Ongoing Rapid Evolution of a Post-Y Region Revealed by Chromosome-Scale Genome Assembly of a Hexaploid Monoecious Persimmon (*Diospyros kaki*)

Ayano Horiuchi,[1] Kanae Masuda,[1] Kenta Shirasawa ![ORCID],[2] Noriyuki Onoue,[3] Naoko Fujita,[1] Koichiro Ushijima ![ORCID],[1] and Takashi Akagi ![ORCID]*,[1,4]

[1]Graduate School of Environmental and Life Science, Okayama University, Okayama, Japan

[2]Department of Frontier Research and Development, Kazusa DNA Research Institute, Chiba, Japan

[3]Institute of Fruit Tree and Tea Science, NARO, Hiroshima, Japan

[4]Japan Science and Technology Agency (JST), PRESTO, Saitama, Japan

Corresponding author: E-mail: takashia@okayama-u.ac.jp.

Associate editor: Melissa Wilson

Abstract

Plants have evolved sex chromosomes independently in many lineages, and loss of separate sexes can also occur. In this study, we assembled a monoecious recently hexaploidized persimmon (*Diospyros kaki*), in which the Y chromosome has lost the maleness-determining function. Comparative genomic analysis of *D. kaki* and its dioecious relatives uncovered the evolutionary process by which the nonfunctional Y chromosome (or $Y^{monoecy}$) was derived, which involved silencing of the sex-determining gene, *OGI*, approximately 2 million years ago. Analyses of the entire X and $Y^{monoecy}$ chromosomes suggested that *D. kaki*'s nonfunctional male-specific region of the Y chromosome (MSY), which we call a post-MSY, has conserved some characteristics of the original functional MSY. Specifically, comparing the functional MSY in *Diospyros lotus* and the nonfunctional "post-MSY" in *D. kaki* indicated that both have been rapidly rearranged, mainly via ongoing transposable element bursts, resembling structural changes often detected in Y-linked regions, some of which can enlarge the nonrecombining regions. The recent evolution of the post-MSY (and possibly also MSYs in dioecious *Diospyros* species) therefore probably reflects these regions' ancestral location in a pericentromeric region, rather than the presence of male-determining genes and/or genes controlling sexually dimorphic traits.

Key words: sex chromosome, genome assembly, monoecy, transposable elements.

Introduction

In contrast to animals, many different plant lineages have independently evolved chromosomal sex determination (or dioecious) systems from functionally hermaphrodite ancestors (Westergaard 1958; Charlesworth 1985; Ming et al. 2011; Henry et al. 2018). A comparative framework can therefore shed light on the diversity of routes by which sex determination and sex chromosomes may evolve. Recent advances in genomic technologies have revealed that the sex-determining factors in different plant lineages differ in the molecular developmental functions involved and also the evolutionary pathways that led to separate sexes (Akagi et al. 2014, 2018, 2019; Harkess et al. 2017, 2020; Müller et al. 2020; Kazama et al. 2022). Some species have single sex-determining genes or small sex-linked regions, whereas other plants have sex chromosomes, distinguished by the presence of physically extensive nonrecombining regions, which sometimes result from recombination suppression involving chromosome rearrangements, creating heteromorphism that is detectable cytologically or in genome sequences. Recombination suppression may sometimes be associated with the evolution of sexual dimorphic phenotypic traits, which can lead to the establishment of sexually antagonistic polymorphisms (Charlesworth and Charlesworth 1978; Rice 1992). Renner and Müller (2021) have questioned whether sex chromosome evolution in different plant lineages shares any common rules and noted that the sizes of the male-specific regions (Y-linked regions, or MSYs) are not related to the ages of different plant sex-determining systems or the gene densities in the MSY, which might reflect the potential to involved in the evolution of sexually dimorphic traits. In kiwifruit (the genus *Actinidia*), sexual dimorphisms are conserved across the genus. Nevertheless, transpositions of the two sex-determining factors have recurrently and independently formed new sex-linked regions with hemizygous MSYs (Akagi et al. 2023). These MSYs contain only three conserved genes, including two sex-determining genes, one of which controls most of the conserved sexual dimorphisms, including

flower numbers per inflorescence, and rates of adult (flowering) nodes/branches (Akagi et al. 2023). This example suggests that MSYs might often evolve by a process different from the one involving the evolution of sexual dimorphism outlined above. One proposal is that a lack of recombination could be the ancestral state for some MSY regions, as plant (and animal) genomes often include extensive pericentromeric regions in which recombination is rare, and in some cases, recombination is also rare in telomeric regions (Charlesworth 2019).

Diploid persimmon species (in the genus *Diospyros*, with 2n = 2x = 30) are dioecious, apart from hermaphroditic mutants/lines (Masuda and Akagi 2022). The sexes are determined by a Y-linked gene, *OGI*, which formed by a duplication of an autosomal counterpart gene, *MeGI*, and expresses a small RNA that suppresses *MeGI* expression (Akagi et al. 2014). *MeGI* encodes a homeodomain ZIP1 (HD-ZIP1) which has neofunctionalized via a lineage-specific duplication to gain roles in both suppressing androecium development and promoting gynecium growth (Yang et al. 2019; Akagi et al. 2020). In contrast, individual trees of the hexaploid Oriental persimmon (*Diospyros kaki*; 2n = 6x = 90) are mainly either gynoecious or wholly monoecious, with occasional production of hermaphroditic flowers on monoecious trees (Masuda et al. 2022; Masuda and Akagi 2022). Genetically female individuals (with hexaplex X: 6A + 6X) are gynoecious, whereas genetically male individuals (carrying at least one Y chromosome) can be monoecious (Akagi et al. 2016a; Masuda et al. 2020). In the *D. kaki* Y-linked region, *OGI* was largely silenced by insertion of a short interspersed nuclear element (SINE)–like sequence named *Kali* into the promoter region and causing loss of the male-determining function (Akagi et al. 2016a). *MeGI* in this species has a novel epigenetic *cis*-regulatory developmental switch that controls its expression pattern, and both male and female flowers can develop, explaining its monoecy (Akagi et al. 2016a, 2023). The Y acts as a $Y^{monoecy}$ factor, or Y^m, adopting terminology similar to Y^h, which is used for the Y of hermaphrodite revertants of dioecious papaya (Wang, Na, et al. 2012).

Here, by chromosome-scale whole-genome assembly of monoecious *D. kaki* cv. Taishu (6A + XXXXYY; Akagi et al. 2016b), we clarify the history of Y^m and the evolution of the former MSY, by comparing Y^m with a functioning Y-linked region in a close diploid relative, the Caucasian persimmon (*Diospyros lotus*).

