CellPress

Trends in Cell Biology

Forum

Towards unraveling the origins of eukaryotic nuclear genome organization

Jolien J.E. van Hooff [1,*]

With 3D genome mapping maturing over the past decade, studies exposed the differences between eukaryotic and prokaryotic genome organization. This raises the question of how the complex eukaryotic genome organization originated. Here, I explore potential pathways to answering this question, guided by our changing understanding of the origins of eukaryotes.

Uncovering the diversity of genome organization across the tree of life

With their intracellular compartments and intricate regulation, eukaryotes are vastly more complex than prokaryotes, and so are their genomes. Eukaryotic nuclear genomes are generally larger and organized into multiple linear chromosomes, while prokaryotic genomes are typically small and comprise a single, circular chromosome (Box 1). How eukaryotic genomes are packed into the nucleus is an important question, because it affects gene regulation and probably also genome evolution. Intricate gene regulation manages eukaryotic cellular complexity, and this complexity part arose through typical eukaryotic genome evolution, harnessing many gene duplications [1]. Consequently, the emergence of eukaryotic 3D genome organization may have contributed to the emergence of eukaryotic cellular complexity. For many eukaryotes, this 3D organization has now been uncovered thanks to the advent of **high-throughput chromosome conformation capture (Hi-C)** (see Glossary) and developments in, often **fluorescence *in situ* hybridization (FISH)**-based, microscopy. Recently, several bacterial and archaeal genome organizations have similarly been scrutinized. Collectively, these studies revealed that (i) across the tree of life, species deploy some shared principles to organize their DNA and (ii) among eukaryotes, and prokaryotes, species do differ in their organizational strategies.

Genome organization of the last eukaryotic common ancestor

Eukaryotes appear to organize their genomes into more hierarchical levels than prokaryotes (Figure 1). How this complexity emerged as eukaryotes diverged from prokaryotes is still enigmatic, partially because we cannot straightforwardly envision how the genome of the **last eukaryotic common ancestor (LECA)** was organized, due to the wide diversity discerned in extant eukaryotes. Here, based on observations in distantly related eukaryotes, I assume LECA had **nucleosomes** regulated by histone post-translational modifications (histone PTMs), **topologically associating domains (TADs)**, **chromatin compartmentalization** in the nucleus, and finally individual chromosomes separated into **chromosome territories** (Figure 1). To improve and refine such a LECA prediction, we need insights into phylogenetically and biologically representative eukaryotes, hence more **protists**. Protists are under-represented in 3D genome characterizations, yet arguably very informative, because LECA itself was one. Therefore, they could provide better models for LECA's genome organization. For example, compared with their multicellular relatives, protists often have somewhat smaller genomes, due to which they may organize their genomes differently. Interesting and putatively amenable protists comprise free-living, heterotrophic organisms such as the amoeboflagellate

Glossary

Asgard archaea: an archaeal superphylum from within which, as most recent literature points to, eukaryotes emerged. This origin specifically pertains to eukaryotes' nuclear genome. Among Asgard archaea, the group that stands out as the most probable candidate for holding eukaryotes is the Heimdallarchaeia [2].
ATAC-seq (Assay for Transposase-Accessible Chromatin with high-throughput sequencing): a genome-wide method to assess the accessibility of DNA and to thereby map the chromatin landscape. ATAC-seq uses a highly active transposase to label and sequence highly accessible regions.
Chromatin: the assembly of DNA and its organizing, compacting proteins.
Chromatin compartmentalization: euchromatic regions tend to localize closely to other euchromatic regions, and the same holds for heterochromatic regions. In eukaryotes, euchromatin is concentrated in the interior of the nucleus, whereas heterochromatic compartments preferentially localize to the nuclear envelope and nucleoli. In animals, these are called 'A' and 'B' compartments, respectively.
Chromosomal interaction domains (CIDs): bacterial and archaeal genome segments within which loci interact more frequently with each other than with loci in other domains, spanning between 10 and 400 kb in length. CIDs are considered the prokaryotic analogs of eukaryotic TADs.
Chromosome territories: a chromosomal architecture in which, during interphase, each chromosome tends to occupy a specific, discrete region within the nucleus.
Euchromatin: open chromatin, not so tightly packed, very accessible to the transcription machinery.
Eukaryogenesis: the entirety of evolutionary events that caused eukaryotic cells to evolve from prokaryotic ancestors.
Eukaryotic signature proteins (ESPs): proteins functioning in eukaryotic cellular processes and conserved in most eukaryotic lineages.
Fluorescence *in situ* hybridization (FISH): imaging of DNA or RNA with fluorescently labeled probes that are complementary to the target sequence. FISH-based techniques, such as oligopaints, optical reconstruction of chromatin architecture (ORCA) and high-throughput, high-resolution, high-coverage, microscopy-based technology (Hi-M) enable visualization of genome regions at a wide range of scales.
Heimdallarchaeia: a phylum of Asgard archaea, to which eukaryotes probably belong [2].
Heterochromatin: tightly packed, condensed chromatin, mostly transcriptionally silenced.
High-throughput chromosome conformation capture (Hi-C): a method to detect 3D genomic interactions across genomic regions. It entails crosslinking of DNA fragments that are in close proximity to one another, cutting the DNA with restriction enzymes, ligating the fragments and high-throughput paired-end sequencing. Hi-C results

