HAL open science

The origin and evolution of sex chromosomes, revealed by sequencing of the Silene latifolia female genome

Jingjing Yue, Marc Krasovec, Yusuke Kazama, Xingtan Zhang, Wangyang Xie, Shencheng Zhang, Xiuming Xu, Baolin Kan, Ray Ming, Dmitry A Filatov

To cite this version:

Jingjing Yue, Marc Krasovec, Yusuke Kazama, Xingtan Zhang, Wangyang Xie, et al.. The origin and evolution of sex chromosomes, revealed by sequencing of the Silene latifolia female genome. Current Biology, 2023, 33 (12), pp.2504-2514.e3. ⟨10.1016/j.cub.2023.05.046⟩. ⟨hal-04269359⟩

HAL Id: hal-04269359

https://hal.science/hal-04269359v1

Submitted on 3 Nov 2023

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Autorisation HAL

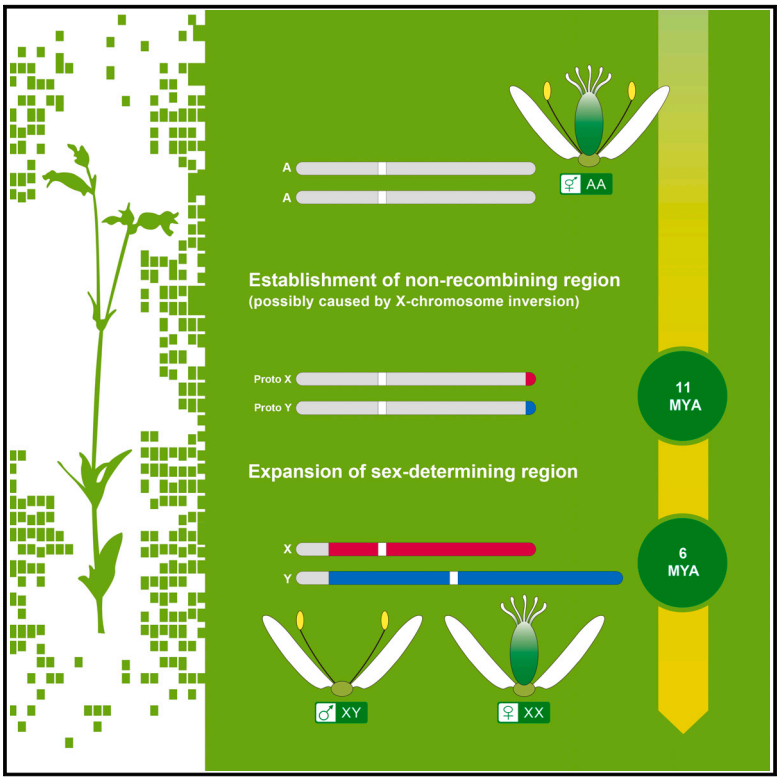
HAL Authorization

Article

Current Biology

The origin and evolution of sex chromosomes, revealed by sequencing of the *Silene latifolia* female genome

Graphical abstract



Highlights

- Genome sequencing reveals sex chromosome evolution in white campion
- The sex-determining region initially evolved in a small, actively recombining region
- The non-recombining region on the Y (NRY) likely evolved due to inversion on the X
- Extensive pericentromeric recombination suppression facilitated NRY expansion

Authors

Jingjing Yue, Marc Krasovec, Yusuke Kazama, ..., Baolin Kan, Ray Ming, Dmitry A. Filatov

Correspondence

rayming@illinois.edu (R.M.), dmitry.filatov@biology.ox.ac.uk (D.A.F.)

In brief

Yue et al. describe the evolution of sex chromosomes in white campion, based on the chromosome-level genome assembly integrated with a genetic map. The sex-determining region (SDR) has likely evolved in an ~15 Mb region, which stopped recombining due to X-linked inversion. SDR expansion was facilitated by extensive pericentromeric recombination suppression.

Yue et al., 2023, Current Biology *33*, 2504–2514
June 19, 2023 © 2023 The Authors. Published by Elsevier Inc.
https://doi.org/10.1016/j.cub.2023.05.046

CellPress

CellPress
OPEN ACCESS

Current Biology

Article

The origin and evolution of sex chromosomes, revealed by sequencing of the *Silene latifolia* female genome

Jingjing Yue,[1,7] Marc Krasovec,[2,3,7] Yusuke Kazama,[4] Xingtan Zhang,[1,5] Wangyang Xie,[1] Shencheng Zhang,[5] Xiuming Xu,[6] Baolin Kan,[1] Ray Ming,[1,*] and Dmitry A. Filatov[2,8,*]
[1]Centre for Genomics and Biotechnology, Fujian Provincial Key Laboratory of Haixia Applied Plant Systems Biology, Key Laboratory of Genetics, Breeding and Multiple Utilization of Crops, Ministry of Education, Fujian Agriculture and Forestry University, Fuzhou 350002, China
[2]Department of Biology, University of Oxford, Oxford OX1 3RB, UK
[3]Sorbonne Université, CNRS, UMR 7232 Biologie Intégrative des Organismes Marins (BIOM), Observatoire Océanologique, 66650 Banyuls-sur-Mer, France
[4]Department of Bioscience and Biotechnology, Fukui Prefectural University, 4-1-1 Kenjojima, Matsuoka, Eiheiji-cho, Fukui 910-1195, Japan
[5]Shenzhen Branch, Guangdong Laboratory for Lingnan Modern Agriculture, Genome Analysis Laboratory of the Ministry of Agriculture, Agricultural Genomics Institute at Shenzhen, Chinese Academy of Agricultural Sciences, Shenzhen 518100, China
[6]Key Laboratory of the Ministry of Education for Coastal and Wetland Ecosystems, College of the Environment and Ecology, Xiamen University, Xiamen 361100, China
[7]These authors contributed equally
[8]Lead contact
*Correspondence: rayming@illinois.edu (R.M.), dmitry.filatov@biology.ox.ac.uk (D.A.F.)
https://doi.org/10.1016/j.cub.2023.05.046

SUMMARY

White campion (*Silene latifolia*, Caryophyllaceae) was the first vascular plant where sex chromosomes were discovered. This species is a classic model for studies on plant sex chromosomes due to presence of large, clearly distinguishable X and Y chromosomes that originated *de novo* about 11 million years ago (mya), but lack of genomic resources for this relatively large genome (~2.8 Gb) remains a significant hurdle. Here we report *S. latifolia* female genome assembly integrated with sex-specific genetic maps of this species, focusing on sex chromosomes and their evolution. The analysis reveals a highly heterogeneous recombination landscape with strong reduction in recombination rate in the central parts of all chromosomes. Recombination on the X chromosome in female meiosis primarily occurs at the very ends, and over 85% of the X chromosome length is located in a massive (~330 Mb) gene-poor, rarely recombining pericentromeric region (Xpr). The results indicate that the non-recombining region on the Y chromosome (NRY) initially evolved in a relatively small (~15 Mb), actively recombining region at the end of the q-arm, possibly as a result of inversion on the nascent X chromosome. The NRY expanded about 6 mya via linkage between the Xpr and the sex-determining region, which may have been caused by expanding pericentromeric recombination suppression on the X chromosome. These findings shed light on the origin of sex chromosomes in *S. latifolia* and yield genomic resources to assist ongoing and future investigations into sex chromosome evolution.

INTRODUCTION

Sex chromosomes are a peculiar part of the genome that evolved independently many times in different groups of organisms.[1,2] In vascular plants they were first discovered exactly a century ago by Kathleen Blackburn,[3] who reported that *Silene latifolia* (then known as *Lychnis alba*) has "an XY pair of chromosomes in the male with a corresponding XX in the female." Despite independent origins of sex chromosomes in different organisms, their properties are quite similar, indicating the generality of evolutionary forces and processes driving their evolution. Non-recombining sex-specific Y (or W) chromosomes are typically genetically degenerate, while X (or Z) chromosomes that actively recombine in the homogametic sex contain hundreds to thousands of functional genes.[4] The contrasting properties of the Y (or W) and X (or Z) chromosomes are striking, given they typically evolve from a pair of autosomes[5] and initially have the same gene composition.[6] This represents a good illustration of how the differences in the recombination landscape can drive evolution of major changes in chromosome structure, density of genes, and repetitive DNA.[7–10]

Evolution of a non-recombining sex-specific region is a key step in sex chromosome evolution, yet how and why recombination cessation on nascent sex chromosomes evolves remains poorly understood.[11–15] Species with very small sex-determining regions (e.g., in fish[16,17] and plants[18–22]) demonstrate that recombination suppression around the sex-determining gene(s) is not obligatory, raising questions about why large, non-recombining regions repeatedly evolved and expanded in size on sex chromosomes of many species independently.[11,14,15,23–27]

 © 2023 The Authors. Published by Elsevier Inc.
This is an open access article under the CC BY license (http://creativecommons.org/licenses/by/4.0/).

CellPress
OPEN ACCESS

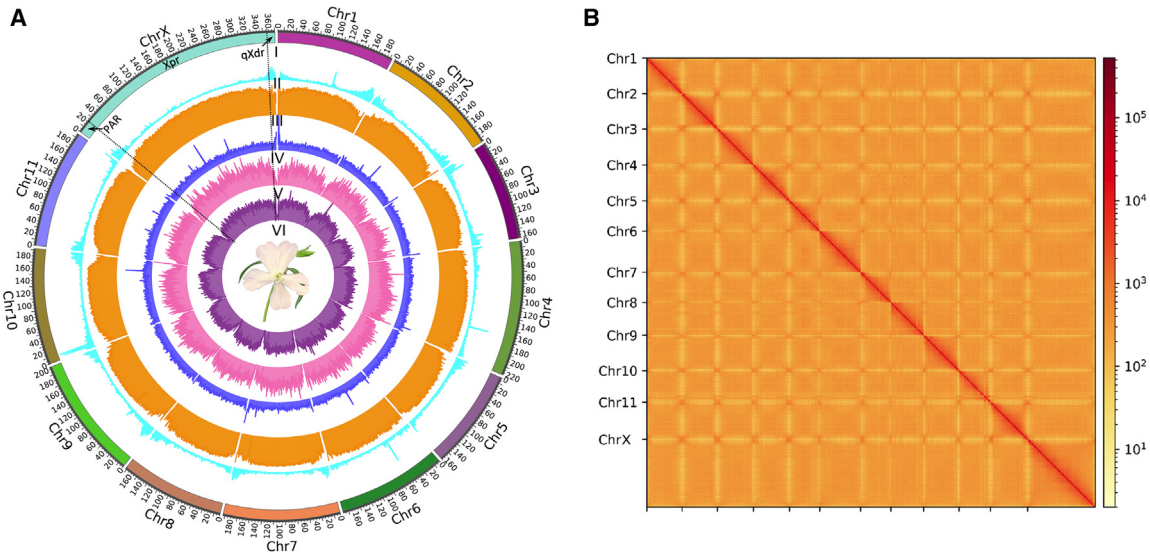


Figure 1. The genome assembly of *S. latifolia* female genome

(A) Rings indicate twelve chromosomes (I), gene density (II), TE density (III), GC density (IV), copia LTR retrotransposon density (V), and gypsy LTR retrotransposon density (VI). The dotted lines show the boundaries between PAR, Xpr, and qXdr on the X chromosome.

