RESEARCH ARTICLE

EVOLUTION

The *Silene latifolia* genome and its giant Y chromosome

Carol Moraga[1,2,3]†, Catarina Branco[1,4,5,6]†, Quentin Rougemont[7]†, Pavel Jedlička[8], Eddy Mendoza-Galindo[9], Paris Veltsos[10], Melissa Hanique[11], Ricardo C. Rodríguez de la Vega[7], Eric Tannier[1,12], Xiaodong Liu[13], Claire Lemaitre[14], Peter D. Fields[15], Corinne Cruaud[16], Karine Labadie[16], Caroline Belser[16], Jerome Briolay[17], Sylvain Santoni[18], Radim Cegan[8], Raquel Linheiro[1,4,5,6], Gabriele Adam[11], Adil El Filali[1], Vinciane Mossion[19], Adnane Boualem[11], Raquel Tavares[4,5,6], Amine Chebbi[20], Richard Cordaux[21], Cécile Fruchard[1], Djivan Prentout[22], Amandine Velt[23], Bruno Spataro[1], Stephane Delmotte[1], Laura Weingartner[24], Helena Toegelová[25], Zuzana Tulpová[25], Petr Cápal[25], Hana Šimková[25], Helena Štorchová[26], Manuela Krüger[26], Oushadee A. J. Abeyawardana[26], Douglas R. Taylor[27], Matthew S. Olson[28], Daniel B. Sloan[15], Sophie Karrenberg[29], Lynda F. Delph[30], Deborah Charlesworth[31], Aline Muyle[1,9], Tatiana Giraud[7], Abdelhafid Bendahmane[11], Alex Di Genova[2,3,32], Mohammed-Amin Madoui[15,33], Roman Hobza[8], Gabriel A. B. Marais[1,4,5,6,34]*

In many species with sex chromosomes, the Y is a tiny chromosome. However, the dioecious plant *Silene latifolia* has a giant ~550-megabase Y chromosome, which has remained unsequenced so far. We used a long- and short-read hybrid approach to obtain a high-quality male genome. Comparative analysis of the sex chromosomes with their homologs in outgroups showed that the Y is highly rearranged and degenerated. Recombination suppression between X and Y extended in several steps and triggered a massive accumulation of repeats on the Y as well as in the nonrecombining pericentromeric region of the X, leading to giant sex chromosomes. Using sex phenotype mutants, we identified candidate sex-determining genes on the Y in locations consistent with their favoring recombination suppression events 11 and 5 million years ago.

Among the multiple paths that the evolution of sex chromosomes can take, some have led to giant Y chromosomes (*1*, *2*). Giant Y chromosomes may result from massive accumulation of repeats, including transposable elements (TEs), but their structure, precise role in sex determination, and evolution remain poorly understood (*3*–*6*). Giant Y chromosomes were first identified in plant species with separate sexes (dioecious plants) (*3*) and also exist in animals (*4*). In the past decade, great advances have been made in studying sex chromosomes using genomics and bioinformatics (*5*, *6*), but no giant plant Y chromosome has yet been assembled.

Silene latifolia (Caryophyllaceae) is a dioecious plant described in the 18th century and studied by many, including Darwin (*7*). Its XY sex-determination system was discovered 100 years ago (*8*). The Y is ~550 Mb, the X is ~400 Mb, and the total haploid genome size is ~2.7 Gb (*9*). Genetic maps show that the X and the Y are largely nonrecombining and share only a single pseudoautosomal region (PAR) (*10*). Recombination has been suppressed progressively, forming groups of X-Y gene pairs with differing synonymous divergence levels, called evolutionary strata (*11*–*13*). The repeat richness (*14*) and size of the *S. latifolia* Y have, however, prevented its assembly so far. Mutants with deletions on the Y chromosome and altered sex phenotypes indicate the presence of three sex-determining regions (*15*, *16*). A candidate gene (*Clavata3*) in one of the regions involved in female sterility has recently been proposed (*17*, *18*), but the other sex-determining genes remain unknown.

