of DNA double-strand breaks that, unfaithfully repaired, could contribute to the bias for the accumulation of mutations in AGRs in *V. dahliae*.

Altogether, the dynamic compartments in the two-speed genome model, such as the AGRs in *V. dahliae*, could emerge by a mutational bias caused by the TE enrichment, the H3K27me3-rich chromatin, and the 3D structure (Fig. 2). Thus, the emergent mutational bias in the AGRs through epigenome characteristics could increase the transient robustness by the capacity to generate abundant mutations needed for evolvability.

The molecular mechanisms behind the plant pathogen evolution in three dimensions

Even though we observed the presence of active TEs, a distinct chromatin profile, and spatial organization that could lead to mutational bias and to the formation and maintenance of the genome compartmentalization in *V. dahliae*, many open questions remain. In the future, it would be crucial to experimentally test the assumption and mechanisms behind mutational bias indicated by the two-speed genome model, that potentially gave rise to the AGRs.

First, to understand the role of TEs in the formation and maintenance of AGRs, one could think of the precise insertion of TEs in distinct regions of the genome by CRISPR/Cas9 (Weber et al. 2023). For instance, the CRISPR/Cas9 manipulation of *Ty1* transposable elements induces double-strand breaks and abundant genomic rearrangements in *S. cerevisiae* (Qi et al. 2023). Thus, the insertion of TEs into specific regions of the *V. dahliae* genome might help to understand the role of TEs in the origin of genomic rearrangements (**Chapter 3**), and the changes in the 3D genome organization (**Chapter 5**) associated with AGRs. Recently, the inactivation of Cas9 by the mutation of its endonuclease domain (dCas9)

showed to be a promising alternative to control DNA methylation and histone modifications in a locus-specific manner without modifying the genetic background (Kwon et al. 2017; Nakamura et al. 2021; Papikian et al. 2019; Pan et al. 2021; Ghoshal et al. 2021). Therefore, to test the role of H3K27me3 and its combination with other histone modifications in the formation and maintenance of AGRs, the manipulation by CRISPR/dCas9 could direct the formation of additional H3K27me3 regions. These experiments could test if H3K27me3 predates the formation of AGRs by influencing the insertion frequencies of TEs (**Chapter 3**), by increasing the number of mutations (**Chapter 4**), and/or by inducing the physical clustering of those regions (**Chapter 5**). Collectively, disentangling the influence of TEs and chromatin on the observations made in this thesis will be necessary to understand the origin of the AGRs.

Second, mutation accumulation experiments should be performed to understand the mutational rates in *V. dahliae* and its AGRs. Mutation accumulation experiments *in vitro* have been reported in different fungi (Hiltunen et al. 2019; Ene et al. 2018; Long et al. 2016; Álvarez-Escribano et al. 2019; Habig et al. 2021; Möller et al. 2019). These experiments have shown that the mutation rates vary between species. For example, the absolute mutation rate in *Marasmius oreades* is 3.8×10^{-12} mutations per site and division, which is ~180 times lower when compared with the *N. crassa* of 6.7×10^{-10} mutations per site and division (de la Peña et al. 2023; Hiltunen et al. 2019). Importantly, these experiments have shown that the mutation rate also varies between regions that differ in chromatin characteristics. For example, in *N. crassa*, the mutation rate in euchromatic regions is 2.46×10^{-10}, while in heterochromatic regions is 2.43×10^{-9}, thus ~10 times higher (de la Peña et al. 2023). Similarly, in *Z. tritici*, the mutation rate in euchromatic regions is 1.96×10^{-10}, while in heterochromatic regions is 1.18×10^{-9}, thus ~15 times higher (Habig et al. 2021). Additionally, mutational rates could vary depending on additional factors, and specifically for pathogens the interaction with the host could alter the mutation rate (Ene et al. 2018; Murray et al. 2021). To date, it remains challenging to perform mutation accumulation experiments of plant pathogens during host interaction *in planta*. For example, *V. dahliae* colonizes the xylem tissue of plants, thus it is challenging to obtain enough fungal biomass after serial infection assays and after long periods of time. Taken together, experimental evolution experiments would help us to uncover the potential role of the mutational bias for the formation and maintenance of the genome compartmentalization of *V. dahliae*.

Everything–Everywhere–All at once

The AGRs in *V. dahliae* were originally discovered by their high degree of presence/absence polymorphisms between *V. dahliae* strains (de Jonge et al. 2013; Faino et al. 2016), and these regions were subsequently further characterized and refined by their distinct epigenomic profile (Cook et al. 2020). Dynamic genome compartments are epigenetically distinct from core genomic regions, which is increasingly recognized as a common pattern in various fungal plant pathogens (Schotanus et al. 2015; Wang et al. 2017; Fokkens et al. 2018; Cook et al. 2020; Soyer et al. 2021; Möller et al. 2023). The work described in this thesis has contributed to a better understanding on the roles of three genomic dimensions in the genome evolution of *V. dahliae*: the architecture of the linear DNA, the chromatin, and the spatial organization of the genome (Fig. 1). Those three dimensions and the interactions between them compose the epigenome that collectively determines the phenotype. For instance, the presence of TEs correlates with genomic rearrangements in the linear DNA (**Chapter 3**) (Faino et al. 2016; de Jonge et al. 2013), with specific chromatin states (**Chapter 4**) (Cook et al. 2020), and also with changes in the 3D organization of the genome (**Chapter 5**). These observations collectively highlight that the study of the epigenome represents a step forward in the understanding of the mutational processes that foster the evolution of fungal plant pathogens.

Plants and fungal plant pathogens are involved in co-evolutionary 'arms races' (Koch et al. 2014; Upson et al. 2018; Frantzeskakis et al. 2019). Plants have evolved immune systems to detect fungal pathogen intrusions and mount defense responses, while fungal plant pathogens evolved a complex molecular tool kit to support host colonization (Rovenich et al. 2014; Cook et al. 2015). This everlasting co-evolutionary 'arms race' is characterized by repeated cycles of adaptation and counter-adaptation of fungal plant pathogens and their hosts (Strotz et al. 2018; McDonald and Stukenbrock 2016). To enable these host adaptation cycles, high levels of genetic variation in fungal plant pathogens are needed (Dong et al. 2015; Möller and Stukenbrock 2017; Upson et al. 2018; Seidl and Thomma 2017). Nowadays, anthropogenic hazards, such as climate change, similarly exert high selection pressure on plant pathogen populations (Fisher et al. 2012; Santini and Ghelardini 2015; Zhan et al. 2018). The epigenome could facilitate the emergence of beneficial mutations through the collective force of the three genomic dimensions (Fig. 2), thereby enabling swift adaptation to highly heterogeneous abiotic and biotic environments. Holistic research on the roles of genetic, epigenetic, and spatial characterization of dynamic genome compartments will aid further understanding of the genome function and evolution in plant pathogenic fungi.

References

Abramo K, Valton A-L, Venev SV, Ozadam H, Fox AN, Dekker J. 2019. A chromosome folding intermediate at the condensin-to-cohesin transition during telophase. *Nat Cell Biol.* **21**: 1393–1402.

Acemel RD, Lupiáñez DG. 2023. Evolution of 3D chromatin organization at different scales. *Curr Opin Genet Dev.* **78**: 102019.

Admire A, Shanks L, Danzl N, Wang M, Weier U, Stevens W, Hunt E, Weinert T. 2006. Cycles of chromosome instability are associated with a fragile site and are increased by defects in DNA replication and checkpoint controls in yeast. *Genes Dev.* **20**: 159–173.

Agrios GN. 2005. *Plant Pathology*. Elsevier.

Aguileta G, Refregier G, Yockteng R, Fournier E, Giraud T. 2009. Rapidly evolving genes in pathogens: Methods for detecting positive selection and examples among fungi, bacteria, viruses and protists. *Infect Genet Evol.* **9**: 656–670.

Akagi Y, Akamatsu H, Otani H, Kodama M. 2009. Horizontal chromosome transfer, a mechanism for the evolution and differentiation of a plant-pathogenic fungus. *Eukaryot Cell* **8**: 1732–1738.

Akdemir KC, Le VT, Chandran S, Li Y, Verhaak RG, Beroukhim R, Campbell PJ, Chin L, Dixon JR, Futreal PA, et al. 2020. Disruption of chromatin folding domains by somatic genomic rearrangements in human cancer. *Nat Genet.* **52**: 294–305.

Allshire RC, Madhani HD. 2018. Ten principles of heterochromatin formation and function. *Nat Rev Mol Cell Biol.* **19**: 229–244.

Alonge M, Wang X, Benoit M, Soyk S, Pereira L, Zhang L, Suresh H, Ramakrishnan S, Maumus F, Ciren D, et al. 2020. Major impacts of widespread structural variation on gene expression and crop improvement in tomato. *Cell* **182**: 145-161 e23.

Álvarez-Escribano I, Sasse C, Bok JW, Na H, Amirebrahimi M, Lipzen A, Schackwitz W, Martin J, Barry K, Gutiérrez G, et al. 2019. Genome sequencing of evolved aspergilli populations reveals robust genomes, transversions in *A. flavus*, and sexual aberrancy in non-homologous end-joining mutants. *BMC Biol.* **17**: 88.

Amselem J, Lebrun MH, Quesneville H. 2015. Whole genome comparative analysis of transposable elements provides new insight into mechanisms of their inactivation in fungal genomes. *BMC Genomics.* **16**: 141.

Amyotte SG, Tan X, Pennerman K, Jimenez-Gasco M del M, Klosterman SJ, Ma L-J, Dobinson KF, Veronese P. 2012. Transposable elements in phytopathogenic *Verticillium spp*.: insights into genome evolution and inter- and intra-specific diversification. *BMC Genomics.* **13**: 314.

Anania C, Acemel RD, Jedamzick J, Bolondi A, Cova G, Brieske N, Kühn R, Wittler L, Real FM, Lupiáñez DG. 2022. *In vivo* dissection of a clustered-CTCF domain boundary reveals developmental principles of regulatory insulation. *Nat Genet* **54**: 1026–1036.

Anania C, Lupiáñez DG. 2020. Order and disorder: abnormal 3D chromatin organization in human disease. *Brief Funct Genomics.* **19**: 128–138.

Anders S, McCarthy DJ, Chen Y, Okoniewski M, Smyth GK, Huber W, Robinson MD. 2013. Count-based differential expression analysis of RNA sequencing data using R and Bioconductor. *Nat Protoc.* **8**: 1765–1786.

Anderson DC, Green GR, Smith K, Selker EU. 2010. Extensive and varied modifications in histone H2B of wild-type and histone deacetylase 1 mutant *Neurospora crassa. Biochemistry* **49**: 5244–5257.

Anderson EC, Frankino PA, Higuchi-Sanabria R, Yang Q, Bian Q, Podshivalova K, Shin A, Kenyon C, Dillin A, Meyer BJ. 2019. X Chromosome domain architecture regulates *Caenorhabditis elegans* lifespan but not dosage compensation. *Dev Cell* **51**: 192-207.e6.

Andersson R, Sandelin A. 2020. Determinants of enhancer and promoter activities of regulatory elements. *Nat Rev Genet.* **21**: 71–87.

Armstrong J, Hickey G, Diekhans M, Fiddes IT, Novak AM, Deran A, Fang Q, Xie D, Feng S, Stiller J. 2020. Progressive Cactus is a multiple-genome aligner for the thousand-genome era. *Nature* **587**: 246–251.

Arzate-Mejía RG, Josué Cerecedo-Castillo A, Guerrero G, Furlan-Magaril M, Recillas-Targa F. 2020. *In situ* dissection of domain boundaries affect genome topology and gene transcription in *Drosophila*. *Nat Commun.* **11**: 894.

Ashe A, Colot V, Oldroyd BP. 2021. How does epigenetics influence the course of evolution? *Philos Trans R Soc Lond B Biol Sci.* **376**: 20200111.

Ba Q, Hei Y, Dighe A, Li W, Maziarz J, Pak I, Wang S, Wagner GP, Liu Y. 2022. Proteotype coevolution and quantitative diversity across 11 mammalian species. *Sci Adv.* **8**: eabn0756.

Badet T, Croll D. 2020. The rise and fall of genes: origins and functions of plant pathogen pangenomes. *Curr Opin Plant Biol* **56**: 65–73.

Badet T, Oggenfuss U, Abraham L, McDonald BA, Croll D. 2020. A 19-isolate reference-quality global pangenome for the fungal wheat pathogen *Zymoseptoria tritici. BMC Biology* **18**: 1-18.

Bai L, Morozov AV. 2010. Gene regulation by nucleosome positioning. *Trends Genet.* **26**: 476–483.

Bailey TL, Johnson J, Grant CE, Noble WS. 2015. The MEME suite. *Nucleic Acids Res.* **43**: W39–W49.

Bailly-Bechet M, Haudry A, Lerat E. 2014. "One code to find them all": a perl tool to conveniently parse RepeatMasker output files. *Mob DNA.* **5**: 1–15.

Baker M. 2011. Making sense of chromatin states. *Nat Methods.* **8**: 717–722.

Banigan EJ, Mirny LA. 2020. Loop extrusion: theory meets single-molecule experiments. *Curr Opin Cell Biol.* **64**: 124–138.

Bannister AJ, Kouzarides T. 2011. Regulation of chromatin by histone modifications. *Cell Res.* **21**: 381–395.

Barrales RR, Forn M, Georgescu PR, Sarkadi Z, Braun S. 2016. Control of heterochromatin localization and silencing by the nuclear membrane protein Lem2. *Genes Dev.* **30**: 133–148.

Barutcu AR, Maass PG, Lewandowski JP, Weiner CL, Rinn JL. 2018. A TAD boundary is preserved upon deletion of the CTCF-rich *Firre* locus. *Nature Commun.* **9**: 1–11.

Basenko EY, Sasaki T, Ji L, Prybol CJ, Burckhardt RM, Schmitz RJ, Lewis ZA. 2015. Genome-wide redistribution of H3K27me3 is linked to genotoxic stress and defective growth. *Proc Natl Acad Sci USA.* **112**: E6339-48.

Bastié N, Chapard C, Dauban L, Gadal O, Beckouët F, Koszul R. 2022. Smc3 acetylation, Pds5 and Scc2 control the translocase activity that establishes cohesin-dependent chromatin loops. *Nat Struct Mol Biol.* **29**: 575–585.

Batada NN, Hurst LD. 2007. Evolution of chromosome organization driven by selection for reduced gene expression noise. *Nature Genet.* **39**: 945–949.

Bauer BW, Davidson IF, Canena D, Wutz G, Tang W, Litos G, Horn S, Hinterdorfer P, Peters J-M. 2021. Cohesin mediates DNA loop extrusion by a "swing and clamp" mechanism. *Cell* **184**: 5448-5464.e22.

Baumann P, Cech TR. 2001. Pot1, the putative telomere end-binding protein in fission yeast and humans. *Science* **292**: 1171–1175.

Beagan JA, Phillips-Cremins JE. 2020. On the existence and functionality of topologically associating domains. *Nat Genet.* **52**: 8–16.

Bedford T, Hartl DL. 2009. Optimization of gene expression by natural selection. *Proc Natl Acad Sci USA* **106**: 1133–1138.

Belton J-M, Lajoie BR, Audibert S, Cantaloube S, Lassadi I, Goiffon I, Baù D, Marti-Renom MA, Bystricky K, Dekker J. 2015. The Conformation of yeast chromosome III is mating type dependent and controlled by the recombination enhancer. *Cell Rep.* **13**: 1855–1867.

Bennetzen JL, Wang H. 2014. The contributions of transposable elements to the structure, function, and evolution of plant genomes. *Annu Rev Plant Biol.* **65**: 505–530.

Bergthorsson U, Andersson DI, Roth JR. 2007. Ohno's dilemma: evolution of new genes under continuous selection. *Proc Natl Acad Sci USA* **104**: 17004–17009.

Bernardi G. 2015. Genome organization and chromosome architecture. *Cold Spring Harb Symp Quant Biol.* **80**: 83–91.

Bertels F, Silander OK, Pachkov M, Rainey PB, van Nimwegen E. 2014. Automated reconstruction of whole-genome phylogenies from short-sequence reads. *Mol Biol Evol* **31**: 1077–1088.

Berthelot C, Muffato M, Abecassis J, Roest Crollius H. 2015. The 3D organization of chromatin explains evolutionary fragile genomic regions. *Cell Rep.* **10**: 1913–1924.

Berthelot C, Villar D, Horvath JE, Odom DT, Flicek P. 2018. Complexity and conservation of regulatory landscapes underlie evolutionary resilience of mammalian gene expression. *Nat Ecol Evol* .**2**: 152–163.

Bewick AJ, Hofmeister BT, Powers RA, Mondo SJ, Grigoriev IV, James TY, Stajich JE, Schmitz RJ. 2019. Diversity of cytosine methylation across the fungal tree of life. *Nat Ecol Evol.* **3**: 479–490.

Bicocca VT, Ormsby T, Adhvaryu KK, Honda S, Selker EU. 2018. ASH1-catalyzed H3K36 methylation drives gene repression and marks H3K27me2/3-competent chromatin. *Elife* **7**. e.41497.

Biggins S. 2013. The composition, functions, and regulation of the budding yeast kinetochore. *Genetics* **194**: 817–846.

Bilaud T, Koering CE, Binet-Brasselet E, Ancelin K, Pollice A, Gasser SM, Gilson E. 1996. The telobox, a Myb-related telomeric DNA binding motif found in proteins from yeast, plants and human. *Nucleic Acids Res.* **24**: 1294–1303.

Bintu B, Mateo LJ, Su J-H, Sinnott-Armstrong NA, Parker M, Kinrot S, Yamaya K, Boettiger AN, Zhuang X. 2018. Super-resolution chromatin tracing reveals domains and cooperative interactions in single cells. *Science* **362:** eaau1783.

Blanco E, González-Ramírez M, Alcaine-Colet A, Aranda S, Di Croce L. 2020. The bivalent genome: characterization, structure, and regulation. *Trends Genet.* **36**: 118–131.

Boddy L, Hiscox J. 2016. Fungal Ecology: Principles and Mechanisms of Colonization and Competition by Saprotrophic Fungi. *Microbiol Spectr* **4**: 4–6.
Bolger AM, Lohse M, Usadel B. 2014. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* **30**: 2114–2120.
Bonett DG. 2006. Confidence interval for a coefficient of quartile variation. *Comput Stat Data Anal.* **50**: 2953–2957.
Bonev B, Cavalli G. 2016. Organization and function of the 3D genome. *Nat Rev Genet.* **17**: 661–678.
Bonfante P, Anca I-A. 2009. Plants, mycorrhizal fungi, and bacteria: a network of interactions. *Annu Rev Microbiol.* **63**: 363–383.
Botstein D, Fink GR. 2011. Yeast: an experimental organism for 21st Century biology. *Genetics* **189**: 695–704.
Boulias K, Greer EL. 2022. Means, mechanisms and consequences of adenine methylation in DNA. *Nat Rev Gene.t* **23**: 411–428.
Boulton SJ, Jackson SP. 1998. Components of the *Ku*-dependent non-homologous end-joining pathway are involved in telomeric length maintenance and telomeric silencing. *EMBO J* **17**: 1819–1828.
Boulton SJ, Jackson SP. 1996. Identification of a *Saccharomyces cerevisiae* Ku80 homologue: roles in DNA double strand break rejoining and in telomeric maintenance. *Nucleic Acids Res.* **24**: 4639–4648.
Bourque G, Burns KH, Gehring M, Gorbunova V, Seluanov A, Hammell M, Imbeault M, Izsvák Z, Levin HL, Macfarlan TS, et al. 2018. Ten things you should know about transposable elements. *Genome Biol.* **19**: 199.
Bouvet GF, Jacobi V, Plourde KV, Bernier L. 2008. Stress-induced mobility of OPHIO1 and OPHIO2, DNA transposons of the Dutch elm disease fungi. *Fungal Genet Biol.* **45**: 565–578.
Bowen NJ, Jordan IK. 2002. Transposable elements and the evolution of eukaryotic complexity. *Curr Issues Mol Biol.* **4**: 65–76.
Bowyer P, Currin A, Delneri D, Fraczek MG. 2022. Telomere-to-telomere genome sequence of the model mould pathogen *Aspergillus fumigatus*. *Nat Commun.* **13**: 5394.
Brachner A, Reipert S, Foisner R, Gotzmann J. 2005. LEM2 is a novel MAN1-related inner nuclear membrane protein associated with A-type lamins. *J Cell Sci.* **118**: 5797–5810.
Brackley CA, Johnson J, Michieletto D, Morozov AN, Nicodemi M, Cook PR, Marenduzzo D. 2018. Extrusion without a motor: a new take on the loop extrusion model of genome organization. *Nucleus* **9**: 95–103.
Brázda V, Bartas M, Bowater RP. 2021. Evolution of diverse strategies for promoter regulation. *Trends Genet.* **37**: 730–744.
Briand N, Collas P. 2020. Lamina-associated domains: peripheral matters and internal affairs. *Genome Biol* **21**: 85.
Bridges CB. 1938. A revised map of the salivary gland X-chromosome of *Drosophila melanogaster*. *J Hered*. **29**: 11-13.
Brosch G, Loidl P, Graessle S. 2008. Histone modifications and chromatin dynamics: a focus on filamentous fungi. *FEMS Microbiol Re.v* **32**: 409–439.
Brown CA, Murray AW, Verstrepen KJ. 2010. Rapid expansion and functional divergence of subtelomeric gene families in yeasts. *Curr Biol.* **20**: 895–903.
Brown EJ, Bachtrog D. 2014. The chromatin landscape of *Drosophila*: comparisons between species, sexes, and chromosomes. *Genome Res.* **24**: 1125–1137.
Bryan E, Warburton M, Webb KM, McLaughlin KA, Spanos C, Ambrosi C, Major V, Baubec T, Rappsilber J, Voigt P. 2021. Nucleosomal asymmetry shapes histone mark binding and promotes poising at bivalent domains. *bioRxiv*. 2021-02.
Buchwalter A, Kaneshiro JM, Hetzer MW. 2019. Coaching from the sidelines: the nuclear periphery in genome regulation. *Nat Rev Genet.* **20**: 39–50.
Buhtz A, Hohe A, Schwarz D, Grosch R. 2017. Effects of *Verticillium dahliae* on tomato root morphology considering plant growth response and defence. *Plant Pathol* **66**: 667–676.
Bupp JM, Martin AE, Stensrud ES, Jaspersen SL. 2007. Telomere anchoring at the nuclear periphery requires the budding yeast Sad1-UNC-84 domain protein Mps3. *J Cell Biol* **179**: 845–854.
Burdukiewicz M, Sidorczuk K, Rafacz D, Pietluch F, Chilimoniuk J, Rödiger S, Gagat P. 2020. Proteomic screening for prediction and design of antimicrobial peptides with AmpGram. *Int J Mol Sci* **21**: 4310.
Burrack LS, Hutton HF, Matter KJ, Clancey SA, Liachko I, Plemmons AE, Saha A, Power EA, Turman B, Thevandavakkam MA, et al. 2016. Neocentromeres provide chromosome segregation accuracy and centromere clustering to multiple loci along a *Candida albicans* chromosome. *PLoS Genet* **12**: e1006317.
Bystricky K, Heun P, Gehlen L, Langowski J, Gasser SM. 2004. Long-range compaction and flexibility of interphase chromatin in budding yeast analyzed by high-resolution imaging techniques. *Proc Natl Acad Sci USA* **101**: 16495–16500.
Camacho C, Coulouris G, Avagyan V, Ma N, Papadopoulos J, Bealer K, Madden TL. 2009. BLAST+: architecture and applications. *BMC Bioinformatics* **10**: 421.

Cambareri EB, Jensen BC, Schabtach E, Selker EU. 1989. Repeat-induced G-C to A-T mutations in Neurospora. *Science* **244**: 1571.

Cambareri EB, Singer MJ, Selker EU. 1991. Recurrence of repeat-induced point mutation (RIP) in *Neurospora crassa*. *Genetics* **127**: 699–710.

Çamdere G, Guacci V, Stricklin J, Koshland D. 2015. The ATPases of cohesin interface with regulators to modulate cohesin-mediated DNA tethering. *Elife* **4**: e.11315.

Cameron DL, Di Stefano L, Papenfuss AT. 2019. Comprehensive evaluation and characterisation of short read general-purpose structural variant calling software. *Nat Commun.* **10**: 3240.

Campos EI, Reinberg D. 2009. Histones: annotating chromatin. *Annu Rev Genet* **43**: 559–599.

Capella-Gutiérrez S, Silla-Martínez JM, Gabaldón T. trimAl: trimming in large-scale phylogenetics analyses.

Carlier F, Li M, Maroc L, Debuchy R, Souaid C, Noordermeer D, Grognet P, Malagnac F. 2021. Loss of EZH2-like or SU(VAR)3–9-like proteins causes simultaneous perturbations in H3K27 and H3K9 tri-methylation and associated developmental defects in the fungus *Podospora anserina*. *Epigenetics Chromatin* **14**: 22.

Carlton PM. 2008. Three-dimensional structured illumination microscopy and its application to chromosome structure. *Chromosome Res* **16**: 351–365.

Carr M, Bensasson D, Bergman CM. 2012. Evolutionary genomics of transposable elements in *Saccharomyces cerevisiae*. *PLoS One* **7**: e50978.

Carr PD, Tuckwell D, Hey PM, Simon L, d'Enfert C, Birch M, Oliver JD, Bromley MJ. 2010. The transposon impala is activated by low temperatures: use of a controlled transposition system to identify genes critical for viability of *Aspergillus fumigatus*. *Eukaryot Cell.* **9**: 438–448.

Casas-Vila N, Scheibe M, Freiwald A, Kappei D, Butter F. 2015. Identification of TTAGGG-binding proteins in *Neurospora crassa*, a fungus with vertebrate-like telomere repeats. *BMC Genomics* **16**: 965.

Castanera R, Lopez-Varas L, Borgognone A, LaButti K, Lapidus A, Schmutz J, Grimwood J, Perez G, Pisabarro AG, Grigoriev IV, et al. 2016. Transposable elements *versus* the fungal genome: impact on whole-genome architecture and transcriptional profiles. *PLoS Genet* **12**: e1006108.

Cavalheiro GR, Pollex T, Furlong EE. 2021. To loop or not to loop: what is the role of TADs in enhancer function and gene regulation? *Curr Opin Genet Dev.* **67**: 119–129.

Cavalli G, Misteli T. 2013. Functional implications of genome topology. *Nat Struct Mol Biol* **20**: 290–299.

Cemel IA, Ha N, Schermann G, Yonekawa S, Brunner M. 2017. The coding and noncoding transcriptome of *Neurospora crassa*. *BMC Genomics* **18**: 978.

Červenák F, Juríková K, Sepšiová R, Neboháčová M, Nosek J, Tomáška L. 2017. Double-stranded telomeric DNA binding proteins: Diversity matters. *Cell Cycle* **16**: 1568–1577.

Červenák F, Sepšiová R, Nosek J, Tomáška Ľ. 2021. Step-by-Step evolution of telomeres: lessons from yeasts. *Genome Biol Evol* **13**: evaa268.

Cha J, Zhou M, Liu Y. 2013. CATP is a critical component of the Neurospora circadian clock by regulating the nucleosome occupancy rhythm at the frequency locus. *EMBO Rep* **14**: 923–930.

Chamary JV, Parmley JL, Hurst LD. 2006. Hearing silence: non-neutral evolution at synonymous sites in mammals. *Nat Rev Genet.* **7**: 98–108.

Chan JE, Kolodner RD. 2011. A genetic and structural study of genome rearrangements mediated by high copy repeat T*y1* elements. *PLoS Genet* **7**: e1002089.

Chan K-L, Roig MB, Hu B, Beckouët F, Metson J, Nasmyth K. 2012. Cohesin's DNA exit gate is distinct from its entrance gate and is regulated by acetylation. *Cell* **150**: 961–974.

Chang L-H, Ghosh S, Noordermeer D. 2020. TADs and their borders: free movement or building a wall? *J Mol Biol* **432**: 643–652.

Chang P-K, Ehrlich KC. 2010. What does genetic diversity of *Aspergillus flavus* tell us about *Aspergillus oryzae*? *Int J Food Microbiol* **138**: 189–199.

Charlesworth B, Jensen JD. 2023. Population genetic considerations regarding evidence for biased mutation rates in *Arabidopsis thaliana*. *Mol Biol Evol* **40**: msac275.

Chavarro-Carrero EA, Vermeulen JP, E Torres D, Usami T, Schouten HJ, Bai Y, Seidl MF, Thomma BPHJ. 2021. Comparative genomics reveals the in planta-secreted *Verticillium dahliae* Av2 effector protein recognized in tomato plants that carry the V2 resistance locus. *Environ Microbiol* **23**: 1941–1958.

Cheatham MR, Rouse MN, Esker PD, Ignacio S, Pradel W, Raymundo R, Sparks AH, Forbes GA, Gordon TR, Garrett KA. 2009. Beyond yield: plant disease in the context of ecosystem services. *Phytopathology* **99**: 1228–1236.

Chen J, Swofford R, Johnson J, Cummings BB, Rogel N, Lindblad-Toh K, Haerty W, Palma F di, Regev A. 2019. A quantitative framework for characterizing the evolutionary history of mammalian gene expression. *Genome Res* **29**: 53–63.