(B) Hi-C chromatin interactions at 1 Mb resolution for twelve chromosomes.

See also Tables S1–S9.

Species where large, non-recombining regions around sex-determining genes evolved recently, such as *S. latifolia*,[28] which is the focus of this study, allow us to study what causes evolution of recombination suppression on sex chromosomes.[15]

Studies in *S. latifolia* have contributed significantly to our understanding of sex chromosome evolution.[29–31] Early work in this species[32] has inspired the development of ideas about how separate sexes and sex chromosomes originate.[33] More recent work brought important clues on Y chromosome degeneration in plants,[28,34–36] sex-biased expression,[37] dosage compensation,[36,38–40] and specific sex-linked gene evolution.[41–45] The *S. latifolia* sex chromosomes evolved *de novo* during or after the transition to separate sexes about 11 million years ago (mya),[28] which is relatively recent compared to mammals[46] or birds,[47] where the sex chromosomes are at least 10 times older. The *de novo* evolution of X and Y chromosomes in *S. latifolia* offers a rare opportunity to study the origination of sex chromosomes once a species evolves separate sexes.[5] The relatively recent origin of sex chromosomes in this species enables the analysis of evolutionary dynamics of sex chromosomes at the early stages of their evolution.

Fragmented assemblies of the *S. latifolia* female[36] and male[28] genomes remain a significant limitation in the work with this species. This fragmentation does not allow one to establish the order and location of genomic contigs and the locations of the recombining and non-recombining regions on the sex chromosomes and autosomes. The relatively large size of the genome (haploid ~2.8 Gb[48,49]) and abundance of repetitive sequences have, so far, precluded the construction of better genome assemblies. In this study, we took advantage of multiple long read sequencing datasets to generate a new female assembly close to chromosome scale and integrate it with genetic maps. This revealed a highly uneven recombination landscape throughout the *S. latifolia* genome, with peripheral recombination in all chromosomes. Extensive pericentromeric recombination suppression (PRS) in females is particularly pronounced on the X chromosome, where it has likely contributed to evolution of sex chromosomes by facilitating or even causing the expansion of the non-recombining region on the Y chromosome (NRY). Furthermore, the PRS on the X chromosome significantly affects the patterns of genetic diversity and divergence in the sex-linked genes that have previously been attributed to Y chromosome degeneration.

RESULTS

Genome sequencing, assembly, and annotation

We used PacBio HiFi sequencing (104.9 Gb, 40× sequence coverage; Table S1) to create a draft assembly of the *S. latifolia* female genome. The initial assembly was 2.64 Gb long (contig N50 = 23.1 Mb; Table S2) with 98.0% genome completeness as assessed by BUSCO (Table S3). This assembly was then integrated with Hi-C data using the ALLHiC pipeline.[50] This yielded 12 pseudo-chromosomes with a total size of 2.45 Gb and an anchor rate of 93.0% (Figure 1A; Table S2). The number of assembled pseudo-chromosomes corresponds to the 12 chromosomes in the *S. latifolia* genome. The accuracy of Hi-C-based pseudo-chromosome construction was evaluated by chromatin contact matrix, which showed a well-organized interaction contact pattern along the diagonals within each pseudo-chromosome (Figure 1B; Table S4). The quality of the female *S. latifolia* assembly was further verified by RNA sequencing (RNA-seq) and genomic short Illumina reads from Krasovec et al.[28] and Papadopulos et al.[36] The results showed that 100% of transcripts were aligned with 99.9% of single base accuracy for the genome (Table S5) and 99.1% of the genomic Illumina reads mapped to the assembly, covering 94.0% of the

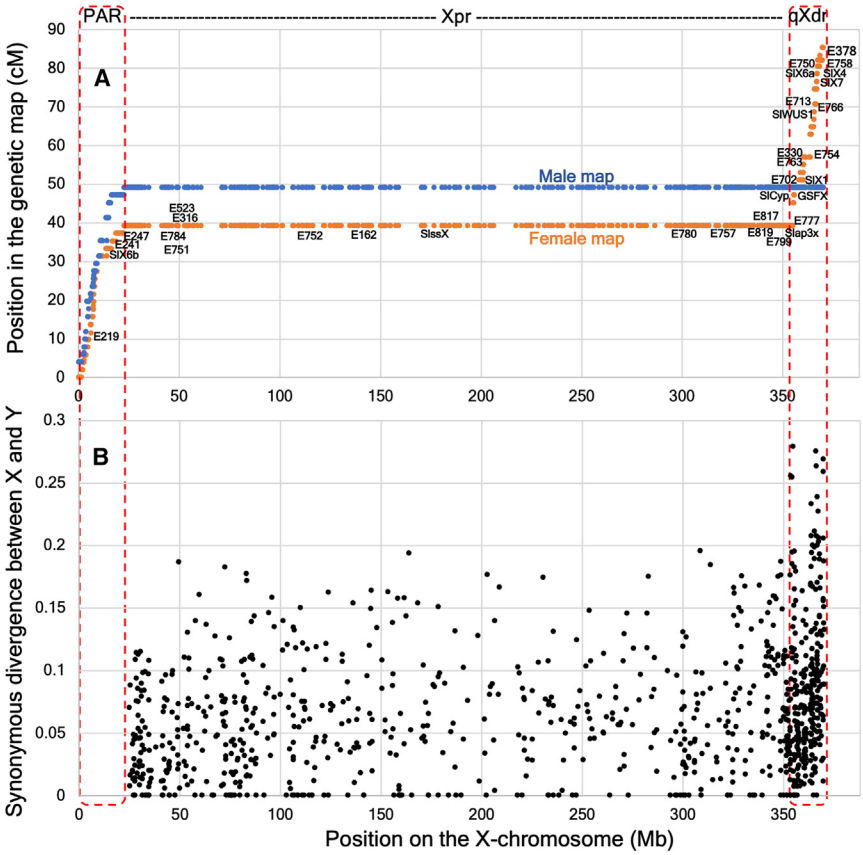


Figure 2. Rarely recombining Xpr region represents most of the X chromosome length in *S. latifolia*

Genetic versus physical position for the X chromosome (A) and synonymous divergence between X- and Y-linked gametologs in the Xpr and qXdr (B). The names of genes in (A) show the locations of sex-linked genes genetically mapped in the previous studies.[35,43,45] The genetic map of the PAR in (A) is longer in the male map due to obligatory sex chromosome pairing in the PAR in male meiosis. See also Figure S1.

genome assembly (Table S6). All these results indicated the high quality of the assembled *S. latifolia* female genome.

The assembled genome was integrated with the previously published RNA-seq-based sex-specific maps,[51] which revealed a good correspondence between the genome sequence and the maps (Figures 2A and S1). In particular, 4,745 and 4,598 markers of the female- and male-specific maps,[51] respectively, were located in the genome, of which 812 and 809 markers were X-linked. This analysis revealed that X-linked markers mapped to the scaffold "Chr12," indicating its correspondence to the X chromosome. Furthermore, the sex-linked genes *SlX1*, *SlX3*, *SlX4*, *SlCyp*, and *SlssX*, previously located on the X chromosome with fluorescent *in situ* hybridization,[52] are all located on the scaffold "Chr12" (Figure 2A), confirming that this scaffold represents the sequence of the X chromosome.

We annotated 37,796 protein-coding genes in the *S. latifolia* female genome with 96.9% annotation completeness assessed by BUSCO (Tables S2 and S7). The average gene density on the X chromosome was significantly lower than that on the autosomes, except Chr4 and Chr6 (Table 1). We annotated 93 microRNAs (miRNAs) in the *S. latifolia* female genome (Table S2). Annotation of repetitive sequences revealed that they represent 82.8% (2.18 Gb) of the genome assembly (Table S8). Long terminal repeat (LTR) retrotransposons (LTRs) were the most abundant transposable elements (TEs), representing 71.72% of the *S. latifolia* female genome (Figure 1A; Table S8). Within the LTRs, Ty3/gypsy elements comprise the most abundant group, composing 35.9% of the genome, followed by Ty1/copia (12.8%) (Table S8). TEs were significantly more abundant on the X chromosome compared to autosomes, except Chr4 and Chr6 (Table 1), comprising 87.1% of the X chromosome, with 13.3% represented by Ty1/copia elements and 37.3% by Ty3/gypsy elements (Figure 1A).

Recombination suppression on the X chromosome

The comparison of the physical and genetic distances along the genomic scaffolds revealed reduced recombination in the central part of all chromosomes, indicating considerable pericentromeric recombination suppression in both sexes (Figure S1). In particular, the female meiosis recombination in the distal regions of the X chromosome is ~2 cM/Mb on average (Figure 2A), while the central region, comprising at least 85% of the X chromosome length, is rarely recombining (~0 cM/Mb). This is consistent with the view that species with large chromosomes tend to have highly peripheral recombination, with large pericentromeric regions lacking recombination.[53,54]

One of the distal actively recombining regions in the X chromosome scaffold contains the genes previously identified[36,51] to be located in the pseudoautosomal region (PAR). As the *S. latifolia* PAR is located in the p-arm of the X chromosome,[55–57] we will refer to this region as the p-arm X distal region (pXdr or PAR; Figure 2, left), while the actively recombining region on the other side of the X chromosome will be referred to as q-arm X distal region (qXdr; Figure 2, right). The rarely recombining region between the pXdr and qXdr will be referred to as the X pericentromeric region (Xpr).

The PAR boundary (as defined in Filatov[58]) is located close to the boundary between pXdr and Xpr. According to the previously published sex-specific genetic map,[51] female recombination rate at the X-linked genes adjacent to the PAR boundary (the "right" and "mid" regions in Filatov[58]) is zero, suggesting that they belong to the rarely recombining Xpr region. On the other hand, the pseudoautosomal genes adjacent to the PAR boundary (the "left" region in Filatov[58]) appear to recombine more actively, suggesting that they are located in the pXdr. Thus, it is likely that the PAR boundary coincides with the pXdr/Xpr

CellPress
OPEN ACCESS

Table 1. The chromosome lengths and the densities of genes and TEs in *S. latifolia* female genome

Chromosome	Contigs	Length (bp)	Genes/Mb	p value[a]	TE%[b]	p value[a]
ChrX	110	370,597,487	12.1		87.1	
Chr1	188	193,566,110	18.3	7.30E−10	82.8	2.20E−39
Chr2	38	196,073,315	18.4	4.70E−06	83.1	8.90E−26
Chr3	34	165,661,342	18.9	4.60E−11	82.9	4.10E−34
Chr4	52	223,726,948	11.9	0.087	87.3	0.0029
Chr5	51	165,362,148	18.4	8.00E−12	82.7	7.20E−47
Chr6	27	179,322,760	11.2	0.037	87.2	3.00E−05
Chr7	57	193,230,414	14.1	0.16	85.1	0.00017
Chr8	119	172,294,220	15.8	5.20E−06	83.8	5.20E−31
Chr9	55	203,056,505	17.3	0.074	83.5	1.10E−09
Chr10	34	192,831,310	17.6	0.034	83.4	9.10E−14
Chr11	48	196,798,467	14.6	0.02	85.1	3.70E−11

[a]Significance of difference from the X chromosome (ChrX) was tested using two-sided Mann-Whitney-Wilcoxon test with p-adjusted Homl correction. 1 Mb window and 0.2 Mb slide steps were used to calculate the TE abundance and gene density.
[b]TE% is the % of DNA sequence represented by transposable elements

boundary, though genetic mapping in several independent crosses is needed to confirm this result.

