Domestication Shapes Recombination Patterns in Tomato

Roven Rommel Fuentes,[1] Dick de Ridder ![ORCID],[1] Aalt D.J. van Dijk,*,[1] and Sander A. Peters ![ORCID]*,[2]

[1]Bioinformatics Group, Wageningen University and Research, Wageningen, The Netherlands
[2]Applied Bioinformatics, Wageningen Plant Research, Wageningen University and Research, Wageningen, The Netherlands

*Corresponding authors: E-mails: aaltjan.vandijk@wur.nl; sander.peters@wur.nl.
Associate editor: Michael Purugganan

Abstract

Meiotic recombination is a biological process of key importance in breeding, to generate genetic diversity and develop novel or agronomically relevant haplotypes. In crop tomato, recombination is curtailed as manifested by linkage disequilibrium decay over a longer distance and reduced diversity compared with wild relatives. Here, we compared domesticated and wild populations of tomato and found an overall conserved recombination landscape, with local changes in effective recombination rate in specific genomic regions. We also studied the dynamics of recombination hotspots resulting from domestication and found that loss of such hotspots is associated with selective sweeps, most notably in the pericentromeric heterochromatin. We detected footprints of genetic changes and structural variants, among them associated with transposable elements, linked with hotspot divergence during domestication, likely causing fine-scale alterations to recombination patterns and resulting in linkage drag.

Key words: domestication, selective sweeps, structural variants, transposable elements, recombination, heterochromatin.

Introduction

Generation of genetic diversity through meiotic recombination has been an integral subject of breeding and genome research following Mendel's work on patterns of inheritance and Morgan's theory of gene linkage and crossing-over (Hunter 2015). Studies on meiotic recombination have indicated that one crossover (CO) per chromosome is obligatory and essential for proper chromosome segregation during prophase I (Jones and Franklin 2006). COs are nonuniformly distributed across plant genomes, mostly located in distal chromosome regions and clustering in hotspots (Mercier et al. 2015; Lambing et al. 2017; Wang and Copenhaver 2018). Characterizing the pattern of recombination has both practical and fundamental relevance, as it enables identification of informative markers, building of genetic maps and genome assemblies, reconstruction of evolutionary histories, association studies on important alleles and profiling linkage drags (Cardon and Abecasis 2003; Wang et al. 2010; Yang et al. 2011; Fransz et al. 2016; Dutta et al. 2017; Marand et al. 2019).

Different methods have been developed to detect recombination events between specific parents, such as chiasmata and recombination nodule counting (Stack et al. 1989; Anderson et al. 2003), pollen genotyping (Drouaud et al. 2013), single nucleotide polymorphism (SNP)-array-based profiling or sequencing of recombinant-inbred line (RIL) populations (Huang et al. 2009; Qi et al. 2009; Wijnker et al. 2013; Demirci et al. 2017), and image analysis of meiotic tetrads (Francis et al. 2007; Lim et al. 2020). These approaches revealed useful information about the recombination landscape of gametes and offspring populations. Still, detection of recombination between all members of a population is impractical and often infeasible, especially for long-generation species. However, analysis of resequencing data of natural populations allows detection of historical recombination rates and associated genomic features, revealing their evolutionary histories. Here, historical recombination refers to the reciprocal exchange of chromosomal segments that has successively occurred between ancestral individuals over multiple generations, in various environments and under changing selective pressures and genetic backgrounds. Recombination rate landscapes and historical CO hotspots in populations of different plants (Choi et al. 2013; Dreissig et al. 2019; Marand et al. 2019; Schwarzkopf et al. 2020), fungi (Stukenbrock and Dutheil 2018), insects (Chan et al. 2012), and mammals (Brunschwig et al. 2012; Stevison et al. 2016; Guo et al. 2018) have previously been subjected to coalescent-based analysis of genetic variation.

Studying historical recombination can help explain the changes that domestication enforced on recombination patterns. Domestication is a process of human-imposed evolution by selection of favorable phenotypes, resulting in genetic modification of wild progenitors to create new forms that meet human needs (Doebley et al. 2006; Yang et al. 2019). In this process, an initial stage of cultivating wild species with desirable traits is followed by a second stage of improvement, further targeting specific traits through selective breeding (Gross and Olsen 2010; Meyer and Purugganan 2013). Strong selection during domestication is accompanied by

1

increased recombination rate in many species (Ross-Ibarra 2004; Moyers et al. 2018). For example, the domesticated population of cacao was reported to have a higher recombination rate compared with wild populations (Schwarzkopf et al. 2020). In addition, the recombination landscape in barley is highly conserved throughout domestication, with fine-scale changes in recombination rate that have been linked to different environmental conditions and with defense-response genes (Dreissig et al. 2019). However, currently, for most plants, information on how domestication shaped the recombination landscape is still scarce.

Given that several genomic features like promoter regions, repeat motifs, nucleosome occupancy, chromatin accessibility, structural variations (SVs), transposable elements (TEs), and nucleotide diversity have been found linked to CO incidence, changes to their patterns due to evolution of a species may have influenced the recombination profiles (Petes 2001; Choi and Henderson 2015; Termolino et al. 2016; Lambing et al. 2017; Dluzewska et al. 2018). Population-level patterns of recombination in different species revealed substantial divergence of hotspots across populations of the same species. Distinct COs hotspots, associated with lineage-specific variation in SV and TE profiles, have been implicated as major drivers in population dynamics (Marand et al. 2019). In rice, potato, and *Arabidopsis thaliana*, specific superfamilies of DNA transposons are abundantly located in recombination-prone regions, which may be explained by nucleosome depletion in DNA transposons (Marand et al. 2017; Choi et al. 2018; Marand et al. 2019).

Domesticated in South America, wild tomato species *Solanum pimpinellifolium* (SP) gave rise to the cherry tomato (*Solanum lycopersicum* var. *cerasiforme*; SLC), which was later improved into the big-fruited tomato (*Solanum lycopersicum* var. *lycopersicum*; SLL) in Mesoamerica (Lin et al. 2014; Blanca et al. 2015; Razifard et al. 2020). Due to domestication and continued selection, current tomato cultivars have lost 95% of the genetic diversity of their wild relatives, pushing breeders to introgress alleles from compatible wild relatives underlying disease resistance, stress tolerance, adaptation to diverse environments, higher yield, and fruit quality (Bai and Lindhout 2007). However, there are some reproductive barriers such as SVs (Soyk et al. 2019) that limit the applications of introgressive hybridization breeding. It was observed that linkage disequilibrium decays over a longer distance in domesticated tomato compared with its wild relatives, implying changes into the recombination patterns (Lin et al. 2014; Zhu et al. 2018). In this study, we address how domestication shaped the recombination patterns specifically for tomato and its related wild species. We investigate CO profiles in tomato populations and its wild relative *S. pimpinellifolium*, using resequencing data assigned into three taxonomic groups: wild tomato (SP), early-domesticated types (SLC), and vintage or heirloom cultivars (SLL). We generated recombination landscapes, identified recombination hotspots, and analyzed genomic features that are associated with the differing recombination patterns between the populations. This provided insights into the factors that contributed to or are associated with the changes in local recombination patterns during

tomato domestication, revealing how domestication has severely constrained the ability of recombination to generate diversity, both for inbred and hybrid crosses. This new data may help the selection of targets for inducing meiotic recombination, cross-checking hotspots in hybrids, identifying tightly linked genes, and defining recombination barriers in hybridization.

Results

Conserved Recombination Landscape between Wild and Vintage Tomato

After domestication, the genetic diversity of tomato has dramatically reduced and linkage disequilibrium decays over a longer distance, indicating changes in recombination patterns (Aflitos et al. 2014; Lin et al. 2014; Tieman et al. 2017; Razifard et al. 2020). Furthermore, there are structural rearrangements between wild and domesticated plants that hamper recombination, manifested by the phenomenon of cosegregation of specific alleles linked to a desired trait (linkage drag). To get more insight into how domestication influences recombination patterns, we profiled the recombination landscape of tomato and wild relatives based on existing resequencing data of 75 accessions from each of the wild (SP), early-domesticated (SLC) and vintage (SLL) populations (fig. 1A; supplementary fig. 1 and table 1, Supplementary Material online).

Consistent with available CO data from recombinant-inbred lines (referred to from here on as COD1; Demirci et al. 2017) and pollen gametes of an interspecific cross (COD2; Fuentes et al. 2020) in tomato, and data from other species (Wijnker et al. 2013; Kianian et al. 2018), the majority of historical recombination in both wild and domesticated tomato occurred in the distal gene-rich euchromatic regions of the chromosomes. More in detail, the recombination landscape of each population correlates with both **COD1** (Spearman's rank correlation; euchromatin, $\rho = 0.32$–0.44; $P < 2.2 \times 10^{-11}$; heterochromatin, $\rho = 0.64$–0.67; $P < 2.2 \times 10^{-16}$) and **COD2** (Spearman's rank correlation; euchromatin, $\rho = 0.31$–0.55; $P < 2.2 \times 10^{-10}$; heterochromatin, $\rho = 0.38$–0.51; $P < 2.2 \times 10^{-16}$). To further verify the consistency of the recombination rates with available data, the population-scaled recombination rate computed using LDhat was converted from ρ/kb to cM/Mb and compared against the EXPIM2012 genetic map generated from a cross between *S. lycopersicum* and *S. pimpinellifolium* (Sim et al. 2012). The correlation between the genetic map and recombination rate estimates of each population is approximately 0.9 (Spearman's rank correlation; $P < 2.2 \times 10^{-16}$), supporting the concordance of population-scaled recombination rates with the genetic map.