Chen J-Y, Klosterman SJ, Hu X-P, Dai X-F, Subbarao KV. 2021a. Key Insights and research prospects at the dawn of the population genomics era for *Verticillium dahliae*. *Annu Rev Phytopathol* **59**: 31–51.
Chen X, Liu H, Chen X, Huang J, Hsiang T, Zheng L. 2021b. ATAC-Seq data for genome-wide profiling of transcription factor binding sites in the rice false smut fungus *Ustilaginoidea virens*. *Mol Plant Microbe Interact.* **34**: 830–834.
Chen X, Schulz-Trieglaff O, Shaw R, Barnes B, Schlesinger F, Källberg M, Cox AJ, Kruglyak S, Saunders CT. 2016. Manta: rapid detection of structural variants and indels for germline and cancer sequencing applications. *Bioinformatics* **32**: 1220–1222.
Chen Y, Ye W, Zhang Y, Xu Y. 2015. High speed BLASTN: an accelerated MegaBLAST search tool. *Nucleic Acids Res* **43**: 7762–7768.
Chen Z, Ponts N. 2020. H2A.Z and chromatin remodelling complexes: a focus on fungi. *Crit Rev Microbiol* **46**: 321–337.
Chiara M, Fanelli F, Mulè G, Logrieco AF, Pesole G, Leslie JF, Horner DS, Toomajian C. 2015. Genome sequencing of multiple isolates highlights subtelomeric genomic diversity within *Fusarium fujikuroi*. *Genome Biol Evol* **7**: 3062–3069.
Chikashige Y, Hiraoka Y. 2001. Telomere binding of the Rap1 protein is required for meiosis in fission yeast. *Curr Biol* **11**: 1618–1623.
Choi JY, Lee YCG. 2020. Double-edged sword: The evolutionary consequences of the epigenetic silencing of transposable elements. *PLoS Genet* **16**: e1008872.
Choudhary MN, Friedman RZ, Wang JT, Jang HS, Zhuo X, Wang T. 2020. Co-opted transposons help perpetuate conserved higher-order chromosomal structures. *Genome Biol* **21**: 16.
Chujo T, Lukito Y, Eaton CJ, Dupont PY, Johnson LJ, Winter D, Cox MP, Scott B. 2019. Complex epigenetic regulation of alkaloid biosynthesis and host interaction by heterochromatin protein I in a fungal endophyte-plant symbiosis. *Fungal Genet Biol.* **125**: 71–83.
Chujo T, Scott B. 2014. Histone H3K9 and H3K27 methylation regulates fungal alkaloid biosynthesis in a fungal endophyte-plant symbiosis. *Mol Microbiol* **92**: 413–434.
Chuong EB, Elde NC, Feschotte C. 2017. Regulatory activities of transposable elements: from conflicts to benefits. *Nat Rev Genet.* **18**: 71–86.
Chuong EB, Elde NC, Feschotte C. 2016. Regulatory evolution of innate immunity through co-option of endogenous retroviruses. *Science* **351**: 1083–1087.
Cingolani P, Platts A, Wang LL, Coon M, Nguyen T, Wang L, Land SJ, Lu X, Ruden DM. 2012. A program for annotating and predicting the effects of single nucleotide polymorphisms, SnpEff: SNPs in the genome of *Drosophila melanogaster* strain w1118; iso-2; iso-3. *Fly* **6**: 80–92.
Ciosk R, Shirayama M, Shevchenko A, Tanaka T, Toth A, Shevchenko A, Nasmyth K. 2000. Cohesin's binding to chromosomes depends on a separate complex consisting of Scc2 and Scc4 proteins. *Mol Cell* **5**: 243–254.
Clutterbuck AJ. 2011. Genomic evidence of repeat-induced point mutation (RIP) in filamentous ascomycetes. *Fungal Genet Biol* **48**: 306–326.
Colabardini AC, Wang F, Miao Z, Pardeshi L, Valero C, de Castro PA, Akiyama DY, Tan K, Nora LC, Silva-Rocha R, et al. 2022. Chromatin profiling reveals heterogeneity in clinical isolates of the human pathogen *Aspergillus fumigatus. PLoS Genet* **18**: e1010001.
Collemare J, O'Connell R, Lebrun M-H. 2019. Nonproteinaceous effectors: the terra incognita of plant-fungal interactions. *New Phytol* **223**: 590–596.
Collemare J, Seidl MF. 2019. Chromatin-dependent regulation of secondary metabolite biosynthesis in fungi: is the picture complete? *FEMS Microbiol Rev* **43**: 591–607.
Comoglio F, Schlumpf T, Schmid V, Rohs R, Beisel C, Paro R. 2015. High-resolution profiling of Drosophila replication start sites reveals a DNA shape and chromatin signature of metazoan origins. *Cell Rep* **11**: 821–834.
Connolly LR, Smith KM, Freitag M. 2013. The *Fusarium graminearum* histone H3 K27 methyltransferase KMT6 regulates development and expression of secondary metabolite gene clusters. *PLoS Genet* **9**: e1003916.
Conrad DF, Hurles ME. 2007. The population genetics of structural variation. *Nat Genet* **39**: S30-6.
Cook DE, Kramer HM, Torres DE, Seidl MF, Thomma BPHJ. 2020. A unique chromatin profile defines adaptive genomic regions in a fungal plant pathogen. *Elife* **9**: e.62208.
Cook DE, Mesarich CH, Thomma BPHJ. 2015. Understanding plant immunity as a surveillance system to detect invasion. *Annu Rev Phytopathol* **53**: 541–563.
Cook LM, Saccheri IJ. 2013. The peppered moth and industrial melanism: evolution of a natural selection case study. *Heredity* **110**: 207–212.
Cook PR. 1999. The organization of replication and transcription. *Science* **284**: 1790–1795.

Cooper JP, Nimmo ER, Allshire RC, Cech TR. 1997. Regulation of telomere length and function by a Myb-domain protein in fission yeast. *Nature* **385**: 744–747.
Corbo M, Damas J, Bursell MG, Lewin HA. 2022. Conservation of chromatin conformation in carnivores. *Proc Natl Acad Sci USA* **119**: e.2120555119.
Cosby RL, Judd J, Zhang R, Zhong A, Garry N, Pritham EJ, Feschotte C. 2021. Recurrent evolution of vertebrate transcription factors by transposase capture. *Science* **371**: eabc6405.
Cournac A, Koszul R, Mozziconacci J. 2016. The 3D folding of metazoan genomes correlates with the association of similar repetitive elements. *Nucleic Acids Res* **44**: 245–255.
Courtney AJ, Ferraro AR, Klocko AD, Lewis ZA. 2020. 1 Chromatin Structure and Function in *Neurospora crassa*. In *Genetics and Biotechnology*, pp. 3–24, Springer International Publishing, Cham.
Cremer T, Cremer C. 2001. Chromosome territories, nuclear architecture and gene regulation in mammalian cells. *Nat Rev Gene.t* **2**: 292–301.
Cremer T, Cremer C. 2006. Rise, fall and resurrection of chromosome territories: a historical perspective. Part II. Fall and resurrection of chromosome territories during the 1950s to 1980s. Part III. Chromosome territories and the functional nuclear architecture: experiments and models from the 1990s to the present. *Eur J Histochem* **50**: 223–272.
Cremer T, Cremer M. 2010. Chromosome territories. *Cold Spring Harb Perspect Biol* **2**: a003889.
Croll D, McDonald BA. 2012. The accessory genome as a cradle for adaptive evolution in pathogens. *PLoS Pathog* **8**: e1002608.
Crombach A, Hogeweg P. 2007. Chromosome rearrangements and the evolution of genome structuring and adaptability. *Mol Biol Evol* **24**: 1130–1139.
Crombach A, Hogeweg P. 2008. Evolution of evolvability in gene regulatory networks. *PLoS Comput Bio.l* **4**: e1000112.
Crombach A, Hogeweg P. 2011. Is RNA-dependent RNA polymerase essential for transposon control? *BMC Syst Biol* **5**: 104.
Cruz-Mireles N, Eisermann I, Garduño-Rosales M, Molinari C, Ryder LS, Tang B, Yan X, Talbot NJ. 2021. The biology of invasive growth by the rice blast fungus *Magnaporthe oryzae*. *Methods Mol Biol* **2356**: 19–40.
Cuypers TD, Hogeweg P. 2012. Virtual genomes in flux: An interplay of neutrality and adaptability explains genome expansion and streamlining. *Genome Biol Evol* **4**: 212–229.
Cuypers TD, Rutten JP, Hogeweg P. 2017. Evolution of evolvability and phenotypic plasticity in virtual cells. *BMC Evol Biol* **17**: 1-16.
Cvijovic I, Good BH, Jerison ER, Desai MM. 2015. Fate of a mutation in a fluctuating environment. *Proc Natl Acad Sci USA* **112**: E5021-8.
Dai Z, Ramesh V, Locasale JW. 2020. The evolving metabolic landscape of chromatin biology and epigenetics. *Nat Rev Genet* **21**: 737–753.
Dauban L, Montagne R, Thierry A, Lazar-Stefanita L, Bastié N, Gadal O, Cournac A, Koszul R, Beckouët F. 2020. Regulation of Cohesin-Mediated Chromosome Folding by Eco1 and Other Partners. *Mol Cell* **77**: 1279-1293. e4.
Davidson IF, Peters J-M. 2021. Genome folding through loop extrusion by SMC complexes. *Nat Rev Mol Cell Biol* **22**: 445–464.
de Jonge R, Bolton MD, Kombrink A, van den Berg GCM, Yadeta KA, Thomma BPHJ. 2013. Extensive chromosomal reshuffling drives evolution of virulence in an asexual pathogen. *Genome Res* **23**: 1271–1282.
de Jonge R, van Esse HP, Maruthachalam K, Bolton MD, Santhanam P, Saber MK, Zhang Z, Usami T, Lievens B, Subbarao KV, et al. 2012. Tomato immune receptor Ve1 recognizes effector of multiple fungal pathogens uncovered by genome and RNA sequencing. *Proc Natl Acad Sci USA* **109**: 5110–5115.
de la Peña MV, Summanen PAM, Liukkonen M, Kronholm I. 2023. Chromatin structure influences rate and spectrum of spontaneous mutations in *Neurospora crassa*. *Genome Res* **33**: 599-611
de Lange T. 2005. Shelterin: the protein complex that shapes and safeguards human telomeres. *Genes Dev* **19**: 2100–2110.
de Wit E, Vos ESM, Holwerda SJB, Valdes-Quezada C, Verstegen MJAM, Teunissen H, Splinter E, Wijchers PJ, Krijger PHL, de Laat W. 2015. CTCF binding polarity determines chromatin looping. *Mol Cell* **60**: 676–684.
Dekker J, Heard E. 2015. Structural and functional diversity of Topologically Associating Domains. *FEBS Lett* **589**: 2877–2884.
Dekker J, Marti-Renom MA, Mirny LA. 2013. Exploring the three-dimensional organization of genomes: interpreting chromatin interaction data. *Nature Rev Genet* **14**: 390–403.
Dekker J, Misteli T. 2015. Long-range chromatin interactions. *Cold Spring Harbor perspectives in biology* **7**: a019356.
Dekker J, Rippe K, Dekker M, Kleckner N. 2002. Capturing chromosome conformation. *Science* **295**: 1306–1311.

Deng W, Wu J, Wang F, Kanoh J, Dehe P-M, Inoue H, Chen J, Lei M. 2015. Fission yeast telomere-binding protein Taz1 is a functional but not a structural counterpart of human TRF1 and TRF2. *Cell Res* **25**: 881–884.

Depotter JRL, Shi-Kunne X, Missonnier H, Liu T, Faino L, van den Berg GCM, Wood TA, Zhang B, Jacques A, Seidl MF, et al. 2019. Dynamic virulence-related regions of the plant pathogenic fungus *Verticillium dahliae* display enhanced sequence conservation. *Mol Ecol* **28**: 3482–3495.

Depotter JRL, van Beveren F, Rodriguez-Moreno L, Kramer HM, Chavarro Carrero EA, van den Berg GCM, Wood TA, Thomma BPHJ, Seidl MF. 2021. The interspecific fungal hybrid *Verticillium longisporum* displays subgenome-specific gene expression. *mBio* **12**: e0149621.

Deschamps J. 2016. Birth and upgrowth of the Hox topological domains during evolution. *Nat Genet* **48**: 227–228.

Deutekom ES, Vosseberg J, van Dam TJP, Snel B. 2019. Measuring the impact of gene prediction on gene loss estimates in Eukaryotes by quantifying falsely inferred absences. *PLoS Comput Biol* **15**: e1007301.

Devos KM, Brown JK, Bennetzen JL. 2002. Genome size reduction through illegitimate recombination counteracts genome expansion in Arabidopsis. *Genome Res.* **12**: 1075–1079.

Dietrich FS, Voegeli S, Kuo S, Philippsen P. 2013. Genomes of *Ashbya* fungi isolated from insects reveal four mating-type loci, numerous translocations, lack of transposons, and distinct gene duplications. *G3 (Bethesda)* **3**: 1225–1239.

Dileep V, Rivera-Mulia JC, Sima J, Gilbert DM. 2015. Large-scale chromatin structure–function relationships during the cell cycle and development: insights from replication timing. *Cold Spring Harb Symp Quant Biol* **80**: 53–63.

Dixon JR, Selvaraj S, Yue F, Kim A, Li Y, Shen Y, Hu M, Liu JS, Ren B. 2012. Topological domains in mammalian genomes identified by analysis of chromatin interactions. *Nature* **485**: 376–380.

Dixon JR, Xu J, Dileep V, Zhan Y, Song F, Le VT, Yardımcı GG, Chakraborty A, Bann DV, Wang Y, et al. 2018. Integrative detection and analysis of structural variation in cancer genomes. *Nat Genet.* **50**: 1388–1398.

Dobin A, Davis CA, Schlesinger F, Drenkow J, Zaleski C, Jha S, Batut P, Chaisson M, Gingeras TR. 2013. STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* **29**: 15–21.

Dong P, Tu X, Chu P-Y, Lü P, Zhu N, Grierson D, Du B, Li P, Zhong S. 2017. 3D Chromatin Architecture of Large Plant Genomes Determined by Local A/B Compartments. *Mol Plant* **10**: 1497–1509.

Dong Q, Wang Y, Qi S, Gai K, He Q, Wang Y. 2018. Histone variant H2A.Z antagonizes the positive effect of the transcriptional activator CPC1 to regulate catalase-3 expression under normal and oxidative stress conditions. *Redox Biol.* **121**: 136–148.

Dong S, Raffaele S, Kamoun S. 2015. The two-speed genomes of filamentous pathogens: waltz with plants. *Curr Opin Genet Dev* **35**: 57–65.

Donnart T, Piednoel M, Higuet D, Bonnivard E. 2017. Filamentous ascomycete genomes provide insights into Copia retrotransposon diversity in fungi. *BMC Genomics* **18**: 1-15.

Dorsett D. 2007. Roles of the sister chromatid cohesion apparatus in gene expression, development, and human syndromes. *Chromosoma* **116**: 1–13.

Draghi J, Wagner GP. 2008. Evolution of evolvability in a developmental model. *Evolution* **62**: 301–315.

Drummond DA, Bloom JD, Adami C, Wilke CO, Arnold FH. 2005a. Why highly expressed proteins evolve slowly. *Proc Natl Acad Sci USA* **102**: 14338–14343.

Drummond DA, Raval A, Wilke CO. 2005b. A single determinant dominates the rate of yeast protein evolution. *Mol Biol Evol* **23**: 327–337.

Duan H, Jones AW, Hewitt T, Mackenzie A, Hu Y, Sharp A, Lewis D, Mago R, Upadhyaya NM, Rathjen JP, et al. 2022. Physical separation of haplotypes in dikaryons allows benchmarking of phasing accuracy in Nanopore and HiFi assemblies with Hi-C data. *Genome Biol* **23**: 84.

Duan Z, Andronescu M, Schutz K, McIlwain S, Kim YJ, Lee C, Shendure J, Fields S, Blau CA, Noble WS. 2010. A three-dimensional model of the yeast genome. *Nature* **465**: 363–367.

Duchaud E, Rusniok C, Frangeul L, Buchrieser C, Givaudan A, Taourit S, Bocs S, Boursaux-Eude C, Chandler M, Charles JF, et al. 2003. The genome sequence of the entomopathogenic bacterium *Photorhabdus luminescens*. *Nature Biotechnology* **21**: 1307–1313.

Dumesic PA, Homer CM, Moresco JJ, Pack LR, Shanle EK, Coyle SM, Strahl BD, Fujimori DG, Yates JR, Madhani HD. 2015. Product Binding Enforces the Genomic Specificity of a Yeast Polycomb Repressive Complex. *Cell* **160**: 204–218.

Dutheil JY, Mannhaupt G, Schweizer G. 2016. A tale of genome compartmentalization: the evolution of virulence clusters in smut fungi. *Genome Biol*. **8**: 681-704

Eagen KP, Aiden EL, Kornberg RD. 2017. Polycomb-mediated chromatin loops revealed by a subkilobase-resolution chromatin interaction map. *Proc Natl Acad Sci USA* **114**: 8764–8769.

Ebert G, Steininger A, Weißmann R, Boldt V, Lind-Thomsen A, Grune J, Badelt S, Heßler M, Peiser M, Hitzler M, et al. 2014. Distribution of segmental duplications in the context of higher order chromatin organisation of human chromosome 7. *BMC Genomics* **15**: 537.

Ebrahimi H, Masuda H, Jain D, Cooper JP. 2018. Distinct "safe zones" at the nuclear envelope ensure robust replication of heterochromatic chromosome regions. *Elife* **7**: e.32911.

Eddy SR. 2011. Accelerated Profile HMM Searches. *PLoS Comput Bio.l* **7**: e1002195.

Edwards RJ, Dong C, Park RF, Tobias PA. 2022. A phased chromosome-level genome and full mitochondrial sequence for the dikaryotic myrtle rust pathogen, *Austropuccinia psidii. bioRxiv*, 2022.04.

Eichler EE, Sankoff D. 2003. Structural dynamics of eukaryotic chromosome evolution. *Science* **301**: 793–797.

Ellinghaus D, Kurtz S, Willhoeft U. 2008. LTRharvest, an efficient and flexible software for de novo detection of LTR retrotransposons. *BMC Bioinformatics* **9**: 18.

Ene IV, Farrer RA, Hirakawa MP, Agwamba K, Cuomo CA, Bennett RJ. 2018. Global analysis of mutations driving microevolution of a heterozygous diploid fungal pathogen. *Proc Natl Acad Sci USA* **115**: E8688–E8697.

Engel SR, Dietrich FS, Fisk DG, Binkley G, Balakrishnan R, Costanzo MC, Dwight SS, Hitz BC, Karra K, Nash RS, et al. 2014. The reference genome sequence of *Saccharomyces cerevisiae*: then and now. *G3(Bethesda)* **4**: 389–398.

Erlendson, Friedman Steven, Freitag Michael. 2017. A matter of scale and dimensions: chromatin of chromosome landmarks in the fungi. *Microbiol Spectr* **5**: 5.4.11.

Ernst J, Kellis M. 2017. Chromatin-state discovery and genome annotation with ChromHMM. *Nat Protoc* **12**: 2478–2492.

Ernst J, Kellis M. 2012. ChromHMM: automating chromatin-state discovery and characterization. *Nat Methods* **9**: 215–216.

Escaramís G, Docampo E, Rabionet R. 2015. A decade of structural variants: description, history and methods to detect structural variation. *Brief Funct Genomics* **14**: 305–314.

Eser U, Chandler-Brown D, Ay F, Straight AF, Duan Z, Noble WS, Skotheim JM. 2017. Form and function of topologically associating genomic domains in budding yeast. *Proc Natl Acad Sci USA* **114**: E3061–E3070.

Fablet M, Vieira C. 2011. Evolvability, epigenetics and transposable elements. *BioMol Concepts* **2**: 333–341.

Faino L, Seidl MF, Datema E, van den Berg GCM, Janssen A, Wittenberg AHJ, Thomma BPHJ. 2015. Single-molecule real-time sequencing combined with optical mapping yields completely finished fungal genome. *MBio* **6**: e.00936-15.

Faino L, Seidl MF, Shi-Kunne X, Pauper M, van den Berg GCM, Wittenberg AHJ, Thomma BPHJ. 2016. Transposons passively and actively contribute to evolution of the two-speed genome of a fungal pathogen. *Genome Res* **26**: 1091–1100.

Falk M, Feodorova Y, Naumova N, Imakaev M, Lajoie BR, Leonhardt H, Joffe B, Dekker J, Fudenberg G, Solovei I. et al. 2019. Heterochromatin drives compartmentalization of inverted and conventional nuclei. *Nature* **570**: 395–399.

Farman ML. 2007. Telomeres in the rice blast fungus *Magnaporthe oryzae*: the world of the end as we know it. *FEMS Microbiol Lett.* **273**: 125–132.

Farman ML, Kim Y-S. 2005. Telomere hypervariability in Magnaporthe oryzae. *Mol Plant Pathol* **6**: 287–298.

Fay JC, Wittkopp PJ. 2008. Evaluating the role of natural selection in the evolution of gene regulation. *Heredity* **100**: 191–199.

Feng G, Leem Y-E, Levin HL. 2012. Transposon integration enhances expression of stress response genes. *Nucleic Acids Res* **41**: 775–789.

Ferraro AR, Ameri AJ, Lu Z, Kamei M, Schmitz RJ, Lewis ZA. 2021. Chromatin accessibility profiling in *Neurospora crassa* reveals molecular features associated with accessible and inaccessible chromatin. *BMC Genomics* **22**: 459.

Feschotte C. 2008. Transposable elements and the evolution of regulatory networks. *Nat Rev Genet.* **9**: 397–405.

Feulner PGD, De-Kayne R. 2017. Genome evolution, structural rearrangements and speciation. *J Evol Biol* **30**: 1488–1490.

Feurtey A, Lorrain C, Croll D, Eschenbrenner C, Freitag M, Habig M, Haueisen J, Möller M, Schotanus K, Stukenbrock EH. 2020. Genome compartmentalization predates species divergence in the plant pathogen genus Zymoseptoria. *BMC Genomics* **21**: 588.

Fiddes IT, Armstrong J, Diekhans M, Nachtweide S, Kronenberg ZN, Underwood JG, Gordon D, Earl D, Keane T, Eichler EE, et al. 2018. Comparative Annotation Toolkit (CAT)—simultaneous clade and personal genome annotation. *Genome Res* **28**: 1029–1038.

Filleton F, Chuffart F, Nagarajan M, Bottin-Duplus H, Yvert G. 2015. The complex pattern of epigenomic variation between natural yeast strains at single-nucleosome resolution. *Epigenetics Chromatin* **8**: 26.

Finn EH, Pegoraro G, Brandao HB, Valton A-L, Oomen ME, Dekker J, Mirny L, Misteli T. 2019. Extensive heterogeneity and intrinsic variation in spatial genome organization. *Cell* **176**: 1502–1515.

Finn RD, Coggill P, Eberhardt RY, Eddy SR, Mistry J, Mitchell AL, Potter SC, Punta M, Qureshi M, Sangrador-Vegas A, et al. 2016. The Pfam protein families database: towards a more sustainable future. *Nucleic Acids Res* **44**: D279-85.

Fisher MC, Gurr SJ, Cuomo CA, Blehert DS, Jin H, Stukenbrock EH, Stajich JE, Kahmann R, Boone C, Denning DW, et al. 2020. Threats posed by the fungal kingdom to humans, wildlife, and agriculture. *MBio* **11**: e.00449-20.

Fisher MC, Henk DA, Briggs CJ, Brownstein JS, Madoff LC, McCraw SL, Gurr SJ. 2012. Emerging fungal threats to animal, plant and ecosystem health. *Nature* **484**: 186–194.

Flavahan WA, Drier Y, Liau BB, Gillespie SM, Venteicher AS, Stemmer-Rachamimov AO, Suvà ML, Bernstein BE. 2016. Insulator dysfunction and oncogene activation in IDH mutant gliomas. *Nature* **529**: 110–114.

Fleetwood DJ, Khan AK, Johnson RD, Young CA, Mittal S, Wrenn RE, Hesse U, Foster SJ, Schardl CL, Scott B. 2011. Abundant degenerate miniature inverted-repeat transposable elements in genomes of epichloid fungal endophytes of grasses. *Genome Biol Evol* **3**: 1253–1264.

Flynn JM, Hubley R, Goubert C, Rosen J, Clark AG, Feschotte C, Smit AF. 2020. RepeatModeler2 for automated genomic discovery of transposable element families. *Proc Natl Acad Sci USA* **117**: 9451–9457.

Fokkens L, Guo L, Dora S, Wang B, Ye K, Sánchez-Rodríguez C, Croll D. 2020. A chromosome-Scale genome assembly for the *Fusarium oxysporum* strain Fo5176 to establish a model Arabidopsis-fungal pathosystem. *G3 (Bethesda)* **10**: 3549–3555.

Fokkens L, Shahi S, Connolly LR, Stam R, Schmidt SM, Smith KM, Freitag M, Rep M. 2018. The multi-speed genome of *Fusarium oxysporum* reveals association of histone modifications with sequence divergence and footprints of past horizontal chromosome transfer events. *bioRxiv* 465070.

Fornes O, Castro-Mondragon JA, Khan A, Van der Lee R, Zhang X, Richmond PA, Modi BP, Correard S, Gheorghe M, Baranašić D. 2020. JASPAR 2020: update of the open-access database of transcription factor binding profiles. *Nucleic Acids Res* **48**: D87–D92.

Fouche S, Badet T, Oggenfuss U, Plissonneau C, Francisco CS, Croll D. 2020. Stress-driven transposable element de-repression dynamics and virulence evolution in a fungal pathogen. *Mol Biol Evol* **37**: 221–239.

Fouché S, Oggenfuss U, Chanclud E, Croll D. 2022. A devil's bargain with transposable elements in plant pathogens. *Trends Genet* **38**: 222–230.

Fouché S, Plissonneau C, Croll D. 2018. The birth and death of effectors in rapidly evolving filamentous pathogen genomes. *Curr Opin Microbiol* **46**: 34–42.

Fradin EF, Thomma BPHJ. 2006. Physiology and molecular aspects of Verticillium wilt diseases caused by *V. dahliae* and *V. albo-atrum*. *Mol Plant Pathol* **7**: 71–86.

Frantzeskakis L, Di Pietro A, Rep M, Schirawski J, Wu CH, Panstruga R. 2020. Rapid evolution in plant-microbe interactions - a molecular genomics perspective. *New Phytol* **225**: 1134–1142.

Frantzeskakis L, Kracher B, Kusch S, Yoshikawa-Maekawa M, Bauer S, Pedersen C, Spanu PD, Maekawa T, Schulze-Lefert P, Panstruga R. 2018. Signatures of host specialization and a recent transposable element burst in the dynamic one-speed genome of the fungal barley powdery mildew pathogen. *BMC Genomics* **19**: 381.

Frantzeskakis L, Kusch S, Panstruga R. 2019. The need for speed: compartmentalized genome evolution in filamentous phytopathogens. *Mol Plant Pathol* **20**: 3–7.

Freitag M. 2017. Histone methylation by SET domain proteins in fungi. *Annu Rev Microbiol* **71**: 413–439.

Freitag M, Hickey PC, Khlafallah TK, Read ND, Selker EU. 2004. HP1 is essential for DNA methylation in neurospora. *Mol Cell* **13**: 427–434.

Freitag M, Williams RL, Kothe GO, Selker EU. 2002. A cytosine methyltransferase homologue is essential for repeat-induced point mutation in *Neurospora crassa*. *Proc Natl Acad Sci USA* **99**: 8802–8807.

Fu L-Y, Zhu T, Zhou X, Yu R, He Z, Zhang P, Wu Z, Chen M, Kaufmann K, Chen D. 2022. ChIP-Hub provides an integrative platform for exploring plant regulome. *Nat Commun* **13**: 3413.

Fudal I, Ross S, Brun H, Besnard A-L, Ermel M, Kuhn M-L, Balesdent M-H, Rouxel T. 2009. Repeat-induced point mutation (RIP) as an alternative mechanism of evolution toward virulence in *Leptosphaeria maculans. Mol Plant Microbe Interact.* **22**: 932–941.

Fudal I, Ross S, Gout L, Blaise F, Kuhn ML, Eckert MR, Cattolico L, Bernard-Samain S, Balesdent MH, Rouxel T. 2007. Heterochromatin-like regions as ecological niches for avirulence genes in the *Leptosphaeria maculans* genome: map-based cloning of AvrLm6. *Mol Plant Microbe Interact* **20**: 459–470.

Fudenberg G, Imakaev M, Lu C, Goloborodko A, Abdennur N, Mirny LA. 2016. Formation of chromosomal domains by loop extrusion. *Cell Rep* **15**: 2038–2049.

Fudenberg G, Pollard KS. 2019. Chromatin features constrain structural variation across evolutionary timescales. *Proc Natl Acad Sci USA* **116**: 2175–2180.

Fuentes RR, Chebotarov D, Duitama J, Smith S, De la Hoz JF, Mohiyuddin M, Wing RA, McNally KL, Tatarinova T, Grigoriev A, et al. 2019. Structural variants in 3000 rice genomes. *Genome Res* **29**: 870–880.

Fujita I, Tanaka M, Kanoh J. 2012. Identification of the functional domains of the telomere protein Rap1 in *Schizosaccharomyces pombe*. *PLoS One* **7**: e49151.

Funabiki H, Hagan I, Uzawa S, Yanagida M. 1993. Cell cycle-dependent specific positioning and clustering of centromeres and telomeres in fission yeast. *J Cell Biol* **121**: 961–976.

Furlan-Magaril M, Rincón-Arano H, Recillas-Targa F. 2009. Sequential chromatin immunoprecipitation protocol: ChIP-reChIP. *Methods Mol Biol* **543**: 253–266.

Gagniuc P, Ionescu-Tirgoviste C. 2012. Eukaryotic genomes may exhibit up to 10 generic classes of gene promoters. *BMC Genomics* **13**: 512.

Galagan JE, Calvo SE, Borkovich KA, Selker EU, Read ND, Jaffe D, FitzHugh W, Ma L-J, Smirnov S, Purcell S, et al. 2003. The genome sequence of the filamentous fungus *Neurospora crassa*. *Nature* **422**: 859–868.

Galagan JE, Selker EU. 2004. RIP: The evolutionary cost of genome defense. *Trends in Genetics* **20**: 417–423.

Galazka JM, Klocko AD, Uesaka M, Honda S, Selker EU, Freitag M. 2016. Neurospora chromosomes are organized by blocks of importin alpha-dependent heterochromatin that are largely independent of H3K9me3. *Genome Res* **26**: 1069–1080.

Galhardo RS, Hastings PJ, Rosenberg SM. 2007. Mutation as a stress response and the regulation of evolvability. *Crit Rev Biochem Mol Biol* **42**: 399–435.

Galindo-Gonzalez L, Mhiri C, Deyholos MK, Grandbastien MA. 2017. LTR-retrotransposons in plants: Engines of evolution. *Gene* **626**: 14–25.

Gan P, Hiroyama R, Tsushima A, Masuda S, Shibata A, Ueno A, Kumakura N, Narusaka M, Hoat TX, Narusaka Y, et al. 2021. Telomeres and a repeat-rich chromosome encode effector gene clusters in plant pathogenic Colletotrichum fungi. *Environ Microbiol* **23**: 6004–6018.

Ganji M, Shaltiel IA, Bisht S, Kim E, Kalichava A, Haering CH, Dekker C. 2018. Real-time imaging of DNA loop extrusion by condensin. *Science* **360**: 102–105.