The genes previously used to define "evolutionary strata" on *S. latifolia* sex chromosomes[36,42,43,51,58–60]—the regions where cessation of X:Y recombination occurred first (stratum 1: *SlX4*, *SlX7*, E766, E713, E758, E750, and E378[35,42,43,60]) and then more recently (stratum2: *SlssX*, *Slap3X*, E799, E817, and E819[35,43,61])—are located in the qXdr and the Xpr, respectively (Figure 2A). This indicates that initial cessation of recombination between the proto-X and proto-Y chromosomes around 11 mya[28] (possibly due to inversion on the X, as explained in the discussion) occurred in an actively recombining qXdr that is relatively small (~15 Mb long), while expansion of the NRY about 6 mya involved inclusion of the massive (~330 Mb) Xpr into the non-recombining male-specific region. Higher X:Y synonymous divergence in qXdr compared to Xpr (Figure 2B; Table 2), reported in the next section, is consistent with this conclusion.

The actively recombining regions at the ends of the X chromosome are gene rich, while the Xpr is gene poor (Figure 1A; Table S9), with an average gene density of 17.8 and 6.6 genes per megabase, respectively. The distribution of TEs showed the opposite pattern, with TEs significantly more abundant in Xpr compared to the ends of the X chromosome (Figure 1A; Table S9). Similar TE depletion and gene enrichment are evident at the ends of most other chromosomes (Figure 1A).

Substitution rates in X- and Y-linked genes

To reconstruct Y-linked gemetologs for the X-linked genes, we employed a previously described approach, which uses male-specific Y-linked SNPs to identify the sequence reads corresponding to the Y chromosome and assemble Y-consensus from these reads.[28,34,36] The accuracy of the Y-reconstruction was confirmed in comparison with the sequences of Y-linked genes previously obtained by manual Sanger sequencing (e.g. *SlY1*, *SlY4*, *DD44Y*, and *SlssY*[61–64]). The resulting Y-linked genes were aligned with X-linked gametologs as well as with their homologs (identified by blast) from a non-dioecious outgroup *Silene uniflora*.[65]

Average synonymous divergence between the X- and Y-linked gametologs (d*S* X:Y) was 6.3% (±0.18%) and 9.1% (±0.42%) in the Xpr and the qXdr, respectively (Table 2; Figure 2B). Consistent with relaxation of selection on the Y chromosome, the non-synonymous to synonymous substitution rate ratios (d*N*/d*S*) for the Y-linked genes were significantly higher compared to their X-linked gametologs in both Xpr and qXdr regions (0.591 versus 0.315 and 0.408 versus 0.267, respectively; Table 2).

The analysis of divergence between the genes in *S. latifolia* female genome and their homologs in outgroup *S. uniflora* (Table 2) revealed that d*N*/d*S* ratios of the X-linked genes in the qXdr are not significantly different from those in the pseudoautosomal genes (0.233 ± 0.0186 and 0.221 ± 0.0170, respectively; t test, NS), while the genes in the Xpr have significantly higher *dN*/*dS* ratios (0.285 ± 0.0123; t test, $p < 0.001$). This is consistent with reduced efficacy of selection in the rarely recombining Xpr on the X chromosome compared to the recombining PAR and qXdr that actively recombines in females.

Gene expression

Y-linked copies were significantly less actively expressed compared to their X-linked gametologs (paired t test, $p < 0.001$), reflecting degeneration of Y-linked genes at the level of gene expression. Overall (X + Y or X + X) expression is female-biased, while this is not the case in the PAR or autosomal genes (Figure 3). The presence of the female bias indicates that the dosage compensation system (if any is present in *S. latifolia*[34,36,38,39,66]) is only partial and does not adequately compensate for reduced expression of Y-linked genes. However, it is possible that the genes with reduced X + Y (compared to X + X) are not dosage sensitive and thus do not need to be compensated. Female bias is significantly stronger in the qXdr compared to the Xpr (t test, $p < 0.001$; Figure 3), indicating that the extent of Y-degeneration in gene expression is stronger for the Y-linked gametologs of qXdr than Xpr genes, possibly because the qXdr is the oldest part of the sex chromosome. Alternatively, weaker female bias in the Xpr compared to the

Table 2. Average substitution rates (±SEs) in the Xpr, qXdr, and PAR genes

	PAR	Xpr	qXdr
No. genes	178	760	198
No. codons	211,704	839,597	254,442
dS X:OG	0.124 ± 0.0068	0.120 ± 0.0042	0.123 ± 0.0070
dS X:Y		0.063 ± 0.0018	0.091 ± 0.0042
dN/*dS* X		0.315 ± 0.0357	0.267 ± 0.0461
dN/*dS* Y		0.591 ± 0.1237[a]	0.408 ± 0.0487[a]
dN/*dS* X:OG	0.221 ± 0.0170	0.285 ± 0.0123	0.233 ± 0.0186

Non-dioecious *S. uniflora* was used as an outgroup (OG).
[a]Paired t tests, $p < 0.001$, revealing significantly higher *dN*/*dS* for Y-linked compared to X-linked genes in both Xpr and qXdr

qXdr is due to differences in female recombination rates in these regions. As Xpr is rarely recombining, the difference in recombination rate (and hence, efficacy of purifying selection keeping genes functional) between the X- and Y-linked genes is smaller compared to that for genes in qXdr that is actively recombining in females.

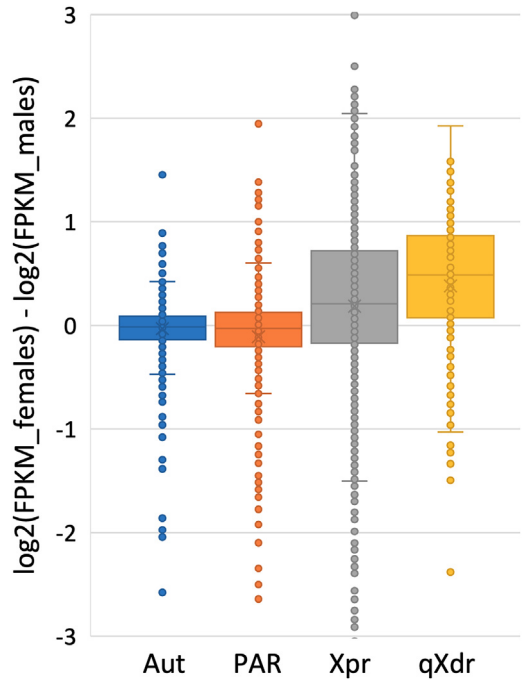


Figure 3. Female-biased expression of the genes in the qXdr and Xpr regions on the sex chromosomes
All comparisons between the groups, except the Aut:PAR comparison, are significant (t tests, $p < 0.001$). The horizontal line within each box shows the median, and the top and bottom of the box indicate the 25th and 75th percentiles, respectively. The error bars show 1.5 times interquartile range that approximately corresponds to two SDs.

DISCUSSION

Lack of reference genome sequence for *S. latifolia* significantly limited the previous work devoted to evolution of sex chromosomes in this interesting system. Here we reported the assembly of the female genome and its integration with genetic maps,[36,51] which shed light on the structure and evolution of sex chromosomes in this species. It revealed that recombination in the *S. latifolia* genome primarily occurs in relatively small regions at the ends of the chromosomes (Figure S1), which is particularly pronounced on the X chromosome (Figure 2A) and may have played a role in sex chromosome evolution, as discussed below. Suppression of recombination in the region around the sex-determining gene(s) is a key step in NRY formation and sex chromosome evolution. Our results help to understand how such recombination suppression evolves, stressing the importance of the pre-existing recombination landscape on nascent sex chromosomes. They also illustrate that the real sequence of events leading to sex chromosome evolution may deviate significantly from the classic "evolutionary strata" scenario[67] and the models of NRY expansion discussed in the literature.[14,24,26,27,68]

Evolution of recombination suppression on sex chromosomes

The size of the qXdr is only about 15 Mb, indicating that recombination suppression between the sex chromosomes in *S. latifolia* males has initially evolved in a relatively small region, not dissimilar to other dioecious plants, such as papaya,[18] persimmon,[19] kiwifruit,[20] asparagus,[21] ginkgo,[22] etc. However, unlike papaya,[69] kiwifruit,[70] and *Rumex*,[71] where the NRY evolved in a rarely recombining pericentromeric region, initial recombination suppression in *S. latifolia* has evolved in the actively recombining region (qXdr), where the *GSFX*—the X-linked gametolog of the putative sex-determining gene *GSFY*[72]—is located (Figure 2A). This is consistent with the comparison of genetic maps in *S. latifolia* and *Silene vulgaris* that revealed the presence of active recombination in this region in both species (green shading in Figure 4). It is interesting that the order of genes in this region is inverted between the two species (Figures 4C and 4D), suggesting that the initial recombination suppression on the proto-Y chromosome in this region was caused by an inversion on the proto-X chromosome. However, without the data from other *Silene* species, it is difficult to test whether this inversion arose in *S. latifolia* or *S. vulgaris* lineages.

Partial genetic maps are available for *Silene otites* and *Silene pseudotites*,[73] which independently evolved separate sexes and homomorphic ZW and XY sex chromosomes corresponding to *S. latifolia* linkage groups LG3 and LG6, respectively. It is interesting that in both cases the sex-determining region corresponds to the central, rarely recombining parts of homologous *S. latifolia* autosomes, which suggests that pre-existing pericentromeric recombination suppression facilitated evolution of sex chromosomes in *S. otites* and *S. pseudotites*, as it was also reported for *Rumex*.[71] Unfortunately, the partial maps available for these *Silene* species focus on LG1, LG3, and LG6, which does not allow us to test whether the inversion in the *S. latifolia* stratum 1 of the X chromosome (Figures 4C and 4D) occurred in *S. latifolia* lineage.