Results and Discussion

Evolutionary Paths of Hexaploid Persimmon and the Silenced Y-Determinant, OGI^m

We assembled *D. kaki* cv. Taishu (2n = 6x = 90, 6A + XXXXYY) whole-genome sequences with PacBio HiFi reads, initially resulting in total 2.39 Gb scaffolds with N50 = 21.2 Mb (N = 37) (supplementary table S1 [which gives the basic genome characterization] and fig. S1,

Supplementary Material online). Further scaffolding using RaGOO/RagTag (Alonge et al. 2019, 2022) with a reference genome of a close relative, *D. lotus* cv. Kunsenshi-male (2A + XY), and integration of allelic sequences yielded 14 chromosome-scale autosomal scaffolds plus the XY pair ("pseudomolecule sequences"), consistent with the basic chromosome number of *Diospyros* species (N = 15). These scaffolds include 36,866 predicted genes covering 94.5% of the eudicot complete core gene set (complete Benchmarking Universal Single-Copy Orthologs; BUSCOs) (fig. 1B and supplementary table S1, Supplementary Material online). All the genome sequences and the annotated data were deposited to the Persimmon Genome Database (http://persimmon.kazusa.or.jp/index.html) and Plant GARDEN (https://plantgarden.jp/en/index). In allopolyploids, if the ancestral chromosome sets are distinct enough, they can be assembled at the haploid (or phased) level, as in diploid species (Hu et al. 2019; Sato et al. 2021). On the other hand, autopolyploids simply increase allelic complexity. Therefore, closely related autosomes are often assembled as consensus sequences and output some duplicated alternative contigs (Kyriakidou et al. 2018), as observed in this study (supplementary table S1, Supplementary Material online). In contrast, ancient haploblocks, such as X and Y chromosomes in *D. kaki*, would be distinguishable, which could be assembled as separate haploblocks (Zhang et al. 2018; Chen et al. 2020).

Synteny analysis of the pseudomolecule sequences and analysis of the distribution of silent divergence (dS) values in putatively homologous gene pairs (see Materials and Methods) suggest that *D. kaki* underwent at least two paleogenome duplication events, producing pairs with dS = 0.62–0.80 and 1.24–1.50; the more recent event suggests a *Diospyros*-specific whole-genome duplication, Dd-α (Akagi et al. 2020), whereas the older one is consistent with the hexaploidization-γ event known to have occurred in the common ancestor of eudicotyledonous plants (Jaillon et al. 2009) (fig. 1). Distributions of the dS values in *D. kaki* allelic sequences show a peak at 0.0141. Part of this diversity includes differences between chromosomal homoeologs created by the very recent hexaploidization event. As this probably involved a strong bottleneck, allelic variation in *D. kaki* should mostly have arisen since the event occurred. Consistent with this expectation that current diversity within *D. kaki* must have been established after the split from its close diploid relative, *Diospyros oleifera*, the raw divergence estimate from this species is slightly higher (median dS = 0.0225, fig. 2A). The divergence of *D. lotus* may have occurred slightly longer ago, as dS is slightly larger (median dS = 0.0301, P = 0.0013 for the dS distribution).

Importantly, the pairwise dS values between the sequences of the SINE transposable element (TE), *Kali*, within the *OGI* sequence (see above) in 12 cultivars from various East Asian areas (supplementary table S3, Supplementary Material online; Akagi et al. 2016b), range up to >0.02 (fig. 2A). These results suggest that this insertion coincided with the divergence of *D. kaki* and *D. oleifera* and may predate the hexaploidization event, as summarized in figure 2B.
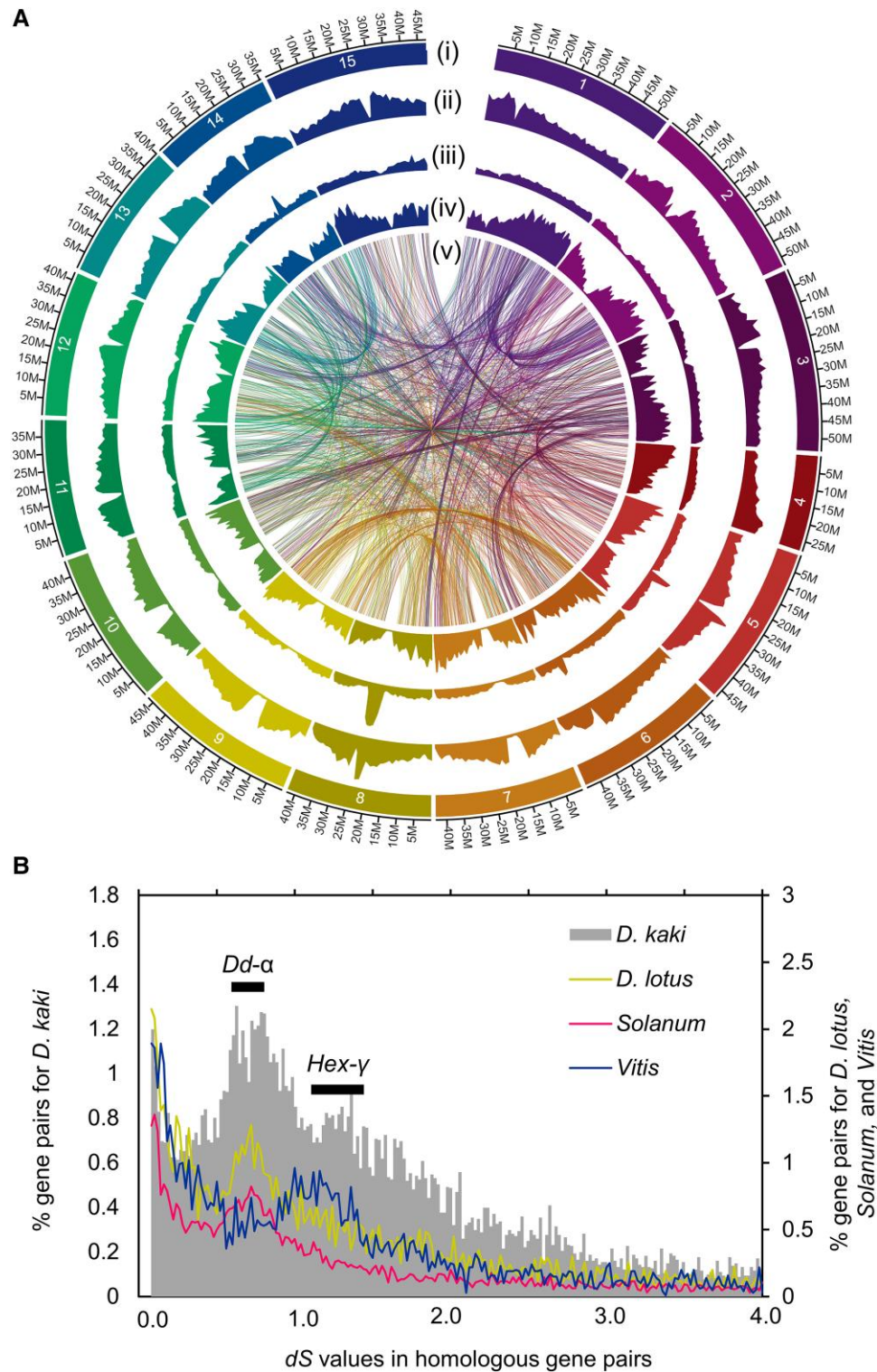
FIG. 1. Characterization of *D. kaki* cv. Taishu draft genome. (*A*) Overview of 15 pseudomolecules with 14 autosomes (Chr 1–14) and Y chromosome (Chr 15). From the outer layers, chromosome numbers (i), relative density of TEs detected by EDTA ((ii) for LTR-type and (iii) for non-LTR type), and relative gene density (iv) were given. In the central area (v), syntenic relationships within the persimmon genome, with gene pairs showing dS = 0.6–0.9, which corresponds a putative whole-genome duplication event, *Dd-α* (see *B*), were indicated. (*B*) Distribution of silent divergence (dS) rates between homologous gene pairs within the *D. kaki*, *D. lotus*, *Solanum*, and *Vitis* genomes. *Diospyros kaki* and *D. lotus* show a consistent peak, corresponding the *Dd-α*, at the same dS value as the *Solanum* genome triplication. Another peak in *D. kaki* (dS = 1.24–1.50) is almost consistent with a peak in the *Vitis* genome, which would correspond the hexaploidization-γ (*Hex-γ*).