Garcia-Luis J, Lazar-Stefanita L, Gutierrez-Escribano P, Thierry A, Cournac A, García A, González S, Sánchez M, Jarmuz A, Montoya A, et al. 2019. FACT mediates cohesin function on chromatin. *Nat Struct Mol Biol* **26**: 970–979.

Garnica DP, Nemri A, Upadhyaya NM, Rathjen JP, Dodds PN. 2014. The ins and outs of rust haustoria. *PLoS Pathog* **10**: e1004329.

Gasch AP, Moses AM, Chiang DY, Fraser HB, Berardini M, Eisen MB. 2004. Conservation and evolution of cis-regulatory systems in ascomycete fungi. *PLoS Biol* **2**: e398.

Gasser SM. 2002. Visualizing chromatin dynamics in interphase nuclei. *Science* **296**: 1412–1416.

Gel B, Díez-Villanueva A, Serra E, Buschbeck M, Peinado MA, Malinverni R. 2016. regioneR: an R/Bioconductor package for the association analysis of genomic regions based on permutation tests. *Bioinformatics* **32**: 289–291.

Gerhart J, Kirschner M. 2007. The theory of facilitated variation. *Proc Natl Acad Sci U S A* **104**: 8582–8589.

Ghavi-Helm Y, Jankowski A, Meiers S, Viales RR, Korbel JO, Furlong EEM. 2019. Highly rearranged chromosomes reveal uncoupling between genome topology and gene expression. *Nat Genet* **51**: 1272–1282.

Ghoshal B, Picard CL, Vong B, Feng S, Jacobsen SE. 2021. CRISPR-based targeting of DNA methylation in *Arabidopsis thaliana* by a bacterial CG-specific DNA methyltransferase. *Proc Natl Acad Sci U S A* **118**: e.2125016118.

Gibriel HAY, Li J, Zhu L, Seidl MF, Thomma BPH. 2019. *Verticillium dahliae* strains that infect the same host plant display highly divergent effector catalogs. *bioRxiv* 528729.

Gindin Y, Valenzuela MS, Aladjem MI, Meltzer PS, Bilke S. 2014. A chromatin structure-based model accurately predicts DNA replication timing in human cells. *Mol Syst Biol* **10**: 722.

Gladyshev E. 2017. Repeat-Induced Point Mutation and Other Genome Defense Mechanisms in Fungi. *The fungal kingdom,* 687-699.

Gligorijević V, Renfrew PD, Kosciolek T, Leman JK, Berenberg D, Vatanen T, Chandler C, Taylor BC, Fisk IM, Vlamakis H, et al. 2021. Structure-based protein function prediction using graph convolutional networks. *Nat Commun* **12**: 3168.

Goerner-Potvin P, Bourque G. 2018. Computational tools to unmask transposable elements. *Nat Rev Genet* **19**: 688–704.

Goffeau A, Barrell BG, Bussey H, Davis RW, Dujon B, Feldmann H, Galibert F, Hoheisel JD, Jacq C, Johnston M, et al. 1996. Life with 6000 genes. *Science* **274**: 546, 563–7.

Gokcumen O, Babb PL, Iskow RC, Zhu Q, Shi X, Mills RE, Ionita-Laza I, Vallender EJ, Clark AG, Johnson WE, et al. 2011. Refinement of primate copy number variationhotspots identifies candidate genomic regions evolving under positive selection. *Genome Biol.* **12**: R52.

Golfier S, Quail T, Kimura H, Brugués J. 2020. Cohesin and condensin extrude DNA loops in a cell cycle-dependent manner. *Elife* **9**: e.53885.
Golicz AA, Bhalla PL, Edwards D, Singh MB. 2020. Rice 3D chromatin structure correlates with sequence variation and meiotic recombination rate. *Commun Biol* **3**: 235.
Gómez-Pérez D, Schmid M, Chaudhry V, Velic A, Maček B, Kemen A, Kemen E. 2022. Proteins released into the plant apoplast by the obligate parasitic protist Albugo selectively repress phyllosphere-associated bacteria. *bioRxiv* 2022.05.
Gong Y, Lazaris C, Sakellaropoulos T, Lozano A, Kambadur P, Ntziachristos P, Aifantis I, Tsirigos A. 2018. Stratification of TAD boundaries reveals preferential insulation of super-enhancers by strong boundaries. *Nature Commun* **9**: 1–12.
Gonzalez-Sandoval A, Gasser SM. 2016. On TADs and LADs: spatial control over gene expression. *Trends in genetics* **32**: 485–495.
Goodwin SB, Ben M'Barek S, Dhillon B, Wittenberg AHJ, Crane CF, Hane JK, Foster AJ, Van der Lee TAJ, Grimwood J, Aerts A, et al. 2011. Finished genome of the fungal wheat pathogen *Mycosphaerella graminicola* reveals dispensome structure, chromosome plasticity, and stealth pathogenesis. *PLoS Gen.* **7**: e1002070.
Gorkin DU, Barozzi I, Zhao Y, Zhang Y, Huang H, Lee AY, Li B, Chiou J, Wildberg A, Ding B, et al. 2020. An atlas of dynamic chromatin landscapes in mouse fetal development. *Nature* **583**: 744–751.
Goto B, Okazaki K, Niwa O. 2001. Cytoplasmic microtubular system implicated in de novo formation of a Rabl-like orientation of chromosomes in fission yeast. *J Cell Sci* **114**: 2427–2435.
Goubert C, Zevallos NA, Feschotte C. 2020. Contribution of unfixed transposable element insertions to human regulatory variation. *Philos Trans R Soc Lond B Biol Sci.* **375**: 20190331.
Grandaubert J, Dutheil JY, Stukenbrock EH. 2019. The genomic determinants of adaptive evolution in a fungal pathogen. *Evol Lett* **3**: 299–312.
Grandaubert J, Lowe RGT, Soyer JL, Schoch CL, Van de Wouw AP, Fudal I, Robbertse B, Lapalu N, Links MG, Ollivier B, et al. 2014. Transposable element-assisted evolution and adaptation to host plant within the *Leptosphaeria maculans-Leptosphaeria biglobosa* species complex of fungal pathogens. *BMC Genomics* **15**: 891.
Green LC, Kalitsis P, Chang TM, Cipetic M, Kim JH, Marshall O, Turnbull L, Whitchurch CB, Vagnarelli P, Samejima K, et al. 2012. Contrasting roles of condensin I and condensin II in mitotic chromosome formation. *J Cell Sci* **125**: 1591–1604.
Gregory TR. 2009. Understanding Natural Selection: Essential Concepts and Common Misconceptions. *Evolution: Education and Outreach* **2**: 156–175.
Greider CW, Blackburn EH. 1989. A telomeric sequence in the RNA of Tetrahymena telomerase required for telomere repeat synthesis. *Nature* **337**: 331–337.
Greider CW, Blackburn EH. 1985. Identification of a specific telomere terminal transferase activity in Tetrahymena extracts. *Cell* **43**: 405–413.
Grewal SIS, Jia S. 2007. Heterochromatin revisited. *Nat Rev Genet* **8**: 35–46.
Groen SC, Ćalić I, Joly-Lopez Z, Platts AE, Choi JY, Natividad M, Dorph K, Mauck WM 3rd, Bracken B, Cabral CLU, et al. 2020. The strength and pattern of natural selection on gene expression in rice. *Nature* **578**: 572–576.
Grund SE, Fischer T, Cabal GG, Antúnez O, Pérez-Ortín JE, Hurt E. 2008. The inner nuclear membrane protein Src1 associates with subtelomeric genes and alters their regulated gene expression. *J Cell Biol* **182**: 897–910.
Gu X, Ye T, Zhang X-R, Nie L, Wang H, Li W, Lu R, Fu C, Du L-L, Zhou J-Q. 2022. Single-chromosome fission yeast models reveal the configuration robustness of a functional genome. *Cell Rep* **40**: 111237.
Gu Z, Eils R, Schlesner M, Ishaque N. 2018. EnrichedHeatmap: an R/Bioconductor package for comprehensive visualization of genomic signal associations. *BMC genomics* **19**: 1–7.
Gu Z, Gu L, Eils R, Schlesner M, Brors B. 2014. circlize Implements and enhances circular visualization in R. *Bioinformatics* **30**: 2811–2812.
Guacci V, Hogan E, Koshland D. 1994. Chromosome condensation and sister chromatid pairing in budding yeast. *J Cell Biol* **125**: 517–530.
Guelen L, Pagie L, Brasset E, Meuleman W, Faza MB, Talhout W, Eussen BH, de Klein A, Wessels L, de Laat W, et al. 2008. Domain organization of human chromosomes revealed by mapping of nuclear lamina interactions. *Nature* **453**: 948–951.
Guo Y, Levin HL. 2010. High-throughput sequencing of retrotransposon integration provides a saturated profile of target activity in *Schizosaccharomyces pombe*. *Genome Res* **20**: 239–248.
Gurung S, Short DPG, Atallah ZK, Subbarao KV. 2014. Clonal expansion of *Verticillium dahliae* in lettuce. *Phytopathology* **104**: 641–649.

Haarhuis JHI, van der Weide RH, Blomen VA, Yáñez-Cuna JO, Amendola M, van Ruiten MS, Krijger PHL, Teunissen H, Medema RH, van Steensel B, et al. 2017. The Cohesin release factor WAPL restricts chromatin loop extension. *Cell* **169**: 693-707.e14.

Haas BJ, Kamoun S, Zody MC, Jiang RHY, Handsaker RE, Cano LM, Grabherr M, Kodira CD, Raffaele S, Torto-Alalibo T, et al. 2009. Genome sequence and analysis of the Irish potato famine pathogen *Phytophthora infestans*. *Nature* **461**: 393–398.

Habig M, Lorrain C, Feurtey A, Komluski J, Stukenbrock EH. 2021. Epigenetic modifications affect the rate of spontaneous mutations in a pathogenic fungus. *Nat Commun* **12**: 5869.

Hagstrom KA, Meyer BJ. 2003. Condensin and cohesin: more than chromosome compactor and glue. *Nat Rev Genet* **4**: 520–534.

Hallgren J, Tsirigos KD, Pedersen MD. 2022. DeepTMHMM predicts alpha and beta transmembrane proteins using deep neural networks. *BioRxiv*. 2022-04.

Hallgrímsson B, Hall BK. 2011. Epigenetics: Linking Genotype and Phenotype in Development and Evolution. *University of California Press.*

Han G-Z. 2019. Origin and evolution of the plant immune system. *New Phytol.* **222**: 70–83.

Hane JK, Williams AH, Taranto AP, Solomon PS, Oliver RP. 2015. Repeat-Induced Point Mutation: A fungal-specific, endogenous mutagenesis process. In *Genetic Transformation Systems in Fungi, Volume 2* (eds. M.A. Van den Berg and K. Maruthachalam), pp. 55–68

Hansen AS, Cattoglio C, Darzacq X, Tjian R. 2018a. Recent evidence that TADs and chromatin loops are dynamic structures. *Nucleus* **9**: 20–32.

Hansen JC, Connolly M, McDonald CJ, Pan A, Pryamkova A, Ray K, Seidel E, Tamura S, Rogge R, Maeshima K. 2018b. The 10-nm chromatin fiber and its relationship to interphase chromosome organization. *Biochem Soc Trans* **46**: 67–76.

Hanson SJ, Cinnéide EÓ, Salzberg LI, Wolfe KH, McGowan J, Fitzpatrick DA, Matlin K. 2021. Genomic diversity, chromosomal rearrangements, and interspecies hybridization in the *Ogataea polymorpha* species complex. *G3 (Bethesda)* **11**: pjkab211.

Harmston N, Ing-Simmons E, Tan G, Perry M, Merkenschlager M, Lenhard B. 2017. Topologically associating domains are ancient features that coincide with Metazoan clusters of extreme noncoding conservation. *Nat Commun* **8**: 441.

Harr JC, Gonzalez-Sandoval A, Gasser SM. 2016. Histones and histone modifications in perinuclear chromatin anchoring: from yeast to man. *EMBO Rep* **17**: 139–155.

Harr JC, Luperchio TR, Wong X, Cohen E, Wheelan SJ, Reddy KL. 2015. Directed targeting of chromatin to the nuclear lamina is mediated by chromatin state and A-type lamins. *J Cell Biol* **208**: 33–52.

Hartmann FE, Croll D. 2017. Distinct trajectories of massive recent gene gains and losses in populations of a microbial eukaryotic pathogen. *Mol Biol Evol* **34**: 2808–2822.

Hartmann FE, Sanchez-Vallet A, McDonald BA, Croll D. 2017. A fungal wheat pathogen evolved host specialization by extensive chromosomal rearrangements. *ISME J* **11**: 1189–1204.

Hassing B, Winter D, Becker Y, Mesarich CH, Eaton CJ, Scott B. 2019. Analysis of *Epichloë festucae* small secreted proteins in the interaction with *Lolium perenne*. *PLoS One* **14**: e0209463.

Hastings PJ, Lupski JR, Rosenberg SM, Ira G. 2009. Mechanisms of change in gene copy number. *Nature Rev Genet* **10**: 551–564.

Hazarika RR, Serra M, Zhang Z, Zhang Y, Schmitz RJ, Johannes F. 2022. Molecular properties of epimutation hotspots. *Nat Plants* **8**: 146–156.

He C, Zhang Z, Li B, Tian S. 2020. The Pattern and Function of DNA methylation in fungal plant pathogens. *Microorganisms* **8**: p227.

Hediger F, Berthiau A-S, van Houwe G, Gilson E, Gasser SM. 2006. Subtelomeric factors antagonize telomere anchoring and Tel1-independent telomere length regulation. *EMBO J* **25**: 857–867.

Hediger F, Neumann FR, Van Houwe G, Dubrana K, Gasser SM. 2002. Live imaging of telomeres: yKu and Sir proteins define redundant telomere-anchoring pathways in yeast. *Curr Biol* **12**: 2076–2089.

Heermann R, Fuchs TM. 2008. Comparative analysis of the *Photorhabdus luminescens* and the *Yersinia enterocolitica* genomes: Uncovering candidate genes involved in insect pathogenicity. *BMC Genomics* **9**: 40.

Heger P, Marin B, Bartkuhn M, Schierenberg E, Wiehe T. 2012. The chromatin insulator CTCF and the emergence of metazoan diversity. *Proc Natl Acad Sci USA* **109**: 17507–17512.

Heinz S, Benner C, Spann N, Bertolino E, Lin YC, Laslo P, Cheng JX, Murre C, Singh H, Glass CK. 2010. Simple combinations of lineage-determining transcription factors prime *cis*-regulatory elements required for macrophage and B cell identities. *Mol Cell* **38**: 576–589.

Helfer S. 2014. Rust fungi and global change. *New Phytol* **201**: 770–780.

Henikoff S, Smith MM. 2015. Histone variants and epigenetics. *Cold Spring Harb Perspect Biol* **7**: a019364.
Henningsen EC, Hewitt T, Dugyala S, Nazareno ES, Gilbert E, Li F, Kianian SF, Steffenson BJ, Dodds PN, Sperschneider J, et al. 2022. A chromosome-level, fully phased genome assembly of the oat crown rust fungus *Puccinia coronata f. sp. avenae*: a resource to enable comparative genomics in the cereal rusts. *G3 (Bethesda)* **12**: jkac149.
Hepp M, Gutierrez JL. 2014. The yeast HMG proteins Hmo1 and Nhp6 exert a differential stimulatory effect on ATP-dependent nucleosome remodeling activity. *FEBS Journal* **281**: 65–783.
Her J, Bunting SF. 2018. How cells ensure correct repair of DNA double-strand breaks. *J Biol Chem* **293**: 10502–10511.
Heun P, Laroche T, Raghuraman MK, Gasser SM. 2001. The positioning and dynamics of origins of replication in the budding yeast nucleus. *J Cell Biol* **152**: 385–400.
Hickey G, Paten B, Earl D, Zerbino D, Haussler D. 2013. HAL: a hierarchical format for storing and analyzing multiple genome alignments. *Bioinformatics* **29**: 1341–1342.
Hill MS, Vande Zande P, Wittkopp PJ. 2021. Molecular and evolutionary processes generating variation in gene expression. *Nat Rev Genet* **22**: 203–215.
Hiltunen M, Grudzinska-Sterno M, Wallerman O, Ryberg M, Johannesson H. 2019. Maintenance of high genome integrity over vegetative growth in the fairy-ring mushroom *Marasmius oreades*. *Curr Biol* **29**: 2758-2765.e6.
Hirano T. 2016. Condensin-based chromosome organization from bacteria to vertebrates. *Cell* **164**: 847–857.
Hirano T. 2012. Condensins: universal organizers of chromosomes with diverse functions. *Genes Dev* **26**: 1659–1678.
Hirano Y, Asakawa H, Sakuno T, Haraguchi T, Hiraoka Y. 2020. Nuclear envelope proteins modulating the heterochromatin formation and functions in fission yeast. *Cells* **9**: 1908.
Ho SS, Urban AE, Mills RE. 2020. Structural variation in the sequencing era. *Nat Rev Genet* **21**: 171–189.
Hocher A, Ruault M, Kaferle P, Descrimes M, Garnier M, Morillon A, Taddei A. 2018. Expanding heterochromatin reveals discrete subtelomeric domains delimited by chromatin landscape transitions. *Genome Res* **28**: 1867–1881.
Hocher A, Taddei A. 2020. Subtelomeres as specialized chromatin domains. *Bioessays* **42**: e1900205.
Hodgkinson A, Eyre-Walker A. 2011. Variation in the mutation rate across mammalian genomes. *Nat Rev Genet* **12**: 756–766.
Hoede C, Arnoux S, Moisset M, Chaumier T, Inizan O, Jamilloux V, Quesneville H. 2014. PASTEC: an automatic transposable element classification tool. *PLoS One* **9**: e91929.
Hoencamp C, Dudchenko O, Elbatsh AMO, Brahmachari S, Raaijmakers JA, van Schaik T, Sedeño Cacciatore Á, Contessoto VG, van Heesbeen RGHP, van den Broek B, et al. 2021. 3D genomics across the tree of life reveals condensin II as a determinant of architecture type. *Science* **372**: 984–989.
Hollister JD, Gaut BS. 2009. Epigenetic silencing of transposable elements: A trade-off between reduced transposition and deleterious effects on neighboring gene expression. *Genome Res* **19**: 1419–1428.
Holwerda S, De Laat W. 2012. Chromatin loops, gene positioning, and gene expression. *Front Genet* **3**: 217.
Honda S, Bicocca VT, Gessaman JD, Rountree MR, Yokoyama A, Yu EY, Selker JML, Selker EU. 2016. Dual chromatin recognition by the histone deacetylase complex HCHC is required for proper DNA methylation in *Neurospora crassa*. *Proc Natl Acad Sci USA* **113**: E6135–E6144.
Honda S, Lewis ZA, Shimada K, Fischle W, Sack R, Selker EU. 2012. Heterochromatin protein 1 forms distinct complexes to direct histone deacetylation and DNA methylation. *Nat Struct Mol Biol* **19**: 471–7, S1.
Honda S, Selker EU. 2008. Direct interaction between DNA methyltransferase DIM-2 and HP1 is required for DNA methylation in *Neurospora crassa*. *Mol Cell Biol* **28**: 6044–6055.
Hsieh T-HS, Weiner A, Lajoie B, Dekker J, Friedman N, Rando OJ. 2015. Mapping nucleosome resolution chromosome folding in yeast by Micro-C. *Cell* **162**: 108–119.
Hu J, Barrett RDH. 2017. Epigenetics in natural animal populations. *J Evol Biol* **30**: 1612–1632.
Huang C, Zhu B. 2018. Roles of H3K36-specific histone methyltransferases in transcription: antagonizing silencing and safeguarding transcription fidelity. *Biophys Rep* **4**: 170–177.
Huang CR, Burns KH, Boeke JD. 2012. Active transposition in genomes. *Annu Rev Genet* **46**: 651–675.
Huang J, Cook DE. 2022. The contribution of DNA repair pathways to genome editing and evolution in filamentous pathogens. *FEMS Microbiol Rev* **46**: fuac035.
Huang L, Li X, Dong L, Wang B, Pan L. 2021. Profiling of chromatin accessibility identifies transcription factor binding sites across the genome of Aspergillus species. *BMC Biol* **19**: 189.
Hua-Van A, Le Rouzic A, Boutin TS, Filée J, Capy P. 2011. The struggle for life of the genome's selfish architects. *Biol Direct* **6**: 19.

Hudson DF, Marshall KM, Earnshaw WC. 2009. Condensin: Architect of mitotic chromosomes. *Chromosome Res* **17**: 131–144.

Huerta-Cepas J, Szklarczyk D, Heller D, Hernández-Plaza A, Forslund SK, Cook H, Mende DR, Letunic I, Rattei T, Jensen LJ, et al. 2019. eggNOG 5.0: a hierarchical, functionally and phylogenetically annotated orthology resource based on 5090 organisms and 2502 viruses. *Nucleic Acids Res* **47**: D309–D314.

Hunter JD. 2007. Matplotlib: A 2D graphics environment. *Comput Sci Eng* **9**: 90–95.

Hurst LD, Pál C, Lercher MJ. 2004. The evolutionary dynamics of eukaryotic gene order. *Nat Rev Genet* **5**: 299–310.

Huynh L, Hormozdiari F. 2019. TAD fusion score: discovery and ranking the contribution of deletions to genome structure. *Genome biology* **20**: 1–13.

Ikeda H, Sone M, Yamanaka S, Yamamoto T. 2017. Structural and spatial chromatin features at developmental gene loci in human pluripotent stem cells. *Nat Commun* **8**: 1616.

Imakaev M, Fudenberg G, McCord RP, Naumova N, Goloborodko A, Lajoie BR, Dekker J, Mirny LA. 2012. Iterative correction of Hi-C data reveals hallmarks of chromosome organization. *Nature Methods* **9**: 999–1003.

Inderbitzin P, Subbarao KV. 2014. Verticillium systematics and evolution: how confusion impedes Verticillium wilt management and how to resolve it. *Phytopathology* **104**: 564–574.

Ingvarsson PK. 2007. Gene expression and protein length influence codon usage and rates of sequence evolution in *Populus tremula*. *Mol Biol Evol* **24**: 836–844.

Irie H, Yamamoto I, Tarumoto Y, Tashiro S, Runge KW, Ishikawa F. 2019. Telomere-binding proteins Taz1 and Rap1 regulate DSB repair and suppress gross chromosomal rearrangements in fission yeast. *PLoS Genet* **15**: e1008335.

Jaillon O, Aury J-M, Brunet F, Petit J-L, Stange-Thomann N, Mauceli E, Bouneau L, Fischer C, Ozouf-Costaz C, Bernot A, et al. 2004. Genome duplication in the teleost fish *Tetraodon nigroviridis* reveals the early vertebrate proto-karyotype. *Nature* **431**: 946–957.

Jain D, Cooper JP. 2010. Telomeric strategies: means to an end. *Annu Rev Genet* **44**: 243–269.

Jamieson K, McNaught KJ, Ormsby T, Leggett NA, Honda S, Selker EU. 2018. Telomere repeats induce domains of H3K27 methylation in *Neurospora*. *Elife* **7**: e.31216.

Jamieson K, Rountree MR, Lewis ZA, Stajich JE, Selker EU. 2013. Regional control of histone H3 lysine 27 methylation in *Neurospora*. *Proc Natl Acad Sci USA* **110**: 6027–6032.

Jamieson K, Wiles ET, McNaught KJ, Sidoli S, Leggett N, Shao Y, Garcia BA, Selker EU. 2016. Loss of HP1 causes depletion of H3K27me3 from facultative heterochromatin and gain of H3K27me2 at constitutive heterochromatin. *Genome Res* **26**: 97–107.

Janevska S, Baumann L, Sieber CMK, Münsterkötter M, Ulrich J, Kämper J, Güldener U, Tudzynski B. 2018. Elucidation of the two H3K36me3 histone methyltransferases Set2 and Ash1 in *Fusarium fujikuroi* unravels their different chromosomal targets and a major impact of Ash1 on genome stability. *Genetics* **208**: 153–171.

Janssen A, Colmenares SU, Karpen GH. 2018. Heterochromatin: Guardian of the Genome. *Annu Rev Cell Dev Biol* **34**: 265–288.

Jeffares DC, Jolly C, Hoti M, Speed D, Shaw L, Rallis C, Balloux F, Dessimoz C, Bahler J, Sedlazeck FJ. 2017. Transient structural variations have strong effects on quantitative traits and reproductive isolation in fission yeast. *Nat Commun* **8**: 14061.

Jenull S, Tscherner M, Mair T, Kuchler K. 2020. ATAC-Seq identifies chromatin landscapes linked to the regulation of oxidative stress in the human fungal pathogen *Candida albicans*. *J Fungi (Basel)* **6**: 182.

Jenuwein T, Allis CD. 2001. Translating the histone code. *Science* **293**: 1074–1080.

Jeppsson K, Sakata T, Nakato R, Milanova S, Shirahige K, Björkegren C. 2022. Cohesin-dependent chromosome loop extrusion is limited by transcription and stalled replication forks. *Sci Adv* **8**: eabn7063.

Jerkovic I, Cavalli G. 2021. Understanding 3D genome organization by multidisciplinary methods. *Nat Rev Mol Cell Biol* **22**: 511–528.

Jia J, Xie Y, Cheng J, Kong C, Wang M, Gao L, Zhao F, Guo J, Wang K, Li G, et al. 2021. Homology-mediated inter-chromosomal interactions in hexaploid wheat lead to specific subgenome territories following polyploidization and introgression. *Genome Biol* **22**: 26.

Jiao Y, Peluso P, Shi J, Liang T, Stitzer MC, Wang B, Campbell MS, Stein JC, Wei X, Chin CS, et al. 2017. Improved maize reference genome with single-molecule technologies. *Nature* **546**: 524–527.

Jin F, Li Y, Dixon JR, Selvaraj S, Ye Z, Lee AY, Yen C-A, Schmitt AD, Espinoza CA, Ren B. 2013. A high-resolution map of the three-dimensional chromatin interactome in human cells. *Nature* **503**: 290–294.

Jin QW, Fuchs J, Loidl J. 2000. Centromere clustering is a major determinant of yeast interphase nuclear organization. *J Cell Sci* **113 (Pt 11)**: 1903–1912.

Jin Y, Hammell M. 2018. Analysis of RNA-Seq Data Using TEtranscripts. *Methods Mol Biol* **1751**: 153–167.

Jin Y, Tam OH, Paniagua E, Hammell M. 2015. TEtranscripts: a package for including transposable elements in differential expression analysis of RNA-seq datasets. *Bioinformatics* **31**: 3593–3599.
Jones JDG, Dangl JL. 2006. The plant immune system. *Nature* **444**: 323–329.
Jones P, Binns D, Chang H-Y, Fraser M, Li W, McAnulla C, McWilliam H, Maslen J, Mitchell A, Nuka G, et al. 2014. InterProScan 5: genome-scale protein function classification. *Bioinformatics* **30**: 1236–1240.
Jørgensen HF, Azuara V, Amoils S, Spivakov M, Terry A, Nesterova T, Cobb BS, Ramsahoye B, Merkenschlager M, Fisher AG. 2007. The impact of chromatin modifiers on the timing of locus replication in mouse embryonic stem cells. *Genome Biol* **8**: R169.
Josephs EB, Lee YW, Stinchcombe JR, Wright SI. 2015. Association mapping reveals the role of purifying selection in the maintenance of genomic variation in gene expression. *Proc Natl Acad Sci USA* **112**: 15390–15395.
Joubert PM, Krasileva KV. 2023. Distinct genomic contexts predict gene presence-absence variation in different pathotypes of a fungal plant pathogen. *bioRxiv*. 2023-02.
Juarez-Reyes A, Castano I. 2019. Chromatin architecture and virulence-related gene expression in eukaryotic microbial pathogens. *Current Genetics* **65**: 435–443.
Jumper J, Evans R, Pritzel A, Green T, Figurnov M, Ronneberger O, Tunyasuvunakool K, Bates R, Žídek A, Potapenko A, et al. 2021. Highly accurate protein structure prediction with AlphaFold. *Nature* **596**: 583–589.
Kabir S, Sfeir A, de Lange T. 2010. Taking apart Rap1: an adaptor protein with telomeric and non-telomeric functions. *Cell Cycle* **9**: 4061–4067.
Kadauke S, Blobel GA. 2009. Chromatin loops in gene regulation. *Biochim Biophys Acta* **1789**: 17–25.
Kainth AS, Chowdhary S, Pincus D, Gross DS. 2021. Primordial super-enhancers: heat shock-induced chromatin organization in yeast. *Trends Cell Biol* **58**: 216–231.
Kaizer H, Connelly CJ, Bettridge K, Viggiani C, Greider CW. 2015. Regulation of telomere length requires a conserved N-terminal domain of Rif2 in *Saccharomyces cerevisiae*. *Genetics* **201**: 573–586.
Kakui Y, Barrington C, Barry DJ, Gerguri T, Fu X, Bates PA, Khatri BS, Uhlmann F. 2020. Fission yeast condensin contributes to interphase chromatin organization and prevents transcription-coupled DNA damage. *Genome Biol* **21**: 272.
Kakui Y, Rabinowitz A, Barry DJ, Uhlmann F. 2017. Condensin-mediated remodeling of the mitotic chromatin landscape in fission yeast. *Nat Genet* **49**: 1553–1557.
Kalitsis P, Zhang T, Marshall KM, Nielsen CF, Hudson DF. 2017. Condensin, master organizer of the genome. *Chromosome Research* **25**: 61–76.
Kamei M, Ameri AJ, Ferraro AR, Bar-Peled Y, Zhao F, Ethridge CL, Lail K, Amirebrahimi M, Lipzen A, Ng V, et al. 2021. IMITATION SWITCH is required for normal chromatin structure and gene repression in PRC2 target domains. *Proc Natl Acad Sci U S A* **118**: e.2010003118.
Kang H, Zhu D, Lin R, Opiyo SO, Jiang N, Shiu SH, Wang GL. 2016. A novel method for identifying polymorphic transposable elements via scanning of high-throughput short reads. *DNA Res* **23**: 241–251.
Kanoh J, Ishikawa F. 2001. spRap1 and spRif1, recruited to telomeres by Taz1, are essential for telomere function in fission yeast. *Curr Biol* **11**: 1624–1630.
Kasowski M, Kyriazopoulou-Panagiotopoulou S, Grubert F, Zaugg JB, Kundaje A, Liu Y, Boyle AP, Zhang QC, Zakharia F, Spacek DV, et al. 2013. Extensive variation in chromatin states across humans. *Science* **342**: 750–752.
Katoh K, Standley DM. 2013. MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Mol Biol Evol* **30**: 772–780.
Kaushal A, Mohana G, Dorier J, Özdemir I, Omer A, Cousin P, Semenova A, Taschner M, Dergai O, Marzetta F. 2021. CTCF loss has limited effects on global genome architecture in *Drosophila* despite critical regulatory functions. *Nature Commun.* **12**: 1–16.
Kawakatsu T, Huang S-SC, Jupe F, Sasaki E, Schmitz RJ, Urich MA, Castanon R, Nery JR, Barragan C, He Y, et al. 2016. Epigenomic diversity in a global collection of *Arabidopsis thaliana* accessions. *Cell* **166**: 492–505.
Kelkar YD, Ochman H. 2012. Causes and consequences of genome expansion in fungi. *Genome Biol Evol* **4**: 13–23.
Kema GHJ, Mirzadi Gohari A, Aouini L, Gibriel HAY, Ware SB, van den Bosch F, Manning-Smith R, Alonso-Chavez V, Helps J, Ben M'Barek S, et al. 2018. Stress and sexual reproduction affect the dynamics of the wheat pathogen effector AvrStb6 and strobilurin resistance. *Nat Genet* **23**: 678.
Kentepozidou E, Aitken SJ, Feig C, Stefflova K, Ibarra-Soria X, Odom DT, Roller M, Flicek P. 2020. Clustered CTCF binding is an evolutionary mechanism to maintain topologically associating domains. *Genome Biol* **21**: 1–19.
Kidwell MG, Lisch DR. 2001. Perspective: Transposable elements, parasitic DNA, and genome evolution. *Evolution* **55**: 1–24.
Kim J, Dean A. 2021. Enhancers navigate the three-dimensional genome to direct cell fate decisions. *Curr Res Struct Biol* **71**: 101–109.