If confirmed that recombination between *S. latifolia* proto-X and Y chromosomes is suppressed by the X-linked inversion, this would contrast with the common assumption in the models that such inversions are Y- rather than X-linked.[24,26,68] Many of the models developed with Y-linked inversions in mind do not

CellPress
OPEN ACCESS

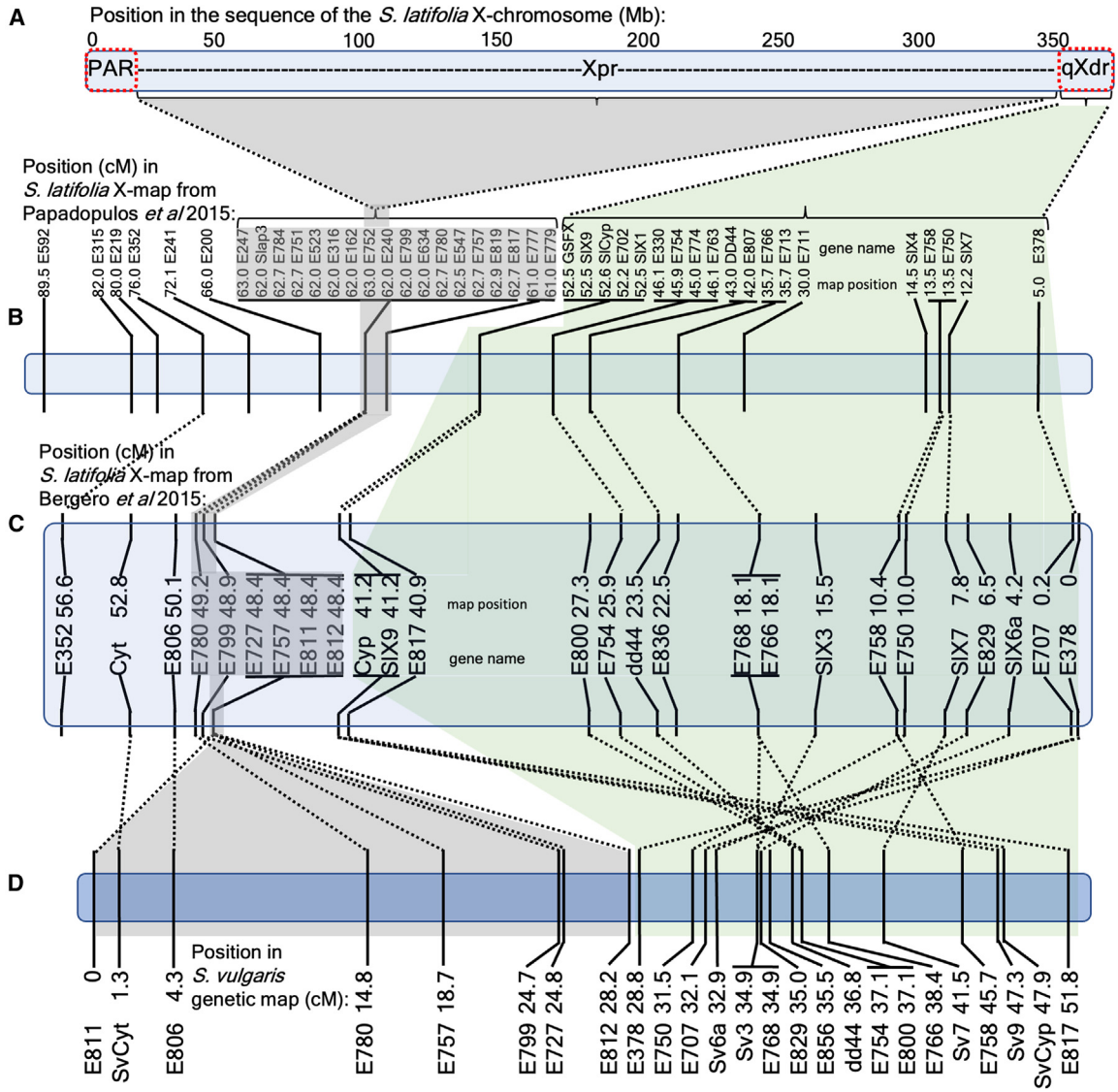


Figure 4. Evolution of recombination on the *S. latifolia* X chromosome

The comparison of the physical (A) and genetic (B–D) maps of the X chromosome in *S. latifolia* (A–C) and its homologous chromosome in non-dioecious *S. vulgaris* (D). (B) shows the genetic positions from Papadopulos et al.[36] for a subset of X-linked genes from this study that correspond to genes previously mapped in Bergero et al.[35,43] (C) and (D) show *S. latifolia* and *S. vulgaris* genetic maps from Bergero et al.[35] The locations of the same gene in different maps are linked with dotted lines. Green and gray shading show the genes and regions corresponding to the qXdr and Xpr, respectively. Note that qXdr is long (i.e., actively recombining) in the genetic maps of both *S. latifolia* and *S. vulgaris*, while Xpr is long only in *S. vulgaris*, and the genetic distance in this physically large region is nearly zero in *S. latifolia*, indicating evolution of suppressed recombination on the X chromosome in this region.

work for X-linked inversions. For example, the deleterious mutations sheltering model[24] requires the inversion to be linked to a permanently heterozygous locus, such as the Y chromosome. However, some of the existing models may still work for X-linked inversions. For example, the crux of the recently proposed early emergence of dosage compensation model[26] is that evolving expression modifiers prevent reversion to X:Y recombination in the region where X:Y recombination stopped due to fixation of an inversion. While that model was developed with Y-linked inversions in mind, it may still work if X:Y recombination is suppressed by an X-linked inversion.

The comparison of the genetic map lengths in the Xpr (Figures 4B and 4C) with the homologous region in *S. vulgaris* (Figure 4D) reveals that recombination suppression is specific to the *S. latifolia* Xpr and thus may have evolved during the evolution of sex chromosomes in this species. It is interesting to speculate that the expansion of the *S. latifolia* NRY to include Xpr may have been driven or facilitated by the evolution of pericentromeric recombination suppression in the central region of the X chromosome. For example, the NRY expansion could have occurred when the pericentromeric recombination suppression on the X chromosome became so extensive that it

reached the sex-determining region. Comparative analysis of genetic and physical distances in a few other *Silene* species will be necessary to test this conjecture.

Tip-biased distribution of recombination on *S. latifolia* chromosomes

Recombination suppression in central regions of the chromosomes is not universal in plants, with a lot of variation across species,[54] but species with large chromosomes (>100 Mb) tend to have peripheral distribution of recombination.[53,54] The concentration of crossovers at the ends of the chromosomes may be explained by mechanistic and/or adaptation-related causes. For example, the mechanistic "telomere-initiation" model postulates that crossovers tend to occur at the ends of chromosomes because the recombination machinery starts at the telomeres and proceeds inward.[53,74] Alternatively, distal distribution of recombination may be selected for to ensure sufficient recombination in gene-dense regions at the ends of large chromosomes.[54] Regardless of whether the cause is mechanistic or adaptation-driven, the peripheral distribution of recombination may be fueled by the tendency of "junk DNA," such as TEs, to accumulate in gene-poor, rarely recombining central chromosome regions[7] that hence expand over time, limiting the actively recombining gene-rich regions to the ends of large chromosomes. This "expanding junkyard" model helps to explain why large chromosomes tend to have peripheral recombination, while on the smaller chromosomes this tendency is weaker, if present at all.[54]

Given the association between the extent of peripheral recombination and the chromosome size,[54] it is interesting to speculate that the expansion of the *S. latifolia* genome may have contributed to evolution of recombination suppression between the X and Y chromosomes. The size of the *S. latifolia* genome (~2.8 Gb) is nearly 3-fold larger compared to non-dioecious outgroups such as *S. vulgaris* (1.13 Gb[48]) or *S. conica* (0.9 Gb[49]). Many species in genus *Dianthus*, which is closely related to *Silene*, have even smaller genomes (e.g., *Dianthus deltoides* genome ~0.45 Gb[75]), making it likely that the large size of *S. latifolia* genome is a derived state. As the chromosome number in diploid *Silene* is mostly conserved (n = 12), tripling of the genome size in the *S. latifolia* lineage (possibly due to accumulation of TEs[76]) must have tripled the size of the chromosomes, which could have exaggerated peripheral recombination, particularly on the X chromosome, which is the second largest (after the Y) chromosome in the *S. latifolia* genome.

Strongly peripheral recombination may have facilitated evolution of recombination suppression between the X and Y chromosomes. This conjecture is in line with the growing body of evidence that genome-wide variation in recombination rate, such as reduced recombination in the heterogametic sex (heterochiasmy),[77–79] or pericentromeric recombination suppression[71] can play a significant role in sex chromosome evolution by facilitating recombination suppression between the X and Y chromosomes. While heterochiasmy is unlikely to be important in *S. latifolia*,[51] the expansion of pericentromeric recombination suppression may have contributed significantly to evolution of sex chromosomes in this species.

The comparisons of genetic distances for autosomal genes in *S. latifolia* and its relatives would allow us to test whether the expansion of pericentromeric recombination suppression is X-specific or if it occurred on all chromosomes. Unfortunately, with the previous efforts focusing on sex chromosomes, little data are available for autosomes. The genetic distances between 14 autosomal genes in LG1 of *S. latifolia* and *S. otites* are similar in both species (Figure 1D in Martin et al.[73]), indicating little change in recombination landscape on that autosome over evolutionary time separating these species. Similarly, little difference in recombination distances between three autosomal genes from LG9 (E534, E526, and E157) was reported for *S. latifolia* and *S. vulgaris*.[43] The similarity of genetic distances for analyzed autosomal genes in *S. latifolia*, *S. otitis*, and *S. vulgaris* suggests that suppression of recombination in the central region of the *S. latifolia* X chromosome is probably specific to that chromosome. However, genetic mapping (ideally complemented with physical location) for more genes in several *Silene* species is necessary to reveal the evolution of recombination suppression.

Extensive pericentromeric recombination suppression on the X chromosome

The rarely recombining Xpr is a large gene-poor region spanning ~330 Mb in the central part of the *S. latifolia* X chromosome (Figure 2). The lack of recombination in this region explains the clustering of genes near the PAR boundary in the genetic maps published previously.[36,43] Furthermore, this explains the apparent contradiction between the cytogenetic evidence that placed *SlX1*, *SlX3*, *SlCyp*, and *DD44X* at the very end of the q-arm on the X chromosome,[52] while the genetic mapping was consistently placing these genes closer to the middle of the X chromosome map.[36,43] Our results reveal that these genes are located in the distal region on the q-arm of the X chromosome, which is consistent with the cytogenetic evidence,[52] while their placement closer to the middle of the genetic maps was caused by the collapse of the physically massive, but rarely recombining Xpr in the genetic maps.

Genes in the qXdr and the Xpr had evolved in very different recombination landscapes before the cessation of recombination between the X and Y chromosomes in these regions. Indeed, the average d*N*/d*S* ratio for the X-linked genes in Xpr is almost 2-fold higher compared to the qXdr (Table 2), indicating less effective purifying selection in the former compared to the latter. Partial relaxation of purifying selection in *S. latifolia* X-linked genes was previously reported,[28] but it was interpreted in the context of lower effective population size of the X-linked compared to autosomal genes rather than reduced recombination on much of the *S. latifolia* X chromosome.

Given the pre-existing recombination suppression in the Xpr, the reduction in the efficacy of selection caused by complete cessation of recombination in the NRY may have been much weaker for the genes in the Xpr compared to the qXdr. This implies that cessation of recombination due to inclusion in the NRY may have had little impact on the already rarely recombining genes in the Xpr. However, given significantly higher d*N*/d*S* for Y-linked compared to X-linked genes in the Xpr (0.591 ± 0.1237 and 0.315 ± 0.0357, respectively; Table 2), inclusion of this region in the NRY resulted in further reduction in the efficacy of selection in Y-linked genes compared to their X-linked gametologs in the Xpr.