Box 1. Eukaryotes' global genome make-up: localized, linear, and large, but not unique?

How eukaryotic genome organization originated is strongly intertwined with how hallmark eukaryotic genome characteristics originated: the nucleus, linear chromosomes, and a large size. These characteristics might have demanded or enabled a sophisticated organization. Thus far, their origins have been mostly addressed theoretically. For example, the nucleus has been suggested to have evolved from an archaeal cell membrane or *de novo*, possibly triggered by a selective pressure to separate transcription from translation, which in turn could be caused by the emergence of spliceosomal introns. The evolution of linear chromosomes has been hypothesized to be correlated to the evolution of meiosis, and to di- or polyploidy: pairing and separating linear homologous chromosomes might be more effective. Eukaryotes' large genomes might result from gene and genome duplications, introns, and transposable elements, possibly caused by random genetic drift. While considered eukaryotic, these characteristics are sporadically observed elsewhere: nucleus-like structures in Atribacterota, linear chromosomes in *Streptomyces*, large and polyploid genomes in, for example, the myxobacterium *Sorangium cellulosum* and halophilic archaea, respectively. Although these features are probably not related to their eukaryotic counterparts, deepening our understanding of why they exist in prokaryotes may strengthen or innovate hypotheses on how they arose in eukaryotes.

Naegleria gruberi (Discoba), the foraminiferan *Reticulomyxa filosa* (Rhizaria), and unicellular relatives of animals, such as *Capsaspora owczarzaki* (Filasterea).

The archaeal roots of the eukaryotic nuclear genome

Recently, seminal discoveries have started to paint an increasingly detailed picture of how the eukaryotic cell evolved, a process called **eukaryogenesis**. Our understanding of eukaryogenesis got boosted, among others, by studies identifying **Asgard archaea** as the closest living ancestors of eukaryotes [2], with eukaryotes embedded within them (Figure 1). Most of today's theories on eukaryogenesis propose an Asgard archaea-related host that engulfed an alphaproteobacteria-related bacterium, where the latter evolved into the mitochondrion [3]. Eukaryotes are, therefore, a merger, which acquired many of their informational genes from the host and many of their metabolic genes from the endosymbiont [1]. The close link between eukaryotes and Asgard archaea is also evidenced by the presence of many **eukaryotic signature proteins (ESPs)** in Asgard archaeal genomes, although their roles in Asgard cells are mostly unknown.

Asgard archaeal genes are hardly characterized because Asgard archaea are very difficult to grow in the laboratory, and therefore, cell biological assays are thus far largely lacking. In fact, they were first identified as eukaryotes' closest relatives based on metagenome assemblies only. Consequently, virtually nothing is known about how Asgard archaea configure, organize, and regulate their genomes. While very challenging, two groundbreaking recent studies successfully cultivated and visualized Asgard archaea [4,5], which did not unveil a nucleus or other intracellular membrane structures. Another study showed that two Asgard archaeal lineages spatially separate their DNA from ribosomes, albeit without signs of a separating membrane [6]. Notably, the DNA-ribosome separation was substantially more pronounced than in *Escherichia coli* and the **TACK** archaeon *Nitrosopumilus maritimus* [6], which confine their chromosome (the **nucleoid**) to a small part of the cell. Hence, in contrast to other characterized prokaryotes, these Asgard archaea might enable this wider separation using a yet unobserved physical barrier, such as a membrane.