Kim KS, Min J-Y, Dickman MB. 2008. Oxalic acid is an elicitor of plant programmed cell death during *Sclerotinia sclerotiorum* disease development. *Mol Plant Microbe Interact* **21**: 605–612.
Kim S, Liachko I, Brickner DG, Cook K, Noble WS, Brickner JH, Shendure J, Dunham MJ. 2017. The dynamic three-dimensional organization of the diploid yeast genome. *Elife* **6**: e.23623.
Kim TH, Dekker J. 2018. ChIP–Quantitative polymerase chain reaction (ChIP-qPCR). *Cold Spring Harb Protoc* **2018**: db.prot082628.
Kimura K, Hirano T. 1997. ATP-dependent positive supercoiling of DNA by 13S condensin: a biochemical implication for chromosome condensation. *Cell* **90**: 625–634.
Kimura M. 1968. Evolutionary rate at the molecular level. *Nature* **217**: 624–626.
Kimura M. 1983. *The Neutral Theory of Molecular Evolution*. Cambridge University Press.
King MC, Lusk CP, Blobel G. 2006. Karyopherin-mediated import of integral inner nuclear membrane proteins. *Nature* **442**: 1003–1007.
King R, Urban M, Hammond-Kosack MCU, Hassani-Pak K, Hammond-Kosack KE. 2015. The completed genome sequence of the pathogenic ascomycete fungus *Fusarium graminearum*. *BMC Genomics* **16**: 544.
Kinkley S, Helmuth J, Polansky JK, Dunkel I, Gasparoni G, Fröhler S, Chen W, Walter J, Hamann A, Chung H-R. 2016. reChIP-seq reveals widespread bivalency of H3K4me3 and H3K27me3 in CD4+ memory T cells. *Nat Commun* **7**: 1–13.
Kiran K, Rawal HC, Dubey H, Jaswal R, Devanna BN, Gupta DK, Bhardwaj SC, Prasad P, Pal D, Chhuneja P, et al. 2016. Draft genome of the wheat rust pathogen (*Puccinia triticina*) unravels genome-wide structural variations during evolution. *Genome Biol Evol* **8**: 2702–2721.
Kitano S, Kurasawa H, Aizawa Y. 2018. Transposable elements shape the human proteome landscape via formation of *cis*-acting upstream open reading frames. *Genes Cells* **23**: 274–284.
Klocko AD, Ormsby T, Galazka JM, Leggett NA, Uesaka M, Honda S, Freitag M, Selker EU. 2016. Normal chromosome conformation depends on subtelomeric facultative heterochromatin in *Neurospora crassa*. *Proc Natl Acad Sci USA* **113**: 15048–15053.
Klocko AD, Summers CA, Glover ML, Parrish R, Storck WK, McNaught KJ, Moss ND, Gotting K, Stewart A, Morrison AM, et al. 2020. Selection and characterization of mutants defective in DNA methylation in. *Genetics* **216**: 671–688.
Klocko AD, Uesaka M, Ormsby T, Rountree MR, Wiles ET, Adhvaryu KK, Honda S, Selker EU. 2019. Nucleosome positioning by an evolutionarily conserved chromatin remodeler prevents aberrant DNA Methylation in *Neurospora*. *Genetics* **211**: 563–578.
Klosterman SJ, Atallah ZK, Vallad GE, Subbarao KV. 2009. Diversity, pathogenicity, and management of Verticillium species. *Annu Rev Phytopathol* **47**: 39–62.
Klosterman SJ, Subbarao KV, Kang S, Veronese P, Gold SE, Thomma BPHJ, Chen Z, Henrissat B, Lee Y-H, Park J, et al. 2011. Comparative genomics yields insights into niche adaptation of plant vascular wilt pathogens. *PLoS Pathog* **7**: e1002137.
Knibbe C, Coulon A, Mazet O, Fayard JM, Beslon G. 2007. A long-term evolutionary pressure on the amount of noncoding DNA. *Mol Biol Evol* **24**: 2344–2353.
Knight PA, Ruiz D. 2013. A fast algorithm for matrix balancing. *IMA J Numer Anal* **33**: 1029–1047.
Koch H, Frickel J, Valiadi M, Becks L. 2014. Why rapid, adaptive evolution matters for community dynamics. *Front Ecol Evol* **2**: 17.
Koenker R, Ng P. 2003. SparseM: A Sparse Matrix Package for R. *J Stat Softw* **8**: 1–9.
Kolasinska-Zwierz P, Down T, Latorre I, Liu T, Liu XS, Ahringer J. 2009. Differential chromatin marking of introns and expressed exons by H3K36me3. *Nat Genet* **41**: 376–381.
Kolesnikova TD, Goncharov FP, Zhimulev IF. 2018. Similarity in replication timing between polytene and diploid cells is associated with the organization of the Drosophila genome. *PLoS One* **13**: e0195207.
Kollmar M. 2015. Polyphyly of nuclear lamin genes indicates an early eukaryotic origin of the metazoan-type intermediate filament proteins. *Sci Rep* **5**: 10652.
Kombrink A, Rovenich H, Shi-Kunne X, Rojas-Padilla E, van den Berg GCM, Domazakis E, de Jonge R, Valkenburg D-J, Sánchez-Vallet A, Seidl MF, et al. 2017. *Verticillium dahliae* LysM effectors differentially contribute to virulence on plant hosts. *Mol Plant Pathol* **18**: 596–608.
Kong M, Cutts EE, Pan D, Beuron F, Kaliyappan T, Xue C, Morris EP, Musacchio A, Vannini A, Greene EC. 2020. Human Condensin I and II Drive Extensive ATP-Dependent Compaction of Nucleosome-Bound DNA. *Mol Cell* **79**: 99-114.e9.
Koonin EV, Wolf YI. 2009. Is evolution Darwinian or/and Lamarckian? *Biol Direct* **4**: 42.
Koreny L, Field MC. 2016. Ancient eukaryotic origin and evolutionary plasticity of nuclear lamina. *Genome Biol Evol* **8**: 2663–2671.

Kramer HM, Cook DE, van den Berg GCM, Seidl MF, Thomma BPHJ. 2021. Three putative DNA methyltransferases of *Verticillium dahliae* differentially contribute to DNA methylation that is dispensable for growth, development and virulence. *Epigenetics Chromatin* **14**: 21.

Kramer HM, Seidl MF, Thomma BPHJ, Cook DE. 2022. Local Rather than Global H3K27me3 dynamics are associated with differential gene expression in *Verticillium dahliae. MBio* **13**: e0356621.

Krassovsky K, Henikoff JG, Henikoff S. 2012. Tripartite organization of centromeric chromatin in budding yeast. *Proc Natl Acad Sci U S A* **109**: 243–248.

Krefting J, Andrade-Navarro MA, Ibn-Salem J. 2018. Evolutionary stability of topologically associating domains is associated with conserved gene regulation. *BMC Biol* **16**: 87.

Krishnan P, Meile L, Plissonneau C, Ma X, Hartmann FE, Croll D, McDonald BA, Sánchez-Vallet A. 2018. Transposable element insertions shape gene regulation and melanin production in a fungal pathogen of wheat. *BMC Biol* **16**: 78.

Kronenberg ZN, Osborne EJ, Cone KR, Kennedy BJ, Domyan ET, Shapiro MD, Elde NC, Yandell M. 2015. Wham: identifying structural variants of biological consequence. *PLoS Comput Biol* **11**: e1004572.

Krueger F. 2015. Trim Galore!: A wrapper around Cutadapt and FastQC to consistently apply adapter and quality trimming to FastQ files, with extra functionality for RRBS data. *Babraham Institute*.

Krumm A, Duan Z. 2019. Understanding the 3D genome: Emerging impacts on human disease. *Semin Cell Dev Biol* **90**: 62–77.

Kruse K, Hug CB, Vaquerizas JM. 2020. FAN-C: a feature-rich framework for the analysis and visualisation of chromosome conformation capture data. *Genome Biol.* **21**: 1–19.

Krylov DM, Wolf YI, Rogozin IB, Koonin EV. 2003. Gene loss, protein sequence divergence, gene dispensability, expression level, and interactivity are correlated in eukaryotic evolution. *Genome Res* **13**: 2229–2235.

Kumar S, Chinnusamy V, Mohapatra T. 2018. Epigenetics of Modified DNA Bases: 5-Methylcytosine and Beyond. *Front Genet* **9**: 640.

Kurdzo EL, Obeso D, Chuong H, Dawson DS. 2017. Meiotic centromere coupling and pairing function by two separate mechanisms in *Saccharomyces cerevisiae*. *Genetics* **205**: 657–671.

Kurtz S, Phillippy A, Delcher AL, Smoot M, Shumway M, Antonescu C, Salzberg SL. 2004. Versatile and open software for comparing large genomes. *Genome Biol* **5**: R12.

Kuzniar A, Maassen J, Verhoeven S, Santuari L, Shneider C, Kloosterman WP, de Ridder J. 2020. sv-callers: a highly portable parallel workflow for structural variant detection in whole-genome sequence data. *PeerJ* **8**: e8214.

Kwon DY, Zhao Y-T, Lamonica JM, Zhou Z. 2017. Locus-specific histone deacetylation using a synthetic CRISPR-Cas9-based HDAC. *Nat Commun* **8**: 15315.

Lai Y, Wang L, Zheng W, Wang S. 2022. Regulatory Roles of Histone Modifications in Filamentous Fungal Pathogens. *J Fungi (Basel)* **8**: 565.

Lamberth C, Rendine S, Sulzer-Mosse S. 2021. Chapter 4 - Agrochemical disease control: The story so far. In *Recent Highlights in the Discovery and Optimization of Crop Protection Products* (eds. P. Maienfisch and S. Mangelinckx), pp. 65–85

Langmead B, Salzberg SL. 2012. Fast gapped-read alignment with Bowtie 2. *Nat Methods* **9**: 357–359.

Laroche T, Martin SG, Gotta M, Gorham HC, Pryde FE, Louis EJ, Gasser SM. 1998. Mutation of yeast Ku genes disrupts the subnuclear organization of telomeres. *Curr Biol* **8**: 653–656.

Larson AG, Elnatan D, Keenen MM, Trnka MJ, Johnston JB, Burlingame AL, Agard DA, Redding S, Narlikar GJ. 2017. Liquid droplet formation by HP1α suggests a role for phase separation in heterochromatin. *Nature* **547**: 236–240.

Laun K, Coggill P, Palmer S, Sims S, Ning Z, Ragoussis J, Volpi E, Wilson N, Beck S, Ziegler A, et al. 2006. The leukocyte receptor complex in chicken is characterized by massive expansion and diversification of immunoglobulin-like loci. *PLOS Genetics* **2**: e73.

Lawrence M, Daujat S, Schneider R. 2016. Lateral Thinking: How histone modifications regulate gene expression. *Trends Genet* **32**: 42–56.

Lawrence TJ, Carper DL, Spangler MK, Carrell AA, Rush TA, Minter SJ, Weston DJ, Labbé JL. 2021. amPEPpy 1.0: a portable and accurate antimicrobial peptide prediction tool. *Bioinformatics* **37**: 2058–2060.

Layer RM, Chiang C, Quinlan AR, Hall IM. 2014. LUMPY: a probabilistic framework for structural variant discovery. *Genome Biol* **15**: R84.

Lazar-Stefanita L, Scolari VF, Mercy G, Muller H, Guérin TM, Thierry A, Mozziconacci J, Koszul R. 2017. Cohesins and condensins orchestrate the 4D dynamics of yeast chromosomes during the cell cycle. *EMBO J* **36**: 2684–2697.

Le Dily F, Bau D, Pohl A, Vicent GP, Serra F, Soronellas D, Castellano G, Wright RHG, Ballare C, Filion G. 2014. Distinct structural transitions of chromatin topological domains correlate with coordinated hormone-induced gene regulation. *Gen dev.* **28**: 2151–2162.

Le Dily F, Beato M. 2015. TADs as modular and dynamic units for gene regulation by hormones. *FEBS letters* **589**: 2885–2892.

Le Rouzic A, Boutin TS, Capy P. 2007. Long-term evolution of transposable elements. *Proc Natl Acad Sci USA* **104**: 19375–19380.

Lê S, Josse J, Husson F. 2008. FactoMineR: An R Package for Multivariate Analysis. *J Stat Softw* **25**: 1–18.

Le SQ, Dang CC, Gascuel O. 2012. Modeling protein evolution with several amino acid replacement matrices depending on site rates. *Mol Biol Evol* **29**: 2921–2936.

Le TN, Miyazaki Y, Takuno S, Saze H. 2015. Epigenetic regulation of intragenic transposable elements impacts gene transcription in *Arabidopsis thaliana*. *Nucleic Acids Res* **43**: 3911–3921.

Lechner J, Carbon J. 1991. A 240 kd multisubunit protein complex, CBF3, is a major component of the budding yeast centromere. *Cell* **64**: 717–725.

Lee YCG, Karpen GH. 2017. Pervasive epigenetic effects of Drosophila euchromatic transposable elements impact their evolution. *Elife* **6**: e25762.

Leem Y-E, Ripmaster TL, Kelly FD, Ebina H, Heincelman ME, Zhang K, Grewal SIS, Hoffman CS, Levin HL. 2008. Retrotransposon Tf1 is targeted to Pol II promoters by transcription activators. *Mol Cell* **30**: 98–107.

Lefrançois P, Auerbach RK, Yellman CM, Roeder GS, Snyder M. 2013. Centromere-like regions in the budding yeast genome. *PLoS Genet* **9**: e1003209.

Leister D. 2004. Tandem and segmental gene duplication and recombination in the evolution of plant disease resistance gene. *Trends in Genetics* **20**: 116–122.

Levy DL, Heald R. 2010. Nuclear size is regulated by importin α and Ntf2 in Xenopus. *Cell* **143**: 288–298.

Lewis ZA, Adhvaryu KK, Honda S, Shiver AL, Knip M, Sack R, Selker EU. 2010a. DNA methylation and normal chromosome behavior in *Neurospora* depend on five components of a histone methyltransferase complex, DCDC. *PLoS Genet* **6**: e1001196.

Lewis ZA, Adhvaryu KK, Honda S, Shiver AL, Selker EU. 2010b. Identification of DIM-7, a protein required to target the DIM-5 H3 methyltransferase to chromatin. *Proc Natl Acad Sci USA* **107**: 8310–8315.

Lewis ZA, Honda S, Khlafallah TK, Jeffress JK, Freitag M, Mohn F, Schübeler D, Selker EU. 2009. Relics of repeat-induced point mutation direct heterochromatin formation in *Neurospora crassa*. *Genome Res* **19**: 427–437.

Lewontin R. 2008. The Genotype/Phenotype Distinction. In *Stanford Encyclopedia of Philosophy*, philpapers.org.

Li B, Oestreich S, de Lange T. 2000. Identification of human Rap1: implications for telomere evolution. *Cell* **101**: 471–483.

Li C-X, Liu L, Zhang T, Luo X-M, Feng J-X, Zhao S. 2022. Three-Dimensional genome map of the filamentous fungus *Penicillium oxalicum*. *Microbiol Spectr* **10**: e0212121.

Li F, Mao G, Tong D, Huang J, Gu L, Yang W, Li G-M. 2013. The histone mark H3K36me3 regulates human DNA mismatch repair through its interaction with MutSα. *Cell* **153**: 590–600.

Li H. 2011. A statistical framework for SNP calling, mutation discovery, association mapping and population genetical parameter estimation from sequencing data. *Bioinformatics* **27**: 2987–2993.

Li H, Durbin R. 2010. Fast and accurate long-read alignment with Burrows–Wheeler transform. *Bioinformatics* **26**: 589–595.

Li H, Durbin R. 2009. Fast and accurate short read alignment with Burrows–Wheeler transform. *Bioinformatics* **25**: 1754–1760.

Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, Marth G, Abecasis G, Durbin R, 1000 Genome Project Data Processing Subgroup. 2009. The Sequence Alignment/Map format and SAMtools. *Bioinformatics* **25**: 2078–2079.

Li L, Lyu X, Hou C, Takenaka N, Nguyen HQ, Ong C-T, Cubeñas-Potts C, Hu M, Lei EP, Bosco G. 2015. Widespread rearrangement of 3D chromatin organization underlies polycomb-mediated stress-induced silencing. *Molecular cell* **58**: 216–231.

Li Y, Haarhuis JHI, Sedeño Cacciatore Á, Oldenkamp R, van Ruiten MS, Willems L, Teunissen H, Muir KW, de Wit E, Rowland BD, et al. 2020. The structural basis for cohesin–CTCF-anchored loops. *Nature* **578**: 472–476.

Li Z, Wang M, Lin K, Xie Y, Guo J, Ye L, Zhuang Y, Teng W, Ran X, Tong Y, et al. 2019. The bread wheat epigenomic map reveals distinct chromatin architectural and evolutionary features of functional genetic elements. *Genome Biol* **20**: 139.

Liang J, Li Y, Dodds PN, Figueroa M, Sperschneider J, Han S, Tsui CKM, Zhang K, Li L, Ma Z, et al. 2022. Haplotype-phased and chromosome-level genome assembly of *Puccinia polysora*, a giga-scale fungal pathogen causing southern corn rust. *Mol Ecol Resour*. 2022-05

Liang Z, Zhang Q, Ji C, Hu G, Zhang P, Wang Y, Yang L, Gu X. 2021. Reorganization of the 3D chromatin architecture of rice genomes during heat stress. *BMC biology* **19**: 1–10.

Liao Y, Zhang X, Chakraborty M, Emerson JJ. 2021a. Topologically associating domains and their role in the evolution of genome structure and function in *Drosophila*. *Genome Res* **31**: 397–410.
Liao Y, Zhang X, Chakraborty M, Emerson JJ. 2021b. Topologically associating domains and their role in the evolution of genome structure and function in Drosophila. *Genome Res* **31**: 397–410.
Lichtman JW, Conchello J-A. 2005. Fluorescence microscopy. *Nat Methods* **2**: 910–919.
Lieberman-Aiden E, van Berkum NL, Williams L, Imakaev M, Ragoczy T, Telling A, Amit I, Lajoie BR, Sabo PJ, Dorschner MO, et al. 2009. Comprehensive mapping of long-range interactions reveals folding principles of the human genome. *Science* **326**: 289–293.
Lin C, Wu Z, Shi H, Yu J, Xu M, Lin F, Kou Y, Tao Z. 2022. The additional PRC2 subunit and Sin3 histone deacetylase complex are required for the normal distribution of H3K27me3 occupancy and transcriptional silencing in *Magnaporthe oryzae*. *New Phytol* **236**: 576–589.
Lin F, Blake DL, Callebaut I, Skerjanc IS, Holmer L, McBurney MW, Paulin-Levasseur M, Worman HJ. 2000. MAN1, an inner nuclear membrane protein that shares the LEM domain with lamina-associated polypeptide 2 and emerin. *J Biol Chem* **275**: 4840–4847.
Lin YL, Gokcumen O. 2019. Fine-Scale Characterization of genomic structural variation in the human genome reveals adaptive and biomedically relevant hotspots. *Genome Biol Evol* **11**: 1136–1151.
Lippman Z, Gendrel A-V, Black M, Vaughn MW, Dedhia N, McCombie WR, Lavine K, Mittal V, May B, Kasschau KD, et al. 2004. Role of transposable elements in heterochromatin and epigenetic control. *Nature* **430**: 471–476.
Lippman Z, Martienssen R. 2004. The role of RNA interference in heterochromatic silencing. *Nature* **431**: 364–370.
Liti G. 2015. The fascinating and secret wild life of the budding yeast *Saccharomyces cerevisiae. Elife* **4**: e05835.
Liu C, Cheng Y-J, Wang J-W, Weigel D. 2017. Prominent topologically associated domains differentiate global chromatin packing in rice from *Arabidopsis*. *Nature plants* **3**: 742–748.
Liu CL, Kaplan T, Kim M, Buratowski S, Schreiber SL, Friedman N, Rando OJ. 2005. Single-nucleosome mapping of histone modifications in *Saccharomyces cerevisiae*. *PLoS Biol* **3**: e328.
Liu H, Zhang J. 2022. Is the mutation rate lower in genomic regions of stronger selective constraints? *Mol Biol Evol* **39**: msac169.
Liu J, Ali M, Zhou Q. 2020a. Establishment and evolution of heterochromatin. *Ann NY Acad Sci* **1476**: 59–77.
Liu Y, Du H, Li P, Shen Y, Peng H, Liu S, Zhou GA, Zhang H, Liu Z, Shi M, et al. 2020b. Pan-Genome of Wild and Cultivated Soybeans. *Cell* **182**: 162-176 e13.
Liu Y, Tian T, Zhang K, You Q, Yan H, Zhao N, Yi X, Xu W, Su Z. 2018. PCSD: a plant chromatin state database. *Nucleic Acids Res* **46**: D1157–D1167.
Lo Presti L, Lanver D, Schweizer G, Tanaka S, Liang L, Tollot M, Zuccaro A, Reissmann S, Kahmann R. 2015. Fungal effectors and plant susceptibility. *Annu Rev Plant Biol* **66**: 513–545.
Loftus BJ, Fung E, Roncaglia P, Rowley D, Amedeo P, Bruno D, Vamathevan J, Miranda M, Anderson IJ, Fraser JA, et al. 2005. The genome of the basidiomycetous yeast and human pathogen *Cryptococcus neoformans*. *Science* **307**: 1321–1324.
Lonfat N, Duboule D. 2015. Structure, function and evolution of topologically associating domains (TADs) at HOX loci. *FEBS Lett* **589**: 2869–2876.
Long H, Behringer MG, Williams E, Te R, Lynch M. 2016. Similar mutation Rates but highly diverse mutation spectra in Ascomycete and Basidiomycete Yeasts. *Genome Biol Evol* **8**: 3815–3821.
Lorrain C, Feurtey A, Möller M, Haueisen J, Stukenbrock E. 2020. Dynamics of transposable elements in recently diverged fungal pathogens: lineage-specific transposable element content and efficiency of genome defences. *G3 (Bethesda)* **11**: p.jkab068
Lorrain C, Gonçalves Dos Santos KC, Germain H, Hecker A, Duplessis S. 2019. Advances in understanding obligate biotrophy in rust fungi. *New Phytol* **222**: 1190–1206.
Lorrain C, Oggenfuss U, Croll D, Duplessis S. 2021. Transposable elements in fungi: coevolution with the host genome shapes, genome architecture, plasticity and adaptation. https://hal.univ-lorraine.fr/hal-03606172.
Lue NF. 2021. Duplex telomere-binding proteins in fungi with canonical telomere repeats: new lessons in the rapid evolution of telomere proteins. *Front Genet* **12**: 638790.
Luger K, Rechsteiner TJ, Flaus AJ, Waye MM, Richmond TJ. 1997. Characterization of nucleosome core particles containing histone proteins made in bacteria. *J Mol Biol* **272**: 301–311.
Lukito Y, Lee K, Noorifar N, Green KA, Winter DJ, Ram A, Hale TK, Chujo T, Cox MP, Johnson LJ, et al. 2021. Regulation of host-infection ability in the grass-symbiotic fungus *Epichloë festucae* by histone H3K9 and H3K36 methyltransferases. *Environ Microbiol* **23**: 2116–2131.
Luo G-Z, Blanco MA, Greer EL, He C, Shi Y. 2015. DNA N6-methyladenine: a new epigenetic mark in eukaryotes? *Nat Rev Mol Cell Biol* **16**: 705–710.

Lupiáñez DG, Spielmann M, Mundlos S. 2016. Breaking TADs: how alterations of chromatin domains result in disease. *Trend Genet* **32**: 225–237.

Luppino JM, Park DS, Nguyen SC, Lan Y, Xu Z, Yunker R, Joyce EF. 2020. Cohesin promotes stochastic domain intermingling to ensure proper regulation of boundary-proximal genes. *Nat Genet* **52**: 840–848.

Lynch M. 2010. Evolution of the mutation rate. *Trends Genet* **26**: 345–352.

Lynch M. 2006. Streamlining and simplification of microbial genome architecture. *Annu Rev Microbiol* **60**: 327–349.

Lynch M. 2007. *The origins of genome architecture*. Sinauer Associates, Inc. Publishers, Sunderland, MA.

Lynch M, Ackerman MS, Gout J-F, Long H, Sung W, Thomas WK, Foster PL. 2016. Genetic drift, selection and the evolution of the mutation rate. *Nat Rev Genet* **17**: 704–714.

Lynch M, Bobay LM, Catania F, Gout JF, Rho M. 2011a. The repatterning of eukaryotic genomes by random genetic drift. *Annu Rev Genomics Hum Genet* **12**: 347–366.

Lynch M, Conery JS. 2003. The origins of genome complexity. *Science* **302**: 1401–1404.

Lynch VJ, Leclerc RD, May G, Wagner GP. 2011b. Transposon-mediated rewiring of gene regulatory networks contributed to the evolution of pregnancy in mammals. *Nat Genet* **43**: 1154–1159.

Lyons E, Pedersen B, Kane J, Alam M, Ming R, Tang H, Wang X, Bowers J, Paterson A, Lisch D. 2008. Finding and comparing syntenic regions among Arabidopsis and the outgroups papaya, poplar, and grape: CoGe with rosids. *Plant physiology* **148**: 1772–1781.

Lyson TA, Guptill A. 2004. Commodity agriculture, civic agriculture and the future of U.S. farming. *Rural Sociol* **69**: 370–385.

Ma J-L, Kim EM, Haber JE, Lee SE. 2003. Yeast Mre11 and Rad1 proteins define a Ku-independent mechanism to repair double-strand breaks lacking overlapping end sequences. *Mol Cell Biol* **23**: 8820–8828.

Ma L-J, van der Does HC, Borkovich KA, Coleman JJ, Daboussi M-J, Di Pietro A, Dufresne M, Freitag M, Grabherr M, Henrissat B, et al. 2010. Comparative genomics reveals mobile pathogenicity chromosomes in *Fusarium*. *Nature* **464**: 367–373.

Macrae TA, Fothergill-Robinson J, Ramalho-Santos M. 2023. Regulation, functions and transmission of bivalent chromatin during mammalian development. *Nat Rev Mol Cell Biol* **24**: 6–26.

Maeshima K, Iida S, Tamura S. 2021. Physical Nature of Chromatin in the Nucleus. *Cold Spring Harb Perspect Biol* **13**: a040675.

Maeshima K, Rogge R, Tamura S, Joti Y, Hikima T, Szerlong H, Krause C, Herman J, Seidel E, DeLuca J, et al. 2016. Nucleosomal arrays self-assemble into supramolecular globular structures lacking 30-nm fibers. *EMBO J* **35**: 1115–1132.

Mahajan A, Yan W, Zidovska A, Saintillan D, Shelley MJ. 2022. Euchromatin activity enhances segregation and compaction of heterochromatin in the cell nucleus. *Physical Review X* **12**: 041033

Makova KD, Hardison RC. 2015. The effects of chromatin organization on variation in mutation rates in the genome. *Nat Rev Genet* **16**: 213–223.

Malik HS, Henikoff S. 2009. Major evolutionary transitions in centromere complexity. *Cell* **138**: 1067–1082.

Malyavko AN, Parfenova YY, Zvereva MI, Dontsova OA. 2014. Telomere length regulation in budding yeasts. *FEBS Lett* **588**: 2530–2536.

Malyavko AN, Petrova OA, Zvereva MI, Dontsova OA. 2019. Functional duplication of Rap1 in methylotrophic yeasts. *Sci Rep* **9**: 7196.

Mank JE. 2017. The transcriptional architecture of phenotypic dimorphism. *Nat Ecol Evol* **1**: 6.

Manuelidis L. 1985. Individual interphase chromosome domains revealed by in situ hybridization. *Hum Genet* **71**: 288–293.

Mao YS, Zhang B, Spector DL. 2011. Biogenesis and function of nuclear bodies. *Trend Genetics* **27**: 295–306.

Marand AP, Eveland AL, Kaufmann K, Springer NM. 2023. *cis*-regulatory elements in plant development, adaptation, and evolution. *Annu Rev Plant Biol*. **74**

Marand AP, Zhao H, Zhang W, Zeng Z, Fang C, Jiang J. 2019. Historical meiotic crossover hotspots fueled patterns of evolutionary divergence in rice. *The Plant Cell* **31**: 645.

Marie-Nelly H, Marbouty M, Cournac A, Flot J-F, Liti G, Parodi DP, Syan S, Guillén N, Margeot A, Zimmer C. 2014. High-quality genome (re) assembly using chromosomal contact data. *Nature Commun.* **5**: 1–10.

Marquardt S. 2001. "Green havoc": Panama disease, environmental change, and labor process in the Central American banana industry. *Am Hist Rev* **106**: 49–80.

Martin T, Fraser HB. 2018. Comparative expression profiling reveals widespread coordinated evolution of gene expression across eukaryotes. *Nat Commun* **9**: 4963.