To compare the recombination landscapes of the three populations in our study, we calculated multiscale correlations, that is, in varying window sizes (supplementary fig. 2, Supplementary Material online), and selected a 1-Mb window size. We found correlations in the range of 0.6–0.7, indicating conservation of the genome-wide recombination landscape despite the differing selection and domestication processes
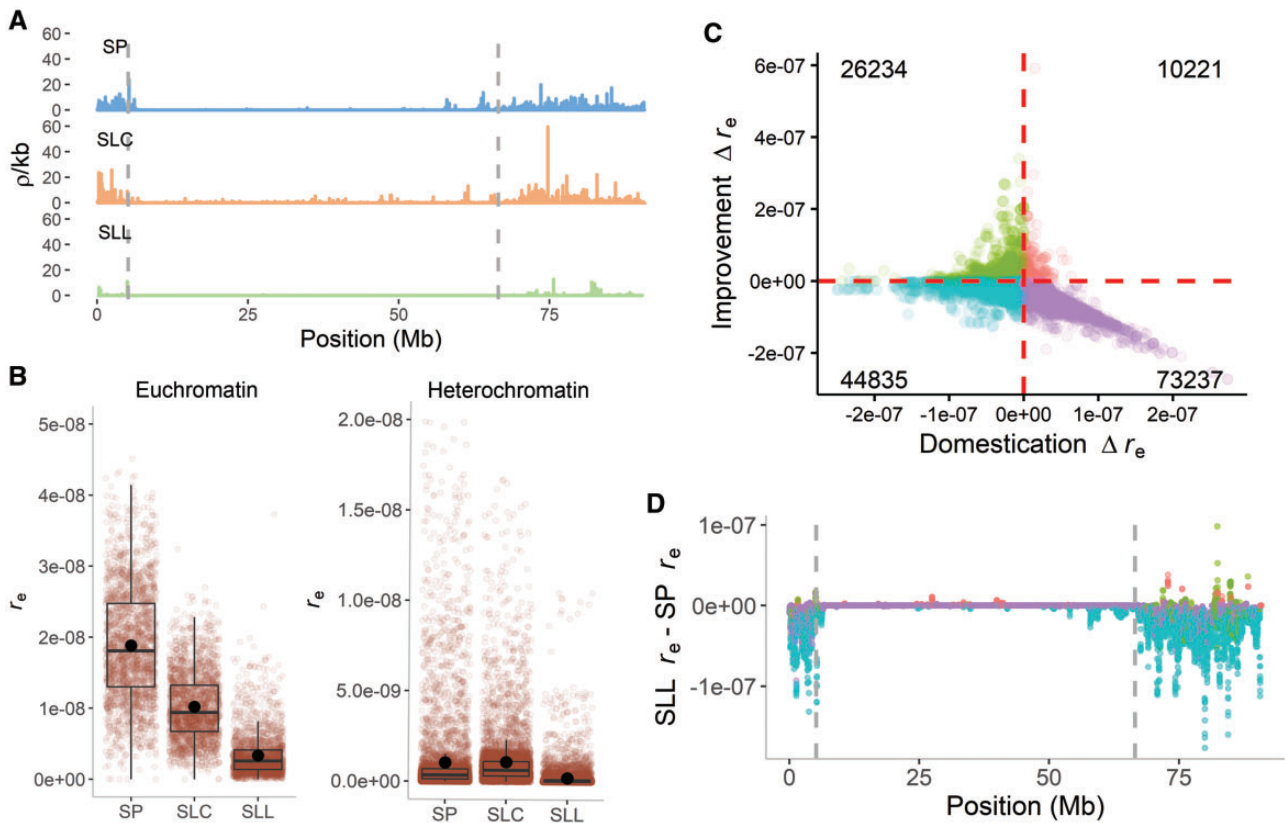
FIG. 1. Recombination landscape and transformation from wild to domesticated tomato. (A) Recombination landscape in chromosome 1 of wild (SP), early-domesticated (SLC), and vintage (SLL) tomato. This ρ/kb landscape is intended to show overall landscape only; r_e is used to compare populations in other analyses. Gray vertical lines mark heterochromatin boundaries. (B) Effective recombination rate (r_e) in 1-Mb windows of both wild and domesticated tomato. (C) Change in effective recombination rate in 50-kb regions during domestication (SLC r_e–SP r_e) and improvement (SLL r_e–SLC r_e). (D) Resulting change in r_e for chromosome 1 after the domestication process or between the wild and vintage population. Gray vertical lines mark the heterochromatin boundaries and the colors correspond to the colors in (C).

(Spearman's rank correlation; $P < 2.2 \times 10^{-16}$, supplementary fig. 3, Supplementary Material online). The higher correlation coefficient between genetic distances and population-scaled recombination rate is influenced by the low marker density in the genetic map, limiting the comparison to the overall landscape. On the other hand, the use of shorter windows when comparing rates between populations or against COD1 and COD2 accounts for local changes in the recombination rates (supplementary fig. 2, Supplementary Material online).

Local Increase of Recombination Rate in Early-Domesticated Tomato

Although the general landscape of recombination is conserved, there are also clear local changes between populations (supplementary fig. 1, Supplementary Material online). Using the ρ value calculated by LDhat, we computed the effective recombination rate (r_e) to account for the difference in effective population size (N_e) of the three groups. We found that the median recombination rate in the SLC population (r_e =9.3×10^{-10}) is higher than the median in both SP (r_e=5.6×10^{-10}) and SLL (r_e=6.8×10^{-11}). Given that the heterochromatin regions have a low recombination rate, we separately analyzed r_e for euchromatic and heterochromatic regions. We determined the borders between euchromatin

and heterochromatin by computing the euchromatin length (μm) from the average length of pachytene chromosome, multiplying it by the euchromatin DNA density (1.54 Mb/μm) and using the length of each euchromatin to identify the heterochromatin boundary. In euchromatin, SP has a significantly higher r_e of 1.8×10^{-8} compared with 9.4×10^{-9} for SLC and 2.6×10^{-9} for SLL. But in the heterochromatin, SLC has a higher median r_e of 6.3×10^{-10} compared with 3.5×10^{-10} and 4.8×10^{-11} for SP and SLL, respectively (fig. 1B; supplementary table 2, Supplementary Material online).

We plotted the change in r_e during the domestication (from SP to SLC) and improvement (from SLC to SLL) stages to detect localized changes of recombination rates (fig. 1C). In the majority of cases, where there is an increase of r_e through domestication (SLC r_e–SP r_e >0) across the whole genome, they are followed by a proportional decrease in r_e during improvement (SLL r_e–SLC r_e <0). The reduction of r_e through domestication is confined mostly to the euchromatic region (fig. 1D; supplementary fig. 4, Supplementary Material online). Although 65% of euchromatin regions have reduced r_e during domestication, over the entire genome more regions (54%) have an increased rate. On the other hand, during the improvement stage, the effective recombination rate in 76% of

the genome was reduced in both euchromatin and heterochromatin.

The local increase of r_e from SP to SLC is consistent with the fact that domestication increases the actual recombination rate in many species, as was demonstrated previously by counting chiasmata per bivalent (Ross-Ibarra 2004; Moyers et al. 2018). This increased recombination is favored during periods of rapid evolutionary change and specifically during domestication (Rees and Dale 1974; Burt and Bell 1987; Otto and Barton 1997; Ross-Ibarra 2004). On the other hand, the significant reduction of effective recombination rate during improvement may be explained by increased inbreeding and homozygosity in the vintage accessions (Moyers et al. 2018); SP and SLC are known to have higher outcrossing rates than SLL (Rick et al. 1978; Rick and Holle 1990). As previously reported, inbreeding results in reduced heterozygosity and effective population size, and longer distance for linkage disequilibrium decay (Allard 1999; Morrell et al. 2003; Kovach et al. 2007). The estimated effective recombination rate may be reduced in inbred species if the homologous chromosomes are identical and no appreciable exchange of alleles is observed after recombination (Moyers et al. 2018). The decreased effective recombination rate has actually been observed in maize improved lines with an estimated 82.3% reduction compared with the wild progenitors (Hufford et al. 2012).

To check for the genetic diversity in each population, we first computed the number of SNPs. The average SNP count per kilobase for SP, SLC, and SLL accessions was 0.097, 0.075, and 0.021, respectively, which indicates a reduction of genetic diversity from wild to heirloom tomato (pairwise Wilcoxon rank-sum test; $P < 8.5 \times 10^{-3}$; supplementary fig. 5, Supplementary Material online). At the individual sample level, SP accessions have more SNPs than SLC accessions, but SLC has more unique SNP sites in the overall whole population. The mean nucleotide diversity measured for SLL (4.4×10^{-4}) and SLC (2.2×10^{-3}) populations is lower than that for SP (3.8×10^{-3}), which is typically associated with domestication syndrome, the distinguishing characteristics between domesticated crop and wild ancestors (Doebley et al. 2006; Bai and Lindhout 2007; Sauvage et al. 2017). Both increased homozygosity and inbreeding contributed to the reduced effectiveness of recombination in the vintage population.

Divergent Hotspots between Populations

Visual inspection of the landscape (supplementary fig. 1, Supplementary Material online; fig. 1A) reveals local peaks of recombination rates throughout the genome of each population. To further investigate this, we detected historical recombination hotspots for each population using sequenceLDhot, reporting 1,784, 2,899, and 667 hotspots for SP, SLC, and SLL, respectively (fig. 2A; supplementary tables 2 and 3, Supplementary Material online) and a total of 5,082 unique hotspots with a median size of 2 kb. Pairs of the three populations have 4–10% of hotspots in common, significantly higher than expected by chance (pairwise Fisher's exact test; $P < 1.9 \times 10^{-5}$; fig. 2B). However, of 181 hotspots shared

between SP and SLC, only four are retained in SLL. This low overlap in hotspots between populations of the same or closely related species was also reported in rice and cocoa (Marand et al. 2019; Schwarzkopf et al. 2020). SLC has 62.5% more unique hotspots than SP in concordance with its increased recombination rate. Out of all the identified hotspots, 84 (1.6%) and 96 (1.8%) overlapped with empirical COs in **COD1** (Fisher's exact test; $P = 6.3 \times 10^{-24}$) and **COD2** (Fisher's exact test; $P = 1.2 \times 10^{-26}$), respectively. Furthermore, 3 of the 23 reported hotspots in Fuentes et al. (2020) match hotspots in SP and SLC. The limited overlap between historical hotspots and COs found in these previous studies is in line with the relatively modest correlation between the population-scaled recombination rate and COs from **COD1 and COD2** reported above. These observations may be explained by the inclusion of different genotypes in the different data set, by the possibly differing CO patterns between the RIL/pollen populations and the natural tomato population, or by the inability of experimental studies to exhaustively sample possible recombination sites. The hotspots mentioned in the succeeding sections refers to recombination hotspots, except for those that explicitly refer to hotspots from previous studies.