The *OGI* gene would then be estimated to have been established (creating a proto-Y) approximately 25 million years ago (Ma) (Akagi et al. 2014), using an estimated rate of 4×10^{-9} substitutions per synonymous site per year in *Arabidopsis* (Beilstein et al. 2010; Wang, Na, et al. 2012), whereas the silence of *OGI* (*OGI^m*) to become a Y^m factor in the *D. kaki* lineage occurred approximately 2 Ma, assuming the same molecular clock rate for the *Kali* sequence.

Genome-wide synteny analysis with MCScanX supported the species divergence order estimated from the dS values, as the *D. oleifera* gene order is more similar to that in the *D. kaki* genome than to the *D. lotus* order (fig. 2C). Note, however, that this might be due to the incompleteness of pseudomolecules in the published *D. lotus* genome database (http://persimmon.kazusa.or.jp/). Genome-wide dS values between *D. kaki* and *D. oleifera* or *D. lotus* in 200-kb sliding windows were mostly
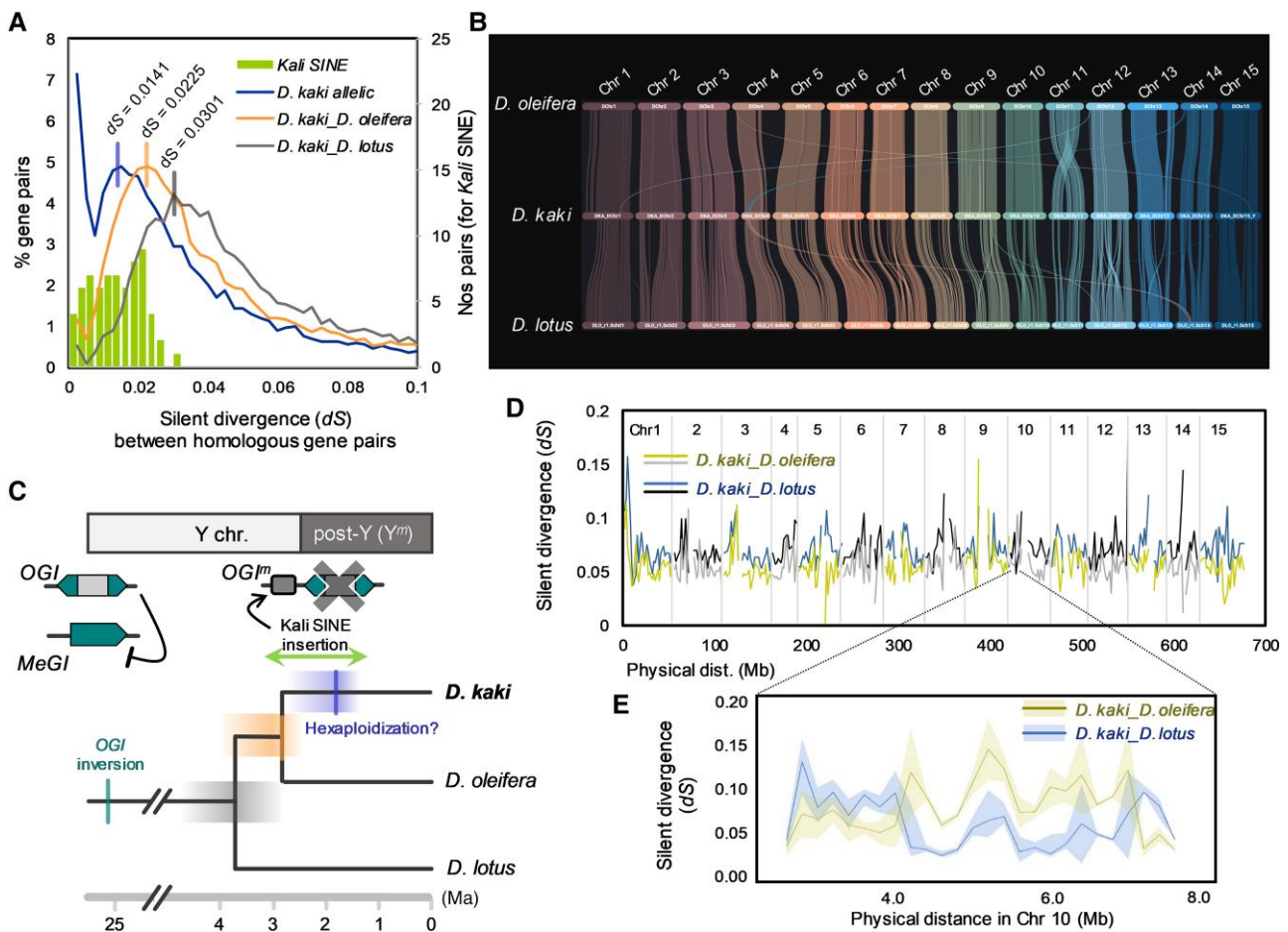
Fig. 2. Species divergence and evolutionary path of the post-Y chromosome in *D. kaki*. (A) Comparison of the distribution of dS value within *D. kaki* allelic sequences, between the orthologs in *D. kaki* and *D. oleifera*, and in *D. kaki* and *D. lotus*. The distribution of dS values in the *Kali*-SINE allelic sequences was given in the bars. (B) Chromosome-scale synteny analysis based on the gene orders in *D. kaki*, *D. oleifera*, and *D. lotus* genomes. (C) Schematic model for the evolution of the post-Y chromosome, including the *OGI^m* silenced by the *Kali*-SINE insertion, with the estimated ages. Ma, million years ago. (D) The 1-Mb bin walking analysis for the dS values between the orthologs in *D. kaki* and *D. oleifera*, and in *D. kaki* and *D. lotus*. Most of the genome showed smaller dS between *D. kaki* and *D. oleifera* than between *D. kaki* and *D. lotus*, which was consistent with the genome-wide dS distribution (given in A). (E) Closing up of a region exhibiting smaller dS between *D. kaki* and *D. lotus* than between *D. kaki* and *D. oleifera*, implying potential introgression from the *D. lotus* genome.

consistent with the phylogenetic relationships of these species (fig. 2D). However, 3–4% of the genomic regions had lower dS values ($P < 0.01$ for each 1-Mb bin, by Student's *t*-tests) between *D. kaki* and *D. lotus* than between *D. kaki* and *D. oleifera* (fig. 2E), suggesting randomness of the coalescent process, or possibly some introgressions from the *D. lotus* lineage after the divergence of *D. kaki/D. oleifera* and *D. lotus* (Hudson and Turelli 2003).

Conservation of MSY Characteristics in the Y^m

Synteny analysis indicated highly conserved gene orders between the *D. kaki* X and Y^m chromosomes, except in the putative MSY (or post-MSY) region that includes the silenced *OGI^mon* (fig. 3A). The post-MSY occupies approximately 1.5 Mb in the pericentromeric region of chromosome 15 (figs. 3A, 3B, and S2, Supplementary Material online), with fragmented syntenic blocks compared with

its X counterpart (fig. 3B). The dS values between X and Y^m alleles in monoecious *D. kaki* (fig. 3C) reached 0.212 in the central region flanking the *OGI* gene (fig. 3D), which is comparable to the dS values in the functional MSY of dioecious *D. lotus* (dS = 0.196 for the locus closest to *OGI*; Akagi et al. 2020) and the dS value between *OGI* and *MeGI* (dS = 0.205) corresponding to the initial establishment of the functional MSY (Akagi et al. 2014). Despite the fact that there are few X-Y allelic pairs of genes in the post-MSY, the dS values between X- and Y-linked sequences clearly decline with increased distance from *OGI^m*, but do not correspond to distinct evolutionary strata (fig. 3D). Strata were also not detected near the functional MSY in dioecious *D. lotus* (Akagi et al. 2020).