Martincorena I, Luscombe NM. 2013. Non-random mutation: the evolution of targeted hypermutation and hypomutation. *Bioessays* **35**: 123–130.

Martincorena I, Seshasayee ASN, Luscombe NM. 2012. Evidence of non-random mutation rates suggests an evolutionary risk management strategy. *Nature* **485**: 95–98.
Martire S, Banaszynski LA. 2020. The roles of histone variants in fine-tuning chromatin organization and function. *Nat Rev Mol Cell Biol* **21**: 522–541.
Mascher M, Gundlach H, Himmelbach A, Beier S, Twardziok SO, Wicker T, Radchuk V, Dockter C, Hedley PE, Russell J, et al. 2017. A chromosome conformation capture ordered sequence of the barley genome. *Nature* **544**: 427–433.
Masel J, Trotter MV. 2010. Robustness and evolvability. *Trends Genet* **26**: 406–414.
Mattick JS, Makunin IV. 2006. Non-coding RNA. *Hum Mol Genet* **15 Spec No 1**: R17-29.
Maxwell PH. 2020. Diverse transposable element landscapes in pathogenic and nonpathogenic yeast models: the value of a comparative perspective. *Mobile DNA* **11**: 16.
McArthur E, Capra JA. 2021. Topologically associating domain boundaries that are stable across diverse cell types are evolutionarily constrained and enriched for heritability. *Am J Hum Genet* **108**: 269–283.
McCarthy FM, Bridges SM, Wang N, Magee GB, Williams WP, Luthe DS, Burgess SC. 2007. AgBase: a unified resource for functional analysis in agriculture. *Nucleic Acids Res* **35**: D599-603.
McClintock B. 1950. The origin and behavior of mutable loci in maize. *Proc Natl Acad Sci USA* **36**: 344–355.
McDonagh A, Fedorova ND, Crabtree J, Yu Y, Kim S, Chen D, Loss O, Cairns T, Goldman G, Armstrong-James D, et al. 2008. Sub-telomere directed gene expression during initiation of invasive aspergillosis. *PLoS Pathog* **4**: e1000154.
McDonald BA, Stukenbrock EH. 2016. Rapid emergence of pathogens in agro-ecosystems: global threats to agricultural sustainability and food security. *Philos Trans R Soc Lond B Biol Sci* **371**: 20160026.
McInnes L, Healy J, Melville J. 2018. Umap: Uniform manifold approximation and projection for dimension reduction. *arXiv* 1802.03426.
McKenna A, Hanna M, Banks E, Sivachenko A, Cibulskis K, Kernytsky A, Garimella K, Altshuler D, Gabriel S, Daly M, et al. 2010. The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Res* **20**: 1297–1303.
McNaught KJ, Wiles ET, Selker EU. 2020. Identification of a PRC2 accessory subunit required for subtelomeric H3K27 methylation in *Neurospora crassa. Mol Cell Biol* **40**: e00003-20.
Medlar AJ, Törönen P, Zosa E, Holm L. 2018. PANNZER 2: Annotate a complete proteome in minutes! *Nucl Acids Res* **43**: 24-29
Meile L, Peter J, Puccetti G, Alassimone J, McDonald BA, Sánchez-Vallet A. 2020. Chromatin dynamics contribute to the spatiotemporal expression pattern of virulence genes in a fungal plant pathogen. *MBio* **11**: 02343-20.
Mekhail K, Moazed D. 2010. The nuclear envelope in genome organization, expression and stability. *Nat Rev Mol Cell Biol* **11**: 317–328.
Meluh PB, Yang P, Glowczewski L, Koshland D, Smith MM. 1998. Cse4p is a component of the core centromere of *Saccharomyces cerevisiae. Cell* **94**: 607–613.
Meng H, Cao Y, Qin J, Song X, Zhang Q, Shi Y, Cao L. 2015. DNA methylation, its mediators and genome integrity. *Int J Biol Sci* **11**: 604–617.
Meng S, Liu Z, Shi H, Wu Z, Qiu J, Wen H, Lin F, Tao Z, Luo C, Kou Y. 2021. UvKmt6-mediated H3K27 trimethylation is required for development, pathogenicity, and stress response in *Ustilaginoidea virens*. *Virulence* **12**: 2972–2988.
Mercy G, Mozziconacci J, Scolari VF, Yang K, Zhao G. 2017. 3D organization of synthetic and scrambled chromosomes. *Science* **355**: eaaf4597
Meyers LA, Bull JJ. 2002. Fighting change with change: adaptive variation in an uncertain world. *Trends Ecol Evol* **17**: 551–557.
Miao VP, Freitag M, Selker EU. 2000. Short TpA-rich segments of the zeta-eta region induce DNA methylation in *Neurospora crassa*. *J Mol Biol* **300**: 249–273.
Milgroom MG, Jiménez-Gasco M del M, Olivares García C, Drott MT, Jiménez-Díaz RM. 2014. Recombination between clonal lineages of the asexual fungus *Verticillium dahliae* detected by genotyping by sequencing. *PLoS One* **9**: e106740.
Miller T, Krogan NJ, Dover J, Erdjument-Bromage H, Tempst P, Johnston M, Greenblatt JF, Shilatifard A. 2001. COMPASS: a complex of proteins associated with a trithorax-related SET domain protein. *Proc Natl Acad Sci U S A* **98**: 12902–12907.
Minh BQ, Schmidt HA, Chernomor O, Schrempf D, Woodhams MD, von Haeseler A, Lanfear R. 2020. IQ-TREE 2: New models and efficient methods for phylogenetic inference in the genomic era. *Mol Biol Evol* **37**: 1530–1534.
Minow MAA, Colasanti J. 2020. Does variable epigenetic inheritance fuel plant evolution? *Genome* **63**: 253–262.

Mirzaei H, Knijnenburg TA, Kim B, Robinson M, Picotti P, Carter GW, Li S, Dilworth DJ, Eng JK, Aitchison JD. 2013. Systematic measurement of transcription factor-DNA interactions by targeted mass spectrometry identifies candidate gene regulatory proteins. *Proc Natl Acad Sci USA* **110**: 3645–3650.

Misteli T. 2007. Beyond the sequence: cellular organization of genome function. *Cell* **128**: 787–800.

Mita P, Boeke JD. 2016. How retrotransposons shape genome regulation. *Curr Opin Genet Dev* **37**: 90–100.

Miyoshi T, Kanoh J, Saito M, Ishikawa F. 2008. Fission yeast Pot1-Tpp1 protects telomeres and regulates telomere length. *Science* **320**: 1341–1344.

Mizuguchi T, Fudenberg G, Mehta S, Belton J-M, Taneja N, Folco HD, FitzGerald P, Dekker J, Mirny L, Barrowman J, et al. 2014. Cohesin-dependent globules and heterochromatin shape 3D genome architecture in *S. pombe*. *Nature* **516**: 432–435.

Mizuguchi T, Taneja N, Matsuda E, Belton J-M, FitzGerald P, Dekker J, Grewal SIS. 2017. Shelterin components mediate genome reorganization in response to replication stress. *Proc Natl Acad Sci U S A* **114**: 5479–5484.

Mohanta TK, Bae H. 2015. The diversity of fungal genome. *Biol Proced Online* **17**: 8.

Möller M, Habig M, Lorrain C, Feurtey A, Haueisen J, Fagundes WC, Alizadeh A, Freitag M, Stukenbrock EH. 2021. Recent loss of the Dim2 DNA methyltransferase decreases mutation rate in repeats and changes evolutionary trajectory in a fungal pathogen. *PLoS Genet* **17**: e1009448.

Möller M, Ridenour JB, Wright DF, Freitag M. 2023. H4K20me3 controls Ash1-mediated H3K36me3 and transcriptional silencing in facultative heterochromatin. *bioRxiv* 2022.11.25.517763.

Möller M, Schotanus K, Soyer JL, Haueisen J, Happ K, Stralucke M, Happel P, Smith KM, Connolly LR, Freitag M, et al. 2019. Destabilization of chromosome structure by histone H3 lysine 27 methylation. *PLoS Genet* **15**: e1008093.

Möller M, Stukenbrock EH. 2017. Evolution and genome architecture in fungal plant pathogens. *Nat Rev Microbiol* **15**: 756–771.

Monroe, Murray KD, Xian W, Carbonell-Bejerano P, Fenster CB, Weigel D. 2022a. Report of mutation biases mirroring selection in Arabidopsis thaliana unlikely to be entirely due to variant calling errors. *bioRxiv* 2022.08.21.504682.

Monroe, Srikant T, Carbonell-Bejerano P, Becker C, Lensink M, Exposito-Alonso M, Klein M, Hildebrandt J, Neumann M, Kliebenstein D, et al. 2022b. Mutation bias reflects natural selection in *Arabidopsis thaliana. Nature* **602**: 101–105.

Montanini B, Chen PY, Morselli M, Jaroszewicz A, Lopez D, Martin F, Ottonello S, Pellegrini M. 2014. Non-exhaustive DNA methylation-mediated transposon silencing in the black truffle genome, a complex fungal genome with massive repeat element content. *Genome Biology* **15**: 16.

Monteiro PT, Oliveira J, Pais P, Antunes M, Palma M, Cavalheiro M, Galocha M, Godinho CP, Martins LC, Bourbon N. 2020. YEASTRACT+: a portal for cross-species comparative genomics of transcription regulation in yeasts. *Nucleic Acids Res* **48**: D642–D649.

Morimoto M, Boerkoel CF. 2013. The role of nuclear bodies in gene expression and disease. *Biology* **2**: 976–1033.

Mukhopadhyay A, Deplancke B, Walhout AJM, Tissenbaum HA. 2008. Chromatin immunoprecipitation (ChIP) coupled to detection by quantitative real-time PCR to study transcription factor binding to DNA in *Caenorhabditis elegans*. *Nat Protoc* **3**: 698–709.

Muller H, Gil J Jr, Drinnenberg IA. 2019a. The impact of centromeres on spatial genome architecture. *Trends Genet* **35**: 565–578.

Muller H, Scolari VF, Agier N, Piazza A, Thierry A, Mercy G, Descorps-Declere S, Lazar-Stefanita L, Espeli O, Llorente B. 2018. Characterizing meiotic chromosomes' structure and pairing using a designer sequence optimized for Hi-C. *Mol Sys Biol* **14**: e8293.

Muller MC, Praz CR, Sotiropoulos AG, Menardo F, Kunz L, Schudel S, Oberhansli S, Poretti M, Wehrli A, Bourras S, et al. 2019b. A chromosome-scale genome assembly reveals a highly dynamic effector repertoire of wheat powdery mildew. *New Phytol* **221**: 2176–2189.

Murayama Y, Uhlmann F. 2014. Biochemical reconstitution of topological DNA binding by the cohesin ring. *Nature* **505**: 367–371.

Murray GGR, Balmer AJ, Herbert J, Hadjirin NF, Kemp CL, Matuszewska M, Bruchmann S, Hossain ASMM, Gottschalk M, Tucker AW, et al. 2021. Mutation rate dynamics reflect ecological change in an emerging zoonotic pathogen. *PLoS Genet* **17**: e1009864.

Musselman CA, Avvakumov N, Watanabe R, Abraham CG, Lalonde M-E, Hong Z, Allen C, Roy S, Nuñez JK, Nickoloff J, et al. 2012. Molecular basis for H3K36me3 recognition by the Tudor domain of PHF1. *Nat Struct Mol Biol* **19**: 1266–1272.

Muszewska A, Hoffman-Sommer M, Grynberg M. 2011. LTR retrotransposons in fungi. *PLoS One* **6**: e29425.

Muszewska A, Steczkiewicz K, Stepniewska-Dziubinska M, Ginalski K. 2019. Transposable elements contribute to fungal genes and impact fungal lifestyle. *Sci Rep* **9**: 4307.
Myler LR, Kinzig CG, Sasi NK, Zakusilo G, Cai SW, de Lange T. 2021. The evolution of metazoan shelterin. *Genes Dev* **35**: 1625–1641.
Nachtomy O, Shavit A, Yakhini Z. 2007. Gene expression and the concept of the phenotype. *Stud Hist Philos Biol Biomed Sci* **38**: 238–254.
Nai Y-S, Huang Y-C, Yen M-R, Chen P-Y. 2020. Diversity of fungal DNA methyltransferases and their association with DNA methylation patterns. *Front Microbiol* **11**: 616922.
Nair N, Shoaib M, Sørensen CS. 2017. Chromatin dynamics in genome stability: roles in suppressing endogenous DNA damage and facilitating DNA Repair. *Int J Mol Sci* **18**: s18071486.
Nakamura M, Gao Y, Dominguez AA, Qi LS. 2021. CRISPR technologies for precise epigenome editing. *Nat Cell Biol* **23**: 11–22.
Nakayashiki H, Nishimoto N, Ikeda K, Tosa Y, Mayama S. 1999. Degenerate MAGGY elements in a subgroup of *Pyricularia grisea*: a possible example of successful capture of a genetic invader by a fungal genome. *Mol Gen Genet* **261**: 958–966.
Nandakumar J, Cech TR. 2013. Finding the end: recruitment of telomerase to telomeres. *Nat Rev Mol Cell Biol* **14**: 69–82.
Naranjo-Ortiz MA, Gabaldón T. 2019a. Fungal evolution: diversity, taxonomy and phylogeny of the Fungi. *Biol Rev Camb Philos Soc* **94**: 2101–2137.
Naranjo-Ortiz MA, Gabaldón T. 2019b. Fungal evolution: major ecological adaptations and evolutionary transitions. *Biol Rev Camb Philos Soc* **94**: 1443–1476.
Nasmyth K. 2001. Disseminating the genome: joining, resolving, and separating sister chromatids during mitosis and meiosis. *Annu Rev Genet* **35**: 673–745.
Naumova N, Imakaev M, Fudenberg G, Zhan Y, Lajoie BR, Mirny LA, Dekker J. 2013. Organization of the mitotic chromosome. *Science* **342**: 948–953.
Nei M, Jin L. 1989. Variances of the average numbers of nucleotide substitutions within and between populations. *Mol Biol Evol* **6**: 290–300.
Nei M, Kumar S. 2000. *Molecular Evolution and Phylogenetics*. Oxford University Press.
Newcomb LL, Hall DD, Heideman W. 2002. AZF1 is a glucose-dependent positive regulator of CLN3 transcription in *Saccharomyces cerevisiae*. *Mol Cell Biol* **22**: 1607–1614.
Nguyen D, Peona V, Unneberg P, Suh A, Jern P, Johannesson H. 2022. Transposon- and genome dynamics in the fungal genus Neurospora: Insights from nearly gapless genome assemblies. *Fungal Genet Rep* **66**: 2020-09
Nguyen L-T, Schmidt HA, von Haeseler A, Minh BQ. 2015. IQ-TREE: a fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. *Mol Biol Evol* **32**: 268–274.
Nichols MH, Corces VG. 2021. Principles of 3D compartmentalization of the human genome. *Cell Rep* **35**: 109330.
Nicolas FE, Torres-Martinez S, Ruiz-Vazquez RM. 2013. Loss and retention of RNA interference in fungi and parasites. *PLoS Pathog* **9**: e1003089.
Nielsen H. 2017a. Predicting secretory proteins with SignalP. In *Protein function prediction*, pp. 59–73, Springer.
Nielsen H. 2017b. Predicting Secretory Proteins with SignalP. *Methods Mol Biol* **1611**: 59–73.
Nielsen PR, Nietlispach D, Mott HR, Callaghan J, Bannister A, Kouzarides T, Murzin AG, Murzina NV, Laue ED. 2002. Structure of the HP1 chromodomain bound to histone H3 methylated at lysine 9. *Nature* **416**: 103–107.
Nierman WC, Pain A, Anderson MJ, Wortman JR, Kim HS, Arroyo J, Berriman M, Abe K, Archer DB, Bermejo C, et al. 2005. Genomic sequence of the pathogenic and allergenic filamentous fungus Aspergillus fumigatus. *Nature* **438**: 1151–1156.
Niu L, Shen W, Shi Z, Tan Y, He N, Wan J, Sun J, Zhang Y, Huang Y, Wang W. 2021. Three-dimensional folding dynamics of the *Xenopus tropicalis* genome. *Nature Genet* 1–13.
Nocetti N, Whitehouse I. 2016. Nucleosome repositioning underlies dynamic gene expression. *Genes Dev* **30**: 660–672.
Nora EP, Goloborodko A, Valton A-L, Gibcus JH, Uebersohn A, Abdennur N, Dekker J, Mirny LA, Bruneau BG. 2017. Targeted degradation of CTCF decouples local insulation of chromosome domains from genomic compartmentalization. *Cell* **169**: 930-944.e22.
Nora EP, Lajoie BR, Schulz EG, Giorgetti L, Okamoto I, Servant N, Piolot T, Van Berkum NL, Meisig J, Sedat J. 2012. Spatial partitioning of the regulatory landscape of the X-inactivation centre. *Nature* **485**: 381–385.
Northcott PA, Lee C, Zichner T, Stütz AM, Erkek S, Kawauchi D, Shih DJH, Hovestadt V, Zapatka M, Sturm D, et al. 2014. Enhancer hijacking activates GFI1 family oncogenes in medulloblastoma. *Nature* **511**: 428–434.

Nottensteiner M, Zechmann B, McCollum C, Huckelhoven R. 2018. A barley powdery mildew fungus non-autonomous retrotransposon encodes a peptide that supports penetration success on barley. *J Exp Bot* **69**: 3745–3758.

Oberbeckmann E, Wolff M, Krietenstein N, Heron M, Ellins JL, Schmid A, Krebs S, Blum H, Gerland U, Korber P. 2019. Absolute nucleosome occupancy map for the *Saccharomyces cerevisiae* genome. *Genome Res* **29**: 1996–2009.

OECD, FAO. 2022. *OECD-FAO agricultural outlook 2022-2031*. Organization for Economic Co-operation and Development (OECD), Paris Cedex, France.

Oggenfuss U, Badet T, Wicker T, Hartmann FE, Singh NK, Abraham LN, Karisto P, Vonlanthen T, Mundt CC, McDonald BA, et al. 2020. A population-level invasion by transposable elements in a fungal pathogen. *Elife* **10**: e69249

Ohno M, Ando T, Priest DG, Kumar V, Yoshida Y, Taniguchi Y. 2019. Sub-nucleosomal genome structure reveals distinct nucleosome folding motifs. *Cell* **176**: 520-534.e25.

Ohno S. 2013. *Evolution by Gene Duplication*. Springer Science & Business Media.

Ohta T, Ina Y. 1995. Variation in synonymous substitution rates among mammalian genes and the correlation between synonymous and nonsynonymous divergences. *J Mol Evol* **41**: 717–720.

Ola M, O'Brien CE, Coughlan AY, Ma Q, Donovan PD, Wolfe KH, Butler G. 2020. Polymorphic centromere locations in the pathogenic yeast *Candida parapsilosis*. *Genome Res* **30**: 684–696.

Olins DE, Olins AL. 2003. Chromatin history: our view from the bridge. *Nat Rev Mol Cell Biol* **4**: 809–814.

Oliver KR, Greene WK. 2009. Transposable elements: Powerful facilitators of evolution. *Bioessays* **31**: 703–714.

Oluwadare O, Highsmith M, Cheng J. 2019. An overview of methods for reconstructing 3-D chromosome and genome structures from Hi-C data. *Biol Proced Online* **21**: 7.

Omrane S, Audeon C, Ignace A, Duplaix C, Aouini L, Kema G, Walker AS, Fillinger S. 2017. Plasticity of the MFS1 promoter leads to multidrug resistance in the wheat pathogen *Zymoseptoria tritici*. *mSphere* **2**: 00393-17.

Ong C-T, Corces VG. 2014. CTCF: an architectural protein bridging genome topology and function. *Nat Rev Genet* **15**: 234–246.

Ono T, Losada A, Hirano M, Myers MP, Neuwald AF, Hirano T. 2003. Differential contributions of condensin I and condensin II to mitotic chromosome architecture in vertebrate cells. *Cell* **115**: 109–121.

Ono T, Yamashita D, Hirano T. 2013. Condensin II initiates sister chromatid resolution during S phase. *J Cell Biol* **200**: 429–441.

Ou S, Su W, Liao Y, Chougule K, Agda JRA, Hellinga AJ, Lugo CSB, Elliott TA, Ware D, Peterson T, et al. 2019. Benchmarking transposable element annotation methods for creation of a streamlined, comprehensive pipeline. *Genome Biol* **20**: 275.

Pais M, Yoshida K, Giannakopoulou A, Pel MA, Cano LM, Oliva RF, Witek K, Lindqvist-Kreuze H, Vleeshouwers VGAA, Kamoun S. 2018. Gene expression polymorphism underpins evasion of host immunity in an asexual lineage of the Irish potato famine pathogen. *BMC Evol Biol* **18**: 93.

Pál C, Papp B, Hurst LD. 2001. Highly expressed genes in yeast evolve slowly. *Genetics* **158**: 927–931.

Palladino F, Laroche T, Gilson E, Axelrod A, Pillus L, Gasser SM. 1993. SIR3 and SIR4 proteins are required for the positioning and integrity of yeast telomeres. *Cell* **75**: 543–555.

Palm W, de Lange T. 2008. How shelterin protects mammalian telomeres. *Annu Rev Genet* **42**: 301–334.

Palmer JM, Stajich J. 2020. Funannotate v1. 8.1: Eukaryotic genome annotation. *Zenodo https://doi org/105281/zenodo*.

Pan C, Sretenovic S, Qi Y. 2021. CRISPR/dCas-mediated transcriptional and epigenetic regulation in plants. *Curr Opin Plant Biol* **60**: 101980.

Papikian A, Liu W, Gallego-Bartolomé J, Jacobsen SE. 2019. Site-specific manipulation of *Arabidopsis* loci using CRISPR-Cas9 SunTag systems. *Nat Commun* **10**: 729.

Papkou A, Guzella T, Yang W, Koepper S, Pees B, Schalkowski R, Barg MC, Rosenstiel PC, Teotonio H, Schulenburg H. 2019. The genomic basis of Red Queen dynamics during rapid reciprocal host-pathogen coevolution. *Proc Natl Acad Sci USA* **116**: 923–928.

Parada LA, McQueen PG, Misteli T. 2004. Tissue-specific spatial organization of genomes. *Genome Biol* **5**: R44.

Parsch J, Ellegren H. 2013. The evolutionary causes and consequences of sex-biased gene expression. *Nat Rev Genet* **14**: 83–87.

Patel L, Kang R, Rosenberg SC, Qiu Y, Raviram R, Chee S, Hu R, Ren B, Cole F, Corbett KD. 2019. Dynamic reorganization of the genome shapes the recombination landscape in meiotic prophase. *Nat Struct Mol Biol* **26**: 164–174.

Payne JL, Wagner A. 2018. The causes of evolvability and their evolution. *Nat Rev Genet* **20**: 24–38.

Pellacani D, Bilenky M, Kannan N, Heravi-Moussavi A, Knapp DJHF, Gakkhar S, Moksa M, Carles A, Moore R, Mungall AJ, et al. 2016. Analysis of normal human mammary epigenomes reveals cell-specific active enhancer states and associated transcription factor networks. *Cell Rep* **17**: 2060–2074.

Penagos-Puig A, Furlan-Magaril M. 2020. Heterochromatin as an Important Driver of Genome Organization. *Front Cell Dev Biol* **8**: 579137.
Peng Z, Oliveira-Garcia E, Lin G, Hu Y, Dalby M, Migeon P, Tang H, Farman M, Cook D, White FF, et al. 2019. Effector gene reshuffling involves dispensable mini-chromosomes in the wheat blast fungus. *PLoS Genet* **15**: e1008272.
Pennacchio LA, Bickmore W, Dean A, Nobrega MA, Bejerano G. 2013. Enhancers: five essential questions. *Nat Rev Genet* **14**: 288–295.
Pentzold C, Kokal M, Pentzold S, Weise A. 2021. Sites of chromosomal instability in the context of nuclear architecture and function. *Cell Mol Life Sci* **78**: 2095–2103.
Pepper JW. 2003. The evolution of evolvability in genetic linkage patterns. *Biosystems* **69**: 115–126.
Peric-Hupkes D, van Steensel B. 2008. Linking cohesin to gene regulation. *Cell* **132**: 925–928.
Petela NJ, Gligoris TG, Metson J, Lee B-G, Voulgaris M, Hu B, Kikuchi S, Chapard C, Chen W, Rajendra E, et al. 2018. Scc2 is a potent activator of cohesin's ATPase that promotes loading by binding Scc1 without Pds5. *Mol Cell* **70**: 1134-1148.e7.
Peter J, De Chiara M, Friedrich A, Yue JX, Pflieger D, Bergstrom A, Sigwalt A, Barre B, Freel K, Llored A, et al. 2018. Genome evolution across 1,011 *Saccharomyces cerevisiae* isolates. *Nature* **556**: 339–344.
Peter M, Kohler A, Ohm RA, Kuo A, Krutzmann J, Morin E, Arend M, Barry KW, Binder M, Choi C, et al. 2016. Ectomycorrhizal ecology is imprinted in the genome of the dominant symbiotic fungus *Cenococcum geophilum*. *Nature Commun* **7**: 12662.
Pfannenstiel BT, Keller NP. 2019. On top of biosynthetic gene clusters: How epigenetic machinery influences secondary metabolism in fungi. *Biotech Adv* **37**: 107345.
Pfeifer B, Wittelsbürger U, Ramos-Onsins SE, Lercher MJ. 2014. PopGenome: an efficient Swiss army knife for population genomic analyses in R. *Mol Biol Evol* **31**: 1929–1936.
Pichugina T, Sugawara T, Kaykov A, Schierding W, Masuda K, Uewaki J, Grand RS, Allison JR, Martienssen RA, Nurse P, et al. 2016. A diffusion model for the coordination of DNA replication in *Schizosaccharomyces pombe*. *Sci Rep* **6**: 18757.
Pidoux AL, Allshire RC. 2004. Kinetochore and heterochromatin domains of the fission yeast centromere. *Chromosome Res* **12**: 521–534.
Piégu B, Bire S, Arensburger P, Bigot Y. 2015. A survey of transposable element classification systems--a call for a fundamental update to meet the challenge of their diversity and complexity. *Mol Phylogenet Evol* **86**: 90–109.
Pigliucci M. 2008. Is evolvability evolvable? *Nat Rev Genet* **9**: 75–82.
Pignatelli M, Serras F, Moya A, Guigó R, Corominas M. 2009. CROC: finding chromosomal clusters in eukaryotic genomes. *Bioinformatics* **25**: 1552–1553.
Pitt CW, Valente LP, Rhodes D, Simonsson T. 2008. Identification and characterization of an essential telomeric repeat binding factor in fission yeast. *J Biol Chem* **283**: 2693–2701.
Plissonneau C, Hartmann FE, Croll D. 2018. Pangenome analyses of the wheat pathogen *Zymoseptoria tritici* reveal the structural basis of a highly plastic eukaryotic genome. *BMC Biol* **16**: 5.
Plissonneau C, Stürchler A, Croll D. 2016. The evolution of orphan regions in genomes of a fungal pathogen of wheat. *mBio* **7**: e01231-16.
Ploetz RC. 2015. Fusarium Wilt of Banana. *Phytopathology* **105**: 1512–1521.
Ploetz RC. 2000. Panama Disease: A classic and destructive disease of Banana. *Plant Health Prog* **1**: 10.
Pokholok DK, Harbison CT, Levine S, Cole M, Hannett NM, Lee TI, Bell GW, Walker K, Rolfe PA, Herbolsheimer E, et al. 2005. Genome-wide map of nucleosome acetylation and methylation in yeast. *Cell* **122**: 517–527.
Polotnianka RM, Li J, Lustig AJ. 1998. The yeast Ku heterodimer is essential for protection of the telomere against nucleolytic and recombinational activities. *Curr Biol* **8**: 831–834.
Ponnusamy S, Alderson NL, Hama H, Bielawski J, Jiang JC, Bhandari R, Snyder SH, Jazwinski SM, Ogretmen B. 2008. Regulation of telomere length by fatty acid elongase 3 in yeast. Involvement of inositol phosphate metabolism and Ku70/80 function. *J Biol Chem* **283**: 27514–27524.
Pope BD, Ryba T, Dileep V, Yue F, Wu W, Denas O, Vera DL, Wang Y, Hansen RS, Canfield TK. 2014. Topologically associating domains are stable units of replication-timing regulation. *Nature* **515**: 402–405.
Poplin R, Chang P-C, Alexander D, Schwartz S, Colthurst T, Ku A, Newburger D, Dijamco J, Nguyen N, Afshar PT. 2018. A universal SNP and small-indel variant caller using deep neural networks. *Nature Biotech***36**: 983–987.
Porto BN, Caixeta ET, Mathioni SM, Vidigal PMP, Zambolim L, Zambolim EM, Donofrio N, Polson SW, Maia TA, Chen C, et al. 2019. Genome sequencing and transcript analysis of *Hemileia vastatrix* reveal expression dynamics of candidate effectors dependent on host compatibility. *PLoS One* **14**: e0215598.
Prendergast JGD, Campbell H, Gilbert N, Dunlop MG, Bickmore WA, Semple CAM. 2007. Chromatin structure and evolution in the human genome. *BMC Evolutionary Biology* **7**: 72.

Previato de Almeida L, Evatt JM, Chuong HH, Kurdzo EL, Eyster CA, Gladstone MN, Gómez-H L, Llano E, Meyer R, Pendas AM, et al. 2019. Shugoshin protects centromere pairing and promotes segregation of nonexchange partner chromosomes in meiosis. *Proc Natl Acad Sci U S A* **116**: 9417–9422.

Price PD, Palmer Droguett DH, Taylor JA, Kim DW, Place ES, Rogers TF, Mank JE, Cooney CR, Wright AE. 2022. Detecting signatures of selection on gene expression. *Nat Ecol Evol* **6**: 1035–1045.

Pugacheva EM, Kubo N, Loukinov D, Tajmul M, Kang S, Kovalchuk AL, Strunnikov AV, Zentner GE, Ren B, Lobanenkov VV. 2020. CTCF mediates chromatin looping via N-terminal domain-dependent cohesin retention. *Proc Natl Acad Sci U S A* **117**: 2020–2031.

Putri GH, Anders S, Pyl PT, Pimanda JE, Zanini F. 2022. Analysing high-throughput sequencing data in Python with HTSeq 2.0. *Bioinformatics* **38**: 2943–2945.