Recombination in the Pericentromeric Heterochromatin

Aside from divergent hotspots, another distinct difference between the recombination landscapes of wild and domesticated tomato is the presence of hotspots in the pericentromeric regions. Previous studies in plants mostly report suppression of recombination in pericentromeric regions, which is largely heterochromatic, and high recombination rates in the distal chromosome regions. However, we observed that hotspots are not confined to the terminal chromosome ends, but also are scattered over the pericentrome. Compared with both SP (51.8%) and SLC (61.5%), only 32.7% of the genome-wide hotspots in vintage tomato (SLL) are located in the pericentromeric regions, which covers 75% of the genome. This rate is comparable to the 39.1% of hotspots distributed in the pericentromere of maize (Pan et al. 2017). Sherman and Stack (1995) actually reported cases of recombination nodules in the pericentromeric heterochromatin of tomato, but these are 20–50× less frequent per unit length of the synaptonemal complex than in euchromatin. This frequency, though, may be different for historical hotspots. Additionally, hotspots in the heterochromatin have generally lower recombination rates than those located in the euchromatin (Wilcoxon rank-sum test; SP, $P < 2 \times 10^{-16}$; SLC, $P < 2 \times 10^{-16}$; SLL, $P = 6.2 \times 10^{-12}$; fig. 2C). To determine whether the observed recombination rates in the heterochromatin are due to the use of a domesticated tomato as reference genome for wild accessions which contains genomic rearrangements, we recomputed recombination rates for the wild population using the *S. pimpinellifolium* genome assembly (Wang et al. 2020) as reference. As shown in supplementary figure 6 (Supplementary Material online), the pericentromeres still exhibit presence of historical COs. It is also important to emphasize that SNPs located in the repeat
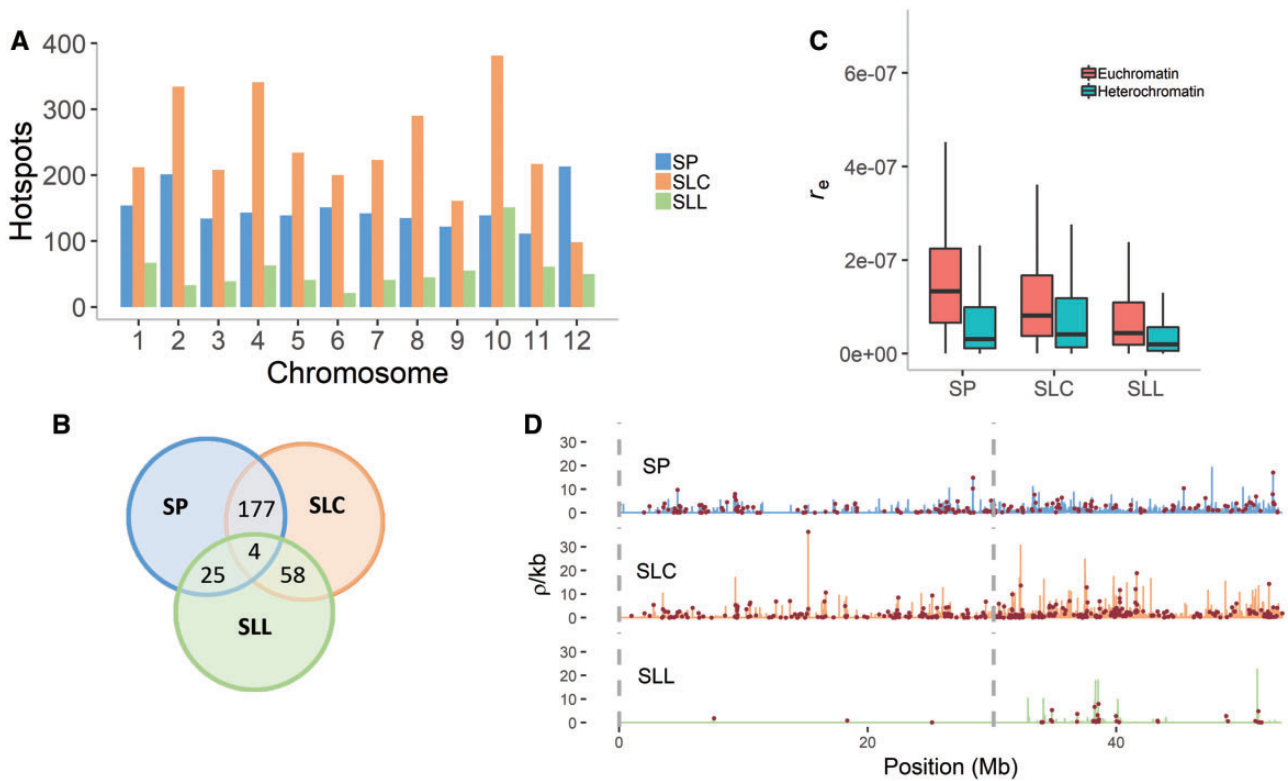
FIG. 2. Historical recombination hotspots. (A) Number of hotspots in each chromosome of the wild and domesticated populations. (B) Small but significant numbers of hotspots are shared between populations. (C) Effective recombination rates of hotspots in euchromatic and heterochromatic regions. (D) Recombination hotspots in chromosome 2. Gray lines mark the heterochromatin boundaries.

or TE regions were excluded in the estimation of recombination rates to avoid issues due to misaligned reads or false-positive SNPs.

To give an example of recombination occurring in the pericentromeric heterochromatin, we show the landscape for chromosome 2 in figure 2D, where both SP and SLC show presence of historical recombination in the heterochromatin. However, the same region is almost devoid of recombination in SLL; hotspots are clustering mostly at the ends of the chromosome, consistent with population data, the genetic map (Sim et al. 2012), RILs (Demirci et al. 2017), and pollen gametes from interspecific crosses (Fuentes et al. 2020). Around 55% and 42% of SP and SLC hotspots in chromosome 2, respectively, are located in the pericentromeric regions in contrast to 9% of SLL, but the number of hotspots per megabase of euchromatin is still higher than in heterochromatin for both SP and SLC. We further examined these heterochromatic hotspots and found that they are close to or within genes (Fisher's exact test; $P = 4.9 \times 10^{-12}$), which suggests that these hotspots may be located in euchromatin islands or accessible regions in the heterochromatin.

Crossover Hotspots in Selective Sweep Genes

The results above confirm that historical COs are nonuniformly distributed over the genome and occur mostly in the distal part of the chromosome. This raises the question how the changing patterns of recombination hotspots may be linked to specific genomic features that evolve during domestication. To test the association of recombination hotspots with gene features, for each population, a permutation test of CO hotspots with sizes below 5 kb was applied. This shows a significant enrichment in promoter regions, defined as 1-kb regions upstream of the transcription start sites, and in gene bodies. Moreover, hotspots are depleted in intergenic regions, further supporting previous reports of recombination mostly occurring near genes. To account for the significant difference of CO distribution in the euchromatin and heterochromatin, we performed enrichment analysis for euchromatin regions only and still found an excess in promoters and gene bodies (fig. 3A). The promoter regions exhibit a 3- to 13-fold increase in COs over the background, which was computed based on a set of 10,000 permutations. Despite the reduction of hotspots in vintage tomato (supplementary table 2, Supplementary Material online), the enrichment in both gene bodies and promoters persists. An excess of COs in promoter and untranslated regions (UTRs) of tomato was previously reported in RILs and pollen data (de Haas et al. 2017; Demirci et al. 2017; Fuentes et al. 2020) and is consistent with observations in other species (Wijnker et al. 2013; Marand et al. 2017; Pan et al. 2017; Choi et al. 2018; Demirci et al. 2018; Kianian et al. 2018). The overrepresentation in the 5′ and 3′ UTRs may be related to epigenetic modifications such as DNA and histone methylation, contrasting open chromatin that may be accessible for the recombination machinery (Eichten et al. 2011; Pan et al. 2017). With the preference of COs to occur near genes, the
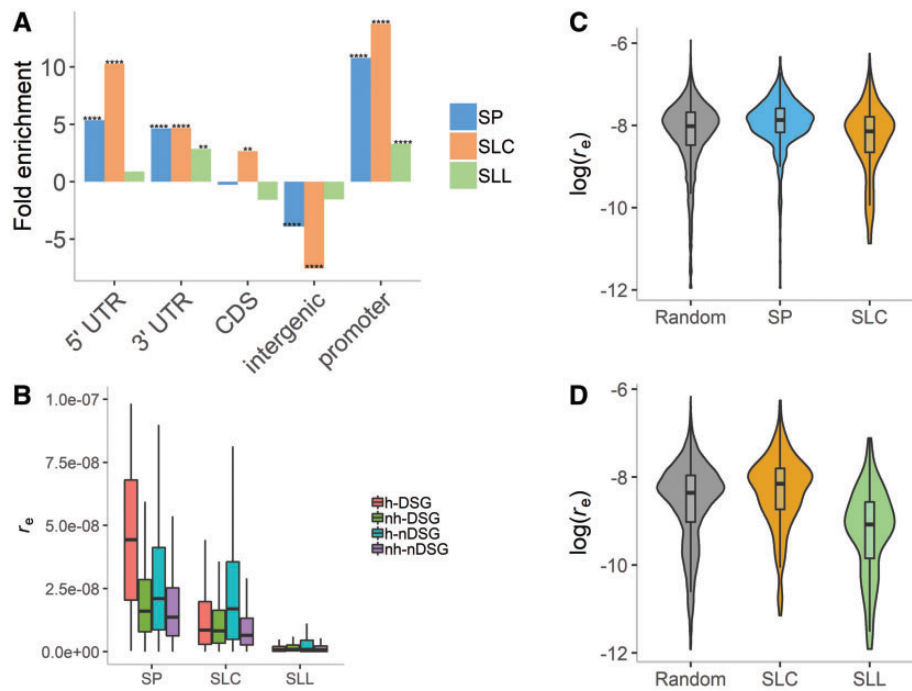
Fig. 3. Recombination hotspots in genes. (A) Enrichment of euchromatin hotspots in UTRs and promoter regions (1-kb upstream of genes) in all three populations. (B) Recombination rates of domestication (DSG) and nondomestication (nDSG) sweep genes overlapping and not overlapping *S. pimpinellifolium* hotspots. **h and nh** mean hotspots and nonhotspots, respectively. (C–D) Recombination rate upstream (<1 kb) of genes with excised promoters due to (C) domestication and (D) improvement.

significant reduction of genes in domesticated tomatoes compared with wild relatives consequently limits the possible sites for recombination (Gao et al. 2019).