Under the two-factor model of the evolution of separate sexes and Y-linked regions (Charlesworth and Charlesworth 1978), hermaphrodite revertant individuals can arise, carrying the so-called "Y^h" chromosomes, and may have been favored by artificial selection, as in the
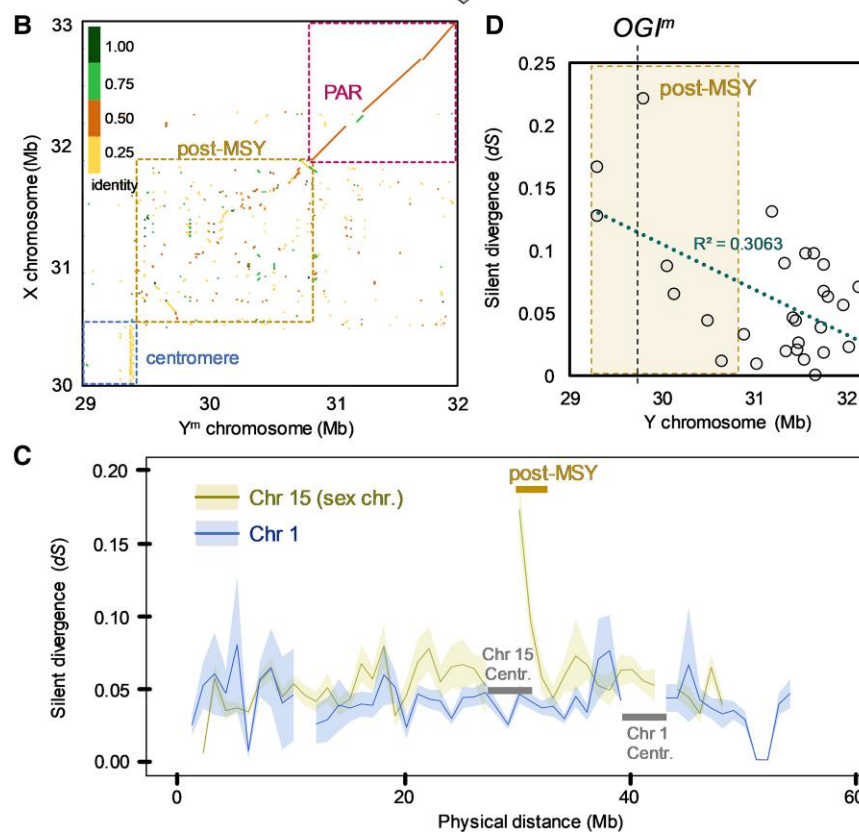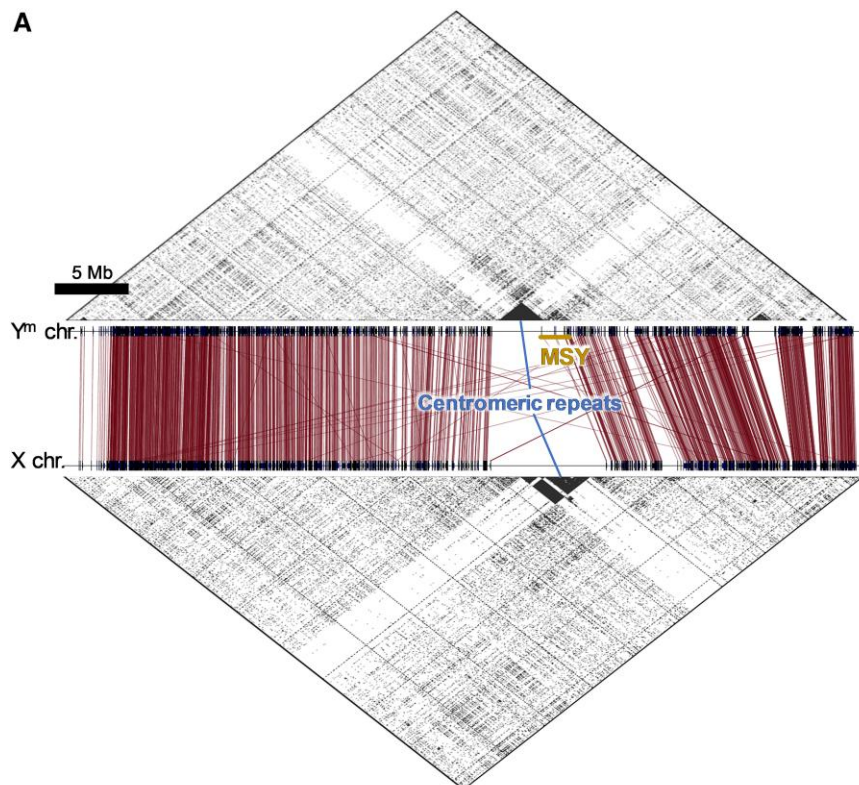
Fig. 3. Comparative analysis of X and Y chromosomes in *D. kaki*. (*A*) Synteny based on the allelic genes between X and Y chromosomes. Putative allelic genes with significant synteny (<1e^{-50} in blastp) were connected with the lines. The top and bottom triangle areas indicated self-synteny dot plots. The post-MSY located flanking the putative centromeric region. (*B*) Synteny between the post-MSY and the counterpart X allelic region. In the post-MSY, fragmental synteny blocks were observed. (*C*) Transition of 1-Mb bin dS values between X and Y allelic genes. For a control, transition of 1-Mb bin dS values among the alleles in Chr 1 (or an autosome) was given. The post-MSY exhibited clear increase in the dS against the X alleles. (*D*) The distribution of the dS values in each gene, in the 3-Mb region surrounding the post-MSY.

domestication of papaya (Wang, Na, et al. 2012; Van Buren et al. 2015, 2016) or in the grapevine (Zhou et al. 2017, 2019). To test the possibility that the persimmon Y^m has also experienced such selection, we analyzed nucleotide diversity (π) and Tajima's *D* values in the Y^m chromosome,

using the genome-wide single nucleotide polymorphisms in resequencing data of five *D. kaki* cultivars (DRA015334 for the sequenced accessions in DDBJ). Recent selection for the Y^m allele should cause a "selective sweep," with decreased diversity in the region. However, we did not detect

5

this (supplementary fig. S3, Supplementary Material online). This could be explained by the fact that all *D. kaki* Y chromosomes so far studied carry *OGI^m* (Akagi et al. 2016a), suggesting that establishment and fixation of Y*^m* could have occurred via a strong naturally occurring bottleneck in population size (as suggested above, when *D. kaki* became hexaploid) or possibly a long ago selective sweep, rather than via recent artificial selection.

Ongoing Rapid Evolution of the Post-MSY

Despite only approximately 3.5 million years having elapsed since the divergence of *D. lotus* and *D. kaki* (see fig. 2B), their MSY regions differ by many rearrangements (figs. 4A and S4A, Supplementary Material online). Overall, only 5 of the 44 genes in the post-MSY are shared with the *D. lotus* MSY, and their physical order is inverted (supplementary fig. S4B, Supplementary Material online), in contrast to the flanking PAR, where most of the genes are shared between *D. lotus* and *D. kaki*.