CellPress
OPEN ACCESS

Furthermore, the female bias in gene expression is significantly weaker for the genes in the Xpr compared to the qXdr (Figure 3). While these findings could be interpreted as a result of more recent inclusion of the Xpr in the NRY compared to the qXdr, they are also consistent with the idea that genetic degeneration, caused by complete recombination suppression, affected the qXdr to a greater degree compared to the Xpr, which was already rarely recombining. Weak Y-degeneration in the Xpr may, at least partly, account for the previously reported slower rate of Y-degeneration of the *S. latifolia* Y chromosome compared to other studied species.[28,34,59] However, the slow Y-degeneration was reported for both evolutionary strata in *S. latifolia*,[28] which cannot be explained by pre-existing recombination suppression as genes in the qXdr undergo frequent recombination on the X chromosome.

Conclusions

The analyses presented above allow us to reconstruct the likely scenario of sex chromosome evolution in *Silene* and assess its implications for our understanding of evolutionary processes on nascent sex chromosomes. The initial recombination suppression, resulting in formation of NRY, occurred around 11 mya[28] at the actively recombining end of the proto-sex-chromosomes, corresponding to the qXdr of the *S. latifolia* X chromosome. This event may have been caused by an inversion on the proto-X chromosome, given the order of genes in the qXdr is inverted compared to the homologous region in *S. vulgaris* (Figures 4C and 4D). A few million years later, this was followed by NRY expansion via inclusion of the massive Xpr into the male-specific region, possibly driven by evolution of strongly peripheral recombination on the X chromosome. The Xpr region continues to recombine in females, albeit at a very low rate. Recombination suppression in the Xpr may have pre-dated or evolved in concert with NRY expansion. Given the homologous region in *S. vulgaris* is recombining normally (Figure 4D), the latter possibility appears plausible. Either way, reduced recombination in Xpr may have facilitated NRY expansion.

This scenario of sex chromosome evolution in *S. latifolia* deviates from the standard "evolutionary strata" model of step-wise NRY expansion,[67] when an actively recombining part of the pseudoautosomal region becomes Y-linked and stops recombining, while its homologous region continues to actively recombine on the X chromosome. Consistent with this model, the NRY in *S. latifolia* evolved in an actively recombining region (qXdr), but contrary to that model, the NRY expansion resulted in sex linkage of a region that may have already been rarely recombining (Xpr), meaning that both X- and Y-linked genes in this region evolve under reduced efficacy of selection. Failing to take into account such a pre-existing recombination landscape may lead to misinterpretation in evolutionary genetic analyses of sequence polymorphism and divergence on young sex chromosomes in general.

The scenario described above is compatible with the classic "two genes" model of dioecy and sex chromosome evolution,[33] with inversion in qXdr preventing recombination between two sex-determining genes, *SPF* and *GSF*. The X-linked homolog of the already isolated sex-determining gene (GSF[72]), along with *SlWUS1*, which was likely involved in evolution of gynoecium suppression,[72] is indeed located in qXdr (Figure 2A), though the *SPF* gene(s) remain to be identified. Our results indicate that the X chromosome may play an active role in NRY evolution, with X- (rather than Y-) linked inversions contributing to evolution of recombination suppression between the X and Y chromosomes, which remains to be considered by models of sex chromosome evolution.

STAR★METHODS

Detailed methods are provided in the online version of this paper and include the following:

- KEY RESOURCES TABLE
- RESOURCE AVAILABILITY
  - Lead contact
  - Materials availability
  - Data and code availability
- EXPERIMENTAL MODEL AND SUBJECT DETAILS
  - Plant material
- METHOD DETAILS
  - Genome sequencing and assembly
  - X-chromosome identification
  - Genome annotation with RNA-seq data
  - Genome filtering and correction
  - miRNA annotation
  - Transposable element (TE) annotation
  - Reconstruction of Y-linked genes
  - Analysis of gene expression
  - Phylogenetic analyses
- QUANTIFICATION AND STATISTICAL ANALYSIS

SUPPLEMENTAL INFORMATION

Supplemental information can be found online at https://doi.org/10.1016/j.cub.2023.05.046.

ACKNOWLEDGMENTS

This work was supported by BBSRC grant BB/P009808/1 to D.A.F., startup funds from a Fujian Agriculture and Forestry University grant to R.M., and JSPS KAKENHI grants JP21KK0128 and JP22H05071 to Y.K.

AUTHOR CONTRIBUTIONS

D.A.F. and R.M. conceived the study; D.A.F., M.K., Y.K., and J.Y. generated the sequence data; J.Y., X.Z., S.Z., W.X., X.X., and B.K. conducted genome assembly and annotation; and D.A.F. integrated the assembly with the genetic map. M.K., J.Y., and D.A.F. analyzed the data and D.A.F. and J.Y. wrote the paper. All authors contributed to editing the manuscript.

DECLARATION OF INTERESTS

The authors declare no competing interests.

Received: January 3, 2023
Revised: April 7, 2023
Accepted: May 19, 2023
Published: June 7, 2023

REFERENCES

1. Bull, J.J. (1983). Evolution of Sex Determining Mechanisms (Benjamin/Cummings Pub. Co., Advanced Book Program).

2. Westergaard, M. (1958). The mechanism of sex determination in dioecious flowering plants. Adv. Genet. *9*, 217–281.
3. Blackburn, K.B. (1923). Sex chromosomes in plants. Nature *112*, 687–688. https://doi.org/10.1038/112687c0.
4. Charlesworth, D. (2008). Sex chromosome origins and evolution. In Evolutionary Genomics and Proteomics, M. Pagel, and A. Pomiankowski, eds. (Sinauer Associates), pp. 207–240.
5. Filatov, D.A. (2005). Evolutionary history of *Silene latifolia* sex chromosomes revealed by genetic mapping of four genes. Genetics *170*, 975–979. https://doi.org/10.1534/genetics.104.037069.
6. Wright, A.E., Dean, R., Zimmer, F., and Mank, J.E. (2016). How to make a sex chromosome. Nat. Commun. *7*, 12087, https://doi.org/10.1038/ncomms12087.
7. Kent, T.V., Uzunović, J., and Wright, S.I. (2017). Coevolution between transposable elements and recombination. Philos. Trans. R. Soc. Lond. B Biol. Sci. *372*, 20160458, https://doi.org/10.1098/rstb.2016.0458.
8. Gaut, B.S., Wright, S.I., Rizzon, C., Dvorak, J., and Anderson, L.K. (2007). Recombination: an underappreciated factor in the evolution of plant genomes. Nat. Rev. Genet. *8*, 77–84. https://doi.org/10.1038/nrg1970.
9. Charlesworth, B., and Charlesworth, D. (2000). The degeneration of Y chromosomes. Philos. Trans. R. Soc. Lond. B Biol. Sci. *355*, 1563–1572. https://doi.org/10.1098/rstb.2000.0717.
10. Charlesworth, D. (2021). The timing of genetic degeneration of sex chromosomes. Philos. Trans. R. Soc. Lond. B Biol. Sci. *376*, 20200093, https://doi.org/10.1098/rstb.2020.0093.
11. Bergero, R., and Charlesworth, D. (2009). The evolution of restricted recombination in sex chromosomes. Trends Ecol. Evol. *24*, 94–102. https://doi.org/10.1016/j.tree.2008.09.010.
12. Charlesworth, D. (2017). Evolution of recombination rates between sex chromosomes. Philos. Trans. R. Soc. Lond. B Biol. Sci. *372*, 20160456, https://doi.org/10.1098/rstb.2016.0456.
13. Charlesworth, D. (2021). When and how do sex-linked regions become sex chromosomes? Evolution *75*, 569–581. https://doi.org/10.1111/evo.14196.
14. Olito, C., and Abbott, J.K. (2023). The evolution of suppressed recombination between sex chromosomes and the lengths of evolutionary strata. Evolution *77*, 1077–1090. https://doi.org/10.1093/evolut/qpad023.
15. Charlesworth, D. (2023). Why and how do Y chromosome stop recombining? J. Evol. Biol. *36*, 632–636. https://doi.org/10.1111/jeb.14137.
16. Koyama, T., Nakamoto, M., Morishima, K., Yamashita, R., Yamashita, T., Sasaki, K., Kuruma, Y., Mizuno, N., Suzuki, M., Okada, Y., et al. (2019). A SNP in a steroidogenic enzyme is associated with phenotypic sex in seriola fishes. Curr. Biol. *29*, 1901–1909.e8. https://doi.org/10.1016/j.cub.2019.04.069.
17. Kuhl, H., Guiguen, Y., Höhne, C., Kreuz, E., Du, K., Klopp, C., Lopez-Roques, C., Yebra-Pimentel, E.S., Ciorpac, M., Gessner, J., et al. (2021). A 180 Myr-old female-specific genome region in sturgeon reveals the oldest known vertebrate sex determining system with undifferentiated sex chromosomes. Philos. Trans. R. Soc. Lond. B Biol. Sci. *376*, 20200089, https://doi.org/10.1098/rstb.2020.0089.
18. Liu, Z., Moore, P.H., Ma, H., Ackerman, C.M., Ragiba, M., Yu, Q., Pearl, H.M., Kim, M.S., Charlton, J.W., Stiles, J.I., et al. (2004). A primitive Y chromosome in papaya marks incipient sex chromosome evolution. Nature *427*, 348–352. https://doi.org/10.1038/nature02228.
19. Akagi, T., Henry, I.M., Tao, R., and Comai, L. (2014). A Y-chromosome-encoded small RNA acts as a sex determinant in persimmons. Science *346*, 646–650. https://doi.org/10.1126/science.1257225.
20. Akagi, T., Pilkington, S.M., Varkonyi-Gasic, E., Henry, I.M., Sugano, S.S., Sonoda, M., Firl, A., McNeilage, M.A., Douglas, M.J., Wang, T., et al. (2019). Two Y-chromosome-encoded genes determine sex in kiwifruit. Nat. Plants *5*, 801–809. https://doi.org/10.1038/s41477-019-0489-6.
21. Harkess, A., Huang, K., van der Hulst, R., Tissen, B., Caplan, J.L., Koppula, A., Batish, M., Meyers, B.C., and Leebens-Mack, J. (2020). Sex determination by two Y-linked genes in garden asparagus. Plant Cell *32*, 1790–1796. https://doi.org/10.1105/tpc.19.00859.
22. Gong, W., and Filatov, D.A. (2022). Evolution of the sex-determining region in *Ginkgo biloba*. Philos. Trans. R. Soc. Lond. B Biol. Sci. *377*, 20210229, https://doi.org/10.1098/rstb.2021.0229.
23. Rice, W.R. (1987). The accumulation of sexually antagonistic genes as a selective agent promoting the evolution of reduced recombination between primitive sex chromosomes. Evolution *41*, 911–914. https://doi.org/10.1111/j.1558-5646.1987.tb05864.x.
24. Jay, P., Tezenas, E., Véber, A., and Giraud, T. (2022). Sheltering of deleterious mutations explains the stepwise extension of recombination suppression on sex chromosomes and other supergenes. PLoS Biol. *20*, e3001698, https://doi.org/10.1371/journal.pbio.3001698.
25. Cavoto, E., Neuenschwander, S., Goudet, J., and Perrin, N. (2018). Sex-antagonistic genes, XY recombination and feminized Y chromosomes. J. Evol. Biol. *31*, 416–427. https://doi.org/10.1111/jeb.13235.
26. Lenormand, T., and Roze, D. (2022). Y recombination arrest and degeneration in the absence of sexual dimorphism. Science *375*, 663–666. https://doi.org/10.1126/science.abj1813.
27. Jeffries, D.L., Gerchen, J.F., Scharmann, M., and Pannell, J.R. (2021). A neutral model for the loss of recombination on sex chromosomes. Philos. Trans. R. Soc. Lond. B Biol. Sci. *376*, 20200096, https://doi.org/10.1098/rstb.2020.0096.
28. Krasovec, M., Chester, M., Ridout, K., and Filatov, D.A. (2018). The mutation rate and the age of the sex chromosomes in *Silene latifolia*. Curr. Biol. *28*, 1832–1838.e4. https://doi.org/10.1016/j.cub.2018.04.069.
29. Charlesworth, D. (2015). Plant contributions to our understanding of sex chromosome evolution. New Phytol. *208*, 52–65. https://doi.org/10.1111/nph.13497.
30. Charlesworth, D. (2016). Plant sex chromosomes. Annu. Rev. Plant Biol. *67*, 397–420. https://doi.org/10.1146/annurev-arplant-043015-111911.
31. Kejnovsky, E., and Vyskot, B. (2010). *Silene latifolia*: the classical model to study heteromorphic sex chromosomes. Cytogenet. Genome Res. *129*, 250–262. https://doi.org/10.1159/000314285.
32. Westergaard, M. (1946). Aberrant Y chromosomes and sex expression in *Melandrium album*. Hereditas *32*, 419–443. https://doi.org/10.1111/j.1601-5223.1946.tb02784.x.
33. Charlesworth, B., and Charlesworth, D. (1978). Model for evolution of dioecy and gynodioecy. Am. Nat. *112*, 975–997. https://doi.org/10.1086/283342.
34. Chibalina, M.V., and Filatov, D.A. (2011). Plant Y chromosome degeneration is retarded by haploid purifying selection. Curr. Biol. *21*, 1475–1479. https://doi.org/10.1016/j.cub.2011.07.045.
35. Bergero, R., Qiu, S., and Charlesworth, D. (2015). Gene loss from a plant sex chromosome system. Curr. Biol. *25*, 1234–1240. https://doi.org/10.1016/j.cub.2015.03.015.
36. Papadopulos, A.S.T., Chester, M., Ridout, K., and Filatov, D.A. (2015). Rapid Y degeneration and dosage compensation in plant sex chromosomes. Proc. Natl. Acad. Sci. USA *112*, 13021–13026. https://doi.org/10.1073/pnas.1508454112.
37. Zemp, N., Tavares, R., Muyle, A., Charlesworth, D., Marais, G.A.B., and Widmer, A. (2016). Evolution of sex-biased gene expression in a dioecious plant. Nat. Plants *2*, 16168, https://doi.org/10.1038/nplants.2016.168.
38. Krasovec, M., Kazama, Y., Ishii, K., Abe, T., and Filatov, D.A. (2019). Immediate dosage compensation is triggered by the deletion of Y-linked genes in *Silene latifolia*. Curr. Biol. *29*, 2214–2221.e4. https://doi.org/10.1016/j.cub.2019.05.060.
39. Muyle, A., Zemp, N., Deschamps, C., Mousset, S., Widmer, A., and Marais, G.A.B. (2012). Rapid de novo evolution of X chromosome dosage compensation in *Silene latifolia*, a plant with young sex chromosomes. PLoS Biol. *10*, e1001308, https://doi.org/10.1371/journal.pbio.1001308.