Knowing that COs preferentially occur near genes and that specific genes are affected by domestication, we examined the relation between recombination hotspots and genes. In detecting association, we included the 2-kb regions flanking both sides of the genes. First, we compared hotspots against R genes, which are known to reside in recombination hotspots (Nieri et al. 2017; Andolfo et al. 2021), and observed a significant overlap (Fisher's exact test; $P = 3.38 \times 10^{-17}$). Afterward, we computed the overlap between hotspots and the selective sweeps or regions with reduced nucleotide diversity previously reported by Lin et al. (2014). We refer to genes in these selective sweeps as domestication (DSG) and improvement (ISG) sweep genes. For heterochromatic DSGs/ISGs, a significant enrichment in both SP and SLC hotspots is observed, whereas there is no association with SLL hotspots (supplementary table 4, Supplementary Material online). Conversely, in the euchromatin, there is a difference between DSGs and ISGs. Euchromatic SP hotspots significantly overlap with both DSGs and ISGs, whereas euchromatic SLC hotspots only show enrichment in ISGs, which implies that many DSGs have lost hotspots after domestication. Lastly, SLL hotspots in the euchromatin overlapped DSGs and ISGs significantly less than expected by chance, indicating that most SLL hotspots are outside selective sweeps.

Many of the hotspots in SP and SLC sweep genes are lost in the SLL population. Nevertheless, it might be that sweep genes still undergo recombination even after hotspots are lost during domestication. To investigate this, we examined

the changes in recombination rates of both sweep and nonsweep genes across these three populations. We computed the effective recombination rates in DSG and ISG genes across the different populations and compared them against the remaining genes (fig. 3B; supplementary fig. 7, Supplementary Material online). Genes with hotspots clearly exhibited an elevated effective recombination rate with even higher rates for sweep genes (DSGs/ISGs) than nonsweep genes. However, the loss of these hotspots during the domestication or improvement process resulted in r_e being reduced to almost the same level as the nonhotspot genes. Interestingly, sweep genes show more reduction in r_e than nonsweep genes, which may reflect how sweep genes were directly affected by tomato domestication. Altogether, supplementary figure 7 (Supplementary Material online) underlines the severity of the decrease in r_e between wild and domesticated tomato genes, which resulted in reduced genetic diversity and forces breeders to introgress alleles from the wild relatives to recover desired traits.

Aside from increasing the frequency of alleles in specific genes, domestication of tomato also resulted in lost or negatively selected promoters during both domestication and improvement stages (Gao et al. 2019). These promoters are considered unfavorable because of their significantly lower frequency in SLC than SP or in SLL than SLC. Given that above it was demonstrated that recombination is associated with promoters, the question is how the loss of these promoters influence recombination. We found that the upstream region (<1 kb) of the genes with promoters under selection during domestication have reduced effective recombination rate in SLC compared with SP (Wilcoxon rank-sum test;

$P < 2\times10^{-16}$) (fig. 3C). Similarly, the upstream region of the genes with promoters under selection during improvement have reduced r_e in SLL compared with SLC (Wilcoxon rank-sum test; $P < 2\times10^{-16}$) (fig. 3D). This analysis reveals that loss of promoters due to domestication affected the recombination rate in the genomic regions from which these promoters were lost, highlighting a specific way in which domestication reduces recombination.

TE- and SV-Associated Hotspots

Domestication and improvement not only influence gene content but have a broader effect on genomic variants (Gao et al. 2019; Alonge et al. 2020). Hence, we finally analyzed the association between hotspots and TEs as well as SVs. Certain TE families show strong association with hotspots (fig. 4A). Both *hAT-Tip100* and *Stowaway* show enrichment, whereas *Tag1, L1, Copia*, and *Gypsy* show strong depletion, based on a permutation test. Most class-I TE or retrotransposons are under-represented in hotspots, except for endogenous retrovirus 1 (*ERV1*) in the SLL population and short interspersed nuclear element (SINE) in all three populations. Moreover, regions with low complexity and simple repeats have an excess of recombination hotspots. Similar to potato, rice, and maize, *Stowaway* and SINE are significantly overrepresented, whereas *Gypsy* and *Copia* are depleted in hotspots (Marand et al. 2017; Pan et al. 2017; Marand et al. 2019). In *Arabidopsis*, DSBs overlapped *Gypsy* and *Copia* elements significantly less than expected by chance (Choi et al. 2018); however, in maize, most DSBs are formed in repetitive regions, predominantly *Gypsy* retrotransposons, but only genic DSBs contribute to CO formation (He et al. 2017). Schwarzkopf et al. (2020) reported that hotspots shared by both domesticated and wild cocoa populations appear to be associated with DNA transposons. Furthermore, Marand et al. (2019) found that the presence of *Stowaway* and *Harbinger* elements in CO hotspots is associated with increased recombination rates, augmented chromatin accessibility, and reduced DNA methylation. Our result is also consistent with the findings on differentially accessible chromatin regions between meiotic cells and somatic cells of tomato (Chouaref J, Tark-Dame M, Koes R, Fransz P, Stam M, unpublished data), corroborating that TE families with an excess of hotspots are accessible in meiotic cells, whereas those with depletion of hotspots are inaccessible. TE families enriched with hotspots are known to be preferentially located in genic regions, but the retrotransposon LTR/ERV1 is particularly interesting because it was also reported to be one of the most transcriptionally active TE families due to its abundance in exonic regions (Mehra et al. 2015). The insertion of retrotransposons in a promoter region or UTR can both regulate gene expression (Alonge et al. 2020; Dominguez et al. 2020) and negatively affect the chance of CO incidence. Similarly, the accumulation of certain DNA transposons that are known to be accessible in meiotic cells may provide new sites for DSBs that can resolve to COs. Our results suggest that TE activities during domestication also influenced the landscape of meiotic recombination (He et al. 2017; Kent et al. 2017; Choi et al. 2018; Underwood and Choi 2019).

Alonge et al. (2020) reported that, compared with SLL and SLC, SP has significantly more SVs relative to the Heinz 1706 reference genome and that the majority of insertions and deletions are associated with *Gypsy* and *Copia* elements. To determine if SVs influence meiotic recombination, we compared the SVs between *S. pimpinellifolium* and *S. lycopersicum* reported in Wang et al. (2020) against the effective recombination rate and hotspots. We found that the length of deletions correlates with recombination rate, specifically deletions longer than 500 bp (fig. 4B), and that recombination hotspots are suppressed in long deletions in the wild population (Fisher's exact test; $P = 2.1\times10^{-6}$). Hotspot suppression is less prominent in deletions with lower allele frequency (supplementary fig. 8, Supplementary Material online). As a specific example, we found a 4-fold reduction of SP hotspots in large deletions (>500 bp) with allele frequency above 0.5 (fig. 4C). Interestingly, the same set of large deletions have low allele frequency and lack hotspot suppression in both SLC and SLL populations, which may be explained by the increased inbreeding and lower heterozygosity of SV regions in domesticated varieties compared with the wild relatives. Knowing the wild ancestral alleles, we can instead look at these regions of low-frequency deletions (relative to Heinz reference) as insertion sites for alleles that fixated in the vintage accessions. Lye and Purugganan (2019) reported that SVs associated with domestication traits have an increased allele frequency in the population due to selection or have become fixated in the population. This suggests that in natural populations, suppression of recombination hotspots in SVs occurs more in outcrossing populations with a certain level of heterozygosity of SVs between compatible individuals (Dluzewska et al. 2018; Wang and Copenhaver 2018). To put this result in a broader perspective, we analyzed the COs detected from the interspecific hybrid between *S. pimpinellifolium* and *S. lycopersium* (**COD2**) and found CO suppression in the large deletions segregating between SP and SLL populations (Fisher's exact test; $P = 1.4\times10^{-5}$; fig. 4D). This suppression of COs in deletions segregating between the parents of hybrid crosses is consistent with results obtained from the natural populations with high levels of outcrossing.

Interestingly, the shorter deletions (<500 bp) overlap recombination hotspots in each population significantly more than expected by chance (Fisher's exact test; SP, $P = 1.1\times10^{-6}$; SLC, $P = 4.9\times10^{-12}$; SLL, $P = 6.4\times10^{-13}$; supplementary fig. 9, Supplementary Material online). These regions with short deletions and hotspots have high homology and are enriched with simple repeats (z-score >11.1) and hAT-Tip100 (z-score >3.1) elements. These small deletions have significant allele-frequency differences between the wild and domesticated populations based on Wang et al. (2020), mostly have low frequency in the vintage population. Moreover, we found that the fixated SLL alleles in these deletion sites are causal mutations in domestication syndrome genes, like *OVATE* (Solyc02g085500), *Lin5* (Solyc09g010080), and *YABBY* (Solyc11g071810) and other genes controlling horticulture traits in tomato (Wang et al. 2020).