The *D. kaki* post-MSY has also accumulated a high density of repetitive sequences, compared with the flanking pseudo-autosomal region (PAR) (fig. 4B). The *D. kaki* post-MSY, but not the corresponding region of the *D. kaki* X chromosome, is enriched with long terminal repeat (LTR)–type TEs, especially the *Gypsy* class, unlike the *D. lotus* MSY (figs. 4C and S5, Supplementary Material online). Both the *D. lotus* MSY and the *D. kaki* post-MSY independently accumulated at least 31 mostly small ($N \geq 3$) clusters of *Gypsy* class TEs (supplementary table S4, Supplementary Material online). Approximately two-thirds of the *Gypsy* TEs in the post-MSY are not in clusters (using the criterion of >80% identity), suggesting that most are independently derived from sources elsewhere in the *D. kaki* genome. However, the largest *Gypsy* cluster (clst. 23) is specific to the *D. kaki* post-MSY (fig. 4D) and probably evolved very recently, after the establishment of the Y*^m*, as pairwise dS values never exceed 0.02 (fig. 4E). Similar recent duplications of specific *Gypsy* TEs are also detected in the *D. lotus* MSY, but on a smaller scale (supplementary fig. S6, Supplementary Material online).

The highly rearranged structures in the *D. lotus* MSY and *D. kaki* post-MSY is not unexpected if these are nonrecombining regions, as chromosomal rearrangements and TE insertions/deletions are not selectively disadvantageous in such regions (because crossing over does not occur, which also prevents deleterious ectopic exchanges between different TE insertions). MSYs in other plant species also show such changes, compared with their X counterparts (Akagi et al. 2019, 2023 for kiwifruit, Harkess et al. 2020 for garden asparagus, Ma et al. 2022 for spinach, and Wang, Na, et al. 2012 for papaya). In *D. lotus*, purifying selection may prevent establishment of some rearrangements, because the MSY contains a functional *OGI* gene, and possibly other factors controlling sexual dimorphisms, but the post-MSY in *D. kaki*, might be under weak or no purifying selection, allowing more insertions of TEs and duplications to occur.

In monoecious *D. kaki* trees, morphological differences between male and female flowers, including inflorescence structure and different flower numbers (supplementary fig. S7, Supplementary Material online), are similar to the sexual dimorphisms in diploid species in the genus *Diospyros* and must thus be independent of the evolution of Y-linked region. Darwin (1877) suggested that some sexual dimorphisms might be direct effects of sex-determining functions, and termed this "compensation," and there is evidence for such pleiotropic effects (in today's terminology) in other persimmon species in the genus *Diospyros* and in kiwifruit (in the genus *Actinidia*) (Akagi and Charlesworth 2019; Akagi et al. 2023). It is interesting that, in both these genera, their MSYs are located within ancestrally repetitive (and therefore probably rarely recombining) pericentromeric or peritelomeric regions (Akagi et al. 2023). Taken together, the recent evolution of the post-MSY in *D. kaki*, which has lost dioecy (and possibly also MSYs in dioecious *Diospyros* species), might reflect these regions' ancestral genomic properties, rather than having evolved male-determining functions and/or sexually dimorphic traits (fig. 5). Still, we cannot exclude several hypotheses for the potential extension of nonrecombining regions independent of sexual dimorphisms, such as evolution of dosage compensation (Lenormand and Roze 2022).

Materials and Methods

Genome Assembly

For whole-genome assembly, *D. kaki* cv. Taishu young leaves were sampled at Grape and Persimmon Research Station, the Institute of Fruit Tree and Tea Science, National Agriculture and Food Research Organization (NARO). Genomic DNA for sequencing was extracted using the Qiagen Genome-tip G100. The genome DNA was sheared into ~20 kb DNA fragments with a g-tube machine (Covaris, Woburn, MA, USA), and a HiFi SMRTbell library was constructed with a SMRTbell Express Template Prep Kit 2.0 (PacBio, Menlo Park, CA, USA). The library DNA was fractionated using the BluePippin (Sage Science, Beverly, MA, USA) to eliminate fragments <20 kb and sequenced with a Four SMRT cell 8M on the Sequel II system (PacBio). The sequence reads were converted into HiFi reads with the CCS pipelines (PacBio, https://ccs.how) and assembled with Hifiasm (Cheng et al. 2021). The obtained contigs were aligned to a reference genome sequence of diploid *D. lotus* male cv. Kunsenshi-male (Akagi et al. 2020; 2n = 2x = 30, 2A + XY) for scaffolding by using RaGOO/RagTag (Alonge et al. 2019, 2022) to build pseudomolecule sequences, finally with manual revisions.

Repeat and Gene Annotations

Repetitive sequences in the assemblies were identified with phRAIDER (Schaeffer et al. 2016) and RepeatMasker (Smit et al. 2015, https://www.repeatmasker.org/) using
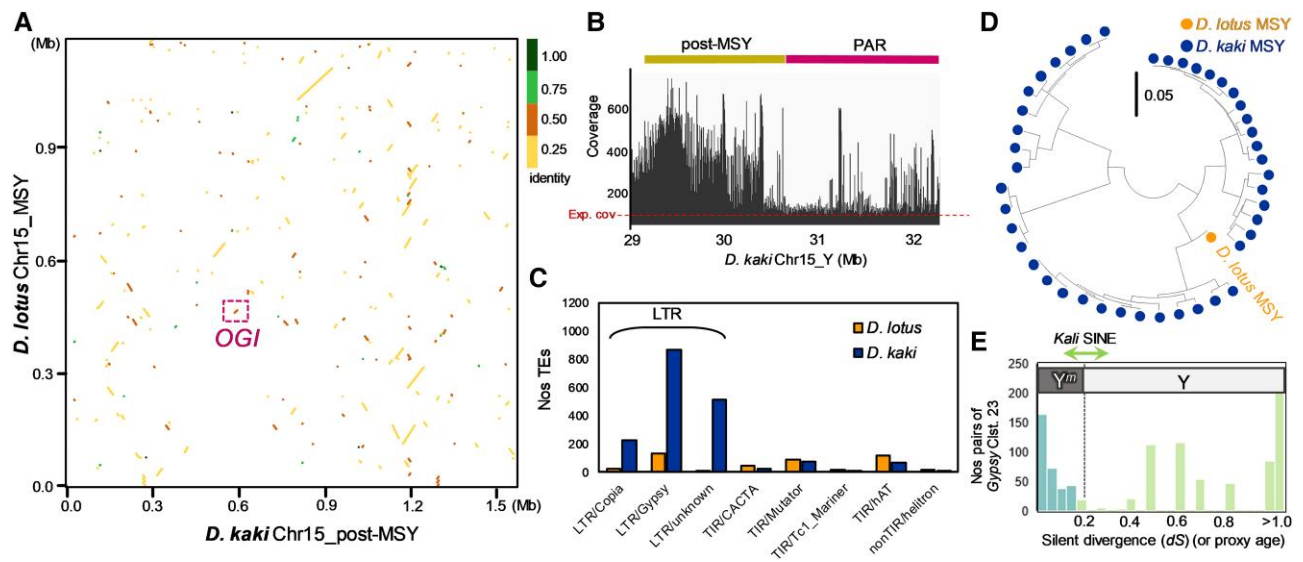
FIG. 4. Comparative analysis of the MSY in *D. lotus* and the post-MSY in *D. kaki*. (*A*) Synteny between the MSY in *D. lotus* and the post-MSY in *D. kaki*. Despite only 3–4 million years after their divergence (see fig. 1*C*), their structures were highly rearranged. (*B*) Mapping of the random gDNA-seq data of cv. Taishu to the post-MSY and the flanking PAR. The sequence coverage suggested that the post-MSY in *D. kaki* was enriched with repetitive sequences. (*C*) Comparison of TE accumulation in the *D. lotus* MSY and the *D. kaki* post-MSY. The LTR-type TEs were highly more enriched in the *D. kaki* post-MSY. (*D*) Phylogenetic analysis of a *Gypsy* cluster 23, which exhibited the *D. kaki* post-MSY–specific recent burst, with the substitution rates <0.05. (*E*) Histogram of the pairwise dS values in the *Gypsy* cluster 23 suggested that dynamic TE bursts occurred even after the *Kali*-SINE insertion, or establishment of the Y*m* chromosome (or post-MSY).
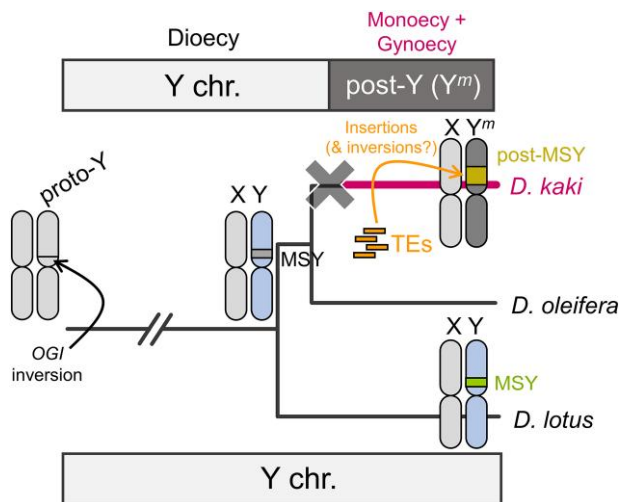