Qi L, Sui Y, Tang X-X, McGinty RJ, Liang X-Z, Dominska M, Zhang K, Mirkin SM, Zheng D-Q, Petes TD. 2023. Shuffling the yeast genome using CRISPR/Cas9-generated DSBs that target the transposable Ty1 elements. *PLoS Genet* **19**: e1010590.

Qi X, Li Y, Honda S, Hoffmann S, Marz M, Mosig A, Podlevsky JD, Stadler PF, Selker EU, Chen JJ-L. 2013. The common ancestral core of vertebrate and fungal telomerase RNAs. *Nucleic Acids Res* **41**: 450–462.

Quinlan AR, Hall IM. 2010. BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics* **26**: 841–842.

Quiroz D, Zhao K, Carbonell-Bejerano P, Monroe G. 2022. Relationship between biased mutagenesis and H3K4me1-targeted DNA repair in plants. bioRxiv 2022-08

Radman-Livaja M, Rando OJ. 2010. Nucleosome positioning: How is it established, and why does it matter? *Dev Biol* **339**: 258–266.

Raffaele S, Farrer RA, Cano LM, Studholme DJ, MacLean D, Thines M, Jiang RHY, Zody MC, Kunjeti SG, Donofrio NM, et al. 2010. Genome evolution following host jumps in the Irish potato famine pathogen lineage. *Science* **330**: 1540–1543.

Raffaele S, Kamoun S. 2012. Genome evolution in filamentous plant pathogens: why bigger can be better. *Nat Rev Microbiol* **10**: 417–430.

Ragab A, Travers A. 2003. HMG-D and histone H1 alter the local accessibility of nucleosomal DNA. *Nucleic Acids Res* **31**: 7083–7089.

Rahnama M, Wang B, Dostart J, Novikova O, Yackzan D, Yackzan A, Bruss H, Baker M, Jacob H, Zhang X, et al. 2021. Telomere roles in fungal genome evolution and adaptation. *Front Genet* **12**: 676751.

Ramírez F, Bhardwaj V, Arrigoni L, Lam KC, Grüning BA, Villaveces J, Habermann B, Akhtar A, Manke T. 2018. High-resolution TADs reveal DNA sequences underlying genome organization in flies. *Nat Commun* **9**: 189.

Ramírez F, Dündar F, Diehl S, Grüning BA, Manke T. 2014. deepTools: a flexible platform for exploring deep-sequencing data. *Nucleic Acids Res* **42**: W187-91.

Ramírez F, Ryan DP, Grüning B, Bhardwaj V, Kilpert F, Richter AS, Heyne S, Dündar F, Manke T. 2016. deepTools2: a next generation web server for deep-sequencing data analysis. *Nucleic Acids Res* **44**: W160-5.

Rao SSP, Huang S-C, Glenn St Hilaire B, Engreitz JM, Perez EM, Kieffer-Kwon K-R, Sanborn AL, Johnstone SE, Bascom GD, Bochkov ID, et al. 2017. Cohesin Loss Eliminates All Loop Domains. *Cell* **171**: 305-320.e24.

Rao SSP, Huntley MH, Durand NC, Stamenova EK, Bochkov ID, Robinson JT, Sanborn AL, Machol I, Omer AD, Lander ES, et al. 2014. A 3D map of the human genome at kilobase resolution reveals principles of chromatin looping. *Cell* **159**: 1665–1680.

Rausch T, Zichner T, Schlattl A, Stütz AM, Benes V, Korbel JO. 2012. DELLY: structural variant discovery by integrated paired-end and split-read analysis. *Bioinformatics* **28**: i333–i339.

Rebollo R, Karimi MM, Bilenky M, Gagnier L, Miceli-Royer K, Zhang Y, Goyal P, Keane TM, Jones S, Hirst M, et al. 2011. Retrotransposon-induced heterochromatin spreading in the mouse revealed by insertional polymorphisms. *PLOS Genetics* **7**: e1002301.

Renz M. 2013. Fluorescence microscopy-a historical and technical perspective. *Cytometry A* **83**: 767–779.

Reusche M, Truskina J, Thole K, Nagel L, Rindfleisch S, Tran VT, Braus-Stromeyer SA, Braus GH, Teichmann T, Lipka V. 2014. Infections with the vascular pathogens *Verticillium longisporum* and *Verticillium dahliae* induce distinct disease symptoms and differentially affect drought stress tolerance of *Arabidopsis thaliana. Environ Exp Bot* **108**: 23–37.

Rhind N, Gilbert DM. 2013. DNA replication timing. *Cold Spring Harb Perspect Biol* **5**: a010132.

Riederer JM, Tiso S, van Eldijk TJB, Weissing FJ. 2022. Capturing the facets of evolvability in a mechanistic framework. *Trends Ecol Evol* **37**: 430–439.

Ristaino JB, Anderson PK, Bebber DP, Brauman KA, Cunniffe NJ, Fedoroff NV, Finegold C, Garrett KA, Gilligan CA, Jones CM, et al. 2021. The persistent threat of emerging plant disease pandemics to global food security. *Proc Natl Acad Sci U S A* **118**: 2022239118.

Rizzo DM, Lichtveld M, Mazet JAK, Togami E, Miller SA. 2021. Plant health and its effects on food safety and security in a One Health framework: four case studies. *One Health Outlook* **3**: 6.
Roach JC, Glusman G, Rowen L, Kaur A, Purcell MK, Smith KD, Hood LE, Aderem A. 2005. The evolution of vertebrate Toll-like receptors. *Proc Natl Acad Sci U S A* **102**: 9577–9582.
Roadmap Epigenomics, Kundaje A, Meuleman W, Ernst J, Bilenky M, Yen A, Heravi-Moussavi A, Kheradpour P, Zhang Z, Wang J, et al. 2015. Integrative analysis of 111 reference human epigenomes. *Nature* **518**: 317–330.
Robinson JT, Thorvaldsdóttir H, Winckler W, Guttman M, Lander ES, Getz G, Mesirov JP. 2011. Integrative genomics viewer. *Nat Biotechnol* **29**: 24–26.
Robinson MD, McCarthy DJ, Smyth GK. 2010. edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics* **26**: 139–140.
Robinson MD, Oshlack A. 2010. A scaling normalization method for differential expression analysis of RNA-seq data. *Genome Biol* **11**: R25.
Rocha EPC, Danchin A. 2004. An analysis of determinants of amino acids substitution rates in bacterial proteins. *Mol Biol Evol* **21**: 108–116.
Rocha PP, Raviram R, Bonneau R, Skok JA. 2015. Breaking TADs: insights into hierarchical genome organization. *Epigenomics* **7**: 523–526.
Rodriguez S, Ward A, Reckard AT, Shtanko Y, Hull-Crew C, Klocko AD. 2022. The genome organization of *Neurospora crassa* at high resolution uncovers principles of fungal chromosome topology. *G3 (Bethesda)* **12**.: jkac053.
Rodríguez-Trelles F, Tarrío R, Ayala FJ. 2003. Evolution of *cis*-regulatory regions versus codifying regions. *Int J Dev Biol* **47**: 665–673.
Roguev A, Schaft D, Shevchenko A, Pijnappel WW, Wilm M, Aasland R, Stewart AF. 2001. The *Saccharomyces cerevisiae* Set1 complex includes an Ash2 homologue and methylates histone 3 lysine 4. *EMBO J* **20**: 7137–7148.
Rojas-Rojas FU, Vega-Arreguín JC. 2021. Epigenetic insight into regulatory role of chromatin covalent modifications in lifecycle and virulence of *Phytophthora*. *Environmental Microbiology Reports* **13**: 445-57
Rolef Ben-Shahar T, Heeger S, Lehane C, East P, Flynn H, Skehel M, Uhlmann F. 2008. Eco1-dependent cohesin acetylation during establishment of sister chromatid cohesion. *Science* **321**: 563–566.
Romero IG, Ruvinsky I, Gilad Y. 2012. Comparative studies of gene expression and the evolution of gene regulation. *Nat Rev Genet* **13**: 505–516.
Rountree MR, Selker EU. 2010. DNA methylation and the formation of heterochromatin in *Neurospora crassa*. *Heredity* **105**: 38–44.
Rousk J, Bååth E. 2011. Growth of saprotrophic fungi and bacteria in soil. *FEMS Microbiol Ecol* **78**: 17–30.
Rouxel T, Grandaubert J, Hane JK, Hoede C, van de Wouw AP, Couloux A, Dominguez V, Anthouard V, Bally P, Bourras S, et al. 2011. Effector diversification within compartments of the *Leptosphaeria maculans* genome affected by Repeat-Induced Point mutations. *Nat Commun* **2**: 202.
Rovenich H, Boshoven JC, Thomma BPHJ. 2014. Filamentous pathogen effector functions: of pathogens, hosts and microbiomes. *Curr Opin Plant Biol* **20**: 96–103.
Rowley MJ, Corces VG. 2018. Organizational principles of 3D genome architecture. *Nature Rev Genet* **19**: 789–800.
Rowley MJ, Nichols MH, Lyu X, Ando-Kuri M, Rivera ISM, Hermetz K, Wang P, Ruan Y, Corces VG. 2017. Evolutionarily Conserved Principles Predict 3D Chromatin Organization. *Mol Cell* **67**: 837-852.e7.
Ryba T, Hiratani I, Lu J, Itoh M, Kulik M, Zhang J, Schulz TC, Robins AJ, Dalton S, Gilbert DM. 2010. Evolutionarily conserved replication timing profiles predict long-range chromatin interactions and distinguish closely related cell types. *Genome Res* **20**: 761–770.
Sabelleck B, Panstruga R. 2018. Novel jack-in-the-box effector of the barley powdery mildew pathogen? *J Exp Bot* **69**: 3511–3514.
Saha A, Wittmeyer J, Cairns BR. 2006. Chromatin remodelling: the industrial revolution of DNA around histones. *Nat Rev Mol Cell Biol* **7**: 437–447.
Sakamoto T, Miura K, Itoh H, Tatsumi T, Ueguchi-Tanaka M, Ishiyama K, Kobayashi M, Agrawal GK, Takeda S, Abe K, et al. 2004. An overview of gibberellin metabolism enzyme genes and their related mutants in rice. *Plant Physiol* **134**: 1642–1653.
Saksouk N, Simboeck E, Déjardin J. 2015. Constitutive heterochromatin formation and transcription in mammals. *Epigenetics Chromatin* **8**: 3.
Sanborn AL, Rao SSP, Huang S-C, Durand NC, Huntley MH, Jewett AI, Bochkov ID, Chinnappan D, Cutkosky A, Li J, et al. 2015. Chromatin extrusion explains key features of loop and domain formation in wild-type and engineered genomes. *Proc Natl Acad Sci U S A* **112**: E6456-65.
Sánchez-Vallet A, Fouché S, Fudal I, Hartmann FE, Soyer JL, Tellier A, Croll D. 2018. The genome biology of effector gene evolution in filamentous plant pathogens. *Annu Rev Phytopathol* **56**: 21–40.

Santini A, Ghelardini L. 2015. Plant pathogen evolution and climate change. *CAB Rev Perspect Agric Vet Sci Nutr Nat Resour* **2015**: 1–8.

Sasaki T, Lynch KL, Mueller CV, Friedman S, Freitag M, Lewis ZA. 2014. Heterochromatin controls γH2A localization in *Neurospora crassa. Eukaryot Cell* **13**: 990–1000.

Sawyer IA, Dundr M. 2016. Nuclear bodies: Built to boost. *J Cell Biol* **213**: 509–511.

Schalbetter SA, Fudenberg G, Baxter J, Pollard KS, Neale MJ. 2019. Principles of meiotic chromosome assembly revealed in *S. cerevisiae*. *Nature Commun* **10**: 1–12.

Schalbetter SA, Goloborodko A, Fudenberg G, Belton J-M, Miles C, Yu M, Dekker J, Mirny L, Baxter J. 2017. SMC complexes differentially compact mitotic chromosomes according to genomic context. *Nat Cell Biol* **19**: 1071–1080.

Schep R, Brinkman EK, Leemans C, Vergara X, van der Weide RH, Morris B, van Schaik T, Manzo SG, Peric-Hupkes D, van den Berg J, et al. 2021. Impact of chromatin context on Cas9-induced DNA double-strand break repair pathway balance. *Mol Cell* **81**: 2216-2230.e10.

Schirawski J, Mannhaupt G, Münch K, Brefort T, Schipper K, Doehlemann G, Di Stasio M, Rössel N, Mendoza-Mendoza A, Pester D, et al. 2010. Pathogenicity determinants in smut fungi revealed by genome comparison. *Science* **330**: 1546–1548.

Schleiffer A, Kaitna S, Maurer-Stroh S, Glotzer M, Nasmyth K, Eisenhaber F. 2003. Kleisins: a superfamily of bacterial and eukaryotic SMC protein partners. *Mol Cell* **11**: 571–575.

Schlichter A, Kasten MM, Parnell TJ, Cairns BR. 2020. Specialization of the chromatin remodeler RSC to mobilize partially-unwrapped nucleosomes. *Elife* **9**: e58130.

Schmidt D, Wilson MD, Ballester B, Schwalie PC, Brown GD, Marshall A, Kutter C, Watt S, Martinez-Jimenez CP, Mackay S, et al. 2010. Five-vertebrate ChIP-seq reveals the evolutionary dynamics of transcription factor binding. *Science* **328**: 1036–1040.

Schnable PS, Ware D, Fulton RS, Stein JC, Wei F, Pasternak S, Liang C, Zhang J, Fulton L, Graves TA, et al. 2009. The B73 maize genome: complexity, diversity, and dynamics. *Science* **326**: 1112–1115.

Schober H, Kalck V, Vega-Palas MA, Van Houwe G, Sage D, Unser M, Gartenberg MR, Gasser SM. 2008. Controlled exchange of chromosomal arms reveals principles driving telomere interactions in yeast. *Genome Res* **18**: 261–271.

Schoenfelder S, Fraser P. 2019. Long-range enhancer–promoter contacts in gene expression control. *Nat Rev Genet* **20**: 437–455.

Schoenfelder S, Sexton T, Chakalova L, Cope NF, Horton A, Andrews S, Kurukuti S, Mitchell JA, Umlauf D, Dimitrova DS, et al. 2010. Preferential associations between co-regulated genes reveal a transcriptional interactome in erythroid cells. *Nat Genet* **42**: 53–61.

Scholthof K-BG. 2007. The disease triangle: pathogens, the environment and society. *Nat Rev Microbiol* **5**: 152–156.

Schotanus K, Soyer JL, Connolly LR. 2015. Histone modifications rather than the novel regional centromeres of *Zymoseptoria tritici* distinguish core and accessory chromosomes. *Epigenetics & Chromatin* 8:41.

Schotanus K, Yadav V, Heitman J. 2021. Epigenetic dynamics of centromeres and neocentromeres in *Cryptococcus deuterogattii*. *PLoS Genet* **17**: e1009743.

Schrader L, Helantera H, Oettler J. 2017. Accelerated evolution of developmentally biased genes in the tetraphenic ant *Cardiocondyla obscurior*. *Mol Biol Evol* **34**: 535–544.

Schrader L, Schmitz J. 2019. The impact of transposable elements in adaptive evolution. *Mol Ecol* **28**: 1537–1549.

Schuster-Böckler B, Lehner B. 2012. Chromatin organization is a major influence on regional mutation rates in human cancer cells. *Nature* **488**: 504–507.

Schweizer G, Münch K, Mannhaupt G, Schirawski J, Kahmann R, Dutheil JY. 2018. Positively selected effector genes and their contribution to virulence in the smut fungus *Sporisorium reilianum*. *Genome Biol Evol* **10**: 629–645.

Schwessinger B, Chen YJ, Tien R, Vogt JK, Sperschneider J, Nagar R, McMullan M, Sicheritz-Ponten T, Sorensen CK, Hovmoller MS, et al. 2020. Distinct life histories impact dikaryotic genome evolution in the rust fungus *Puccinia striiformis* causing stripe rust in wheat. *Genome Biol Evol* **12**: 597–617.

Schwessinger B, Sperschneider J, Cuddy WS, Garnica DP, Miller ME, Taylor JM, Dodds PN, Figueroa M, Park RF, Rathjen JP. 2018. A near-complete haplotype-phased genome of the dikaryotic wheat stripe rust fungus *Puccinia striiformis f. sp. tritici* reveals high interhaplotype diversity. *mBio* **9**: 29463659.

Seidl MF, Cook DE, Thomma BPHJ. 2016. Chromatin biology impacts adaptive evolution of filamentous plant pathogens. *PLoS Pathog* **12**: e1005920.

Seidl MF, Faino L, Shi-Kunne X, van den Berg GC, Bolton MD, Thomma BP. 2015. The genome of the saprophytic fungus *Verticillium tricorpus* reveals a complex effector repertoire resembling that of its pathogenic relatives. *Mol Plant Microbe Interact* **28**: 362–373.

Seidl MF, Kramer HM, Cook DE, Fiorin GL, van den Berg GCM, Faino L, Thomma BPHJ. 2020. Repetitive elements contribute to the diversity and evolution of centromeres in the fungal genus Verticillium. *MBio* **11**: e01714-20.
Seidl MF, Thomma B. 2014. Sex or no sex: evolutionary adaptation occurs regardless. *Bioessays*. **36**: 335-345
Seidl MF, Thomma BPHJ. 2017. Transposable elements direct the coevolution between plants and microbes. *Trends Genet* **33**: 842–851.
Selker E. 1990. Premeiotic instability of repeated sequences in *Neurospora crassa*. *Annu Rev Genet* **24**: 579–613.
Selker EU, Stevens JN. 1987. Signal for DNA methylation associated with tandem duplication in *Neurospora crassa*. *Mol Cell Biol* **7**: 1032–1038.
Selker EU, Tountas NA, Cross SH, Margolin BS, Murphy JG, Bird AP, Freitag M. 2003. The methylated component of the *Neurospora crassa* genome. *Nature* **422**: 893–897.
Selosse M-A, Schneider-Maunoury L, Martos F. 2018. Time to re-think fungal ecology? Fungal ecological niches are often prejudged. *New Phytol* **217**: 968–972.
Sentmanat MF, Elgin SCR. 2012. Ectopic assembly of heterochromatin in *Drosophila melanogaster* triggered by transposable elements. *Proc Natl Acad Sci U S A* **109**: 14104–14109.
Sepsiova R, Necasova I, Willcox S, Prochazkova K, Gorilak P, Nosek J, Hofr C, Griffith JD, Tomaska L. 2016. Evolution of telomeres in *Schizosaccharomyces pombe* and its possible relationship to the diversification of telomere binding proteins. *PLoS One* **11**: e0154225.
Sexton T, Cavalli G. 2015. The role of chromosome domains in shaping the functional genome. *Cell* **160**: 1049–1059.
Sexton T, Yaffe E, Kenigsberg E, Bantignies F, Leblanc B, Hoichman M, Parrinello H, Tanay A, Cavalli G. 2012. Three-dimensional folding and functional organization principles of the *Drosophila* genome. *Cell* **148**: 458–472.
Shaban M, Miao Y, Ullah A, Khan AQ, Menghwar H, Khan AH, Ahmed MM, Tabassum MA, Zhu L. 2018. Physiological and molecular mechanism of defense in cotton against *Verticillium dahliae*. *Plant Physiol Biochem* **125**: 193–204.
Shastry BS. 2009. SNPs: impact on gene function and phenotype. *Methods Mol Biol* **578**: 3–22.
Shatkin AJ, Tatum EL. 1959. Electron microscopy of *Neurospora crassa* mycelia. *J Biophys Biochem Cytol* **6**: 423–426.
Shi-Kunne X, Faino L, van den Berg GCM, Thomma BPHJ, Seidl MF. 2018. Evolution within the fungal genus Verticillium is characterized by chromosomal rearrangement and gene loss. *Environ Microbiol* **20**: 1362–1373.
Short DPG, Gurung S, Hu X, Inderbitzin P, Subbarao KV. 2014. Maintenance of sex-related genes and the co-occurrence of both mating types in *Verticillium dahliae*. *PLoS One* **9**: e112145.
Siepel A, Bejerano G, Pedersen JS, Hinrichs AS, Hou M, Rosenbloom K, Clawson H, Spieth J, Hillier LW, Richards S. 2005. Evolutionarily conserved elements in vertebrate, insect, worm, and yeast genomes. *Genome Res* **15**: 1034–1050.
Sims RJ 3rd, Reinberg D. 2009. Processing the H3K36me3 signature. *Nat Genet* **41**: 270–271.
Sinzelle L, Izsvák Z, Ivics Z. 2009. Molecular domestication of transposable elements: from detrimental parasites to useful host genes. *Cell Mol Life Sci* **66**: 1073–1093.
Skibbens RV. 2019. Condensins and cohesins - one of these things is not like the other! *J Cell Sci* **132**: jcs.220491.
Slattery MG, Liko D, Heideman W. 2006. The function and properties of the Azf1 transcriptional regulator change with growth conditions in *Saccharomyces cerevisiae*. *Eukaryotic cell* **5**: 313–320.
Slotkin RK, Martienssen R. 2007. Transposable elements and the epigenetic regulation of the genome. *Nat Rev Genet* **8**: 272–285.
Smith KM, Dobosy JR, Reifsnyder JE, Rountree MR, Anderson DC, Green GR, Selker EU. 2010. H2B- and H3-specific histone deacetylases are required for DNA methylation in *Neurospora crassa*. *Genetics* **186**: 1207–1216.
Smith KM, Galazka JM, Phatale PA, Connolly LR, Freitag M. 2012. Centromeres of filamentous fungi. *Chromosome research* **20**: 635–656.
Smogorzewska A, de Lange T. 2004. Regulation of telomerase by telomeric proteins. *Annu Rev Biochem* **73**: 177–208.
Smolle M, Workman JL. 2013. Transcription-associated histone modifications and cryptic transcription. *Biochim Biophys Acta* **1829**: 84–97.
Snelders NC, Boshoven JC, Song Y, Schmitz N. 2023. A highly polymorphic effector protein promotes fungal virulence through suppression of plant-associated Actinobacteria. *New Phytol* **237**: 944-958
Snelders NC, Kettles GJ, Rudd JJ, Thomma BPHJ. 2018. Plant pathogen effector proteins as manipulators of host microbiomes? *Mol Plant Pathol* **19**: 257–259.
Snelders NC, Rovenich H, Petti GC, Rocafort M, van den Berg GCM, Vorholt JA, Mesters JR, Seidl MF, Nijland R, Thomma BPHJ. 2020. Microbiome manipulation by a soil-borne fungal plant pathogen using effector proteins. *Nat Plants* **6**: 1365–1374.
Sniegowski PD, Gerrish PJ, Johnson T, Shaver A. 2000. The evolution of mutation rates: separating causes from consequences. *Bioessays* **22**: 1057–1066.

Solovei I, Thanisch K, Feodorova Y. 2016. How to rule the nucleus: divide et impera. *Curr Opin Cell Biol* **40**: 47–59.
Song R, Li J, Xie C, Jian W, Yang X. 2020. An overview of the molecular genetics of plant resistance to the Verticillium Wilt pathogen V*erticillium dahliae*. *Int J Mol Sci* **21**: ijms21031120.
Soyer JL, Clairet C, Gay EJ, Lapalu N, Rouxel T, Stukenbrock EH, Fudal I. 2021. Genome-wide mapping of histone modifications during axenic growth in two species of *Leptosphaeria maculans* showing contrasting genomic organization. *Chromosome Res* **29**: 219–236.
Soyer JL, El Ghalid M, Glaser N, Ollivier B, Linglin J, Grandaubert J, Balesdent M-H, Connolly LR, Freitag M, Rouxel T, et al. 2014. Epigenetic control of effector gene expression in the plant pathogenic fungus *Leptosphaeria maculans*. *PLoS Genet* **10**: e1004227.
Soyer JL, Grandaubert J, Haueisen J, Schotanus K, Stukenbrock EH. 2019. In planta chromatin immunoprecipitation in Zymoseptoria tritici reveals chromatin-based regulation of putative effector gene expression. *bioRxiv* 544627.
Sperschneider J, Dodds PN. 2022. EffectorP 3.0: Prediction of apoplastic and cytoplasmic effectors in fungi and oomycetes. *Mol Plant Microbe Interact* **35**: 146–156.
Sperschneider J, Dodds PN, Gardiner DM, Singh KB, Taylor JM. 2018. Improved prediction of fungal effector proteins from secretomes with EffectorP 2.0. *Mol Plant Pathol* **19**: 2094–2110.
Sperschneider J, Gardiner DM, Thatcher LF, Lyons R, Singh KB, Manners JM, Taylor JM. 2015. Genome-Wide analysis in three Fusarium pathogens identifies rapidly evolving chromosomes and genes associated with pathogenicity. *Genome Biol Evol* **7**: 1613–1627.
Sperschneider J, Jones AW, Nasim J, Xu B, Jacques S, Zhong C, Upadhyaya NM, Mago R, Hu Y, Figueroa M, et al. 2021. The stem rust fungus *Puccinia graminis f. sp. tritici* induces centromeric small RNAs during late infection that are associated with genome-wide DNA methylation. *BMC Biol* **19**: 203.
Spielmann M, Lupiáñez DG, Mundlos S. 2018. Structural variation in the 3D genome. *Nature Rev Genet* **19**: 453–467.
Stajic D, Jansen LET. 2021. Empirical evidence for epigenetic inheritance driving evolutionary adaptation. *Philos Trans R Soc Lond B Biol Sci* **376**: 20200121.
Stajich JE, Berbee ML, Blackwell M, Hibbett DS, James TY, Spatafora JW, Taylor JW. 2009. The fungi. *Curr Biol* **19**: R840-5.
Stam R, Münsterkötter M, Pophaly SD, Fokkens L, Sghyer H, Güldener U, Hückelhoven R, Hess M. 2018. A New Reference Genome Shows the One-Speed Genome Structure of the Barley Pathogen *Ramularia collo-cygni*. *Genome Biol Evol* **10**: 3243–3249.
Stamatakis A. 2014. RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics* **30**: 1312–1313.
Stanke M, Keller O, Gunduz I, Hayes A, Waack S, Morgenstern B. 2006. AUGUSTUS: ab initio prediction of alternative transcripts. *Nucleic Acids Res* **34**: W435-9.
Starnes JH, Thornbury DW, Novikova OS, Rehmeyer CJ, Farman ML. 2012. Telomere-targeted retrotransposons in the rice blast fungus *Magnaporthe oryzae*: agents of telomere instability. *Genetics* **191**: 389–406.
Steenwyk JL, Buida TJ 3rd, Li Y, Shen X-X, Rokas A. 2020. ClipKIT: A multiple sequence alignment trimming software for accurate phylogenomic inference. *PLoS Biol* **18**: e3001007.
Steinberg G, Gurr SJ. 2020. Fungi, fungicide discovery and global food security. *Fungal Genet Biol* **144**: 103476.
Steinberg-Neifach O, Lue NF. 2015. Telomere DNA recognition in Saccharomycotina yeast: potential lessons for the co-evolution of ssDNA and dsDNA-binding proteins and their target sites. *Front Genet* **6**: 162.
Steinbiss S, Willhoeft U, Gremme G, Kurtz S. 2009. Fine-grained annotation and classification of *de novo* predicted LTR retrotransposons. *Nucleic Acids Res* **37**: 7002–7013.
Stern DL, Frankel N. 2013. The structure and evolution of *cis*-regulatory regions: the shavenbaby story. *Philos Trans R Soc Lond B Biol Sci* **368**: 20130028.
Stewart MN, Dawson DS. 2004. Potential roles for centromere pairing in meiotic chromosome segregation. *Cell Cycle* **3**: 1232–1234.
Stone JR, Wray GA. 2001. Rapid evolution of *cis*-regulatory sequences via local point mutations. *Mol Biol Evol* **18**: 1764–1770.
Strahl BD, Allis CD. 2000. The language of covalent histone modifications. *Nature* **403**: 41–45.
Strotz LC, Simões M, Girard MG, Breitkreuz L, Kimmig J, Lieberman BS. 2018. Getting somewhere with the Red Queen: chasing a biologically modern definition of the hypothesis. *Biol Lett* **14**: 2017.0734.
Struhl K, Segal E. 2013. Determinants of nucleosome positioning. *Nat Struct Mol Biol* **20**: 267–273.
Stuart T, Eichten SR, Cahn J, Karpievitch YV, Borevitz JO, Lister R. 2016. Population scale mapping of transposable element diversity reveals links to gene regulation and epigenomic variation. *eLife* **5**: e60.

Studt L, Janevska S, Arndt B, Boedi S, Sulyok M, Humpf H-U, Tudzynski B, Strauss J. 2017. Lack of the COMPASS component Ccl1 reduces H3K4 trimethylation levels and affects transcription of secondary metabolite genes in two plant–pathogenic Fusarium species. *Front Microbiol* **7**: 02144.

Studt L, Rösler SM, Burkhardt I, Arndt B, Freitag M, Humpf HU, Dickschat JS, Tudzynski B. 2016. Knock-down of the methyltransferase Kmt6 relieves H3K27me3 and results in induction of cryptic and otherwise silent secondary metabolite gene clusters in *Fusarium fujikuroi*. *Environ Microbiol* **18**: 4037–4054.

Stukenbrock EH. 2013. Evolution, selection and isolation: a genomic view of speciation in fungal plant pathogens. *New Phytol* **199**: 895–907.

Stukenbrock EH, Jørgensen FG, Zala M, Hansen TT, McDonald BA, Schierup MH. 2010. Whole-genome and chromosome evolution associated with host adaptation and speciation of the wheat pathogen *Mycosphaerella graminicola. PLoS Genet* **6**: e1001189.

Stukenbrock EH, McDonald BA. 2008. The origins of plant pathogens in agro-ecosystems. *Annu Rev Phytopathol* **46**: 75–100.

Sugiyama T, Cam HP, Sugiyama R, Noma K-I, Zofall M, Kobayashi R, Grewal SIS. 2007. SHREC, an effector complex for heterochromatic transcriptional silencing. *Cell* **128**: 491–504.

Sun L, Zhang Y, Zhang Z, Zheng Y, Du L, Zhu B. 2016. Preferential Protection of Genetic Fidelity within Open Chromatin by the Mismatch Repair Machinery. *J Biol Chem* **291**: 17692–17705.

Sun Z, Zhang Y, Jia J, Fang Y, Tang Y, Wu H, Fang D. 2020. H3K36me3, message from chromatin to DNA damage repair. *Cell Biosci* **10**: 9.

Sundaram V, Wysocka J. 2020. Transposable elements as a potent source of diverse *cis*-regulatory sequences in mammalian genomes. *Philos Trans R Soc Lond B Biol Sci* **375**: 20190347.