Interestingly, Asgard archaea have larger genomes than other archaea [2], which may require a more sophisticated organization to properly regulate transcription and replication. For example, maybe Asgard archaea harbor eukaryote-like **chromatin** compartmentalization (Figure 1), which was in fact observed in the TACK archaeon *Sulfolobus* [7]. If so, compartmentalization may already have existed

in a 2D contact map of chromatin interactions across a chromosome region, a whole chromosome or a whole genome, depending on the resolution. From this contact map, a 3D genome structure can be modeled.

Lamina attraction: eukaryotic heterochromatin tends to cluster at the periphery of the nucleus. One explanatory hypothesis holds that nuclear lamina attract heterochromatin to the nuclear envelope.

Last eukaryotic common ancestor (LECA): the most recent ancestor of all eukaryotic lineages. LECA was probably a unicellular, biflagellated eukaryote ('protist'), which had all characteristic intracellular complexity, such as the intricate endomembrane system, mitochondria, an intricate cytoskeleton, spliceosomal introns, and the capacity to execute meiosis and phagocytosis.

Nucleoid: the prokaryotic chromosome, typically compacted by nucleoid-associated proteins (NAPs), and occupying only a fraction of the cellular volume. In contrast to the eukaryotic nucleus, nucleoids are not surrounded by a nuclear membrane.

Nucleosomes: the primary building blocks of eukaryotic chromatin; the nucleosome consists of ~147 bp of DNA wrapped around an octamer of histone proteins. Post-translational modifications (PTMs) of histone proteins affect the accessibility of the DNA to proteins such as transcription factors.

Protists: unicellular eukaryotes, found in all major branches of the eukaryotic tree of life. Over the past decade, many new kingdom-level protist lineages were discovered, illustrating that they probably represent a large and partially unexplored biodiversity.

Rabl-like: a chromosomal architecture in which, during interphase, the centromeres and/or telomeres of different chromosomes cluster in the nucleus.

TACK: an archaeal superphylum composed of, among others, Thaumarchaeota, Aigarchaeota, Crenarchaeota, and Korarchaeota, and which forms a sister clade to Asgard archaea (including eukaryotes, Figure 1). Early on, before the discovery of Asgard archaea, several TACK lineages had been shown to possess some ESPs.

Topologically associating domains (TADs): genome segments within which loci interact more frequently than with those in other domains, including neighboring chromatin regions. Hence, TADs, which were revealed through Hi-C, are discerned through blocks of many contacts in Hi-C contact maps, with clear borders. TADs span tens to hundreds of kilobases and have been observed in various eukaryotes. TADs probably act as regulatory units and replication domains.