FIG. 5. Model for the ongoing evolution of the post-Y chromosome. Since the establishment of the *OGI* inversion, the common ancestor of *D. lotus* and *D. kaki* had formed a putative MSY. Immediately after the divergence of *D. lotus* and *D. kaki*, *OGI* was silenced to establish the post-Y chromosome in the *D. kaki* ancestral lineage. The post-MSY has been continuously extended and rearranged via active TE, which might be more rapid than the MSY in dioecious *D. lotus*.

de novo repeat libraries built with RepeatModeler2 (Flynn et al. 2020). Repeat elements were classified into nine types with RepeatMasker: SINEs, long interspersed nuclear elements (LINEs), LTR elements, DNA elements, small RNA, satellites, simple repeats, low-complexity repeats, and unclassified.

Protein-coding genes were predicted from the repeat-masked genome sequences. Gene prediction was conducted with the Braker2 pipeline, trained with 22 Illumina short-read mRNA-seq data from a variety of organs (fruit flesh, Maeda et al. 2019; flower buds and flowers, Masuda et al. 2022; and young leaf/flower primordia, Akagi et al. 2016a), in various developmental stages, according to the previous pipeline (Shirasawa et al. 2022). The completeness of the assemblies was assessed using the BUSCO score (Simão et al. 2015).

Synteny Analyses

Chromosome-scale sequence synteny was evaluated with D-genies (Cabanettes and Klopp 2018) for dot plot visualization (http://dgenies.toulouse.inra.fr/). Two whole-genome sequence (fasta) files were aligned with Minimap2 with the D-genies default conditions. Large-scale synteny based on the gene orders was examined with MCScanX (Wang, Tang, et al. 2012), in which the detected collinearity was visualized using SynVisio (https://synvisio.github.io/#/). All-versus-all BLASTP analyses were performed among the protein sequences in the *D. lotus*, *D. oleifera*, and *D. kaki* genomes, with an *e* value cutoff of $<1e^{-40}$. Syntenic blocks were constructed using MCScanX, with BLASTP and gff files, after preprocessing to be suitable for MCScanX. Intragenomic collinearity was evaluated by all-versus-all BLASTP analyses ($<e^{-50}$ in blastp), for the whole genes, followed by the selection with threshold values of silent divergence (dS; described later) and visualization with the Circa software (https://omgenomics.com/circa/). Short-range syntenic blocks (or

repetitive blocks within a chromosome) were identified with Mummer4 (Marçais et al. 2018), using the nucmer program, with the --maxmatch argument (with minimum length of the syntenic block = 25 bp).

Detection of Genetic Diversity and Age Estimation

For genetic diversity within the paralogs in the interspecific comparisons, gene pairs between *D. kaki* and *D. lotus* or *D. kaki* and *D. oleifera*, which exhibited significant sequence similarity ($<e^{-50}$ in blastp), were subjected to in-codon frame alignment using their protein and nucleotide sequences with Pal2Nal and MAFFT ver. 7 under the L-INS-i model (Katoh and Standley 2013). The resultant alignments were subjected to MEGA X (Kumar et al. 2018) to estimate the Jukes and Cantor corrected values of silent divergence (dS). For genetic diversity within the genome-wide *D. kaki* alleles, the whole genes in the pseudomolecule sequences (14 + XY chromosome) were aligned to the predicted genes in the initial scaffolds excluding the components of the pseudomolecule sequences, followed by detection of dS values, as described. For detection of dS values surrounding the *Kali*-SINE, we sequenced the <2-kb PCR products flanking the *Kali*-SINE and the mutated *OGI*, in 12 cultivars (Akagi et al. 2016a, 2016b). To estimate the divergence time between the gene pairs, we adopted an estimated rate of 4×10^{-9} substitutions per synonymous site per year, according to the previous reports (Beilstein et al. 2010; Wang, Na, et al. 2012).

Characterization of the Genomic Context Around the MSY

For the construction of Illumina genomic libraries, we used approximately 0.5 μg of genomic DNA of cv. Taishu. The libraries were constructed using the KAPA HyperPlus Library Preparation Kit (KAPA Biosystems) and sequenced using Illumina's HiSeq4000 (150-bp paired-end reads). All Illumina sequencing was conducted at the Vincent J. Coates Genomics Sequencing Laboratory at UC Berkeley, and the raw reads were processed using custom Python scripts developed in the Comai laboratory and available online (http://comailab.genomecenter.ucdavis.edu/index.php/), as previously described (Akagi et al. 2014). The preprocessed reads were aligned to the whole genomic scaffolds of each species, with Burrows–Wheeler Aligner (BWA)-mem (Li and Durbin 2009), allowing up to 12% mismatches. The mapped reads were visualized with Integrative Genomics Viewer (IGV) (Robinson et al. 2011).

De Novo TE Annotations and Evolution

TEs in the genomes of *D. kaki* and *D. lotus* were detected with the Extensive de novo TE Annotator (EDTA) pipeline that integrates structure- and homology-based approaches for TE identification, including LTRharvest, LTR_FINDER_parallel, LTR_retriever, Generic Repeat Finder, TIR-Learner, MITE-Hunter, and HelitronScanner,

with extra basic and advanced filters (Ou and Jiang 2019). Clustering of TEs within the *Gypsy* family was conducted with cd-hit-est in the Cd-hit (Cluster Database at High Identity with Tolerance) (Li and Godzik 2006), with the -c 0.8 (>80% sequence identity) option.