Supek F, Bošnjak M, Škunca N, Šmuc T. 2011. REVIGO summarizes and visualizes long lists of gene ontology terms. *PLoS One* **6**: e21800.

Szabo Q, Bantignies F, Cavalli G. 2019. Principles of genome folding into topologically associating domains. *Sci Adv* **5**: eaaw1668.

Szabo Q, Jost D, Chang J-M, Cattoni DI, Papadopoulos GL, Bonev B, Sexton T, Gurgo J, Jacquier C, Nollmann M. 2018. TADs are 3D structural units of higher-order chromosome organization in *Drosophila*. *Sci Adv* **4**: eaar8082.

Szathmáry E, Smith JM. 1995. The major evolutionary transitions. *Nature* **374**: 227–232.

Szitenberg A, Cha S, Opperman CH, Bird DM, Blaxter ML, Lunt DH. 2016. Genetic Drift, not life history or RNAi, determine long-term evolution of transposable elements. *Genome Biol Evol* **8**: 2964–2978.

Szostak JW, Blackburn EH. 1982. Cloning yeast telomeres on linear plasmid vectors. *Cell* **29**: 245–255.

Taberlay PC, Achinger-Kawecka J, Lun ATL, Buske FA, Sabir K, Gould CM, Zotenko E, Bert SA, Giles KA, Bauer DC, et al. 2016. Three-dimensional disorganization of the cancer genome occurs coincident with long-range genetic and epigenetic alterations. *Genome Res* **26**: 719–731.

Taddei A, Gasser SM. 2012. Structure and function in the budding yeast nucleus. *Genetics* **192**: 107–129.

Takahashi-Nakaguchi A, Sakai K, Takahashi H, Hagiwara D, Toyotome T, Chibana H, Watanabe A, Yaguchi T, Yamaguchi M, Kamei K, et al. 2018. *Aspergillus fumigatus* adhesion factors in dormant conidia revealed through comparative phenotypic and transcriptomic analyses. *Cell Microbiol* **20**: e12802.

Talbert PB, Henikoff S. 2017. Histone variants on the move: substrates for chromatin dynamics. *Nat Rev Mol Cell Biol* **18**: 115–126.

Talbert PB, Henikoff S. 2010. Histone variants--ancient wrap artists of the epigenome. *Nat Rev Mol Cell Biol* **11**: 264–275.

Tanabe H, Müller S, Neusser M, von Hase J, Calcagno E, Cremer M, Solovei I, Cremer C, Cremer T. 2002. Evolutionary conservation of chromosome territory arrangements in cell nuclei from higher primates. *Proc Natl Acad Sci U S A* **99**: 4424–4429.

Tanizawa H, Kim K-D, Iwasaki O, Noma K-I. 2017. Architectural alterations of the fission yeast genome during the cell cycle. *Nat Struct Mol Biol* **24**: 965–976.

Team, R. 2019. Core R: A Language and Environment for Statistical Computing, Version 3.5. 3. *Vienna: R Foundation for Statistical Computing*.

Team RC. 2013. R: A language and environment for statistical computing.

Tempel S. 2012. Using and understanding RepeatMasker. *Methods Mol Biol* **859**: 29–51.

Testa AC, Oliver RP, Hane JK. 2016. OcculterCut: A comprehensive survey of AT-Rich regions in fungal genomes. *Genome Biol Evol* **8**: 2044–2064.

Therizols P, Duong T, Dujon B, Zimmer C, Fabre E. 2010. Chromosome arm length and nuclear constraints determine the dynamic relationship of yeast subtelomeres. *Proc Natl Acad Sci U S A* **107**: 2025–2030.

Thomma, Nürnberger T, Joosten MHAJ. 2011. Of PAMPs and effectors: the blurred PTI-ETI dichotomy. *Plant Cell* **23**: 4–15.

Thon MR, Pan H, Diener S, Papalas J, Taro A, Mitchell TK, Dean RA. 2006. The role of transposable element clusters in genome evolution and loss of synteny in the rice blast fungus *Magnaporthe oryzae*. *Genome Biol* **7**: R16.

Tjong H, Gong K, Chen L, Alber F. 2012. Physical tethering and volume exclusion determine higher-order genome organization in budding yeast. *Genome Res* **22**: 1295–1305.

Tobias PA, Edwards RJ, Surana P, Mangelson H, Inácio V, do Céu Silva M, Várzea V, Park RF, Batista D. 2022. A chromosome-level genome resource for studying virulence mechanisms and evolution of the coffee rust pathogen *Hemileia vastatrix*. *bioRxiv* 2022.07.29.502101.

Tolhuis B, Palstra RJ, Splinter E, Grosveld F, de Laat W. 2002. Looping and interaction between hypersensitive sites in the active beta-globin locus. *Mol Cell* **10**: 1453–1465.

Torosin NS, Anand A, Golla TR, Cao W, Ellison CE. 2020. 3D genome evolution and reorganization in the *Drosophila melanogaster* species group. *PLoS Genetics* **16**: e1009229.

Torres DE, Martin Kramer H, Tracanna V, Fiorin GL, Cook DE, Seidl MF, Thomma BPH. 2023. Three-dimensional chromatin organization promotes genome evolution in a fungal plant pathogen. *bioRxiv* 2023.04.04.535574.

Torres DE, Oggenfuss U, Croll D, Seidl MF. 2020. Genome evolution in fungal plant pathogens: looking beyond the two-speed genome model. *Fungal Biol Rev* **34**: 136–143.

Torres DE, Thomma BPHJ, Seidl MF. 2021. Transposable elements contribute to genome dynamics and gene expression variation in the fungal plant pathogen *Verticillium dahliae*. *Genome Biol Evol* **13**: evab135.

Tralamazza SM, Nanchira Abraham L, Reyes-Avila CS, Corrêa B, Croll D. 2022. Histone H3K27 Methylation perturbs transcriptional robustness and underpins dispensability of highly conserved genes in fungi. *Mol Biol Evol* **39**: msab323.

Trojer P, Reinberg D. 2007. Facultative Heterochromatin: Is There a Distinctive Molecular Signature? *Molecular Cell* **28**: 1–13. http://dx.doi.org/10.1016/j.molcel.2007.09.011.

Tsochatzidou M, Malliarou M, Papanikolaou N, Roca J, Nikolaou C. 2017. Genome urbanization: clusters of topologically co-regulated genes delineate functional compartments in the genome of *Saccharomyces cerevisiae*. *Nucleic Acids Res* **45**: 5818–5828.

Tsukahara S, Kobayashi A, Kawabe A, Mathieu O, Miura A, Kakutani T. 2009. Bursts of retrotransposition reproduced in Arabidopsis. *Nature* **461**: 423–426.

Tsushima A, Gan P, Kumakura N, Narusaka M, Takano Y, Narusaka Y, Shirasu K. 2019. Genomic plasticity mediated by transposable elements in the plant pathogenic fungus *Colletotrichum higginsianum*. *Genome Biol Evol* **11**: 1487–1500.

Tzur YB, Wilson KL, Gruenbaum Y. 2006. SUN-domain proteins: "Velcro" that links the nucleoskeleton to the cytoskeleton. *Nat Rev Mol Cell Biol* **7**: 782–788.

Uhlmann F, Lottspeich F, Nasmyth K. 1999. Sister-chromatid separation at anaphase onset is promoted by cleavage of the cohesin subunit Scc1. *Nature* **400**: 37–42.

Uphoff N. 2012. Supporting food security in the 21st century through resource-conserving increases in agricultural production. *Agriculture & Food Security* **1**: 1–12.

Upson JL, Zess EK, Białas A, Wu C-H, Kamoun S. 2018. The coming of age of EvoMPMI: evolutionary molecular plant–microbe interactions across multiple timescales. *Curr Opin Plant Biol* **44**: 108–116.

Usami T, Momma N, Kikuchi S, Watanabe H, Hayashi A, Mizukawa M, Yoshino K, Ohmori Y. 2017. Race 2 of *Verticillium dahliae* infecting tomato in Japan can be split into two races with differential pathogenicity on resistant rootstocks. *Plant Pathol* **66**: 230–238.

Vallad GE, Subbarao KV. 2008. Colonization of resistant and susceptible lettuce cultivars by a green fluorescent protein-tagged isolate of *Verticillium dahliae*. *Phytopathology* **98**: 871–885.

Valton A-L, Dekker J. 2016. TAD disruption as oncogenic driver. *Curr Opin Genet Dev* **36**: 34–40.

Valton A-L, Venev SV, Mair B, Khokhar ES, Tong AHY, Usaj M, Chan K, Pai AA, Moffat J, Dekker J. 2022. A cohesin traffic pattern genetically linked to gene regulation. *Nat Struct Mol Biol* **29**: 1239–1251.

Van Bortle K, Nichols MH, Li L, Ong C-T, Takenaka N, Qin ZS, Corces VG. 2014. Insulator function and topological domain border strength scale with architectural protein occupancy. *Genome Biol* **15**: 1–18.

van Dam P, Fokkens L, Ayukawa Y, van der Gragt M, Ter Horst A, Brankovics B, Houterman PM, Arie T, Rep M. 2017. A mobile pathogenicity chromosome in Fusarium oxysporum for infection of multiple cucurbit species. *Sci Rep* **7**: 9042.

van de Vossenberg BTLH, Prodhomme C, van Arkel G, van Gent-Pelzer MPE, Bergervoet M, Brankovics B, Przetakiewicz J, Visser RGF, van der Lee TAJ, Vossen JH. 2019a. The *Synchytrium endobioticum* AvrSen1 triggers a hypersensitive response in Sen1 potatoes while natural variants evade detection. *Mol Plant-Microbe Int* **32**: 1536–1546.

van de Vossenberg BTLH, Warris S, Nguyen HDT, van Gent-Pelzer MPE, Joly DL, van de Geest HC, Bonants PJM, Smith DS, Lévesque CA, van der Lee TAJ. 2019b. Comparative genomics of chytrid fungi reveal insights into the obligate biotrophic and pathogenic lifestyle of *Synchytrium endobioticum*. *Scientific Reports* **9**: 8672.
van der Walt S, Colbert SC, Varoquaux G. 2011. The NumPy array: a structure for efficient numerical computation. *Comp Sci Eng* **13**: 22–30.
van Kan JAL, Shaw MW, Grant-Downton RT. 2014. Botrytis species: relentless necrotrophic thugs or endophytes gone rogue? *Mol Plant Pathol* **15**: 957–961.
van Steensel B, Belmont AS. 2017. Lamina-Associated Domains: links with chromosome architecture, heterochromatin, and gene repression. *Cell* **169**: 780–791.
van Wersch S, Li X. 2019. Stronger When Together: Clustering of Plant NLR Disease resistance Genes. *Trends Plant Sci* **24**: 688–699.
van Wyk S, Harrison CH, Wingfield BD, De Vos L, van der Merwe NA, Steenkamp ET. 2019. The RIPper, a web-based tool for genome-wide quantification of Repeat-Induced Point (RIP) mutations. *PeerJ* **7**: e7447.
Vanhoenacker E, Sandell L, Roze D. 2018. Stabilizing selection, mutational bias, and the evolution of sex. *Evolution* **72**: 1740–1758.
Van't Hof AE, Campagne P, Rigden DJ, Yung CJ, Lingley J, Quail MA, Hall N, Darby AC, Saccheri IJ. 2016. The industrial melanism mutation in British peppered moths is a transposable element. *Nature* **534**: 102–105.
Veltri D, Kamath U, Shehu A. 2018. Deep learning improves antimicrobial peptide recognition. *Bioinformatics* **34**: 2740–2747.
Venner S, Feschotte C, Biemont C. 2009. Dynamics of transposable elements: towards a community ecology of the genome. *Trends Genet* **25**: 317–323.
Viljoen A, Ma L-J, Molina AB. 2020. CHAPTER 8: Fusarium Wilt (Panama Disease) and Monoculture in Banana Production: Resurgence of a Century-Old Disease. In *Emerging Plant Diseases and Global Food Security*, *Epidemiology*, pp. 159–184, The American Phytopathological Society.
Villar D, Berthelot C, Aldridge S, Rayner TF, Lukk M, Pignatelli M, Park TJ, Deaville R, Erichsen JT, Jasinska AJ, et al. 2015. Enhancer evolution across 20 mammalian species. *Cell* **160**: 554–566.
Villaseñor R, Baubec T. 2021. Regulatory mechanisms governing chromatin organization and function. *Curr Opin Cell Biol* **70**: 10–17.
Villaseñor R, Pfaendler R, Ambrosi C, Butz S, Giuliani S, Bryan E, Sheahan TW, Gable AL, Schmolka N, Manzo M, et al. 2020. ChromID identifies the protein interactome at chromatin marks. *Nat Biotechnol* **38**: 728–736.
Vitte C, Bennetzen JL. 2006. Analysis of retrotransposon structural diversity uncovers properties and propensities in angiosperm genome evolution. *Proc Natl Acad Sci U S A* **103**: 17638–17643.
Voigt P, LeRoy G, Drury WJ 3rd, Zee BM, Son J, Beck DB, Young NL, Garcia BA, Reinberg D. 2012. Asymmetrically modified nucleosomes. *Cell* **151**: 181–193.
Vu H, Ernst J. 2022. Universal annotation of the human genome through integration of over a thousand epigenomic datasets. *Genome Biol* **23**: 9.
Waddington. 1942. Canalization of development and the inheritance of acquired characters *Nature* **150**: 563–565.
Wagner A. 2008. Robustness and evolvability: a paradox resolved. *Proc Biol Sci* **275**: 91–100.
Wagner A. 2013. *Robustness and Evolvability in Living Systems*. Princeton University Press.
Wagner A. 2005. Robustness, evolvability, and neutrality. *FEBS Lett* **579**: 1772–1778.
Wagner GP, Altenberg L. 1996. Perspective: Complex adaptations and the evolution of evolvability *Evolution* **50**: 967–976.
Wagner N, Krohne G. 2007. LEM-Domain proteins: new insights into lamin-interacting proteins. *Int Rev Cytol* **261**: 1–46.
Wako T, Yoshida A, Kato J, Otsuka Y, Ogawa S, Kaneyoshi K, Takata H, Fukui K. 2020. Human metaphase chromosome consists of randomly arranged chromatin fibres with up to 30-nm diameter. *Sci Rep* **10**: 8948.
Walkowiak S, Gao L, Monat C, Haberer G, Kassa MT, Brinton J, Ramirez-Gonzalez RH, Kolodziej MC, Delorean E, Thambugala D, et al. 2020. Multiple wheat genomes reveal global variation in modern breeding. *Nature* **588**: 277–283.
Wang F, Sethiya P, Hu X, Guo S, Chen Y, Li A, Tan K, Wong KH. 2021a. Transcription in fungal conidia before dormancy produces phenotypically variable conidia that maximize survival in different environments. *Nat Microbiol* **6**: 1066–1081.
Wang J, Jia ST, Jia S. 2016. New Insights into the Regulation of Heterochromatin. *Trends Genet* **32**: 284–294.
Wang J, Lawry ST, Cohen AL, Jia S. 2014. Chromosome boundary elements and regulation of heterochromatin spreading. *Cell Mol Life Sci* **71**: 4841–4852.
Wang J, Reddy BD, Jia S. 2015a. Rapid epigenetic adaptation to uncontrolled heterochromatin spreading. *eLife* **4**: e06179.

Wang L, Chen H, Li J, Shu H, Zhang X, Wang Y, Tyler BM, Dong S. 2020a. Effector gene silencing mediated by histone methylation underpins host adaptation in an oomycete plant pathogen. *Nucleic Acids Res* **48**: 1790–1799.

Wang L, Sun Y, Sun X, Yu L, Xue L, He Z, Huang J, Tian D, Hurst LD, Yang S. 2020b. Repeat-induced point mutation in Neurospora crassa causes the highest known mutation rate and mutational burden of any cellular life. *Genome Biol* **21**: 142.

Wang M, Li J, Wang P, Liu F, Liu Z, Zhao G, Xu Z, Pei L, Grover CE, Wendel JF, et al. 2021b. Comparative genome analyses highlight transposon-mediated genome expansion and the evolutionary architecture of 3D genomic folding in cotton. *Mol Biol Evol* **38**: 3621–3636.

Wang M, Wang P, Lin M, Ye Z, Li G, Tu L, Shen C, Li J, Yang Q, Zhang X. 2018. Evolutionary dynamics of 3D genome architecture following polyploidization in cotton. *Nature Plants* **4**: 90–97.

Wang Q, Jiang C, Wang C, Chen C, Xu J-R, Liu H. 2017. Characterization of the two-speed subgenomes of *Fusarium graminearum* reveals the fast-speed subgenome specialized for adaption and infection. *Front Plant Sci* **8**: 140.

Wang R, Mozziconacci J, Bancaud A, Gadal O. 2015b. Principles of chromatin organization in yeast: relevance of polymer models to describe nuclear organization and dynamics. *Curr Opin Cell Biol* **34**: 54–60.

Waskom M, Botvinnik O, O'Kane D, Hobson P, Lukauskas S, Gemperline DC, Augspurger T, Halchenko Y, Cole JB, Warmenhoven J. 2017. Mwaskom/Seaborn: V0. 8.1 (September 2017). *Zenodo*.

Waterfield NR, Daborn PJ, ffrench-Constant RH. 2002. Genomic islands in Photorhabdus. *Trends Microbiol* **10**: 541–545.

Weber T, Blin K, Duddela S, Krug D, Kim HU, Bruccoleri R, Lee SY, Fischbach MA, Müller R, Wohlleben W. 2015. antiSMASH 3.0—a comprehensive resource for the genome mining of biosynthetic gene clusters. *Nucleic Acids Res* **43**: W237–W243.

Weber VM, Doucet AJ, Cristofari G. 2023. Precise and scarless insertion of transposable elements by Cas9-mediated genome engineering. *Methods Mol Biol* **2607**: 329–353.

Weiner A, Hsieh T-HS, Appleboim A, Chen HV, Rahat A, Amit I, Rando OJ, Friedman N. 2015. High-resolution chromatin dynamics during a yeast stress response. *Mol Cell* **58**: 371–386.

Wells JL, Pryce DW, McFarlane RJ. 2006. Homologous chromosome pairing in *Schizosaccharomyces pombe*. *Yeast* **23**: 977–989.

Wells JN, Feschotte C. 2020. A field guide to eukaryotic transposable elements. *Annu Rev Genet* **54**: 539–561.

West AG, Fraser P. 2005. Remote control of gene transcription. *Hum Mol Genet* **14 Spec No 1**: R101-11.

Westermann S, Drubin DG, Barnes G. 2007. Structures and functions of yeast kinetochore complexes. *Annu Rev Biochem* **76**: 563–591.

Wicker T, Sabot F, Hua-Van A, Bennetzen JL, Capy P, Chalhoub B, Flavell A, Leroy P, Morgante M, Panaud O, et al. 2007. A unified classification system for eukaryotic transposable elements. *Nat Rev Genet* **8**: 973–982.

Wickham H. 2011. Ggplot2. *Wiley Interdiscip Rev Comput Stat* **3**: 180–185.

Wickham H. 2016. *ggplot2: elegant graphics for data analysis*. springer.

Wiles ET, McNaught KJ, Kaur G, Selker JML, Ormsby T, Aravind L, Selker EU. 2020. Evolutionarily ancient BAH-PHD protein mediates Polycomb silencing. *Proc Natl Acad Sci U S A* **117**: 11614–11623.

Wiles ET, Mumford CC, McNaught KJ, Tanizawa H, Selker EU. 2022. The ACF chromatin-remodeling complex is essential for Polycomb repression. *Elife* **11**. http://dx.doi.org/10.7554/eLife.77595.

Wiles ET, Selker EU. 2017. H3K27 methylation: a promiscuous repressive chromatin mark. *Curr Opin Genet Dev* **43**: 31–37.

Wilson GA. 2007. *Multifunctional Agriculture: A Transition Theory Perspective*. CABI.

Wilson RA, Talbot NJ. 2009. Under pressure: investigating the biology of plant infection by *Magnaporthe oryzae. Nat Rev Microbiol* **7**: 185–195.

Winter DJ, Ganley ARD, Young CA, Liachko I, Schardl CL, Dupont PY, Berry D, Ram A, Scott B, Cox MP. 2018. Repeat elements organise 3D genome structure and mediate transcription in the filamentous fungus *Epichlöe festucae*. *PLoS Genet* **14**: e1007467.

Wittkopp PJ, Kalay G. 2011. *Cis*-regulatory elements: molecular mechanisms and evolutionary processes underlying divergence. *Nat Rev Genet* **13**: 59–69.

Wolfe KH, Sharp PM, Li WH. 1989. Mutation rates differ among regions of the mammalian genome. *Nature* **337**: 283–285.

Wolff J, Rabbani L, Gilsbach R, Richard G, Manke T, Backofen R, Grüning BA. 2020. Galaxy HiCExplorer 3: a web server for reproducible Hi-C, capture Hi-C and single-cell Hi-C data analysis, quality control and visualization. *Nucleic Acids Res* **48**: W177–W184.

Wong H, Marie-Nelly H, Herbert S, Carrivain P, Blanc H, Koszul R, Fabre E, Zimmer C. 2012. A predictive computational model of the dynamic 3D interphase yeast nucleus. *Curr Biol* **22**: 1881–1890.

Wood V, Gwilliam R, Rajandream M-A, Lyne M, Lyne R, Stewart A, Sgouros J, Peat N, Hayles J, Baker S, et al. 2002. The genome sequence of *Schizosaccharomyces pombe. Nature* **415**: 871–880.
Wray GA. 2007. The evolutionary significance of *cis*-regulatory mutations. *Nat Rev Genet* **8**: 206–216.
Wray GA, Hahn MW, Abouheif E, Balhoff JP, Pizer M, Rockman MV, Romano LA. 2003. The evolution of transcriptional regulation in eukaryotes. *Mol Biol Evol* **20**: 1377–1419.
Wright D, Schaeffer SW. 2022. The relevance of chromatin architecture to genome rearrangements in. *Philos Trans R Soc Lond B Biol Sci* **377**: 20210206.
Wright S, Finnegan D. 2001. Genome evolution: Sex and the transposable element. *Current Biology* **11**: R296–R299.
Wyka SA, Mondo SJ, Liu M, Dettman J, Nalam V, Broders KD. 2020. Whole genome comparisons of ergot fungi reveals the divergence and evolution of species within the genus Claviceps are the result of varying mechanisms driving genome evolution and host range expansion. *bioRxiv* 2020.04.13.039230.
Xhemalce B, Dawson MA, Bannister AJ. 2012. Histone modifications. In *Epigenetic regulation and epigenomics* (ed. R.A. Meyers), pp. 657–703.
Xia C, Huang L, Huang J, Zhang H, Huang Y, Benhamed M, Wang M, Chen X, Zhang M, Liu T, et al. 2022. Folding features and dynamics of 3D genome architecture in plant fungal pathogens. *Microbiol Spectr* **10**: e0260822.
Xia C, Wang M, Yin C, Cornejo OE, Hulbert SH, Chen X. 2018. Genomic insights into host adaptation between the wheat stripe rust pathogen (*Puccinia striiformis f. sp. tritici*) and the barley stripe rust pathogen (*Puccinia striiformis f. sp. hordei*). *BMC Genomics* **19**: 664.
Xie J-H, Tang W, Lu G-D, Hong Y-H, Zhong Z-H, Wang Z, Zheng H-K. 2023. Histone H3K27me3 methylation regulates the expression of secreted proteins distributed at fast-evolving regions through transcriptional repression of transposable elements. *J Integr Agric*. **0.**
Xie T, Zhang F-G, Zhang H-Y, Wang X-T, Hu J-H, Wu X-M. 2019. Biased gene retention during diploidization in Brassica linked to three-dimensional genome organization. *Nat Plants* **5**: 822–832.
Xiong L, Adhvaryu KK, Selker EU, Wang Y. 2010. Mapping of lysine methylation and acetylation in core histones of *Neurospora crassa. Biochemistry* **49**: 5236–5243.
Xu B, Wang H, Wright S, Hyle J, Zhang Y, Shao Y, Niu M, Fan Y, Rosikiewicz W, Djekidel MN, et al. 2021. Acute depletion of CTCF rewires genome-wide chromatin accessibility. *Genome Biol* **22**: 244.
Xue J, Chen H, Wu J, Takeuchi M, Inoue H, Liu Y, Sun H, Chen Y, Kanoh J, Lei M. 2017. Structure of the fission yeast *S. pombe* telomeric Tpz1-Poz1-Rap1 complex. *Cell Res* **27**: 1503–1520.
Yadav RK, Matsuda A, Lowe BR, Hiraoka Y, Partridge JF. 2021. Subtelomeric chromatin in the fission yeast. *Microorganisms* **9**: 9091977.
Yadav V, Sreekumar L, Guin K, Sanyal K. 2018a. Five pillars of centromeric chromatin in fungal pathogens. *PLoS Pathog* **14**: e1007150.
Yadav V, Sun S, Billmyre RB, Thimmappa BC, Shea T, Lintner R, Bakkeren G, Cuomo CA, Heitman J, Sanyal K. 2018b. RNAi is a critical determinant of centromere evolution in closely related fungi. *Proc Natl Acad Sci U S A* **115**: 3108–3113.
Yan K, Zhang Z, Yang J, McLaughlin SH, Barford D. 2018. Architecture of the CBF3-centromere complex of the budding yeast kinetochore. *Nat Struct Mol Biol* **25**: 1103–1110.
Yang C, Ma L, Xiao D, Liu X, Jiang X, Ying Z, Lin Y. 2021. Chromosome-scale assembly of the *Sparassis latifolia* genome obtained using long-read and Hi-C sequencing. *G3(Bethesda)* **11**: jkab173.
Yang S, Li W, Qi S, Gai K, Chen Y, Suo J, Cao Y, He Y, Wang Y, He Q. 2014. The highly expressed methionine synthase gene of *Neurospora crassa* is positively regulated by its proximal heterochromatic region. *Nucleic Acids Res* **42**: 6183–6195.
Yang Y, Zhang Y, Ren B, Dixon JR, Ma J. 2019. Comparing 3D genome organization in multiple species using Phylo-HMRF. *Cell systems* **8**: 494–505.
Yao G, Chen W, Sun J, Wang X, Wang H, Meng T, Zhang L, Guo L. 2023. Gapless genome assembly of *Fusarium verticillioides*, a filamentous fungus threatening plant and human health. *Sci Data* **10**: 229.
Yao R-W, Wang Y, Chen L-L. 2019. Cellular functions of long noncoding RNAs. *Nat Cell Biol* **21**: 542–551.
Yardımcı GG, Ozadam H, Sauria MEG, Ursu O, Yan K-K, Yang T, Chakraborty A, Kaul A, Lajoie BR, Song F. 2019. Measuring the reproducibility and quality of Hi-C data. *Genome biology* **20**: 1–19.
Yasuhara JC, DeCrease CH, Wakimoto BT. 2005. Evolution of heterochromatic genes of Drosophila. *Proc Natl Acad Sci U S A* **102**: 10958–10963.
Yeaman S. 2013. Genomic rearrangements and the evolution of clusters of locally adaptive loci. *Proc Natl Acad Sci U S A* **110**: E1743–E1751.
Yildirir G, Sperschneider J, Malar C M, Chen ECH, Iwasaki W, Cornell C, Corradi N. 2022. Long reads and Hi-C sequencing illuminate the two-compartment genome of the model arbuscular mycorrhizal symbiont *Rhizophagus irregularis. New Phytol* **233**: 1097–1107.

Yona AH, Manor YS, Herbst RH, Romano GH, Mitchell A, Kupiec M, Pilpel Y, Dahan O. 2012. Chromosomal duplication is a transient evolutionary solution to stress. *Proc Natl Acad Sci U S A* **109**: 21010–21015.

Yoshida K, Saunders DG, Mitsuoka C, Natsume S, Kosugi S, Saitoh H, Inoue Y, Chuma I, Tosa Y, Cano LM, et al. 2016. Host specialization of the blast fungus *Magnaporthe oryzae* is associated with dynamic gain and loss of genes linked to transposable elements. *BMC Genomics* **17**: 370.

Yoshida MM, Kinoshita K, Aizawa Y, Tane S, Yamashita D, Shintomi K, Hirano T. 2022. Molecular dissection of condensin II-mediated chromosome assembly using in vitro assays. *Elife* **11**: e.78984.

Young RS, Hayashizaki Y, Andersson R, Sandelin A, Kawaji H, Itoh M, Lassmann T, Carninci P, FANTOM Consortium, Bickmore WA, et al. 2015. The frequent evolutionary birth and death of functional promoters in mouse and human. *Genome Res* **25**: 1546–1557.

Yu EY, Hsu M, Holloman WK, Lue NF. 2018. Contributions of recombination and repair proteins to telomere maintenance in telomerase-positive and negative*Ustilago maydis*. *Mol Microbiol* **107**: 81–93.

Yu EY, Zahid SS, Ganduri S, Sutherland JH, Hsu M, Holloman WK, Lue NF. 2020. Structurally distinct telomere-binding proteins in *Ustilago maydis* execute non-overlapping functions in telomere replication, recombination, and protection. *Commun Biol* **3**: 777.

Yu G, Wang L-G, Han Y, He Q-Y. 2012. clusterProfiler: an R package for comparing biological themes among gene clusters. *OMICS* **16**: 284–287.

Yu H, Zhang L, Shang X, Peng B, Li Y, Xiao S, Tan Q, Fu Y. 2022. Chromosomal genome and population genetic analyses to reveal genetic architecture, breeding history and genes related to cadmium accumulation in *Lentinula edodes*. *BMC Genomics* **23**: 120.

Yu H-G, Koshland D. 2005. Chromosome morphogenesis: condensin-dependent cohesin removal during meiosis. *Cell* **123**: 397–407.

Yuan J, Sun H, Wang Y, Li L, Chen S, Jiao W, Jia G, Wang L, Mao J, Ni Z, et al. 2022. Open chromatin interaction maps reveal functional regulatory elements and chromatin architecture variations during wheat evolution. *Genome Biol* **23**: 34.

Yue J-X, Li J, Aigrain L, Hallin J, Persson K, Oliver K, Bergström A, Coupland P, Warringer J, Lagomarsino MC, et al. 2017. Contrasting evolutionary genome dynamics between domesticated and wild yeasts. *Nat Genet* **49**: 913–924.

Zaccaron AZ, Chen L-H, Samaras A, Stergiopoulos I. 2022. A chromosome-scale genome assembly of the tomato pathogen *Cladosporium fulvum* reveals a compartmentalized genome architecture and the presence of a dispensable chromosome. *Microb Genom* **8**: mgen.0.000819.