To study the repeat-rich Y chromosome, we used Oxford Nanopore Technologies (ONT) sequencing and generated a new marker-dense genetic map to obtain a chromosome-scale *S. latifolia* genome assembly. We also used high-quality genome assemblies of closely related nondioecious *Silene* species as outgroups to make inferences about the evolution of the *S. latifolia* sex chromosomes. We compared the epigenetics (DNA methylation and small RNAs) of the X and the Y. To identify individual candidate sex-determining genes, we sequenced mutants with Y deletions for three sex phenotypes [hermaphrodites and asexuals with early or intermediate or with late abnormalities in anther development, described in (*16*)], and generated expression data at two critical stages in male and female flower development.

The structure and gene content of the sex chromosomes

To assemble the complex *S. latifolia* genome, we used the sequencing, assembly and annotation strategy detailed in the supplementary text, section S1; Table 1; tables S1 to S4; and figs. S1 to S5 (*19*). Our assemblies of the X and the Y chromosomes are high quality. In systems in which the X and Y are differentiated by many single-nucleotide polymorphisms (SNPs) and insertions-deletions (indels), female/male read-depth ratios using stringent mapping parameters are expected to be ~1, ~2, and ~0 for

[1]Laboratoire Biométrie et Biologie Evolutive (LBBE), CNRS/Université Claude Bernard Lyon 1, Villeurbanne, France. [2]Instituto de Ciencias de la Ingeniería, Universidad de O'Higgins, Rancagua, Chile. [3]Centro UOH de Bioingenieria (CUBI), Universidad de O'Higgins, Rancagua, Chile. [4]CIBIO, Centro de Investigação em Biodiversidade e Recursos Genéticos, InBIO, Laboratório Associado, Campus de Vairão, Universidade do Porto, Vairão, Portugal. [5]Departamento de Biologia, Faculdade de Ciências, Universidade do Porto, Porto, Portugal. [6]BIOPOLIS Program in Genomics, Biodiversity and Land Planning, CIBIO, Campus de Vairão, Vairão, Portugal. [7]Université Paris-Saclay, CNRS, AgroParisTech, Laboratoire Ecologie Systématique et Evolution, UMR 8079, Bâtiment 680, Gif-sur-Yvette, France. [8]Department of Plant Developmental Genetics, Institute of Biophysics of the Czech Academy of Sciences, Brno, Czech Republic. [9]Centre d'Ecologie Fonctionnelle et Evolutive (CEFE), University of Montpellier, CNRS, EPHE, IRD, Montpellier, France. [10]Ecology, Evolution and Genetics Research Group, Biology Department, Vrije Universiteit Brussel, Brussels, Belgium. [11]Institute of Plant Sciences Paris-Saclay (IPS2), Université Paris-Saclay, CNRS, INRAE, Université d'Évry, Gif-sur-Yvette, France. [12]Inria Lyon Research Center, Villeurbanne, France. [13]Department of Biology, University of Copenhagen, Copenhagen, Denmark. [14]Institut de Recherche en Informatique et Systèmes Aléatoires (IRISA), Université de Rennes, Inria, CNRS, Rennes, France. [15]Department of Biology, Colorado State University, Fort Collins, CO, USA. [16]Genoscope, Institut François Jacob, CEA, CNRS, Université d'Évry, Université Paris-Saclay, Évry, France. [17]Développement de Techniques et Analyse Moléculaire de la Biodiversité (DTAMB), Université Claude Bernard Lyon 1, Campus de la Doua, Villeurbanne, France. [18]Genomic Platform, Amélioration Génétique et Adaptation des Plantes Méditerranéennes et Tropicales (AGAP), Université de Montpellier, CIRAD, INRAE, Montpellier, France. [19]Department of Ecology and Genetics, Division of Plant Ecology and Evolution, Uppsala University, Uppsala, Sweden. [20]Efor, Grosspeter Tower (Spaces), Basel, Switzerland. [21]Évolution Génomes Comportement Écologie, Université Paris-Saclay, CNRS, IRD, Gif-sur-Yvette, France. [22]Department of Biological Sciences, Columbia University, New York, NY, USA. [23]Santé de la Vigne et Qualité du Vin (SVQV), INRAE, Colmar, France. [24]University of Louisville School of Medicine, Undergraduate Medical Education, Louisville, KY, USA. [25]Institute of Experimental Botany of the Czech Academy of Sciences, Olomouc, Czech Republic. [26]Plant Reproduction Laboratory, Institute of Experimental Botany, Czech Academy of Sciences, Prague, Czech Republic. [27]Department of Biology, University of Virginia, Charlottesville, VA, USA. [28]Department of Biological Sciences, Texas Tech University, Lubbock, TX, USA. [29]Department of Ecology and Genetics, Division of Plant Ecology and Evolution, Uppsala University, Uppsala, Sweden. [30]Department of Biology, Indiana University, Bloomington, IN, USA. [31]Institute of Ecology and Evolution, School of Biological Sciences, University of Edinburgh, Ashworth Laboratories, Edinburgh, UK. [32]Center for Mathematical Modeling, UMI-CNRS 2807, Santiago, Chile. [33]Service d'Etude des Prions et des Infections Atypiques (SEPIA), Institut François Jacob, Commissariat à l'Energie Atomique et aux Energies Alternatives (CEA), Université Paris Saclay, Fontenay-aux-Roses, France. [34]GreenUPorto–Sustainable Agrifood Production Research Centre, Departamento de Biologia, Faculdade de Ciências, Universidade do Porto, Vairão, Portugal.
*Corresponding author. Email: gabriel.marais@cnrs.fr
†These authors contributed equally to this work.