As we reported above, we identified some TE families that associate with hotspots and we speculated that TE activities
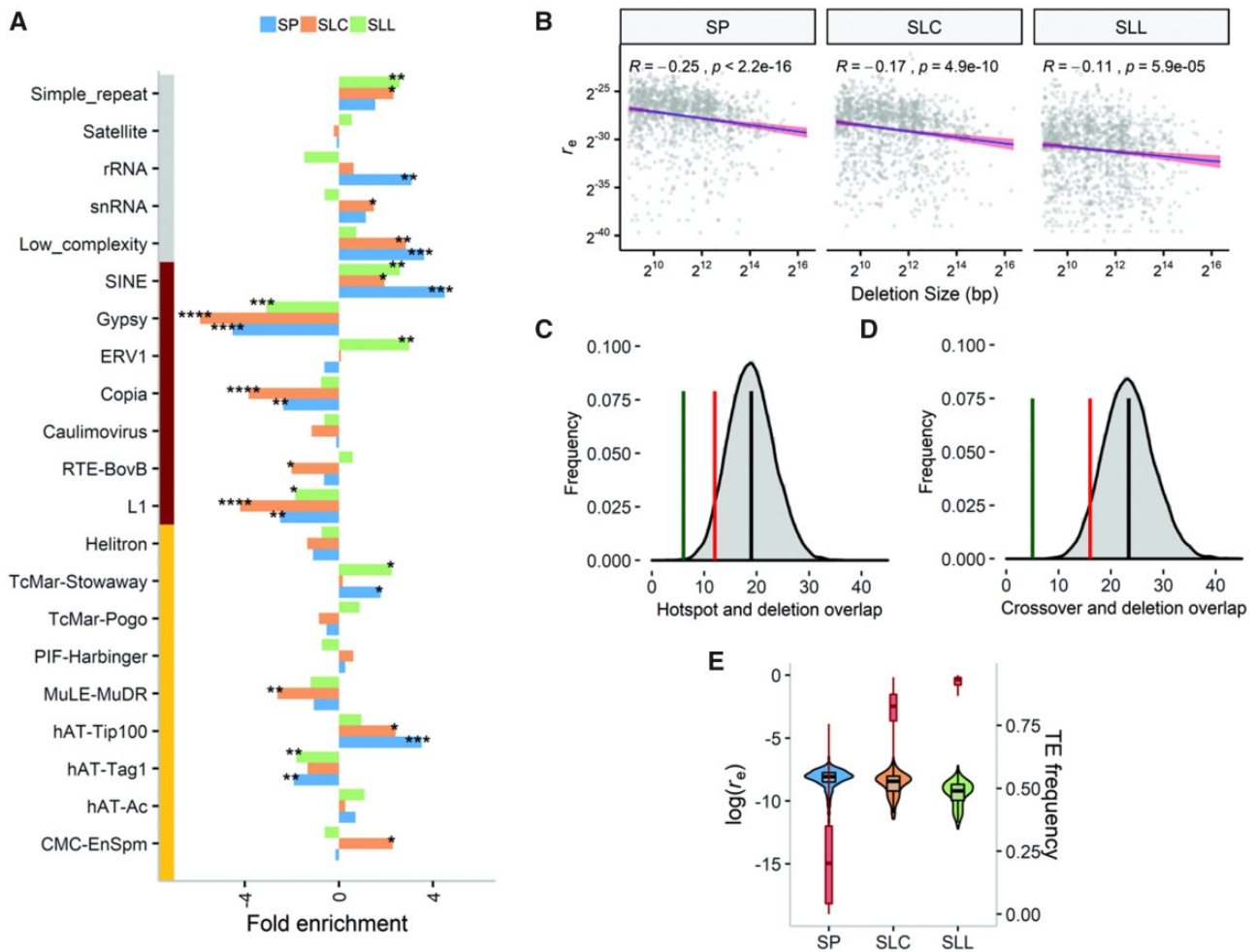
FIG. 4. Recombination and genomic variants. Using permutation tests, we identified (A) specific TE families with an excess or depletion of recombination hotspots. TE families are grouped into repeat elements (gray), retrotransposons (brown), and DNA transposons (yellow). (B) Scatter plots of effective recombination rate and deletion size ($n = 1,255$) per population. (C–D) Significance of overlap between (C) hotspots and deletions in SP and (D) empirical COs and deletions segregating between SP and SLL. The black and red vertical lines indicate the average number of overlaps found in 10,000 permutation sets and the number of overlaps at $P = 0.05$, respectively. The green vertical line indicates the observed number of overlaps. (E) Recombination rates (violin) and allele frequencies (red boxplot) of *Gypsy*, *Copia*, and *L1* elements.

during domestication affects recombination patterns. Comparing TEs with deletions, we selected TE elements that show a significant difference in frequency between the wild and domesticated populations, mostly *Gypsy*, *Copia*, and *L1* elements. As shown in figure 4E, the frequencies of these TE elements increased during domestication (fewer deletion alleles), whereas recombination rates in these elements significantly declined. This is consistent with the suppression of hotspots in these specific TE families that became fixed in the vintage population (fig. 4A). The result indicates that despite the lower deletion frequency or lower heterozygosity in a region, genomic content such as presence of specific TEs can influence whether recombination is suppressed or not.

Discussion

Throughout domestication, the general recombination landscape of both wild and domesticated tomato remains conserved, consistent with the known phenomenon in tomato

and other plant species that recombination rates are significantly higher in distal parts of chromosomes (Mercier et al. 2015; Lambing et al. 2017; Wang and Copenhaver 2018). Effective recombination rates are in agreement with empirical CO data derived from RILs and pollen gametes, and with genetic distances from the EXPIM2012 linkage map. Despite the conservation of the recombination landscape, we observed local increases in recombination rates across the chromosomes of early-domesticated tomato (SLC), which reflects the expected increase of recombination due to domestication (Ross-Ibarra 2004; Moyers et al. 2018). Concomitantly, some regions of the gene-rich euchromatin showed a reduction of recombination. After the domestication stage, further reduction of effective recombination rates is observed in vintage (SLL) tomato. This might either be due to low diversity in highly homozygous regions limiting the detection of actual recombination or because the recombination has actually decreased during the improvement stage (Otto and Barton 1997; Moyers et al. 2018). Nevertheless, the effective recombination rates reported here allowed us to identify genomic

regions that still recombine and generate diversity in vintage tomato.

We also identified recombination hotspots and found a small but significant overlap between the three populations, representing conserved recombination sites apparently not affected by domestication. The number of hotspots in vintage tomato is four times less than in its progenitor, revealing genomic sites in domesticated tomato preferentially maintained with lower if not completely suppressed recombination. Previous studies have reported that the pericentromeric heterochromatin of tomato displays extremely low recombination compared with the distal chromosome regions (Demirci et al. 2017; Fuentes et al. 2020), but we identified historical recombination hotspots near or within genes located in the heterochromatin, hinting on their accessibility as possible euchromatin islands within heterochromatin. These heterochromatin hotspots have a lower recombination rate than those located in euchromatin, which could explain why it can be hard to observe them in previous studies. Furthermore, compared with SP and SLC, the majority of SLL hotspots are located in the euchromatin. Our results on the divergence of hotspots between tomato populations suggests highly dynamic recombination patterns, likely as a result of the intense selection in domestication (Otto and Barton 1997; Schwarzkopf et al. 2020). Despite the use of interspecific tomato crosses in several studies, genetic shuffling through meiosis still remains strongly confined to the ends of the chromosomes (Demirci et al. 2017; Fuentes et al. 2020). Interestingly, Dreissig et al. (2019) reported an increase in pericentromeric COs, although they have been observed for crosses of domesticated and wild barley accessions. The low recombination rate in 75% of the tomato genome limits possibilities of generating diversity or breaking fixed haplotype blocks in either or both wild and domesticated tomato. However, despite the very limited sites for recombination in the domesticated tomato, detecting recombination hotspots in the wild genomes may help reveal factors that led to their divergence and ways to restore hotspots and genetic diversity in tomato.

The recombination hotspots found in wild and domesticated tomatoes are enriched in gene bodies and promoter regions and depleted in intergenic regions, which agrees with previous reports on tomato COs (Demirci et al. 2017; Fuentes et al. 2020). This implies that the reduction in the number of genes and promoters from wild to vintage tomato limits possible hotspot locations. Gao et al. (2019) reported hundreds of genes and promoters excised as a result of domestication and improvement, reducing recombination rate in specific genomic regions of vintage tomato and increasing the spans of haplotype blocks. The profile of recombination hotspots in the wild population reveals genomic regions that provide candidate targets for inducing meiotic recombination or loss-of-function mutations to domesticate wild plants or reintroduce desirable traits, or serve as guide for cross-checking hotspots from experimental populations (Li et al. 2018; Zsögön et al. 2018).

Comparing with the known selective sweep genes, we discovered that they overlap with recombination hotspots in both the wild and early-domesticated tomato significantly more than expected by chance. The hotspot-associated selective sweep genes in the SLC population indicated selection for increased recombination rate during domestication, which was then followed by a loss of hotspots after subsequent improvement. Otto and Barton (1997) proposed that higher recombination rates during domestication were favored to elevate the fixation rate of adaptive or beneficial alleles. These manifested as reduced nucleotide diversity in sweep regions of the SLL population that historically comprised hotspots. The increased recombination rate and number of hotspots in SLC may also relate to adaptive evolution in response to changing environments during domestication. Similar observations have been made for rice genes that showed higher recombination in response to stresses (Si et al. 2015). In addition, Razifard et al. (2020) identified an overrepresentation of defense-response genes in the selective sweeps of an SLC population, which agrees with our observation of excess recombination hotspots in R genes. Unlike SP and SLC hotspots which significantly overlap selective sweeps, SLL hotspots are mostly located outside the sweeps, showing the divergence of SLL hotspots from regions under selection during tomato domestication and improvement. The reduced genetic diversity in sweep regions of the highly inbred SLL population could have resulted from the loss of hotspots or reduced recombination incidence. Alternatively, the low diversity, possibly due to a selection bottleneck, may have led to the low effectivity of recombination in these sweep regions. Our results support an evolution of effective recombination related to the fixation of alleles in selective sweeps genes.