CellPress
OPEN ACCESS

40. Muyle, A., Shearn, R., and Marais, G.A. (2017). The evolution of sex chromosomes and dosage compensation in plants. Genome Biol. Evol. *9*, 627–645. https://doi.org/10.1093/gbe/evw282.

41. Krasovec, M., Nevado, B., and Filatov, D.A. (2018). A comparison of selective pressures in plant X-linked and autosomal genes. Genes *9*, 234. https://doi.org/10.3390/genes9050234.

42. Bergero, R., Forrest, A., Kamau, E., and Charlesworth, D. (2007). Evolutionary strata on the X chromosomes of the dioecious plant *Silene latifolia*: evidence from new sex-linked genes. Genetics *175*, 1945–1954. https://doi.org/10.1534/genetics.106.070110.

43. Bergero, R., Qiu, S., Forrest, A., Borthwick, H., and Charlesworth, D. (2013). Expansion of the pseudo-autosomal region and ongoing recombination suppression in the *Silene latifolia* sex chromosomes. Genetics *194*, 673–686. https://doi.org/10.1534/genetics.113.150755.

44. Qiu, S., Bergero, R., Forrest, A., Kaiser, V.B., and Charlesworth, D. (2010). Nucleotide diversity in *Silene latifolia* autosomal and sex-linked genes. Proc. Biol. Sci. *277*, 3283–3290. https://doi.org/10.1098/rspb.2010.0606.

45. Qiu, S., Bergero, R., Guirao-Rico, S., Campos, J.L., Cezard, T., Gharbi, K., and Charlesworth, D. (2016). RAD mapping reveals an evolving, polymorphic and fuzzy boundary of a plant pseudoautosomal region. Mol. Ecol. *25*, 414–430. https://doi.org/10.1111/mec.13297.

46. Bellott, D.W., Hughes, J.F., Skaletsky, H., Brown, L.G., Pyntikova, T., Cho, T.J., Koutseva, N., Zaghlul, S., Graves, T., Rock, S., et al. (2014). Mammalian Y chromosomes retain widely expressed dosage-sensitive regulators. Nature *508*, 494–499. https://doi.org/10.1038/nature13206.

47. Zhou, Q., Zhang, J., Bachtrog, D., An, N., Huang, Q., Jarvis, E.D., Gilbert, M.T.P., and Zhang, G. (2014). Complex evolutionary trajectories of sex chromosomes across bird taxa. Science *346*, 1246338. https://doi.org/10.1126/science.1246338.

48. Siroký, J., Lysák, M.A., Dolezel, J., Kejnovský, E., and Vyskot, B. (2001). Heterogeneity of rDNA distribution and genome size in *Silene* spp. Chromosome Res. *9*, 387–393.

49. Williams, A.M., Itgen, M.W., Broz, A.K., Carter, O.G., and Sloan, D.B. (2021). Long-read transcriptome and other genomic resources for the angiosperm *Silene noctiflora*. G3 (Bethesda) *11*, jkab189, https://doi.org/10.1093/g3journal/jkab189.

50. Zhang, X., Zhang, S., Zhao, Q., Ming, R., and Tang, H. (2019). Assembly of allele-aware, chromosomal-scale autopolyploid genomes based on Hi-C data. Nat. Plants *5*, 833–845. https://doi.org/10.1038/s41477-019-0487-8.

51. Filatov, D.A. (2023). Heterochiasmy and sex chromosome evolution in *Silene*. Genes *14*, 543. https://doi.org/10.3390/genes14030543.

52. Howell, E.C., Armstrong, S.J., and Filatov, D.A. (2011). Dynamic gene order on the *Silene latifolia* Y chromosome. Chromosoma *120*, 287–296. https://doi.org/10.1007/s00412-011-0311-3.

53. Haenel, Q., Laurentino, T.G., Roesti, M., and Berner, D. (2018). Meta-analysis of chromosome-scale crossover rate variation in eukaryotes and its significance to evolutionary genomics. Mol. Ecol. *27*, 2477–2497. https://doi.org/10.1111/mec.14699.

54. Brazier, T., and Glémin, S. (2022). Diversity and determinants of recombination landscapes in flowering plants. PLoS Genet. *18*, e1010141, https://doi.org/10.1371/journal.pgen.1010141.

55. Armstrong, S.J., and Filatov, D.A. (2008). A cytogenetic view of sex chromosome evolution in plants. Cytogenet. Genome Res. *120*, 241–246. https://doi.org/10.1159/000121073.

56. Lengerova, M., Moore, R.C., Grant, S.R., and Vyskot, B. (2003). The sex chromosomes of *Silene latifolia* revisited and revised. Genetics *165*, 935–938. https://doi.org/10.1093/genetics/165.2.935.

57. Kazama, Y., Sugiyama, R., Suto, Y., Uchida, W., and Kawano, S. (2006). The clustering of four subfamilies of satellite DNA at individual chromosome ends in *Silene latifolia*. Genome *49*, 520–530. https://doi.org/10.1139/g05-130.

58. Filatov, D.A. (2022). Recent expansion of the non-recombining sex-linked region on *Silene latifolia* sex chromosomes. J. Evol. Biol. *35*, 1696–1708. https://doi.org/10.1111/jeb.14063.

59. Bergero, R., and Charlesworth, D. (2011). Preservation of the Y transcriptome in a 10-million-year-old plant sex chromosome system. Curr. Biol. *21*, 1470–1474. https://doi.org/10.1016/j.cub.2011.07.032.

60. Nicolas, M., Marais, G., Hykelova, V., Janousek, B., Laporte, V., Vyskot, B., Mouchiroud, D., Negrutiu, I., Charlesworth, D., and Monéger, F. (2005). A gradual process of recombination restriction in the evolutionary history of the sex chromosomes in dioecious plants. PLoS Biol. *3*, e4. https://doi.org/10.1371/journal.pbio.0030004.

61. Filatov, D.A. (2005). Substitution rates in a new *Silene latifolia* sex-linked gene. Mol. Biol. Evol. *22*, 402–408. https://doi.org/10.1093/molbev/msi003.

62. Delichère, C., Veuskens, J., Hernould, M., Barbacar, N., Mouras, A., Negrutiu, I., and Monéger, F. (1999). *SlY1*, the first active gene cloned from a plant Y chromosome, encodes a WD-repeat protein. EMBO J *18*, 4169–4179. https://doi.org/10.1093/emboj/18.15.4169.

63. Atanassov, I., Delichère, C., Filatov, D.A., Charlesworth, D., Negrutiu, I., and Monéger, F. (2001). Analysis and evolution of two functional Y-linked loci in a plant sex chromosome system. Mol. Biol. Evol. *18*, 2162–2168.

64. Moore, R.C., Kozyreva, O., Lebel-Hardenack, S., Siroky, J., Hobza, R., Vyskot, B., and Grant, S.R. (2003). Genetic and functional analysis of DD44, a sex-linked gene from the dioecious plant *Silene latifolia*, provides clues to early events in sex chromosome evolution. Genetics *163*, 321–334. https://doi.org/10.1093/genetics/163.1.321.

65. Papadopulos, A.S.T., Helmstetter, A.J., Osborne, O.G., Comeault, A.A., Wood, D.P., Straw, E.A., Mason, L., Fay, M.F., Parker, J., Dunning, L.T., et al. (2021). Rapid parallel adaptation to anthropogenic heavy metal pollution. Mol. Biol. Evol. *38*, 3724–3736. https://doi.org/10.1093/molbev/msab141.