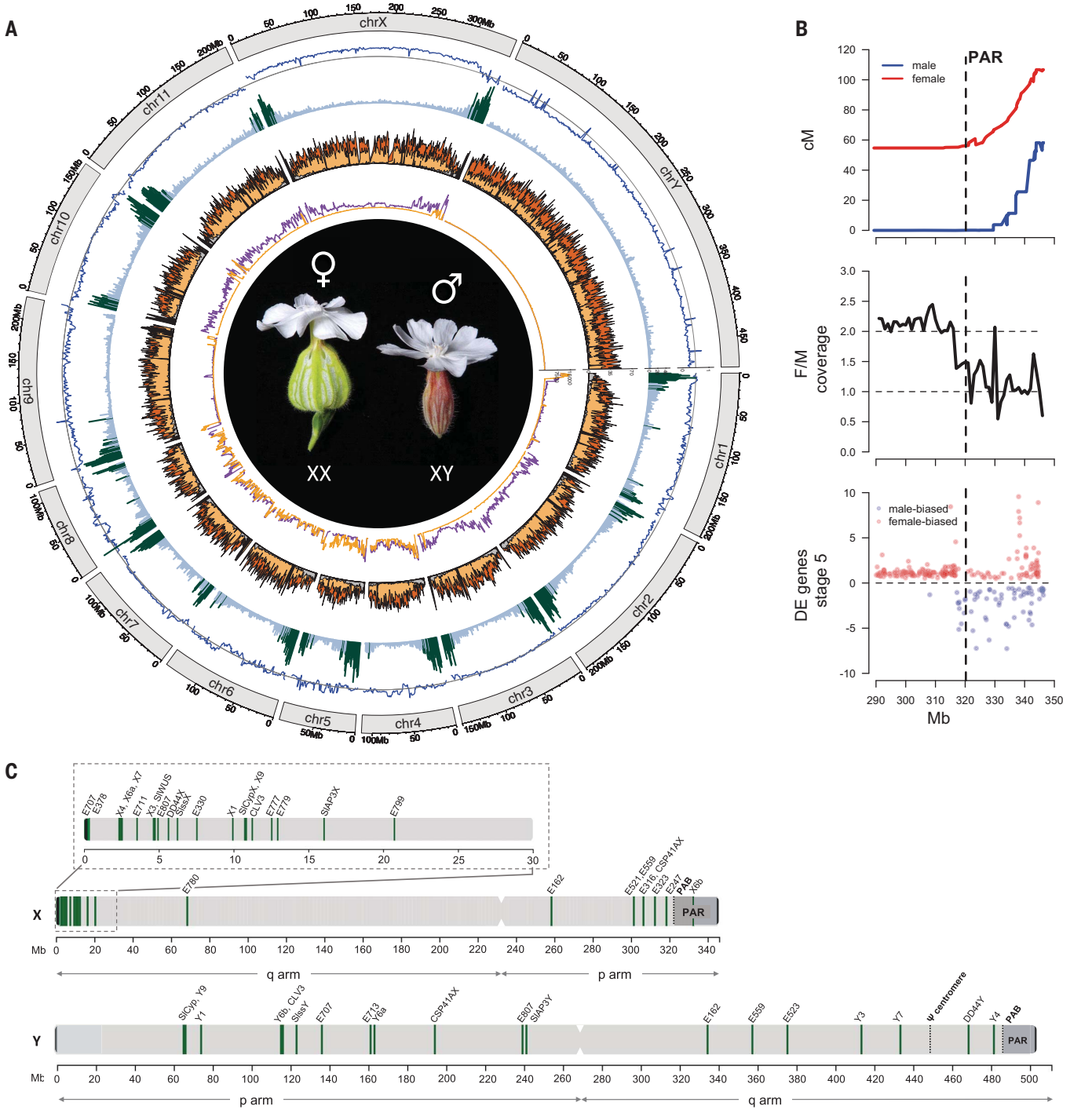

Fig. 1. Assembly of the *S. latifolia* male genome. (**A**) Circos plot of the *S. latifolia* male genome. Circles from the outside to the inside correspond to the following: (i) Coverage ratio (female/male). (ii) Gene density, where regions with high densities (exceeding average density plus 1 standard deviation) are highlighted in dark green. (iii) Density of repeat families; subtelomere- and centromere-associated satellite elements (black) and LTR elements—Ty3/Gypsy (orange), Ty1/Copia (yellow), LINE (violet), and Helitron (gray). (iv) SNP density in the male (orange) and female (purple), consistent with the sequenced male being highly homozygous (although some chromosomes show heterozygosity) and with male and female being full siblings. chr, chromosome. (**B**) Zoom-in on the X chromosome, showing recombination in males (blue) and females (red) defining the pseudoautosomal boundary at position 321 Mb on the X chromosome (top). The boundary was confirmed by other features, such as the female/male (F/M) sequence coverage ratio (middle) and a change in significant differential expression (DE) between male and female flowers (stage 5) (bottom); the full analysis of differential gene expression is shown in fig. S18. All panels have data summarized in 1-Mb windows. (**C**) Structure of the sex chromosomes, p and q arms, centromeres, and the PAR and pseudoautosomal boundary (PAB) are depicted. The sex-linked genes characterized in previous work are also shown. A zoom-in on the q arm of the X is showing a number of those genes. [Photo credits: Paris Veltsos and Lynda Delph]

the autosomes, X chromosome, and Y chromosome, respectively, as observed in our data (Fig. 1A). A smaller set of experimentally validated sex-linked genes compiled previously (*20*, *21*) also mapped as expected to their previously assigned X or Y positions (Fig. 1C). The X chromosome sequence obtained is 346 Mb, with a distribution of genes and repeats (in particular centromere- and subtelomere-associated repeats) as expected for a metacentric chromosome (Fig. 1A). Sex-specific recombination data, the female/male read coverage ratio, and differential expression identified the PAR (Fig. 1, B and C). This PAR is a small, gene-rich region of 25 Mb with 1286 genes (51 genes per megabase versus the X-chromosomal mean of 10 genes per megabase). Our X assembly is very similar to that of a recently published *S. latifolia* female genome (*13*) (fig. S6). The Y chromosome assembly is 485 Mb long, excluding the PAR. Y-specific centromeric repeats locate the centromere and a remnant of a former centromere [called a pseudocentromere (*22*)] at the expected locations (Fig. 1A). The Y assembly roughly agrees with the deletion-based map of the Y chromosome (*16*) given the uncertainties in the map (fig. S3).