In all three populations, recombination hotspots were found to be positively (e.g., *Stowaway*, SINE) and negatively (e.g., *Gypsy*, *Copia*) associated with specific families of TEs. TE families with excess recombination hotspots were preferentially located in genic regions (Mehra et al. 2015). During meiosis, these specific TEs probably are accessible to recombination complexes, whereas most retrotransposons have closed chromatin status (Chouaref J, Tark-Dame M, Koes R, Fransz P, Stam M, unpublished data). Aside from TEs, we also identified association of hotspots with SVs. We showed that large deletions suppress recombination in SP populations, but this depends on the frequency of the deletion variant in the population. Since most large deletions have lower frequency in both SLC and SLL populations, we do not observe hotspot suppression by deletions in these populations. Reduced recombination in SVs combined with geographic isolation can lead to the development of alleles that are incompatible with distantly related haplotypes (Bomblies and Weigel 2007; Jiao and Schneeberger 2020). As an example, CO suppression is observed in a hybrid cross between SP and SLL parental accessions, implying that SVs can cause linkage drag that constrains introgression of specific alleles from the wild relatives while excluding unwanted alleles (Taagen et al. 2020). Rowan et al. (2019) provided several hypotheses as to why COs are suppressed in SV regions, such as absence of repair template, COs in SVs creating inviable gametes, DSB in SVs preferentially resolving to NCOs, constrained interaction with

the central element and homologous chromosomes, and lastly the DNA methylation in SVs.

On the other hand, we found enrichment of hotspots in short deletions across all populations, which we hypothesize is not due to mutations inducing hotspots but more likely due to mutations formed by nonallelic homologous recombination in the hotspot regions with tracts of homology (Escaramís et al. 2015; Balachandran and Beck 2020). These small causal mutations, that are located in many horticulture trait genes, may have originated from the wild progenitor and may confer adaptive domestication traits that promote their fixation in the vintage population (Lye and Purugganan 2019). Previous studies already reported recurrent SVs in regions with high recombination rate (McVean 2010; Liu et al. 2012; Rowan et al. 2019; Badet et al. 2021), although there are other mechanisms that generate more SVs (Escaramís et al. 2015). SVs are reported to have key roles in the evolution of domesticated species and are associated with postdomestication traits (Lye and Purugganan 2019). Thus, the divergence of TEs and SVs, both associated with recombination rates and hotspots, also contributes to the differing patterns of recombination during domestication. Identifying these divergent sequences between populations can allow us to determine genomic regions that are unlikely to recombine, or alleles that remain tightly linked in the hybrid crosses. Our results further confirm that domestication influenced local genomic contents, which in turn affected the recombination patterns in tomato.

Materials and Methods

Computing Recombination Rates and Detecting Hotspots

We collected resequencing data from previous studies under accession numbers PRJEB5235 (Aflitos et al. 2014), PRJNA454805 (Razifard et al. 2020), PRJNA353161 (Tieman et al. 2017), and PRJNA259308 (Lin et al. 2014), and based on previous population analyses (Tieman et al. 2017; Zhu et al. 2018; Gao et al. 2019; Alonge et al. 2020; Razifard et al. 2020), we selected accessions that were unanimously assigned to the S. pimpinellifolium (SP), S. lycopersicum var. cerasiforme (SLC), or S. lycopersicum var. lycopersicum (SLL) taxonomic groups. Most accessions were sequenced using Illumina HiSeq2000 except those from Razifard which utilized Illumina NextSeq. Modern processing cultivars and hybrids were excluded to avoid obscuring the recombination patterns in each of these three major groups. From the combined list of accessions, we selected 81 SPs, 140 SLCs, and 136 SLLs.

Reads were trimmed using Trimmomatic (Bolger et al. 2014) and aligned against the tomato Heinz 1706 (SL4.0) reference genome (Hosmani et al. 2019) using bwa mem (Li 2013). Accessions with read and physical coverage below $5\times$ and 70%, respectively, were discarded. Filtering left 75 SP, 122 SLC, and 117 SLL samples. SNPs were identified per accession using GATK HaplotypeCaller (Poplin et al. 2017) with default parameters and then we randomly selected 75 accessions per population for GATK joint-genotyping and hard-filtering. Using bedtools (Quinlan and Hall 2010) and

bcftools (Danecek et al. 2011), we selected biallelic SNPs located outside repeat or TE regions, with a minimum allele frequency of 0.05, less than 10% missing data, and consisting of at most one heterozygous genotype. Heterozygous genotypes were then converted to missing data. All missing calls were imputed and phased using Beagle v. 5.1 (window=5, overlap=2, iterations=30, err=0.001, burnin=10) (Browning et al. 2018).

A widely used method to infer historical recombination rate is LDhat (Auton and McVean 2007), which implements coalescent resampling with a Bayesian reversible-jump Markov Chain Monte Carlo (rjMCMC) algorithm to estimate population-scaled recombination rates ($\rho = 4 N_e r_e$) from pairs of SNPs. We estimated recombination rate using the LDhat v2.2 interval program per window of 5,000 SNPs, overlapping by 500 SNPs. LDhat was run with 20 million iterations, sampling every 2,000 iterations, using a mutation rate of 0.001 and the first 2,000 samples for burn-in. We subsequently detected recombination hotspots using sequenceLDhot (Fearnhead 2006) with nonoverlapping 1-kb windows, a 500-bp step size, and background recombination rate set as the median rate in a 50-kb window centered at each 1-kb window. Hotspots with rates greater than 10 times and less than 200 times the background and with likelihood ratio (LR) above the 95th percentile are reported. We merged hotspots within 500 bp of each other and used the highest LR and recombination rate for merged intervals. Only hotspots with at least a ten times increase in recombination rate, based on the ratio of LDhat-computed ρ and the background ρ, were reported as the final set of CO hotspots.

To compare recombination rates between populations and account for potential differences in effective population sizes (N_e), we first computed the nucleotide diversity (θ_w, Watterson's theta), using LDhat convert and used the neutral mutation rate (μ) of 1×10^{-8} (Baer et al. 2007; Lin et al. 2014; Moyers et al. 2018) to estimate the $4N_e$ ($4N_e=\theta_w/\mu$). The estimated $4N_e$ was then used to calculate the effective recombination rate per generation ($r_e=\rho/4N_e$) for each population.

Comparison with Genetic Maps

For validating the recombination map, we compared median recombination rates against the CO frequencies from tomato RILs (Demirci et al. 2017) and pollen gametes (Fuentes et al. 2020) using a 50-kb sliding window and performed Spearman's rank correlation test. Further correlation testing was done by comparing against EXPIM2012 genetic map (Sim et al. 2012), using the method from Choi et al. (2013). Between every pair of adjacent SNPs in the genetic map, the population-scaled recombination rate (ρ/kb) was converted to cM/Mb. We also compared the recombination landscapes between populations using multiscale correlations by calculating the average recombination rates in varying window sizes, by randomly sampling these windows 10,000 times, and subsequently calculating Spearman's correlation coefficients.

Compute Euchromatin Regions

The border between euchromatin and heterochromatin in each chromosome was computed according to Stack et al. (2009) and Demirci et al. (2017). Using Table 1 of Sherman and Stack (1992), we calculated the average length and heterochromatin length (μm) of each pachytene chromosome. Euchromatin length was calculated by subtracting heterochromatin length from each arm length and multiplying by the euchromatin DNA density (1.54 Mb/μm). Then, we determined the euchromatin-heterochromatin boundary based on the length of each euchromatic region. We also computed the heterochromatic region boundaries per chromosome relative to the Heinz 1706 (SL4.0) reference genome.

Supplementary Material

Supplementary data are available at *Molecular Biology and Evolution* online.

Acknowledgments

Data Availability

Scripts are available at https://github.com/rrfuentes/histo-recom.git.

References

Aflitos S, Schijlen E, de Jong H, de Ridder D, Smit S, Finkers R, Wang J, Zhang G, Li N, Mao L, 100 Tomato Genome Sequencing Consortium, et al. 2014. Exploring genetic variation in the tomato (*Solanum* section *lycopersicon*) clade by whole-genome sequencing. *Plant J.* 80(1):136–148.

Allard RW. 1999. History of plant population genetics. *Annu Rev Genet.* 33:1–27.

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(1):145–161.e123.

Anderson LK, Doyle GG, Brigham B, Carter J, Hooker KD, Lai A, Rice M, Stack SM. 2003. High-resolution crossover maps for each bivalent of *Zea mays* using recombination nodules. *Genetics* 165(2):849–865.

Andolfo G, D'Agostino N, Frusciante L, Ercolano MR. 2021. The tomato interspecific NB-LRR gene arsenal and its impact on breeding strategies. *Genes (Basel)* 12(2):184.

Auton A, McVean G. 2007. Recombination rate estimation in the presence of hotspots. *Genome Res.* 17(8):1219–1227.

Badet T, Fouche S, Hartmann FE, Zala M, Croll D. 2021. Machine-learning predicts genomic determinants of meiosis-driven structural variation in a eukaryotic pathogen. *Nat Commun.* 12(1):3551.

Baer CF, Miyamoto MM, Denver DR. 2007. Mutation rate variation in multicellular eukaryotes: causes and consequences. *Nat Rev Genet.* 8(8):619–631.

Bai Y, Lindhout P. 2007. Domestication and breeding of tomatoes: what have we gained and what can we gain in the future? *Ann Bot.* 100(5):1085–1094.

Balachandran P, Beck CR. 2020. Structural variant identification and characterization. *Chromosome Res.* 28(1):31–47.

Blanca J, Montero-Pau J, Sauvage C, Bauchet G, Illa E, Díez MJ, Francis D, Causse M, van der Knaap E, Cañizares J. 2015. Genomic variation in tomato, from wild ancestors to contemporary breeding accessions. *BMC Genomics* 16:257.

Bolger AM, Lohse M, Usadel B. 2014. Trimmomatic: a flexible trimmer for illumina sequence data. *Bioinformatics* 30(15):2114–2120.