For the evolutionary topology of some *Gypsy* clusters, we aligned their sequences with MAFFT ver. 7 with the L-INS-i model, followed by manual pruning using SeaView. The alignments were concatenated, and all sites, including missing and gap data, were used to construct phylogenic trees with the maximum likelihood (ML) method by using iqtree2 (Minh et al. 2020), with automatically optimized parameters.

Accession Numbers and Construction of Database

All of the genome sequences and the annotated data were deposited to the Persimmon Genome Database (http://persimmon.kazusa.or.jp/index.html) and Plant GARDEN (https://plantgarden.jp/en/index). All PacBio genome sequencing data have been deposited in the DDBJ database: Short Read Archives (SRA) database (BioProject ID PRJDB14984, BSEG01000001–BSEG01000016).

Supplementary Material

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

Author Contributions

T.A. conceived the study. A.H., K.M., and T.A. designed the experiments. A.H., K.M., K.S., and T.A. conducted the experiments. A.H., K.M., K.S., and T.A. analyzed the data. N.O., N.F., K.U., and T.A. contributed to plant resources and facilities. A.H. and T.A. drafted the manuscript. All authors approved the manuscript.

References

Akagi T, Charlesworth D. 2019. Pleiotropic effects of sex-determining genes in the evolution of dioecy in two plant species. *Proc R Soc B.* **286**:20191805.

Akagi T, Henry IM, Kawai T, Comai L, Tao R. 2016a. Epigenetic regulation of the sex determination gene *MeGI* in polyploid persimmon. *Plant Cell* **28**:2905–2915.

Akagi T, Henry IM, Ohtani H, Morimoto T, Beppu K, Kataoka I, Tao R. 2018. A Y-encoded suppressor of feminization arose via lineage-specific duplication of a cytokinin response regulator in kiwifruit. *Plant Cell* **30**:780–795.

Akagi T, Henry IM, Tao R, Comai L. 2014. A Y-chromosome–encoded small RNA acts as a sex determinant in persimmons. *Science* **346**: 646–650.

Akagi T, Kawai T, Tao R. 2016b. A male determinant gene in diploid dioecious *Diospyros*, *OGI*, is required for male flower production in monoecious individuals of Oriental persimmon (*D. kaki*). *Sci Hort.* **213**:243–251.

Akagi T, Pilkington SM, Varkonyi-Gasic E, Henry IM, Sugano SS, Sonoda M, Firl A, McNeilage MA, Douglas MJ, Wang T, et al. 2019. Two Y-chromosome-encoded genes determine sex in kiwifruit. *Nat Plants.* **5**:801–809.

Akagi T, Shirasawa K, Nagasaki H, Hirakawa H, Tao R, Comai L, Henry IM. 2020. The persimmon genome reveals clues to the evolution of a lineage-specific sex determination system in plants. *PLoS Genet.* **16**:e1008566.

Akagi T, Varkonyi-Gasic E, Shirasawa K, Catanach A, Henry IM, Mertten D, Datson P, Masuda K, Fujita N, Kuwada E, et al. 2023. Recurrent neo-sex chromosome evolution in kiwifruit. *Nat Plants.* **9**:393–402.

Alonge M, Lebeigle L, Kirsche M, Jenike K, Ou S, Aganezov S, Wang X, Lippman ZB, Schatz MC, Soyk S. 2022. Automated assembly scaffolding using RagTag elevates a new tomato system for high-throughput genome editing. *Genome Biol.* **23**:258.

Alonge M, Soyk S, Ramakrishnan S, Wang X, Goodwin S, Sedlazeck FJ, Lippman ZB, Schatz MC. 2019. RaGOO: fast and accurate reference-guided scaffolding of draft genomes. *Genome Biol.* **20**:224.

Beilstein MA, Nagalingum NS, Clements MD, Manchester SR, Mathews S. 2010. Dated molecular phylogenies indicate a Miocene origin for *Arabidopsis thaliana*. *Proc Natl Acad Sci USA.* **107**:18724–18728.

Cabanettes F, Klopp C. 2018. D-GENIES: dot plot large genomes in an interactive, efficient and simple way. *PeerJ* **6**:4958.

Charlesworth D. 1985. Distribution of dioecy and self-incompatibility in angiosperms. In: Greenwood PJ and Slatkin M, editors. *Evolution: essays in honour of John Maynard Smith.* Cambridge: Cambridge University Press. p. 237–268.

Charlesworth D. 2019. Young sex chromosomes in plants and animals. *New Phytol.* **224**:1095–1107.

Charlesworth B, Charlesworth D. 1978. A model for the evolution of dioecy and gynodioecy. *Am Nat.* **112**:975–997.

Chen H, Zeng Y, Yang Y, Huang L, Tang B, Zhang H, Hao F, Liu W, Li Y, Liu Y, et al. 2020. Allele-aware chromosome-level genome assembly and efficient transgene-free genome editing for the autotetraploid cultivated alfalfa. *Nat Commun.* **11**:2494.

Cheng H, Concepcion GT, Feng X, Zhang H, Li H. 2021. Haplotype-resolved de novo assembly using phased assembly graphs with hifiasm. *Nat Methods.* **18**:170–175.

Darwin CR. 1877. *The different forms of flowers on plants of the same species.* London: John Murray.

Flynn JM, Hubley R, Goubert C, Rosen J, Clark AG, Feschotte C, Smit AF. 2020. Repeatmodeler2 for automated genomic discovery of transposable element families. *Proc Natl Acad Sci USA.* **117**: 9451–9457.

Harkess A, Huang K, Van der Hulst R, Tissen B, Caplan JL, Koppula A, Batish M, Meyers BC, Leebens-Mack J. 2020. Sex determination by two Y-linked genes in garden asparagus. *Plant Cell* **32**: 1790–1796.

Harkess A, Zhou J, Xu C, Bowers JE, Van der Hulst R, Ayyampalayam S, Mercati F, Riccardi P, McKain MR, Kakrana A, et al. 2017. The asparagus genome sheds light on the origin and evolution of a young Y chromosome. *Nat Commun.* **8**:1279.

Henry IM, Akagi T, Tao R, Comai L. 2018.One hundred ways to invent the sexes: theoretical and observed paths to dioecy in plants. *Annu Rev Plant Biol.* **69**:553–575.

Hu Y, Chen J, Fang L, Zhang Z, Ma W, Niu Y, Ju L, Deng J, Zhao T, Lian J, et al. 2019. Reference genome sequences of two cultivated allotetraploid cottons, *Gossypium hirsutum* and *Gossypium barbadense*. *Nat Genet.* **51**: 224–229.

Hudson RR, Turelli M. 2003. Stochasticity overrules the "three-times rule": genetic drift, genetic draft, and coalescence times for nuclear loci versus mitochondrial DNA. *Evolution* **57**:182–190.

Jaillon O, Aury JM, Wincker P. 2009. "Changing by doubling", the impact of whole genome duplications in the evolution of eukaryotes. *Comptes Rendus Biol.* **332**:241–253.

Katoh K, Standley DM. 2013. MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Mol Biol Evol.* **30**:772–780.

Kazama Y, Kitoh M, Kobayashi I, Ishii K, Krasovec M, Yasui Y, Abe T, Kawano S, Filatov DA. 2022. A *CLAVATA3*-like gene acts as a gynoecium suppression function in white campion. *Mol Bio Evol.* **39**:msac195.