Evolution of the sex chromosomes

The high-quality assembly of the X and Y chromosomes illuminates the evolution of recombination suppression between the sex chromosomes of *S. latifolia*. We detected three evolutionary strata and extensive rearrangements on the sex chromosomes, especially the Y. To analyze rearrangements and estimate synonymous site divergence (d_S) between the X and Y copies, we used gametologs (X-Y gene pairs) that also have 1:1 orthologs in the two nondioecious outgroup species, *Silene conica* and *Silene vulgaris*. A change-point analysis of these X-Y d_S values of 401 gametolog pairs, based on the gene rank in the X chromosome assembly, divided the nonrecombining region into four adjoining X chromosome sectors with different means (Fig. 2, A to C, and fig. S7A) and defined three evolutionary strata—S1, S2 and S3—based on the different d_S levels (Fig. 2, A to C). A larger set of 598 gametologs (X-Y gene pairs without requiring orthologs in the outgroups) gave similar results (fig. S7B). Strata S1 and S2 are in good agreement with previously defined strata (fig. S7, C and D); S3 has not been detected before. It splits the oldest stratum, S1, in two, owing to a lower mean X-Y d_S compared with its two flanking regions, S1a and S1bc (Fig. 2C and fig. S7D; S1b and S1c are associated with two different inversions). Strata S3 and S2 have similar mean X-Y d_S, but S3 has much higher X-Y synteny compared with S2, and the two strata were formed differently. Strata S1 and S3 are small regions, both located within the first 27 Mb of the X chromosome q arm. Stratum S2, on the other hand, is very large and includes most of the X chromosome, including the pericentromeric regions on both arms. Using a molecular clock approach, we estimate that strata S2 and S3 both evolved most recently—5.4 [95% confidence interval (4.4, 6.5)] and 4.4 (3.3, 5.5) million years ago (Ma), respectively—whereas stratum S1 arose 11.8 (10.6, 13.1) Ma, which is inferred to be when the *S. latifolia* sex chromosomes originated (*23*).

Gene order comparisons between *S. latifolia* and two nondioecious close relatives, *S. conica* and *S. vulgaris* used as outgroups, revealed large syntenic blocks with some rearrangements (fig. S8). The *S. latifolia* Y chromosome is most rearranged compared with the X or the homologous chromosomal blocks in either outgroup, with the notable exception of the S3 stratum, which includes 4 Mb of X-Y synteny (Fig. 2A and figs. S8 and S9). The *S. latifolia* X shows homology with chromosome 5 of *S. conica*, and smaller parts of chromosomes 1, 2, and 6, and with four *S. vulgaris* scaffolds (1, 3, 6, and 16) (Fig. 2A and fig. S8). Reconstruction of the rearrangements between the X, the Y, and the homologous chromosomal blocks in the outgroups (figs. S9 and S10) suggests that stratum S1 may have evolved by two inversions early in the evolution of the sex chromosomes, one on the X encompassing S1a to S1b and one on the Y including S1c.