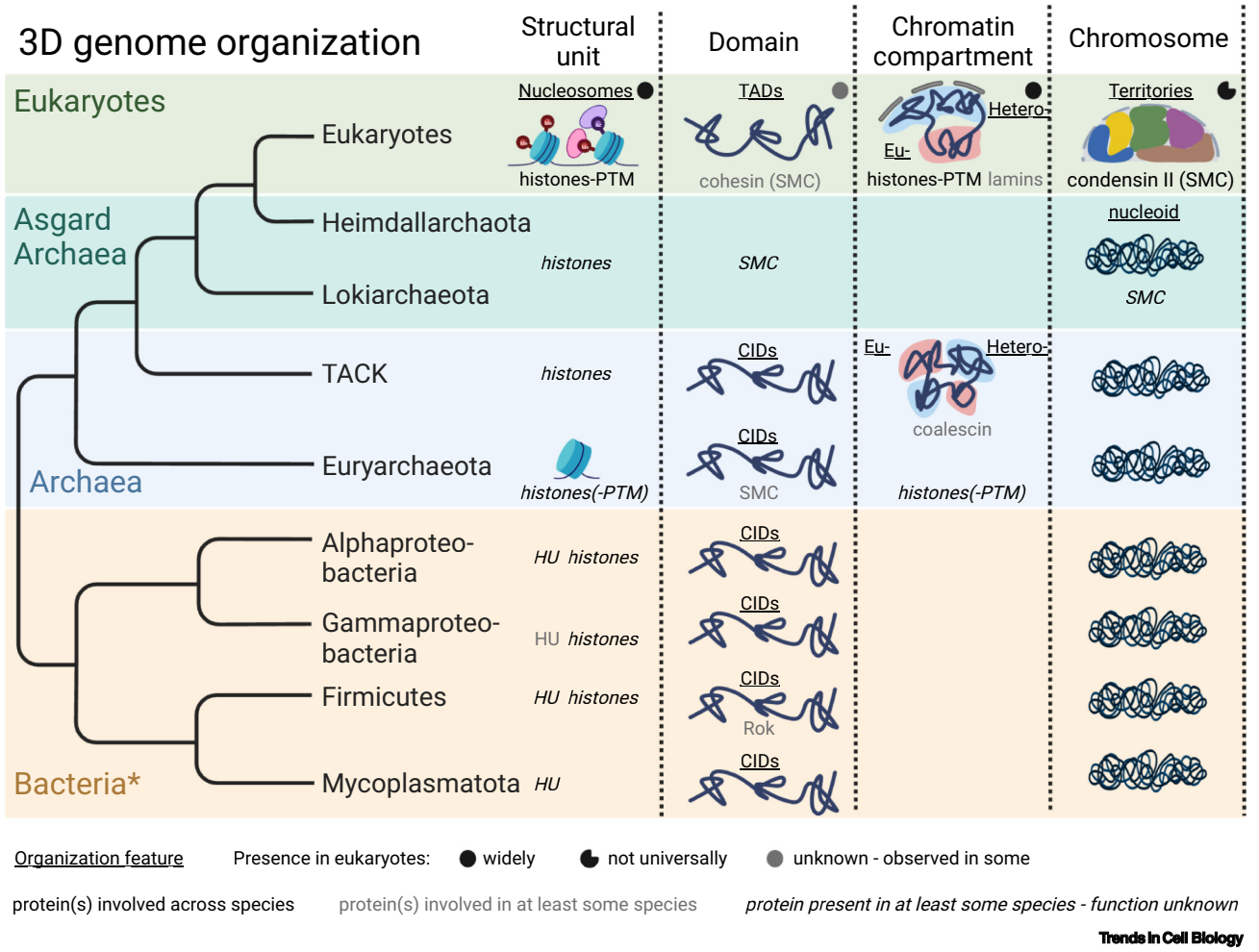

Figure 1. Organizational features of eukaryotic, archaeal, and bacterial genomes and the putative proteins involved. Simplified species tree of clades for which data on 3D genome organization is available, or species that might be informative for the evolution of eukaryotic 3D organization (Asgard archaea). Organizational features at different scales are displayed from small to large (left to right), insofar as they are relevant to eukaryotes. Importantly, the absence of a given feature illustration does not necessarily mean the feature is absent from this clade: it just has not been observed (yet). Moreover, protein presence is mostly predicted from bioinformatics studies without any functional characterization; therefore, such a presence does not imply the presence of a particular organizational feature (as indicated by the italicized text). *Various bacteria also harbor higher-order organization features, that is, larger than CIDs. Yet, these are different from eukaryotic chromatin compartments as they do not separate euchromatin and heterochromatin. For example, *Escherichia coli* (Gammaproteobacteria) has so-called macrodomains, which are enabled by SMC (MukBEF). Created with BioRender.com. Abbreviations: CIDs, chromosomal interaction domains; PTM, post-translational modification; TACK, Thaumarchaeota, Aigarchaeota, Crenarchaeota, and Korarchaeota; TADs, topologically associating domains.

in the common ancestor of Asgard archaea and eukaryotes. When cultivation methods for Asgard archaea continue to improve, and alongside strategies are developed to subject them to Hi-C, **Assay for Transposase-Accessible Chromatin with high-throughput sequencing (ATAC-seq)**, transcriptomics, and chromatin imaging, we might be able to assess if they have and compartmentalize **euchromatin** and **heterochromatin**. Such techniques will allow us to study Asgard archaeal chromosome organization and regulation from the gene to the chromosome level. Particularly if applied to diverse lineages of **Heimdallarchaeia**, they could pave the way for delineating the evolution of genome organization during eukaryogenesis.

Tracing the prokaryotic roots of eukaryotic genome-shaping proteins

The increasing number of archaeal genome sequences allows us to scrutinize in depth the origins of proteins involved in eukaryotic genome organization. These proteins were probably derived from the archaea-related host, but may also have evolved *de novo*. The proteins shaping

the basic structural units of eukaryotic chromatin, histones, have long been known to be present in archaea, and shape archaeal chromatin [8]. Recently, histones were also found in bacteria and may also organize their DNA [9,10]. If and how Asgard archaeal histones organize DNA is not clear, yet they probably have DNA-binding capacity [8]. Recently, PTMs were detected in the histones of some Euryarchaeota (Figure 1), but their large-scale usage for chromatin regulation might embody a eukaryotic innovation [11]. Histone PTMs demarcate euchromatin and heterochromatin, and therefore also underlie chromatin compartmentalization, potentially in collaboration with lamina-associated proteins that might attract heterochromatin to the nuclear envelope, a process referred to as **lamina attraction** (Figure 1). To my knowledge, lamins have not been detected in prokaryotes and are probably a eukaryotic innovation that accompanied the origination of the eukaryotic nuclear envelope (Box 1).