Bomblies K, Weigel D. 2007. Hybrid necrosis: autoimmunity as a potential gene-flow barrier in plant species. *Nat Rev Genet.* 8(5):382–393.

Browning BL, Zhou Y, Browning SR. 2018. A one-penny imputed genome from next-generation reference panels. *Am J Hum Genet.* 103(3):338–348.

Brunschwig H, Levi L, Ben-David E, Williams RW, Yakir B, Shifman S. 2012. Fine-scale maps of recombination rates and hotspots in the mouse genome. *Genetics* 191(3):757–764.

Burt A, Bell G. 1987. Mammalian chiasma frequencies as a test of two theories of recombination. *Nature* 326(6115):803–805.

Cardon LR, Abecasis GR. 2003. Using haplotype blocks to map human complex trait loci. *Trends Genet.* 19(3):135–140.

Chan AH, Jenkins PA, Song YS. 2012. Genome-wide fine-scale recombination rate variation in drosophila melanogaster. *PLoS Genet.* 8(12):e1003090.

Choi K, Henderson IR. 2015. Meiotic recombination hotspots—a comparative view. *Plant J.* 83(1):52–61.

Choi K, Zhao X, Kelly KA, Venn O, Higgins JD, Yelina NE, Hardcastle TJ, Ziolkowski PA, Copenhaver GP, Franklin FCH, et al. 2013. *Arabidopsis* meiotic crossover hot spots overlap with H2A.Z nucleosomes at gene promoters. *Nat Genet.* 45(11):1327–1336.

Choi K, Zhao X, Tock AJ, Lambing C, Underwood CJ, Hardcastle TJ, Serra H, Kim J, Cho HS, Kim J, et al. 2018. Nucleosomes and DNA methylation shape meiotic DSB frequency in *Arabidopsis thaliana* transposons and gene regulatory regions. *Genome Res.* 28(4):532–546.

Danecek P, Auton A, Abecasis G, Albers C, Banks E, DePristo M, Handsaker RE, Lunter G, Marth GT, Sherry ST, et al.; 1000 Genomes Project Analysis Group. 2011. The variant call format and VCFtools. *Bioinformatics* 27(15):2156–2158.

de Haas LS, Koopmans R, Lelivelt CLC, Ursem R, Dirks R, Velikkakam James G. 2017. Low-coverage resequencing detects meiotic recombination pattern and features in tomato RILs. *DNA Res.* 24(6):549–558.

Demirci S, Peters SA, de Ridder D, van Dijk ADJ. 2018. DNA sequence and shape are predictive for meiotic crossovers throughout the plant kingdom. *Plant J.* 95(4):686–699.

Demirci S, van Dijk ADJ, Sanchez Perez G, Aflitos SA, de Ridder D, Peters SA. 2017. Distribution, position and genomic characteristics of crossovers in tomato recombinant inbred lines derived from an interspecific cross between *Solanum lycopersicum* and *Solanum pimpinellifolium*. *Plant J.* 89(3):554–564.

Dluzewska J, Szymanska M, Ziolkowski PA. 2018. Where to cross over? Defining crossover sites in plants. *Front Genet.* 9:609.

Doebley JF, Gaut BS, Smith BD. 2006. The molecular genetics of crop domestication. *Cell* 127(7):1309–1321.

Dominguez M, Dugas E, Benchouaia M, Leduque B, Jimenez-Gomez JM, Colot V, Quadrana L. 2020. The impact of transposable elements on tomato diversity. *Nat Commun.* 11(1):4058.

Dreissig S, Mascher M, Heckmann S. 2019. Variation in recombination rate is shaped by domestication and environmental conditions in barley. *Mol Biol Evol.* 36(9):2029–2039.

Drouaud J, Khademian H, Giraut L, Zanni V, Bellalou S, Henderson IR, Falque M, Mezard C. 2013. Contrasted patterns of crossover and non-crossover at *Arabidopsis thaliana* meiotic recombination hotspots. *PLoS Genet.* 9(11):e1003922.

Dutta R, Mainsah J, Yatskiv Y, Chakrabortty S, Brennan P, Khuder B, Qiu S, Fedorova L, Fedorov A. 2017. Intricacies in arrangement of SNP haplotypes suggest "great admixture" that created modern humans. *BMC Genomics* 18(1):433.

Eichten SR, Swanson-Wagner RA, Schnable JC, Waters AJ, Hermanson PJ, Liu S, Yeh CT, Jia Y, Gendler K, Freeling M, et al. 2011. Heritable epigenetic variation among maize inbreds. *PLoS Genet.* 7(11):e1002372.

Escaramís G, Docampo E, Rabionet R. 2015. A decade of structural variants: description, history and methods to detect structural variation. *Brief Funct Genomics*. 14(5):305–314.

Fearnhead P. 2006. SequenceLDhot: detecting recombination hotspots. *Bioinformatics* 22(24):3061–3066.

Francis KE, Lam SY, Harrison BD, Bey AL, Berchowitz LE, Copenhaver GP. 2007. Pollen tetrad-based visual assay for meiotic recombination in *Arabidopsis*. *Proc Natl Acad Sci U S A*. 104(10):3913–3918.

Fransz P, Linc G, Lee CR, Aflitos SA, Lasky JR, Toomajian C, Ali H, Peters J, van Dam P, Ji X, et al. 2016. Molecular, genetic and evolutionary analysis of a paracentric inversion in *Arabidopsis thaliana*. *Plant J*. 88(2):159–178.

Fuentes R, Hesselink T, Nieuwenhuis R, Bakker L, Schijlen E, van Dooijeweert W, Diaz Trivino S, de Haan JR, Sanchez Perez G, Zhang X, et al. 2020. Meiotic recombination profiling of interspecific hybrid F1 tomato pollen by linked read sequencing. *Plant J*. 102(3):480–492.

Gao L, Gonda I, Sun H, Ma Q, Bao K, Tieman DM, Burzynski-Chang EA, Fish TL, Stromberg KA, Sacks GL, et al. 2019. The tomato pan-genome uncovers new genes and a rare allele regulating fruit flavor. *Nat Genet*. 51(6):1044–1051.

Gross BL, Olsen KM. 2010. Genetic perspectives on crop domestication. *Trends Plant Sci*. 15(9):529–537.

Guo J, Chen H, Yang P, Lee YT, Wu M, Przytycka TM, Kwoh CK, Zheng J. 2018. LDSplitdDBb: a database for studies of meiotic recombination hotspots in MHC using human genomic data. *BMC Med Genomics*. 11(Suppl 2):27.

He Y, Wang M, Dukowic-Schulze S, Zhou A, Tiang CL, Shilo S, Sidhu GK, Eichten S, Bradbury P, Springer NM, et al. 2017. Genomic features shaping the landscape of meiotic double-strand-break hotspots in maize. *Proc Natl Acad Sci U S A*. 114(46):12231–12236.

Hosmani PS, Flores-Gonzalez M, van de Geest H, Maumus F, Bakker LV, Schijlen E, van Haarst J, Cordewener J, Sanchez-Perez G, Peters S, et al. 2019. An improved de novo assembly and annotation of the tomato reference genome using single-molecule sequencing, Hi-C proximity ligation and optical maps. bioRxiv. https://doi.org/10.1101/767764.

Huang X, Feng Q, Qian Q, Zhao Q, Wang L, Wang A, Guan J, Fan D, Weng Q, Huang T, et al. 2009. High-throughput genotyping by whole-genome resequencing. *Genome Res*. 19(6):1068–1076.

Hufford MB, Xu X, van Heerwaarden J, Pyhajarvi T, Chia JM, Cartwright RA, Elshire RJ, Glaubitz JC, Guill KE, Kaeppler SM, et al. 2012. Comparative population genomics of maize domestication and improvement. *Nat Genet*. 44(7):808–811.

Hunter N. 2015. Meiotic recombination: the essence of heredity. *Cold Spring Harb Perspect Biol*. 7(12):a016618.

Jiao WB, Schneeberger K. 2020. Chromosome-level assemblies of multiple *Arabidopsis* genomes reveal hotspots of rearrangements with altered evolutionary dynamics. *Nat Commun*. 11(1):989.

Jones GH, Franklin FC. 2006. Meiotic crossing-over: obligation and interference. *Cell* 126(2):246–248.

Kent TV, Uzunovic J, Wright SI. 2017. Coevolution between transposable elements and recombination. *Philos Trans R Soc Lond B Biol Sci*. 372(1736):20160458.

Kianian PMA, Wang M, Simons K, Ghavami F, He Y, Dukowic-Schulze S, Sundararajan A, Sun Q, Pillardy J, Mudge J, et al. 2018. High-resolution crossover mapping reveals similarities and differences of male and female recombination in maize. *Nat Commun*. 9(1):2370.

Kovach MJ, Sweeney MT, McCouch SR. 2007. New insights into the history of rice domestication. *Trends Genet*. 23(11):578–587.

Lambing C, Franklin FCH, Wang C-JR. 2017. Understanding and manipulating meiotic recombination in plants. *Plant Physiol*. 173(3):1530–1542.

Li H. 2013. Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. arXiv:1303.3997.

Li T, Yang X, Yu Y, Si X, Zhai X, Zhang H, Dong W, Gao C, Xu C. 2018. Domestication of wild tomato is accelerated by genome editing. *Nat Biotechnol*. 36(12):1160–1163.

Lim EC, Kim J, Park J, Kim EJ, Kim J, Park YM, Cho HS, Byun D, Henderson IR, Copenhaver GP, et al. 2020. Deeptetrad: high-throughput image analysis of meiotic tetrads by deep learning in *Arabidopsis thaliana*. *Plant J*. 101(2):473–483.

Lin T, Zhu G, Zhang J, Xu X, Yu Q, Zheng Z, Zhang Z, Lun Y, Li S, Wang X, et al. 2014. Genomic analyses provide insights into the history of tomato breeding. *Nat Genet*. 46(11):1220–1226.

Liu P, Carvalho CM, Hastings PJ, Lupski JR. 2012. Mechanisms for recurrent and complex human genomic rearrangements. *Curr Opin Genet Dev*. 22(3):211–220.