66. Muyle, A., Zemp, N., Fruchard, C., Cegan, R., Vrana, J., Deschamps, C., Tavares, R., Hobza, R., Picard, F., Widmer, A., and Marais, G.A.B. (2018). Genomic imprinting mediates dosage compensation in a young plant XY system. Nat. Plants *4*, 677–680. https://doi.org/10.1038/s41477-018-0221-y.

67. Lahn, B.T., and Page, D.C. (1999). Four evolutionary strata on the human X chromosome. Science *286*, 964–967. https://doi.org/10.1126/science.286.5441.964.

68. Ironside, J.E. (2010). No amicable divorce? Challenging the notion that sexual antagonism drives sex chromosome evolution. Bioessays *32*, 718–726. https://doi.org/10.1002/bies.200900124.

69. Iovene, M., Yu, Q., Ming, R., and Jiang, J. (2015). Evidence for emergence of sex-determining gene(s) in a centromeric region in *Vasconcellea parviflora*. Genetics *199*, 413–421. https://doi.org/10.1534/genetics.114.173021.

70. Pilkington, S.M., Tahir, J., Hilario, E., Gardiner, S.E., Chagné, D., Catanach, A., McCallum, J., Jesson, L., Fraser, L.G., McNeilage, M.A., et al. (2019). Genetic and cytological analyses reveal the recombination landscape of a partially differentiated plant sex chromosome in kiwifruit. BMC Plant Biol. *19*, 172. https://doi.org/10.1186/s12870-019-1766-2.

71. Rifkin, J.L., Beaudry, F.E.G., Humphries, Z., Choudhury, B.I., Barrett, S.C.H., and Wright, S.I. (2021). Widespread recombination suppression facilitates plant sex chromosome evolution. Mol. Biol. Evol. *38*, 1018–1030. https://doi.org/10.1093/molbev/msaa271.

72. Kazama, Y., Kitoh, M., Kobayashi, T., Ishii, K., Krasovec, M., Yasui, Y., Abe, T., Kawano, S., and Filatov, D.A. (2022). A *CLAVATA3*-like gene acts as a gynoecium suppression function in white campion. Mol. Biol. Evol. *39*, msac195, https://doi.org/10.1093/molbev/msac195.

73. Martin, H., Carpentier, F., Gallina, S., Godé, C., Schmitt, E., Muyle, A., Marais, G.A.B., and Touzet, P. (2019). Evolution of young sex chromosomes in two dioecious sister plant species with distinct sex determination systems. Genome Biol. Evol. *11*, 350–361. https://doi.org/10.1093/gbe/evz001.

74. Otto, S.P., and Payseur, B.A. (2019). Crossover interference: shedding light on the evolution of recombination. Annu. Rev. Genet. *53*, 19–44. https://doi.org/10.1146/annurev-genet-040119-093957.

75. Siljak-Yakovlev, S., Muratović, E., Bogunić, F., Bašić, N., Šolić, M.E., Boškailo, A., Bourge, M., Valentin, N., Stevanović, V., Garnatje, T., et al. (2020). Genome size of Balkan flora: a database (GeSDaBaF) and C-values for 51 taxa of which 46 are novel. Plant Syst. Evol. *306*, 40, https://doi.org/10.1007/s00606-020-01670-y.

76. Filatov, D.A., Howell, E.C., Groutides, C., and Armstrong, S.J. (2009). Recent spread of a retrotransposon in the *Silene latifolia* genome, apart from the Y chromosome. Genetics *181*, 811–817. https://doi.org/10.1534/genetics.108.099267.

77. Bergero, R., Gardner, J., Bader, B., Yong, L., and Charlesworth, D. (2019). Exaggerated heterochiasmy in a fish with sex-linked male coloration polymorphisms. Proc. Natl. Acad. Sci. USA *116*, 6924–6931. https://doi.org/10.1073/pnas.1818486116.

78. Rifkin, J.L., Hnatovska, S., Yuan, M., Sacchi, B.M., Choudhury, B.I., Gong, Y., Rastas, P., Barrett, S.C.H., and Wright, S.I. (2022). Recombination landscape dimorphism and sex chromosome evolution in the dioecious plant *Rumex hastatulus*. Philos. Trans. R. Soc. Lond. B Biol. Sci. *377*, 20210226, https://doi.org/10.1098/rstb.2021.0226.

79. Sardell, J.M., and Kirkpatrick, M. (2020). Sex differences in the recombination landscape. Am. Nat. *195*, 361–379. https://doi.org/10.1086/704943.

80. Kazama, Y., Sugiyama, R., Matsunaga, S., Shibata, F., Uchida, W., Hizume, M., and Kawano, S. (2003). Organization of the KpnI family of chromosomal distal-end satellite DNAs in *Silene latifolia*. J. Plant Res. *116*, 317–326. https://doi.org/10.1007/s10265-003-0106-6.

81. Cheng, H., Concepcion, G.T., Feng, X., Zhang, H., and Li, H. (2021). Haplotype-resolved de novo assembly using phased assembly graphs with hifiasm. Nat. Methods *18*, 170–175. https://doi.org/10.1038/s41592-020-01056-5.

82. Servant, N., Varoquaux, N., Lajoie, B.R., Viara, E., Chen, C.J., Vert, J.P., Heard, E., Dekker, J., and Barillot, E. (2015). HiC-Pro: an optimized and flexible pipeline for Hi-C data processing. Genome Biol. *16*, 259. https://doi.org/10.1186/s13059-015-0831-x.

83. Altschul, S.F., Gish, W., Miller, W., Myers, E.W., and Lipman, D.J. (1990). Basic local alignment search tool. J. Mol. Biol. *215*, 403–410. https://doi.org/10.1016/S0022-2836(05)80360-2.

84. Zhang, H., Song, L., Wang, X., Cheng, H., Wang, C., Meyer, C.A., Liu, T., Tang, M., Aluru, S., Yue, F., et al. (2021). Fast alignment and preprocessing of chromatin profiles with Chromap. Nat. Commun. *12*, 6566. https://doi.org/10.1038/s41467-021-26865-w.

85. Dudchenko, O., Batra, S.S., Omer, A.D., Nyquist, S.K., Hoeger, M., Durand, N.C., Shamim, M.S., Machol, I., Lander, E.S., Aiden, A.P., and Aiden, E.L. (2017). *De novo* assembly of the *Aedes aegypti* genome using Hi-C yields chromosome-length scaffolds. Science *356*, 92–95. https://doi.org/10.1126/science.aal3327.

86. Durand, N.C., Robinson, J.T., Shamim, M.S., Machol, I., Mesirov, J.P., Lander, E.S., and Aiden, E.L. (2016). Juicebox provides a visualization system for Hi-C contact maps with unlimited zoom. Cell Syst. *3*, 99–101. https://doi.org/10.1016/j.cels.2015.07.012.

87. Wolff, J., Bhardwaj, V., Nothjunge, S., Richard, G., Renschler, G., Gilsbach, R., Manke, T., Backofen, R., Ramírez, F., and Grüning, B.A. (2018). Galaxy HiCExplorer: a web server for reproducible Hi-C data analysis, quality control and visualization. Nucleic Acids Res. *46*, W11–W16. https://doi.org/10.1093/nar/gky504.

88. Simão, F.A., Waterhouse, R.M., Ioannidis, P., Kriventseva, E.V., and Zdobnov, E.M. (2015). BUSCO: assessing genome assembly and annotation completeness with single-copy orthologs. Bioinformatics *31*, 3210–3212. https://doi.org/10.1093/bioinformatics/btv351.

89. Langmead, B., Trapnell, C., Pop, M., and Salzberg, S.L. (2009). Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. Genome Biol. *10*, R25. https://doi.org/10.1186/gb-2009-10-3-r25.

90. Yang, X., and Li, L. (2011). miRDeep-P: a computational tool for analyzing the microRNA transcriptome in plants. Bioinformatics *27*, 2614–2615. https://doi.org/10.1093/bioinformatics/btr430.

91. Ou, S., Su, W., Liao, Y., Chougule, K., Agda, J.R.A., Hellinga, A.J., Lugo, C.S.B., Elliott, T.A., Ware, D., Peterson, T., et al. (2019). Benchmarking transposable element annotation methods for creation of a streamlined, comprehensive pipeline. Genome Biol. *20*, 275. https://doi.org/10.1186/s13059-019-1905-y.

92. Li, H., and Durbin, R. (2010). Fast and accurate long-read alignment with Burrows-Wheeler transform. Bioinformatics *26*, 589–595. https://doi.org/10.1093/bioinformatics/btp698.

93. Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G., and Durbin, R.; 1000 Genome Project Data Processing Subgroup (2009). The sequence alignment/map format and SAMtools. Bioinformatics *25*, 2078–2079. https://doi.org/10.1093/bioinformatics/btp352.

94. Li, B., and Dewey, C.N. (2011). RSEM: accurate transcript quantification from RNA-seq data with or without a reference genome. BMC Bioinf. *12*, 323. https://doi.org/10.1186/1471-2105-12-323.

95. Yang, Z. (2007). PAML 4: phylogenetic analysis by maximum likelihood. Mol. Biol. Evol. *24*, 1586–1591. https://doi.org/10.1093/molbev/msm088.

96. Yang, Z., and Nielsen, R. (1998). Synonymous and nonsynonymous rate variation in nuclear genes of mammals. J. Mol. Evol. *46*, 409–418. https://doi.org/10.1007/pl00006320.

97. Xie, T., Zheng, J.F., Liu, S., Peng, C., Zhou, Y.M., Yang, Q.Y., and Zhang, H.Y. (2015). *De novo* plant genome assembly based on chromatin interactions: a case study of *Arabidopsis thaliana*. Mol. Plant *8*, 489–492. https://doi.org/10.1016/j.molp.2014.12.015.

98. Yue, J., VanBuren, R., Liu, J., Fang, J., Zhang, X., Liao, Z., Wai, C.M., Xu, X., Chen, S., Zhang, S., et al. (2022). SunUp and Sunset genomes revealed impact of particle bombardment mediated transformation and domestication history in papaya. Nat. Genet. *54*, 715–724. https://doi.org/10.1038/s41588-022-01068-1.

99. Hu, X.S., and Filatov, D.A. (2016). The large-X effect in plants: increased species divergence and reduced gene flow on the *Silene* X-chromosome. Mol. Ecol. *25*, 2609–2619. https://doi.org/10.1111/mec.13427.

100. Edgar, R.C. (2004). MUSCLE: a multiple sequence alignment method with reduced time and space complexity. BMC Bioinformatics *5*, 113. https://doi.org/10.1186/1471-2105-5-113.