Eukaryotic genomes are organized into TADs, and TAD formation is supported by the structural maintenance of chromosomes (SMC) complex cohesin in some species (Figure 1). A recent seminal study revealed that another SMC complex, condensin II, has a conserved role in forming chromosome territories [12]. Species lacking condensin II instead form a **Rabl-like** chromosome configuration. Whereas eukaryotes bear four different SMC complexes, most prokaryotes probably contain only one (bacteria) or two (archaea) [13], and the expansion of the SMC complex inventory in eukaryogenesis likely played a role in layering genome organization.

Of course, the presence, or absence, of eukaryotic genome-shaping proteins does not imply the presence, or absence, of a particular organizational feature. For example, a TAD/**chromosomal interaction domain (CID)**-like organization could be widespread (Figure 1), yet the proteins involved might be different, implying analogous rather than homologous features.

Towards evolutionary cell biology of genome organization

To uncover the origins of eukaryotic nuclear genome organization, we first need to reconstruct LECA's genome organization. This requires a broader sampling of current-day eukaryotes, particularly protists, combined with an enhanced understanding of the underlying molecular mechanisms across species, to distinguish homologous from analogous features. This may also involve genetics and cellular biology in nonmodel eukaryotes. Second, we need a model of the starting point of the transition: genome organization of the Asgard archaea-related ancestor of eukaryotes. Building this model will be facilitated by characterizing genome organization in current-day Asgard archaea. Ultimately, understanding the evolution and diversity of 3D genome organization across the tree of life will spark further exploration of how it affects gene regulation and evolution.

Acknowledgments

I thank Geert Kops, Michael Seidl, and Thijs Ettema for their feedback. This work was supported by a personal fellowship from the Dutch Research Council (NWO VI. Veni.212.099).

Declaration of interests

None declared by the author.

[1]Laboratory of Microbiology, Wageningen University and Research, 6708 WE Wageningen, The Netherlands

*Correspondence:
jolien.vanhooff@wur.nl (J.J.E. van Hooff).

https://doi.org/10.1016/j.tcb.2023.07.008

References

1. Vosseberg, J. *et al.* (2021) Timing the origin of eukaryotic cellular complexity with ancient duplications. *Nat. Ecol. Evol.* 5, 92–100
2. Eme, L. *et al.* (2023) Inference and reconstruction of the heimdallarchaeial ancestry of eukaryotes. *Nature* 618, 992–999
3. Gabaldón, T. (2021) Origin and early evolution of the eukaryotic cell. *Annu. Rev. Microbiol.* 75, 631–647
4. Imachi, H. *et al.* (2020) Isolation of an archaeon at the prokaryote–eukaryote interface. *Nature* 577, 519–525
5. Rodrigues-Oliveira, T. *et al.* (2022) Actin cytoskeleton and complex cell architecture in an Asgard archaeon. *Nature* 613, 332–339
6. Avcı, B. *et al.* (2022) Spatial separation of ribosomes and DNA in Asgard archaeal cells. *ISME J.* 16, 606–610
7. Takemata, N. *et al.* (2019) Physical and functional compartmentalization of archaeal chromosomes. *Cell* 179, 165–179.e18
8. Henneman, B. *et al.* (2018) Structure and function of archaeal histones. *PLoS Genet.* 14, e1007582
9. Alva, V. and Lupas, A.N. (2019) Histones predate the split between bacteria and archaea. *Bioinformatics* 35, 2349–2353
10. Hocher, A. *et al.* (2023) Histone-organized chromatin in bacteria. *bioRxiv* Published online January 26, 2023. https://doi.org/10.1101/2023.01.26.525422
11. Grau-Bové, X. *et al.* (2022) A phylogenetic and proteomic reconstruction of eukaryotic chromatin evolution. *Nat. Ecol. Evol.* 6, 1007–1023
12. Hoencamp, C. *et al.* (2021) 3D genomics across the tree of life reveals condensin II as a determinant of architecture type. *Science* 372, 984–989
13. Yoshinaga, M. and Inagaki, Y. (2021) Ubiquity and origins of structural maintenance of chromosomes (SMC) proteins in eukaryotes. *Genome Biol. Evol.* 13, evab256