Zadoks JC. 2017. On social and political effects of plant pest and disease epidemics. *Phytopathology* **107**: 1144–1148.

Zahid S, Aloe S, Sutherland JH, Holloman WK, Lue NF. 2022. Ustilago maydis telomere protein Pot1 harbors an extra N-terminal OB fold and regulates homology-directed DNA repair factors in a dichotomous and context-dependent manner. *PLoS Genet* **18**: e1010182.

Zanne AE, Powell JR, Flores-Moreno H, Kiers ET, van 't Padje A, Cornwell WK. 2020. Finding fungal ecological strategies: Is recycling an option? *Fungal Ecol* **46**: 100902.

Zeder MA. 2015. Core questions in domestication research. *Proc Natl Acad Sci U S A* **112**: 3191–3198.

Zeilinger S, Gupta VK, Dahms TES, Silva RN, Singh HB, Upadhyay RS, Gomes EV, Tsui CK-M, Nayak S C. 2016. Friends or foes? Emerging insights from fungal interactions with plants. *FEMS Microbiol Rev* **40**: 182–207.

Zemach A, McDaniel IE, Silva P, Zilberman D. 2010. Genome-wide evolutionary analysis of eukaryotic DNA methylation. *Science* **328**: 916–919.

Zemach A, Zilberman D. 2010. Evolution of eukaryotic DNA methylation and the pursuit of safer sex. *Curr Biol* **20**: R780-5.

Zeng Z, Zhang W, Marand AP, Zhu B, Buell CR, Jiang J. 2019. Cold stress induces enhanced chromatin accessibility and bivalent histone modifications H3K4me3 and H3K27me3 of active genes in potato. *Genome Biol* **20**: 123.

Zenk F, Zhan Y, Kos P, Löser E, Atinbayeva N, Schächtle M, Tiana G, Giorgetti L, Iovino N. 2021. HP1 drives de novo 3D genome reorganization in early Drosophila embryos. *Nature* **593**: 289–293.

Zentner GE, Henikoff S. 2013. Regulation of nucleosome dynamics by histone modifications. *Nat Struct Mol Biol* **20**: 259–266.

Zhan J, Ericson L, Burdon JJ. 2018. Climate change accelerates local disease extinction rates in a long-term wild host-pathogen association. *Glob Chang Biol* **24**: 3526–3536.

Zhang H, Yohe T, Huang L, Entwistle S, Wu P, Yang Z, Busk PK, Xu Y, Yin Y. 2018. dbCAN2: a meta server for automated carbohydrate-active enzyme annotation. *Nucleic Acids Res* **46**: W95–W101.

Zhang J, Yang J-R. 2015. Determinants of the rate of protein sequence evolution. *Nat Rev Genet* **16**: 409–420.

Zhang Q, Guan P, Zhao L, Ma M, Xie L, Li Y, Zheng R, Ouyang W, Wang S, Li H, et al. 2021a. Asymmetric epigenome maps of subgenomes reveal imbalanced transcription and distinct evolutionary trends in *Brassica napus. Mol Plant* **14**: 604–619.

Zhang W, Huang J, Cook DE. 2021b. Histone modification dynamics at H3K27 are associated with altered transcription of in planta induced genes in *Magnaporthe oryzae*. *PLoS Genet* **17**: e1009376.

Zhang Y, Li T, Preissl S, Amaral ML, Grinstein JD, Farah EN, Destici E, Qiu Y, Hu R, Lee AY. 2019. Transcriptionally active HERV-H retrotransposons demarcate topologically associating domains in human pluripotent stem cells. *Nat Genet* **51**: 1380–1388.

Zhang Y, McCord RP, Ho Y-J, Lajoie BR, Hildebrand DG, Simon AC, Becker MS, Alt FW, Dekker J. 2012. Spatial organization of the mouse genome and its role in recurrent chromosomal translocations. *Cell* **148**: 908–921.

Zhao K, Kong D, Jin B, Smolke CD, Rhee SY. 2021. A novel bivalent chromatin associates with rapid induction of camalexin biosynthesis genes in response to a pathogen signal in Arabidopsis. *Elife* **10**: e.69508.

Zhao L, Xie L, Zhang Q, Ouyang W, Deng L, Guan P, Ma M, Li Y, Zhang Y, Xiao Q, et al. 2020. Integrative analysis of reference epigenomes in 20 rice varieties. *Nat Commun* **11**: 2658.

Zhao Y, Garcia BA. 2015. Comprehensive catalog of currently documented histone modifications. *Cold Spring Harb Perspect Biol* **7**: a025064.

Zheng J, Guo N, Wagner A. 2020. Selection enhances protein evolvability by increasing mutational robustness and foldability. *Science* **370**: abb5962.

Zhimulev IF, Zykova TY, Goncharov FP, Khoroshko VA, Demakova OV, Semeshin VF, Pokholkova GV, Boldyreva LV, Demidova DS, Babenko VN, et al. 2014. Genetic organization of interphase chromosome bands and interbands in *Drosophila melanogaster*. *PLoS One* **9**: e101631.

Zhong X. 2016. Comparative epigenomics: a powerful tool to understand the evolution of DNA methylation. *New Phytol* **210**: 76–80.

Zhong Z, Marcel TC, Hartmann FE, Ma X, Plissonneau C, Zala M, Ducasse A, Confais J, Compain J, Lapalu N, et al. 2017. A small secreted protein in *Zymoseptoria tritici* is responsible for avirulence on wheat cultivars carrying the Stb6 resistance gene. *New Phytol* **214**: 619–631.

Zhou S, Liu X, Sun W, Zhang M, Yin Y, Pan S, He D, Shen M, Yang J, Zheng Q, et al. 2021. The COMPASS-like complex modulates fungal development and pathogenesis by regulating H3K4me3-mediated targeted gene expression in *Magnaporthe oryzae*. *Mol Plant Pathol* **22**: 422–439.

Zhou Z, Dang Y, Zhou M, Li L, Yu C-H, Fu J, Chen S, Liu Y. 2016. Codon usage is an important determinant of gene expression levels largely through its effects on transcription. *Proc Natl Acad Sci U S A* **113**: E6117–E6125.

Zhu D, Wen Y, Yao W, Zheng H, Zhou S, Zhang Q, Qu L-J, Chen X, Wu Z. 2023. Distinct chromatin signatures in the Arabidopsis male gametophyte. *Nat Genet*. **55**: 706-720.

Zhu Q, Ramakrishnan M, Park J, Belden WJ. 2019. Histone H3 lysine 4 methyltransferase is required for facultative heterochromatin at specific loci. *BMC Genomics* **20**: 350.

Zinder JC, Olinares PDB, Svetlov V, Bush MW, Nudler E, Chait BT, Walz T, de Lange T. 2022. Shelterin is a dimeric complex with extensive structural heterogeneity. *Proc Natl Acad Sci U S A* **119**: e2201662119.

Zofall M, Smith DR, Mizuguchi T, Dhakshnamoorthy J, Grewal SIS. 2016. Taz1-Shelterin promotes facultative heterochromatin assembly at chromosome-internal sites containing late replication origins. *Mol Cell* **62**: 862–874.

Zuin J, Dixon JR, van der Reijden MIJA, Ye Z, Kolovos P, Brouwer RWW, van de Corput MPC, van de Werken HJG, Knoch TA, van IJcken WFJ, et al. 2014. Cohesin and CTCF differentially affect chromatin architecture and gene expression in human cells. *Proc Natl Acad Sci U S A* **111**: 996–1001.

Zuo W, Chen G, Gao Z, Li S, Chen Y, Huang C, Chen J, Chen Z, Lei M, Bian Q. 2021. Stage-resolved Hi-C analyses reveal meiotic chromosome organizational features influencing homolog alignment. *Nat Commun* **12**: 5827.

Acknowledgements

"…When was life ever truly ours?,
When are we ever what we are?
We are ill-reputed, nothing more than vertigo –
and emptiness,
a frown in the mirror, horror and vomit,
life is never truly ours, it always belong to the others,
Life is no one's,
we all are life?, bread of the sun for the others,
The other that we all are?,
when I am another, my acts are more mine –
when they are the acts of others,
in order to be I must be another,
leave myself, search for myself in the others,
The others that don't exist if I don't exists,
the others that give me total existence,
I am not, there is no I, we are always us..."

Octavio Paz, Sunstone

"…¿ la vida, cuando fue de veras nuestra?,
¿cuándo somos de veras lo que somos?,
Bien mirado no somos, nunca somos a solas sino –
vértigo y vacío,
Muecas en el espejo, horror y vómito,
Nunca la vida es nuestra, –
es de los otros,
La vida no es de nadie,
¿todos somos la vida?, pan de sol para los otros,
¿los otros todos que nosotros somos?,
Soy otro cuando soy, los actos míos son mas mios –
sino tambien de todos,
Para que pueda ser he de ser otro,
Salir de mi, buscarme entre los otros,
Los otros que no son si yo no existo,
Los otros que me dan plena existencia,
No soy, no hay yo, siempre somos nosotros…"

Octavio Paz, Piedra de Sol

To Michael Seidl. I believe that few people have a profound impact on you that inspires you to become like them in the future, and you certainly are that person for me. To you, for being the pleasant, thoughtful, inspiring, and gentle mentor you were. Your mentor's guidance transcended the words and plots printed in these chapters. You encouraged me, empowered me with freedom, and supported me no matter how sunny or dark my days were. You have my full gratitude.

To Bart Thomma who undeniably helped shape my academic path. You are an example of excellence and cutting-edge science. I wish one day I will make justice to those ingredients you shared with me.

A Michael Seidl. Creo que pocas personas marcan nuestro camino tan profundamente, que te inspiran a ser un poco como ellos en el futuro, tú eres esa persona para mí. A ti, por ser ese mentor agradable, reflexivo, inspirador y gentil que fuiste. Tu guía superó las palabras y gráficas impresas en estos capítulos. Me alentaste, me diste libertad y me apoyaste sin importar lo soleado u oscuro de mis días. Tienes toda mi gratitud.

A Bart Thomma quien sin duda marcó mi camino académico. Eres un ejemplo de excelencia y la ciencia de frontera. Ojalá algún día le haga justicia a esos ingredientes que me compartiste.

To Jan Kees van Amerongen. You care about our systems and computers as much as you do for the big tree in your office. Our rocket science wouldn't be possible without your help and support.

To Laura and Sam. You made this journey more sweet, interesting, and pleasant. Separated by a couple of years, our talks and laughs nurtured many spaces during my journey, not only within these pages. I am happy to have you in my life and to share these last moments with you.

To the volume I of this journey, which was in Wageningen. The place where kind PIs like Francine, Jan and Gert made me feel at home since day one. The place in which, all the PhDs and postdocs like Xiaoqian, Sander, Nelia, Gabriel, Sergio, Einar, Marteen, Nick, and Edgar, who shared similar goals and struggles, made me realize that this is not a solitary path, but a collective one.

To the volume II of this journey, which was in Utrecht. The place where I was lucky to share the *queer* office with Tina, Sarah and later on Laura. The place where I shared lunch, beers, and rainy or sunny days with Bastiaan, Daniel(s), Bas, and Monica, and where I was incredibly supported by my colleagues in crime Xin, Petros, and Anouk.

To all my extended family outside the boundaries of the University, without whom my days and nights would not have been the same. To you Gore, German, Jurrian, Silvana, Luz, Francisco, Ruth and Gerard. Without the supporting fabric that we knitted together, these pages would not have been possible.

A Jan Kees van Amerongen. A ti, que te preocupas tanto por nuestros sistemas y computadoras como lo haces por el arbolóte de tu oficina. Nuestra *rocket science* no sería posible sin tu ayuda y apoyo.

A Laura y Sam. Ustedes hicieron este viaje más rico, interesante y placentero. Separados por un par de años, nuestras pláticas y risas enriquecieron muchas partes de mi camino, no solo estas páginas. Estoy muy feliz de tenerlos en mi vida y compartir estos últimos momentos con ustedes.

Al volumen I de este camino, que fue en Wageningen. El lugar donde PI's como Francine, Jan y Gert me hicieron sentir como en casa. El lugar en el que los PhDs y posdocs como Xiaoqian, Sander, Nelia, Gabriel, Sergio, Einar, Marteen, Nick y Edgar, quienes compartíamos metas y problemas similares, me hicieron darme cuenta de que este no es un camino solitario, sino más bien uno colectivo.

Al volumen II de este viaje, que fue en Utrecht. El lugar donde tuve la suerte de compartir la *queer office* con Tina, Sarah y luego Laura. El lugar donde compartí el almuerzo, cervezas, días lluviosos y soleados con Bastiaan, Daniel(s), Bas y Mónica, y donde tuve el apoyo de mis grandes compañeros Xin, Petros y Anouk.

A toda mi familia extendida fuera de la Universidad, sin quienes mis días y mis noches no hubiesen sido lo mismo. A ustedes Gore, Germán, Jurrian, Silvana, Luz, Francisco, Ruth y Gerard. Sin ese tejido social, de amistad y de soporte que tejimos juntos, estas páginas tampoco hubiesen sido posibles.

It is sometimes possible to choose your family, but very few times in life you are fortunate enough to be together across multiple continents.

To you Marcela. You cannot imagine how grateful I am for sharing this journey with you. Within all the solitude and struggles that an immigrant can experience, I have been so lucky to have my best friend close to me.

To all my family for their cumulative effort, love, and care. From my grandparents to my aunts and uncles. Especially, my aunts Carmela, Irma, Maria, Lolis, and Carmen. To my sister Karla and my brother Carlos, the sky is the limit.

Every time that I was home sick I sang in my mind 'el oso carpintero', which immediately brought my mom to me. We made this journey together, you physically in Mexico, but always close to me, always supporting me, and always here. These pages are also part of your effort, sweat, and unreachable attitude that I admire so much. There are no words to express my love for you. When you read this, I hope we sit next to each other, and you sing me once more that song.

Finally, not only these pages but a large part of my journey in The Netherlands would not have been possible in this way without the tireless shoulder and boundless love of José. You have nurtured me with generous love, affection, happiness, and acceptance. All of those ingredients have contributed in many ways to the printed words and plots here, but even more, it is thanks to them that I become the person I am today.
All my love.

To all of *us*.

En ocasiones tienes la oportunidad de elegir a tu familia, pero muy pocas veces en la vida tienes la suerte de estar con ellos a través de varios continentes.

A ti Marcela. No te imaginas lo agradecido que estoy por compartir este viaje contigo. Dentro de toda la soledad y las dificultades de ser un inmigrante, he tenido mucha suerte de tener a mi mejor amiga cerca de mí. Te quiero mucho.

A toda mi familia por su esfuerzo, amor y cuidado. Desde mis abuelos hasta mis tías y tíos. En especial, a mis tías Carmela, Irma, María, Lolis y Carmen. A mi hermana Karla y mi hermano Carlos, el cielo es el límite y solo tenemos una vida para sacarle jugo.

Cada vez que extrañaba México cantaba 'el oso carpintero' y eso inmediatamente me acercaba a mi mamá. Este camino lo hicimos juntos, tú físicamente en México, siempre apoyándome, pero siempre aquí cerca, en mi mente y en mi corazón. Estas páginas también son parte de tu esfuerzo, tu sudor y esa actitud inalcanzable que admiro tanto. No hay palabras para expresar todo el amor que te tengo. Espero que cuando leas esto, estemos sentados uno al lado del otro y me cantes una vez más esa canción.

Finalmente, no solo estas páginas, sino gran parte de mi camino por Holanda, no hubiese sido de esta manera posible sin el incansable hombro y el gran amor de José. Me has alimentado con amor, cariño, felicidad y aceptación. Todos esos ingredientes han contribuido de muchas maneras a las palabras y gráficas impresas aquí, pero aún más, gracias a ellos me he convertido en la persona que soy ahora.
Todo mi amor.

A todos *nosotros*.

About the author

Written by Marcela Aragón and José Hopkins

David E. Torres Sanchez was born on September 5th of 1990 in Mexico City. From a young age, plants captivated David's heart, an interest nurtured by his grandmother, Carmen. He remembers helping her tender her lush garden, where he learned the secrets of the medicinal remedies hidden within the greenery, as she was a bit of an ethnobotanical herself. His wonders with plants were also nurtured by David's other grandma, Virginia. With her, he learned how to transform plants into delicious dishes, developing another one of his passions: cuisine. He didn't know back then, but looking into the mysteries of plants paved his life's journey.

As he grew older, David thought his path lay in the realm of natural sciences, but a brief detour into medicine showed him that healing humans wasn't quite his calling. Instead, he found himself drawn to the world of Biology, motivated to discover how the human body works with the idea of going into cancer research. It was until David took his first course in plants at the Faculty of Sciences of the National Autonomous University of Mexico (UNAM) that he re-discovered how magical, mystical, and musical plants were when compared to mundane humans. He found it intriguing that despite being sessile organisms, plants adapt to a myriad of environments and conquer almost everywhere on the planet.

This infatuation drove him to sign up for all courses available on plants, and he became a part of the Laboratory of Plant Development to develop a project proposal for a BSc thesis. He eagerly worked on a proposal involving viruses, epigenetics, and plants. However, after all the work, something else caught his attention. Under the mentorship of Dr. Judith Marquez-Guzman, David embarked on an adventure that involved scaling 10-meter-tall cacti at night in Tehuacán-Cuicatlán Biosphere Reserve. This daring escapade wasn't just about exhilarating encounters with bats and climbing not-so-safe ladders but a quest to understand the intricate process of flower evolution and plant speciation of this columnar cacti tribe.

After completing his BSc thesis, David was still fascinated by epigenetics and plants. When looking around, he noticed that most of what was done in this field had to do with plant-pathogen interactions. Hence, he ventured into an MSc in Phytopathology

at the Postgraduate College (COLPOS), supported by a grant from the National Council of Science and Technology of Mexico (CONACyT). Leaving behind the big city for the first time, he moved to Texcoco, affectionally known as "Tex-Mex."

But life's twists and turns had more in store for David. After various questionable landlords and failed attempts at doing yoga, David ended up not working with epigenetics. Instead, he studied the potential of using fungi as a biological control for chrysanthemum white rust. Under the supervision of Dr. Reyna Rojas-Martínez, David sampled chrysanthemum leaves co-infected with the pathogen and the beneficial fungi and managed to isolate, identify, and characterize the identity and chemistry of the biological control agents. As part of this research, David obtained a grant to do a research stay in the Laboratory of Organic Chemistry from the University of Milan under the supervision of Dr. Gemma Assante and Dr. Sabrina DallaValle. Here, he analyzed the chemistry of fungal secondary metabolites and learned how to brew coffee in a moka pot and cook polenta as comfort food on cold days (skills that would be fundamental for his survival later in his PhD).

Once he finished his MSc thesis, eager to prove to himself and the world that it was possible to get an "actual job" as a Biologist, David left the big city again. He moved to León - an industrial city in Guanajuato - where he worked as a researcher for a company that produces hydroponic leafy greens. He started doing efficacy trials of agrochemicals on different varieties of spinach and lettuce but a year later became an executive project manager for two and a half years. Along the way, what caught his attention was that he had to deal constantly with fungal diseases even though lettuces were grown in a closed and super-controlled system. This observation sparked an intriguing thought. He reasoned that pathogens must have evolutionary strategies adapting relatively fast to infect plants under these circumstances. Once again, the answer was probably the realm of epigenetics.

Fueled by a burning desire to study epigenetics and tired of his executive lifestyle, he decided to quit his job and pursue a PhD in exactly this topic. To his luck, he found Dr. Michael Seidl, whose research at (back then) Wageningen University aligned with his vision. With Michael's guidance, David wrote a PhD proposal that secured funding from CONACyT. Once again, he left the big city and the comfort of nice weather and delicious tacos and moved to the Netherlands.

And so, what followed is the story of this thesis. A journey through the mysteries of genome evolution that, without David knowing, began when he was a little boy gardening and cooking with his grandmothers in Mexico. A drive that surely will guide him in his future steps beyond this thesis as he will continue fruitfully working with plant-pathogen interactions from an evolutionary perspective but now switching the angle to his true love: plants.

Education Statement of the Graduate School
Experimental Plant Sciences

Issued to: David Eduardo Torres Sánchez
Date: 18 September 2023
Group: Laboratory of Phytopathology
University: Wageningen University and Research

1) Start-Up Phase	date	cp
► **First presentation of your project**		
Chromatin dynamics and effector evolution	14 Dec 2018	1.5
► **Writing or rewriting a project proposal**		
Chromatin dynamics impacts effector evolution in the tomato wilt pathogen *Verticillium dahliae*	11 Jan 2019	6.0
► **Writing a review or book chapter**		
David E. Torres & Ursula Oggenfuss *et al.* Genome evolution in fungal plant pathogens: looking beyond the two-speed genome model, *Fungal Biology Reviews*, Volume 34, Issue 3, 2020, Pages 136-143, doi.org/10.1016/j.fbr.2020.07.001.	01 Sep 2020	6.0
► **MSc courses**		
Subtotal Start-Up Phase		13.5

2) Scientific Exposure	date	cp
► **EPS PhD days**		
EPS PhD days 'Get2Gether' 2020, Soest, NL	10-11 Feb 2020	0.6
EPS PhD days 'Get2Gether' 2021, Online	01-02 Feb 2021	0.4
► **EPS theme symposia**		
EPS Theme 2 Symposium & Willie Commelin Scholten Day "Interactions between plants and biotic agents", Wageningen, NL	01 Feb 2019	0.3
EPS Theme 2 Symposium & Willie Commelin Scholten Day "Interactions between plants and biotic agents", Online	09 Feb 2021	0.2
EPS Theme 2 Symposium & Willie Commelin Scholten Day "Interactions between plants and biotic agents", Online	08 Feb 2022	0.2
EPS Theme 4 Symposium "Genome Biology", Online	11 Dec 2020	0.2
EPS Theme 4 Symposium "Genome Biology", Online	17 Jan 2022	0.3
EPS Theme 4 Symposium "Genome Biology and Gene Regulation", Amsterdam, NL	06 Dec 2022	0.3
► **Lunteren Days and other national platforms**		
Annual meeting "Experimental Plant Sciences", Lunteren, NL	08-09 Apr 2019	0.6
6th UBC symposium "Bioinformatics for Life", Utrecht, NL	07 Oct 2019	0.3
3rd Netherlands Society of Evolutionary Biology Meeting, Online	19-20 Apr 2021	0.4
18th Dutch Chromatin Meeting, Online	12 Nov 2021	0.3
19th Dutch Chromatin Meeting, Leiden, NL	27 Oct 2022	0.3
Science for Life Conference, Utrecht, NL	14 Nov 2022	0.3
► **Seminars (series), workshops and symposia**		
Seminar: Nick Talbot, 'Investigating the biology of plant infection by the rice blast fungus'	02 Apr 2019	0.1
Seminar: Daniel Croll, 'Parasites within parasites : how transposable elements drive the evolution of plant pathogenic fungi'	21 Jun 2019	0.1
Seminar: Eva Stukenbrock, 'Causes and consequences of chromosome instability in a fungal plant patrhogen'	13 Dec 2019	0.1
Seminar: Iñaki Ruiz Trillo, 'Leave elegance to the tailor, the truth about the microbial origin of the animals'	05 Apr 2023	0.1
Workshop: Snakemake - Johannes Köster, Online	13 May 2020	0.3
Symposium: 34th Meeting of the Fusarum working group, Utrecht, NL	30 Oct 2019	0.3
Symposium: FEMS Yeast Research Webinar - Genomic insights into yeast diversity and evolution, Online	28 Jan 2021	0.2
Symposium: Viral and Immune dynamics, Utrecht, NL	19 Oct 2022	0.2
Symposium: Host-Microbe Genetics Meeting, Utrecht, NL	21 Oct 2022	0.3
Symposium: The rise and fall of complex life, Utrecht, NL	24 Jan 2023	0.2
Symposium: Immunological lessons from effector repertoires, Utrecht, NL	25 Apr 2023	0.2
► **Seminar plus**		
Discussion Seminar - Masterclass Phytopathology - Nick Talbot (The Sainsbury Laboratory, UK)	02 Apr 2019	0.1
Discussion Seminar - Masterclass Phytopathology - Eva Stukenbrock (Max Planck Institute of Evolutionary Biology, DE)	13 Dec 2019	0.1
► **International symposia and congresses**		
15th European Congress of Fungal Genetics, Rome, IT	17-20 Feb 2020	1.2
31th Fungal Genetics Conference, Asilomar, Pacific Groove, California, US	15-20 Mar 2022	1.5
► **Presentations**		
Poster: 'Understanding the origin, diversity and evolution of the causal agent of Panama disease', 15th European Congress of Fungal Genetics, Rome, IT	17 Feb 2020	1.0
Poster: 'Transposable elements drive genome dynamics in the plant pathogenic fungus *Verticillium dahliae* ', 15th European Congress of Fungal Genetics, Rome, IT	19 Feb 2020	1.0
Poster: ' The 3D chromatin organization contributes to the evolution of the fungal plant pathogen *Verticillium dahliae* ', 18th Dutch Chromatin Meeting, Online	12 Nov 2021	1.0
Poster: 'Gene expression divergence correlates with histone modifications in the fungal plant pathogen *Verticillium dahliae* ', 31th Fungal Genetics Conference, Pacific Groove, California, US	17 Mar 2022	1.0
Talk: 'Global epidemic of Panama disease is caused by a new fungal species originating from South-east Asia', 34th Meeting of the Fusarium working group, Utrecht, NL	30 Oct 2019	1.0
Talk: 'Evolutionary constraints of the 3D chromatin organization in a plant pathogenic fungus', Utrecht Bioinformatics Center seminar, Utrecht, NL	01 Nov 2021	1.0
Talk: 'The three-dimensional chromatin organization contributes to the evolution of the fungal plant pathogen *Verticillium dahliae* ', EPS Theme 4 Symposium "Genome Biology", Online	17 Jan 2022	1.0
Talk: 'Three-dimensional chromatin organization determines the evolution of adaptive genomic regions in the plant pathogen *Verticillium dahliae* ', 31th Fungal Genetics Conference, Pacific Groove, California, US	18 Mar 2022	1.0
Talk: 'Chromatin organization underpins the evolutionary dynamics of a soil-borne plant pathogen', Host-Microbe Genetics Meeting, Utrecht, NL	21 Oct 2022	1.0
► **IAB interview**		
► **Excursions**		
Subtotal Scientific Exposure		18.7

3) In-Depth Studies	date	cp
► **Advanced scientific courses & workshops**		
EPS course 'Transcription Factors and Transcriptional Regulation, Wageningen, NL	10-12 Dec 2018	1.0
Software Carpentry workshop 'Computational Skills for Life Scientists', Amsterdam, NL	14-15 Jan 2019	0.6
BioSB course 'Machine Learning for Bioinformatics and Systems Biology', Online	05-09 Oct 2020	1.5
EMBO workshop 'Evo-Chromo: Evolutionary approaches to research in chromatin', Aarhus, DK	11-14 May 2022	1.1
► **Journal club**		
Evolution journal club - Theoretical Biology and Bioinformatics, Utrecht University, Utrecht, NL	Apr 2020- Aug 2021	1.1
► **Individual research training**		
Subtotal In-Depth Studies		5.3

CONTINUED ON NEXT PAGE

4) Personal Development	*date*	*cp*
► **General skill training courses**		
EPS Introduction Course, Wageningen, NL	29 Oct 2019	0.3
WGS course 'Presenting with Impact', Online	03-17 May 2021	1.0
UU course 'This thing called Science', Online	Jan - Dec 2021	2.0
WGS course 'Scientific Writing', Online	15 Jun - 28 Sep 2022	2.0
► **Organisation of meetings, PhD courses or outreach activities**		
Organisation of Seminar & Masterclass Phythopathology - Nick Talbot (The Sainsbury Laboratory, UK)	02 Apr 2019	0.5
Organisation of Seminar & Masterclass Phythopathology - Eva Stukenbrock (Max Planck Institute of Evolutionary Biology, DE)	13 Dec 2019	0.3
► **Membership of EPS PhD Council**		
Subtotal Personal Development		6.1

TOTAL NUMBER OF CREDIT POINTS*	**43.6**

Herewith the Graduate School declares that the PhD candidate has complied with the educational requirements set by the Educational Committee of EPS with a minimum total of 30 ECTS credits.

** A credit represents a normative study load of 28 hours of study.*

The research described in this thesis was financially supported by Consejo Nacional de Ciencia y Tecnologia (CONACyT, Mexico), scholarship number 2018-000009-01EXTF-00188.
Financial support from Wageningen University and Research for printing this thesis is gratefully acknowledged.

Cover design by Jose Hopkins Brocq and David E. Torres Sánchez.

```
#
##
# nucleosome distance between them
pdf("nucleosomes.distance.3.strains.pdf",width=6, height=5)
ggplot(distribution,aes(x=distance,y=..scaled..,color=strain)) +
  geom_density(size=1) +
  #scale_y_sqrt(labels=percent) +
  ylab('Frequency (%)') + xlab('Distance (bp)') +
  scale_x_sqrt(breaks=c(0,1000,5000,10000,20000)) +
  #scale_x_continuous(breaks=c(1000,5000,10000,15000,20000)) +
  scale_color_manual(values = c("#228EE3", "#9A5BD4", "#62dc1c")) +
  theme_bw() +
  theme(axis.text = element_text(size=12,color='black'),
        panel.border = element_rect(colour = "black", fill=NA, size=1),panel.grid = element_blank())
####
##
#
#
#
```

```
#
##
import#
#
#
```

plant pathogen evolution plant
pathogen evolution plant
pathogen evolution plant
pathogen evolution plant
pathogen evolution plant
pathogen evolution plant
pathogen evolution plant
pathogen evolution plant
pathogen evolution plant
pathogen evolution plant
pathogen evolution plant
pathogen evolution plant
pathogen evolution plant
pathogen evolution plant
pathogen evolution plant
pathogen evolution plant
pathogen evolution plant
pathogen evolution plant
pathogen evolution plant
pathogen evolution plant
pathogen evolution plant
pathogen evolution plant
pathogen evolution plant
pathogen evolution plant
pathogen evolution plant

```
# nucleosome sizes
pdf("nucleosomes.3.strains.pdf",width=6, height=5
ggplot(distribution,aes(x=size,y=..scaled..,color=strain)) +
geom_density(size=1) +
#scale_y_sqrt(labels=percent) +
ylab('Frequency (%)') + xlab('Size (bp)') +
scale_x_continuous(breaks=c(0,150,300,450,600,750,900),expand = c(0,1,0)) +
```