Kumar S, Stecher G, Li M, Knyaz C, Tamura K. 2018. MEGA X: molecular evolutionary genetics analysis across computing platforms. *Mol Biol Evol.* **35**:1547.

Kyriakidou M, Tai HH, Anglin NL, Ellis D, Strömvik MV. 2018. Current strategies of polyploid plant genome sequence assembly. *Front Plant Sci.* **9**:1660.

Lenormand T, Roze D. 2022. Y recombination arrest and degeneration in the absence of sexual dimorphism. *Science* **375**:663–666.

Li H, Durbin R. 2009. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* **25**:1754–1760.

Li W, Godzik A. 2006. Cd-hit: a fast program for clustering and comparing large sets of protein or nucleotide sequences. *Bioinformatics* **22**: 1658–1659.

Ma J, Ali S, Saleem MH, Mumtaz S, Yasin G, Ali B, Al-Ghamdi AA, Elshikh MS, Vodnar DC, Marc RA et al. 2022. Short-term responses of spinach (*Spinacia oleracea* L.) to the individual and combinatorial effects of nitrogen, phosphorus and potassium and silicon in the soil contaminated by boron. *Front Plant Sci.* **13**:983156.

Maeda H, Akagi T, Onoue N, Kono A, Tao R. 2019. Evolution of lineage-specific gene networks underlying the considerable fruit shape diversity in persimmon. *Plant Cell Physiol.* **60**: 2464–2477.

Marçais G, Delcher AL, Phillippy AM, Coston R, Salzberg SL, Zimin A. 2018. MUMmer4: a fast and versatile genome alignment system. *PLoS Comput Biol.* **14**:e1005944.

Masuda K, Akagi T. 2022. Sexual system and its evolution. In: *The persimmon genome.* Cham: Springer. p. 97–107.

Masuda K, Ikeda Y, Matsuura T, Kawakatsu T, Tao R, Kubo Y, Ushijima K, Henry IM, Akagi T. 2022. Reinvention of hermaphroditism via activation of a RADIALIS-like gene in hexaploid persimmon. *Nat Plants.* **8**:217–224.

Masuda K, Yamamoto E, Shirasawa K, Onoue N, Kono A, Ushijima K, Kubo Y, Tao R, Henry IM, Akagi T. 2020. Genome-wide study on the polysomic genetic factors conferring plasticity of flower sexuality in hexaploid persimmon. *DNA Res.* **27**:dsaa012.

Ming R, Bendahmane A, Renner SS. 2011. Sex chromosomes in land plants. *Annu Rev Plant Biol.* **62**:485–514.

Minh BQ, Schmidt HA, Chernomor, O, Schrempf D, Woodhams MD, Von Haeseler A, Lanfear R. 2020. IQ-TREE 2: new models and efficient methods for phylogenetic inference in the genomic era. *Mol Biol Evol.* **37**:1530–1534.

Müller NA, Kersten B, Leite Montalvão AP, Mähler N, Bernhardsson C, Bräutigam K, Carracedo Lorenzo Z, Hoenicka H, Kumar V, Mader M, et al. 2020. A single gene underlies the dynamic evolution of poplar sex determination. *Nat Plants.* **6**:630–637.

Ou S, Jiang N. 2019. LTR_FINDER_parallel: parallelization of LTR_FINDER enabling rapid identification of long terminal repeat retrotransposons. *Mob DNA.* **10**:1–3.

Renner SS, Müller NA. 2021. Plant sex chromosomes defy evolutionary models of expanding recombination suppression and genetic degeneration. *Nat Plants.* **7**:392–402.

Rice WR. 1992. Sexually antagonistic genes: experimental evidence. *Science* **256**:1436–1439.

Robinson JT, Thorvaldsdóttir H, Winckler W, Guttman M, Lander ES, Getz G, Mesirov JP. 2011. Integrative Genomics Viewer. *Nat Biotechnol.* **29**:24–26.

Sato K, Abe F, Mascher M, Haberer G, Gundlach H, Spannagl M, Shirasawa K, Isobe S. 2021. Chromosome-scale genome assembly of the transformation-amenable common wheat cultivar 'Fielder'. *DNA Res.* **28**:dsab008.

Schaeffer CE, Figueroa ND, Liu X, Karro JE. 2016. Phraider: pattern-hunter based rapid ab initio detection of elementary repeats. *Bioinformatics* **32**:i209–i215.

Shirasawa K, Ueta S, Murakami K, Abdelrahman M, Kanno A, Isobe S. 2022. Chromosome-scale haplotype-phased genome assemblies of the male and female lines of wild asparagus (*Asparagus kiusianus*), a dioecious plant species. *DNA Res.* **29**:dsac002.

Simão FA, Waterhouse RM, Ioannidis P, Kriventseva EV, Zdobnov EM. 2015. BUSCO: assessing genome assembly and annotation completeness with single-copy orthologs. *Bioinformatics* **31**:3210–3212.

Smit AFA, Hubley R, Green P. 2015. *RepeatMasker Open*-4.0. 2013–2015. http://www.repeatmasker.org.

VanBuren R, Wai CM, Zhang J, Han J, Arro J, Lin Z, Liao Z, Yu Q, Wang ML, Zee F *et al.* 2016. Extremely low nucleotide diversity in the X-linked region of papaya caused by a strong selective sweep. *Genome Biol.* **17**:1–11.

VanBuren R, Zeng F, Chen C, Zhang J, Wai CM, Han J, Aryal R, Gschwend AR, Wang J, Na JK *et al.* 2015. Origin and domestication of papaya Y^h chromosome. *Genome Res.* **25**:524–533.

Wang J, Na JK, Yu Q, Gschwend AR, Han J, Zeng F, Aryal R, VanBuren R, Murray JE, Zhang W, *et al.* 2012. Sequencing papaya X and Yh chromosomes reveals molecular basis of incipient sex chromosome evolution. *Proc Natl Acad Sci USA.* **109**:13710–13715.

Wang Y, Tang H, DeBarry JD, Tan X, Li J, Wang X, Lee T, Jin H, Marler B, Guo H, *et al.* 2012. *MCScanx*: a toolkit for detection and evolutionary analysis of gene synteny and collinearity. *Nucleic Acids Res.* **40**:e49.

Westergaard M. 1958. The mechanism of sex determination in dioecious flowering plants. *Adv Genet.* **9**:217–281.

Yang HW, Akagi T, Kawakatsu T, Tao R. 2019. Gene networks orchestrated by *MeGI*: a single-factor mechanism underlying sex determination in persimmon. *Plant J.* **98**:97–111.

Zhang J, Zhang X, Tang H, Zhang Q, Hua X, Ma X, Zhu F, Jones T, Zhu X, Bowers J, *et al.* 2018. Allele-defined genome of the autopolyploid sugarcane *Saccharum spontaneum* L. *Nat Genet.* **50**:1565–1573.

Zhou Y, Massonnet M, Sanjak JS, Cantu D, Gaut BS. 2017. Evolutionary genomics of grape (*Vitis vinifera* ssp. *vinifera*) domestication. *Proc Natl Acad Sci USA.* **114**:11715–11720.

Zhou Y, Minio A, Massonnet M, Solares E, Lv Y, Beridze T, Cantu D, Gaut BS. 2019. The population genetics of structural variants in grapevine domestication. *Nat Plants.* **5**:965–979.