Stratum S3 is the only region of extended synteny between the X and Y chromosomes. Together with its slightly lower d_S mean compared with that of the other strata, this suggests that S3 is the most recent stratum that we could detect (Fig. 2, fig. S7D, and fig. S9). Comparisons between *S. latifolia* X and the homologous chromosomal blocks in the outgroups suggest that S3 genes were ancestrally located among S1 genes (fig. S9). We therefore propose that S3 (initially within the X region that evolved to become the S1 stratum) was lost from the Y and later regained by a recent duplicative translocation from the X, resetting the X-Y d_S to zero about 4.4 Ma. Stratum S2 is more rearranged than S3, consistent with being older than S3, and probably arose through a different mechanism. Reconstruction of the rearrangements between the X, the Y, and the outgroups (fig. S11) indicates that S2 could not have formed by a single rearrangement. We found several inversions, some of them pericentric, as previously suggested (*16*, *22*). Many may have occurred after recombination stopped, possibly mediated by the high abundance of repeats.

Our repeat analysis revealed very high TE densities—mainly the long terminal repeat (LTR) retrotransposons *Copia* and *Gypsy*—on the Y but also, to a lesser extent, on the X compared with the autosomes (Table 1, table S4, and Fig. 1). The two *S. latifolia* sex chromosomes are 4 to 5.5 times as large as the *S. conica* chromosome 5, whereas there is only a twofold increase for the *S. latifolia* autosomes compared with their *S. conica* homologs. In eukaryotes, the

Table 1. Statistics for the male genome and sex chromosomes of *S. latifolia*. All metrics were calculated for the total genome assembly as well as specifically for the sex chromosomes. Contigs were obtained using N1 (a single N opens a gap). MITE, miniature inverted-repeat transposable element; LINE, long interspersed nuclear element.

Genome metrics	Whole genome	X chromosome	Y chromosome
Total assembled size (bp)	2,716,527,704	346,484,273	486,334,681
Number of contigs	1545	59	37
N50 (bp)	18,434,108	29,020,133	36,745,246
N90 (bp)	6,521,844	11,745,531	27,237,198
Number of scaffolds	912	21	11
N50 (bp)	200,709,446	50,520,191	59,479,783
N90 (bp)	141,278,959	21,428,974	44,285,091
Largest length size (bp)	237,716,014	133,616,578	135,552,442
Gaps (%)	1.44	6.56	0.59
Anchored of the total sequences (anchoring rate)	2,575,517,323 (94.8%)		
Annotated protein-coding genes	35,436	3520	2301
Mean gene length (bp)	4155	4332	4045
BUSCO score of annotated protein-coding genes from all scaffolds*	C: 92.5% [S: 78.1%, D: 14.4%], F: 0.7%, M: 6.8%, *n*: 425		
Identified repeats	79.20%	77.75%	81.24%
Annotated repeats	61.18%	58%	65%
LTR retrotransposons	54.42%	52.7%	60.55%
DNA transposons	4.11%	3.11%	3.22%
MITE	0.14%	0.098%	0.068%
LINE	0.48%	0.37%	0.55%
Satellite repeats	2.04%	2.014%	1.18%

*C, complete; S, single copy; D, duplicated; F, fragmented; M, missing; *n*, total number of BUSCO genes.