CellPress
OPEN ACCESS

STAR★METHODS

KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Deposited data		
Silene latifolia female genome assembly	This study	GWH: PRJCA014197
Silene latifolia PacBio sequences	This study	SRA: PRJNA952727
Silene latifolia Hi-C data	This study	SRA: PRJNA952727
Experimental models: Organisms/strains		
Silene latifolia plants	Kazama et al.[80]	Inbred K-line
Software and algorithms		
CCS v6.4	N/A	https://github.com/PacificBiosciences/ccs
Hifiasm	Cheng et al.[81]	https://github.com/chhylp123/hifiasm
ALLHiC v0.9.13	Zhang et al.[50]	https://github.com/tangerzhang/ALLHiC/wiki
HiC-Pro	Servant et al.[82]	https://github.com/nservant/HiC-Pro
BLASTN	Altschul et al.[83]	https://blast.ncbi.nlm.nih.gov/Blast.cgi
Chromap v0.2.4	Zhang et al.[84]	https://github.com/haowenz/chromap
3D-DNA v180922	Dudchenko et al.[85]	https://github.com/aidenlab/3d-dna
Juicebox Assembly Tools	Durand et al.[86]	https://github.com/aidenlab/Juicebox
HiC Explorer v3.7.2	Wolff et al.[87]	https://hicexplorer.readthedocs.io/en/latest/
BUSCO	Simao et al.[88]	https://busco.ezlab.org/
GETA v1.0	N/A	https://github.com/chenlianfu/geta
GetaFilter.pl	N/A	https://github.com/LengFeng00/biotree.club.git
IGV-GSAman	N/A	https://gitee.com/CJchen/IGV-sRNA
Bowtie	Langmead et al.[89]	https://bowtie-bio.sourceforge.net/manual.shtml
miRDeep-P	Yang et al.[90]	https://github.com/rajewsky-lab/mirdeep2
EDTA v1.9.6	Ou et al.[91]	https://github.com/topics/edta
BWA-MEM	Li and Durbin.[92]	https://github.com/lh3/bwa
Samtools v1.7	Li et al.[93]	https://github.com/samtools/samtools
RSEM	Li et al.[94]	https://github.com/deweylab/RSEM
PAML	Yang et al.[95]	http://abacus.gene.ucl.ac.uk/software/paml.html
branch model	Yang et al.[96]	https://github.com/Bumblebee-Project/Bumblebee/wiki/Branching-Model

RESOURCE AVAILABILITY

Lead contact

Further information and requests for resources and reagents should be directed to and will be fulfilled by the lead contact, Dmitry A. Filatov (dmitry.filatov@biology.ox.ac.uk)

Materials availability

This study did not generate new unique reagents.

Data and code availability

The *S. latifolia* female genome assembly and annotation have been deposited to Genome Warehouse (GWH) database in BIG data Center (https://ngdc.cncb.ac.cn/gwh/) under accession number GWHCBIJ00000000 and BioProject accession PRJCA014197. The PacBio and Hi-C data have been uploaded to NCBI Sequence Read Archive (SRA) database as Bioproject PRJNA952727.

EXPERIMENTAL MODEL AND SUBJECT DETAILS

Plant material

S. latifolia seedlings for the highly inbred K-line[80] were grown in the greenhouse of the Centre for Genomics and Biotechnology at Fujian Agriculture and Forestry University (FAFU) at 22°C under long-day photoperiod (16h of artificial light).

METHOD DETAILS

Genome sequencing and assembly

PacBio library construction and sequencing of the female genome

DNA for sequencing was extracted from young leaves of female *S. latifolia* plants. Genomic DNA was sheared and size-selection was carried out using BluePippin system. SMRTbell libraries were constructed according to the protocol from PacBio. Subsequently, 104.9 Gb of sequence data were generated with the PacBio Sequel IIe System. PacBio high-quality HiFi reads was generated from subreads by CCS software (v6.4, https://github.com/PacificBiosciences/ccs). HiFi reads were assembled using hifiasm software[81] with default parameters, which yielded a 2.64 Gb long contig-level genome assembly (Table S2).

Hi-C library construction and sequencing

For Hi-C scaffolding, young leaves of *S. latifolia* female plants from the inbred K-line[80] were used to construct the Hi-C libraries by ANOROAD company according to previously published methods.[97] The Illumina HiSeq X Ten platform was used to sequence the chimeric fragments representing the original cross-linked fragments after constructing the paired-end sequencing libraries. Hi-C reads were uniquely mapped to the contig assemblies and reads within 500 bp regions of HindIII restriction sites were retained for further analysis. A total of 134.8 Gb raw sequencing Hi-C reads were used to assemble the chromosome-level genome using ALLHiC (v0.9.8, https://github.com/tangerzhang/ALLHiC). Hi-C sequencing data were assessed using HiC-Pro[82] and the results showed a high proportion of validated reads (59.2%) (Table S4).

Chromosome assembly by integration with Hi-C scaffolding and genetic map

The chromosome level genome assembly was integrated with the previously published RNA-seq-based high-density genetic maps,[36,51] but mostly the sex-specific maps from the recent study[51] were used in the analyses as they contained nearly three times more genes than the older sex-average map.[36] We aligned the sequences of the markers to the *S. latifolia* female contig genome using BLASTN,[83] retaining BLAST hits with >97% identity longer than 100 bp. As the maps were constructed with transcriptome sequencing,[36,51] the markers represent actual expressed genes rather than often repetitive non-coding regions, which facilitated finding of these markers in the genome sequence. Due to presence of introns in the genomic sequence of the genetically mapped genes, each gene usually had several adjacent blast hits, corresponding to different exons. The lowest position of the blast hit for each gene was regarded as the genomic position of the particular marker. The correspondence of the positions of markers in the genetic map and genome sequence was checked manually. The contigs were divided into different groups according to the markers of RNA-seq based map, and the group information was combined with the ALLHiC pipeline.[50] Finally, the contigs were linked into 12 pseudo-chromosomes in the *S. latifolia* female genome.

The integration of the chromosome level assembly with the genetic map revealed minor inconsistencies, which were corrected as follows. The Hi-C reads were aligned to the contigs using the Chromap software (v0.2.4).[84] Subsequently the run-assembly-visualizer.sh script from 3D-DNA pipeline (v180922)[85] was used to obtain input files for Juicebox Assembly Tools (v2.20.00).[86] Then the Juicebox Assembly Tools was used to get the corrected chromosomes.

Assessment of accuracy and completeness of the chromosome assembly

The chromatin contact matrix was used to assess the accuracy of chromosome assemblies based on Hi-C by using HiC Explorer (v3.7.2).[87] The accuracy and completeness of the assembly was assessed by BUSCO,[88] and mapping of RNA-seq and genomic Illumina paired end reads. The single base accuracy and the genome coverage rate by the short reads indicated the accuracy of our *S. latifolia* genome assembly (Tables S5 and S6).

X-chromosome identification

The X chromosome was identified by the presence of X-linked genes from the genetic maps[36,51] as well as the X-linked genes identified previously, such as *SlX1*,[62] *DD44X*,[64] *SlX3*,[60] *SlX4*,[63] *SlssX*,[61] *SlCypX*[42] and the 'E-genes' from Bergero et al.[35,43] (E777, E757, E780, E799, E330 etc).

Genome annotation with RNA-seq data

RNA-seq data from the previous study[36] were used for *S. latifolia* female genome annotation. Protein-coding genes were annotated based on a previously published approach.[98] Briefly, RNA-seq transcripts together with homologous proteins were inputted into GETA pipeline (v.1.0, https://github.com/chenlianfu/geta). All parameters were set to the original defaults and the false-positive gene models were filtered by set Pfam database path.

Genome filtering and correction

To improve the gene annotation of *S. latifolia* female genome, firstly, the perl script GetaFilter.pl (https://github.com/LengFeng00/biotree.club.git) was used to filter genes that did not meet the standard. The genes were required to meet at least one of the following

conditions: FPKM>3, blastp hits with >30% identity and >100 alignment length, containing at least one Pfam domain and Augustus_-transcript Support_percentage > 50. Secondly, the IGV-GSAman software (https://gitee.com/CJchen/IGV-sRNA) was used to manually correct and filter out the genes that included TE sequences in coding regions. Finally, the completeness of genome annotation was evaluated by BUSCO.

miRNA annotation

For miRNA annotation, the plant miRNAs downloaded from the publicly available database miRBase (most recent access 5 July 2018, http://www.mirbase.org/) were aligned to *S. latifolia* female genome using bowtie[89] as described previously.[98] The mapping results were filtered with PERL script filter_alignments.pl from the miRDP1.3 package.[90] The miRDeep-P program,[90] a plant-seecific scoring system explicitly anticipating miRNAs for plants, was used to trim, filter and identification of new miRNAs.

Transposable element (TE) annotation

The EDTA v1.9.6 *de novo* annotation tool[91] was used to annotate LTR, terminal inverted repeat, and Helitron elements. For EDTA, the following parameters were used in addition to defaults: –step all, –species others, –sensitive 0, –anno 1, and –threads 4. Sequences with multiple paralogs were mapped back to the genome and manually extended to determine the full-length boundary of each TE. A total of 15,560 full-length, representative Copia and Gypsy copies were successfully annotated. The significance of difference in TE abundance between autosomes and the X chromosome was tested with R package ggpubr (version 0.4.0). Significance was tested using the two-sided Mann-Whitney-Wilcoxon test with multiple comparison. P.adjust homl method was used to adjust the P values.

Reconstruction of Y-linked genes

To reconstruct Y-linked gametologs for the X-linked genes we followed segregation-based approach described previously.[28,34,36] In particular, using previously published sequence data from genetic crosses, we called single nucleotide polymorphisms (SNPs) for parents and progeny, including 10 F1 progeny with genome sequence data available[28] and 50 F2 progeny with transcriptome data available.[36] Trimmed RNAseq reads were mapped against the reference genome with BWA mem v0.7.17[92] and sorted with Samtools v1.7.[93] Then, SNP calling was done with Samtools mpileup (options: -d 1000 -q 20 -Q 20) and sites filtered with bcftools filter 1.7.

The analysis of SNP segregation in these genetic crosses allowed us to identify Y-linked SNPs that are always inherited from father to sons and never to daughters. The sequence reads containing the Y-SNPs along with their paired reads were separated and assembled into contigs. For this Y-reconstruction procedure we focused on coding regions, while (often repetitive) non-coding regions were excluded from analysis. As described previously,[28,34,36] this approach allows accurate reconstruction of Y-linked homologs of X-linked genes in *S. latifolia*. The accuracy of the Y-reconstruction was confirmed using the sequences of previously published Y-linked genes (*SlY1*,[62] *DD44Y*,[64] *SlY3*,[60] *SlY4*,[63] *SlssY*[61] and *SlY7*[42]).

Analysis of gene expression

For gene expression analysis we used previously published RNA-seq data from Chibalina and Filatov,[34] Papadopulos et al.,[36] and Hu and Filatov.[99] The analysis of gene expression was conducted with RSEM package[94] with default options.

Phylogenetic analyses

To analyse substitution rates in sex-linked genes we created three-sequence alignments including X- and Y-linked gametologs as well as a sequence of homologous gene from an outgroup species. As an outgroup we used non-dioecious *Silene uniflora* for which a fragmented genome assembly is already available.[65] We used CDS sequences of X-linked genes to blast-search the *S. uniflora* genomes. The best-matching homologs were aligned with the X- and Y-linked genes using muscle.[100] These three-sequence alignments were used for analysis of synonymous and non-synonymous substitution rates in X- and Y-linked genes with codeml from PAML package.[95] Substitution rates were estimated using branch model[96] allowing for separate substitution rate for each branch of the phylogeny.

QUANTIFICATION AND STATISTICAL ANALYSIS

The Wilcoxon tests were performed using R ggpubr package (version 0.4.0) and the P values were adjusted with p.adjust homl correction in R. Paired t-tests were done in Excel.