Lye ZN, Purugganan MD. 2019. Copy number variation in domestication. *Trends Plant Sci*. 24(4):352–365.

Marand AP, Jansky SH, Zhao H, Leisner CP, Zhu X, Zeng Z, Crisovan E, Newton L, Hamernik AJ, Veilleux RE, et al. 2017. Meiotic crossovers are associated with open chromatin and enriched with stowaway transposons in potato. *Genome Biol*. 18(1):203.

Marand AP, Zhao H, Zhang W, Zeng Z, Fang C, Jiang J. 2019. Historical meiotic crossover hotspots fueled patterns of evolutionary divergence in rice. *Plant Cell*. 31(3):645–662.

McVean G. 2010. What drives recombination hotspots to repeat DNA in humans? *Philos Trans R Soc Lond B Biol Sci*. 365(1544):1213–1218.

Mehra M, Gangwar I, Shankar R. 2015. A deluge of complex repeats: the solanum genome. *PLoS One*. 10(8):e0133962.

Mercier R, Mezard C, Jenczewski E, Macaisne N, Grelon M. 2015. The molecular biology of meiosis in plants. *Annu Rev Plant Biol*. 66:297–327.

Meyer RS, Purugganan MD. 2013. Evolution of crop species: genetics of domestication and diversification. *Nat Rev Genet*. 14(12):840–852.

Morrell PL, Lundy KE, Clegg MT. 2003. Distinct geographic patterns of genetic diversity are maintained in wild barley (*Hordeum vulgare* ssp. *spontaneum*) despite migration. *Proc Natl Acad Sci U S A*. 100(19):10812–10817.

Moyers BT, Morrell PL, McKay JK. 2018. Genetic costs of domestication and improvement. *J Hered*. 109(2):103–116.

Nieri D, Di Donato A, Ercolano MR. 2017. Analysis of tomato meiotic recombination profile reveals preferential chromosome positions for NB-LRR genes. *Euphytica* 213(9):206.

Otto SP, Barton NH. 1997. The evolution of recombination: removing the limits to natural selection. *Genetics* 147(2):879–906.

Pan Q, Deng M, Yan J, Li L. 2017. Complexity of genetic mechanisms conferring nonuniformity of recombination in maize. *Sci Rep*. 7(1):1205.

Petes TD. 2001. Meiotic recombination hot spots and cold spots. *Nat Rev Genet*. 2(5):360–369.

Poplin R, Ruano-Rubio V, DePristo MA, Fennell TJ, Carneiro MO, Van der Auwera GA, Kling DE, Gauthier LD, Levy-Moonshine A, Roazen D, et al. 2017. Scaling accurate genetic variant discovery to tens of thousands of samples. bioRxiv.

Qi J, Wijeratne AJ, Tomsho LP, Hu Y, Schuster SC, Ma H. 2009. Characterization of meiotic crossovers and gene conversion by whole-genome sequencing in *Saccharomyces cerevisiae*. *BMC Genomics* 10:475.

Quinlan AR, Hall IM. 2010. Bedtools: a flexible suite of utilities for comparing genomic features. *Bioinformatics* 26(6):841–842.

Razifard H, Ramos A, Della Valle AL, Bodary C, Goetz E, Manser EJ, Li X, Zhang L, Visa S, Tieman D, et al. 2020. Genomic evidence for complex domestication history of the cultivated tomato in Latin America. *Mol Biol Evol*. 37(4):1118–1132.

Rees H, Dale PJ. 1974. Chiasmata and variability in *Lolium* and *Festuca* populations. *Chromosoma* 47(3):335–351.

Rick CM, Holle M. 1990. Andean *Lycopersicon esculentum* var. *cerasiforme*: genetic variation and its evolutionary significance. *Econ Bot*. 44(S3):69–78.

Rick CM, Holle M, Thorp RW. 1978. Rates of cross-pollination in *Lycopersicon pimpinellifolium*: impact of genetic variation in floral characters. *Pl Syst Evol*. 129(1–2):31–44.

Ross-Ibarra J. 2004. The evolution of recombination under domestication: a test of two hypotheses. *Am Nat*. 163(1):105–112.

Rowan BA, Heavens D, Feuerborn TR, Tock AJ, Henderson IR, Weigel D. 2019. An ultra high-density *Arabidopsis thaliana* crossover map that

refines the influences of structural variation and epigenetic features. *Genetics* 213(3):771–787.

Sauvage C, Rau A, Aichholz C, Chadoeuf J, Sarah G, Ruiz M, Santoni S, Causse M, David J, Glemin S. 2017. Domestication rewired gene expression and nucleotide diversity patterns in tomato. *Plant J.* 91(4):631–645.

Schwarzkopf EJ, Motamayor JC, Cornejo OE. 2020. Genetic differentiation and intrinsic genomic features explain variation in recombination hotspots among cocoa tree populations. *BMC Genomics* 21(1):332.

Sherman JD, Stack SM. 1992. Two-dimensional spreads of synaptonemal complexes from solanaceous plants. V. Tomato (*Lycopersicon esculentum*) karyotype and idiogram. *Genome* 35(2):354–359.

Sherman JD, Stack SM. 1995. Two-dimensional spreads of synaptonemal complexes from solanaceous plants. VI. High-resolution recombination nodule map for tomato (*Lycopersicon esculentum*). *Genetics* 141(2):683–708.

Si W, Yuan Y, Huang J, Zhang X, Zhang Y, Zhang Y, Tian D, Wang C, Yang Y, Yang S. 2015. Widely distributed hot and cold spots in meiotic recombination as shown by the sequencing of rice F2 plants. *New Phytol.* 206(4):1491–1502.

Sim SC, Durstewitz G, Plieske J, Wieseke R, Ganal MW, Van Deynze A, Hamilton JP, Buell CR, Causse M, Wijeratne S, et al. 2012. Development of a large SNP genotyping array and generation of high-density genetic maps in tomato. *PLoS One.* 7(7):e40563.

Soyk S, Lemmon ZH, Sedlazeck FJ, Jimenez-Gomez JM, Alonge M, Hutton SF, Van Eck J, Schatz MC, Lippman ZB. 2019. Duplication of a domestication locus neutralized a cryptic variant that caused a breeding barrier in tomato. *Nat Plants.* 5(5):471–479.

Stack SM, Anderson LK, Sherman JD. 1989. Chiasmata and recombination nodules in *Lilium longiflorum*. *Genome* 32(3):486–498.

Stack SM, Royer SM, Shearer LA, Chang SB, Giovannoni JJ, Westfall DH, White RA, Anderson LK. 2009. Role of fluorescence in situ hybridization in sequencing the tomato genome. *Cytogenet Genome Res.* 124(3-4):339–350.

Stevison LS, Woerner AE, Kidd JM, Kelley JL, Veeramah KR, McManus KF, Great Ape Genome P, Bustamante CD, Hammer MF, Wall JD, Great Ape Genome Project 2016. The time scale of recombination rate evolution in great apes. *Mol Biol Evol.* 33(4):928–945.

Stukenbrock EH, Dutheil JY. 2018. Fine-scale recombination maps of fungal plant pathogens reveal dynamic recombination landscapes and intragenic hotspots. *Genetics* 208(3):1209–1229.

Taagen E, Bogdanove AJ, Sorrells ME. 2020. Counting on crossovers: controlled recombination for plant breeding. *Trends Plant Sci.* 25(5):455–465.

Termolino P, Cremona G, Consiglio MF, Conicella C. 2016. Insights into epigenetic landscape of recombination-free regions. *Chromosoma* 125(2):301–308.

Tieman D, Zhu G, Resende MF Jr, Lin T, Nguyen C, Bies D, Rambla JL, Beltran KS, Taylor M, Zhang B, et al. 2017. A chemical genetic roadmap to improved tomato flavor. *Science* 355(6323):391–394.

Underwood CJ, Choi K. 2019. Heterogeneous transposable elements as silencers, enhancers and targets of meiotic recombination. *Chromosoma* 128(3):279–296.

Wang J, de Villena FP, Moore KJ, Wang W, Zhang Q, McMillan L. 2010. Genome-wide compatible SNP intervals and their properties. *ACM Int Conf Bioinform Comput Biol (2010).* 2010:43–52.

Wang X, Gao L, Jiao C, Stravoravdis S, Hosmani PS, Saha S, Zhang J, Mainiero S, Strickler SR, Catala C, et al. 2020. Genome of *Solanum pimpinellifolium* provides insights into structural variants during tomato breeding. *Nat Commun.* 11(1):5817.

Wang Y, Copenhaver GP. 2018. Meiotic recombination: mixing it up in plants. *Annu Rev Plant Biol.* 69:577–609.

Wijnker E, Velikkakam James G, Ding J, Becker F, Klasen JR, Rawat V, Rowan BA, de Jong DF, de Snoo CB, Zapata L, et al. 2013. The genomic landscape of meiotic crossovers and gene conversions in *Arabidopsis thaliana*. *Elife* 2:e01426.

Yang H, Wang JR, Didion JP, Buus RJ, Bell TA, Welsh CE, Bonhomme F, Yu AH, Nachman MW, Pialek J, et al. 2011. Subspecific origin and haplotype diversity in the laboratory mouse. *Nat Genet.* 43(7):648–655.

Yang Z, Li G, Tieman D, Zhu G. 2019. Genomics approaches to domestication studies of horticultural crops. *Hortic Plant J.* 5(6):240–246.

Zhu G, Wang S, Huang Z, Zhang S, Liao Q, Zhang C, Lin T, Qin M, Peng M, Yang C, et al. 2018. Rewiring of the fruit metabolome in tomato breeding. *Cell* 172(1–2):249–261.e212.

Zsögön A, Čermák T, Naves ER, Notini MM, Edel KH, Weinl S, Freschi L, Voytas DF, Kudla J, Peres LEP. 2018. De novo domestication of wild tomato using genome editing. *Nat Biotechnol.* 36(12):1211–1216.