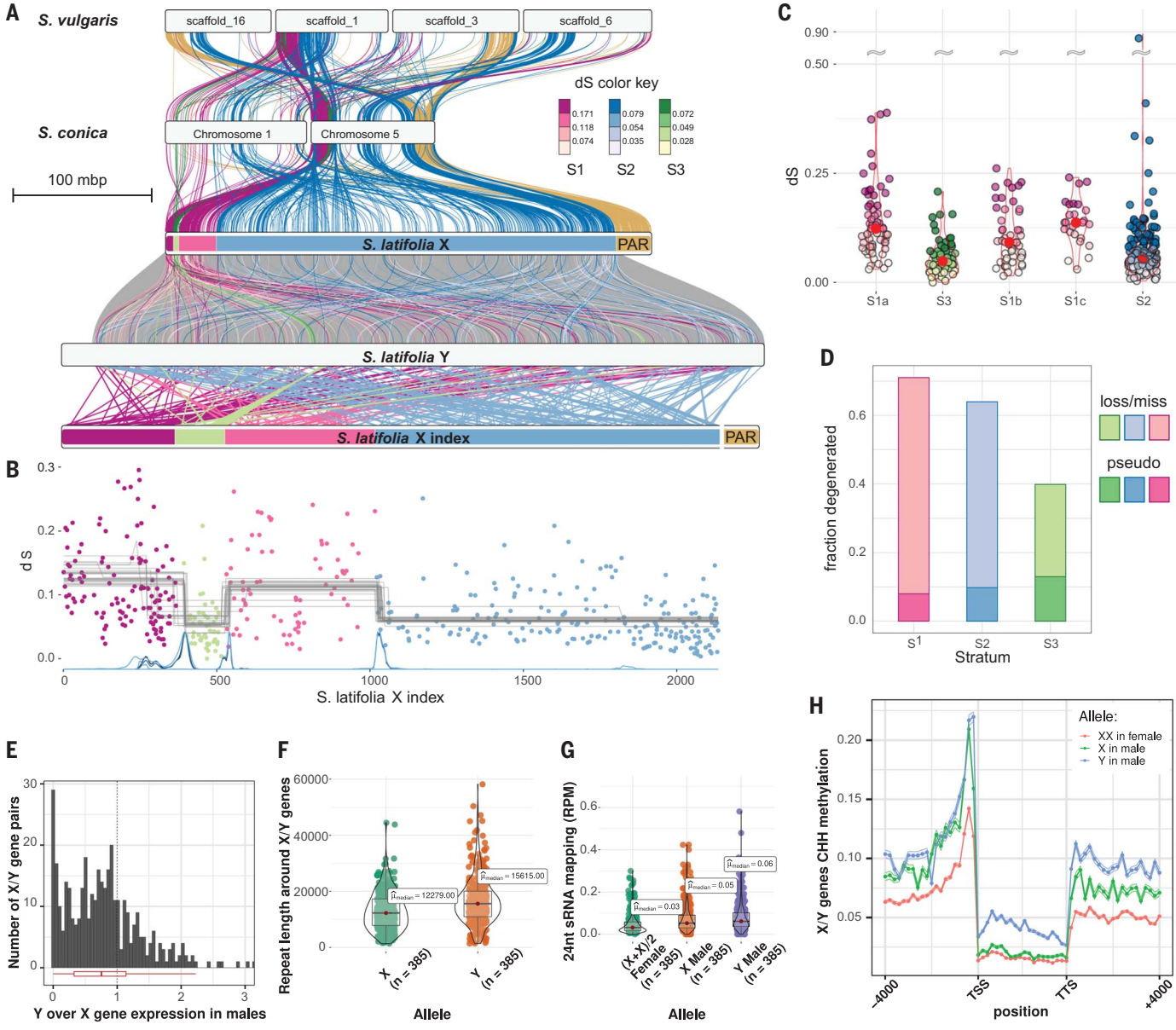

Fig. 2. The evolution of the *S. latifolia* sex chromosomes. (**A**) Syntenic relationships between *S. latifolia* gametologs and autosomal genes in *S. conica* and *S. vulgaris*. Links between homologous chromosomal blocks in the outgroups and *S. latifolia* X are colored by evolutionary strata. Links between *S. latifolia* X and Y are colored by synonymous divergence bins (inset). Synteny between *S. latifolia* Y coordinates and *S. latifolia* X gene rank is shown at the bottom. The *S. latifolia* X chromosome tracks are colored by median synonymous divergence of the strata. The PAR region on *S. latifolia* Y, not shown, is placed to the right of the chromosome track. (**B**) Change-point analysis of d_S along the X chromosome using 401 one-to-one gametologs with d_S values < 0.3. The *x* axis shows gene ranking on the X. Lines at the bottom show the density of the posterior distribution of the change-point locations. Gray lines show the average d_S of the inferred blocks. (**C**) Distribution of synonymous divergence values between *S. latifolia* male X and Y gametologs per stratum. (**D**) Fraction of gene losses and pseudogenization on *S. latifolia* sex chromosomes per stratum. Gene losses on Y were inferred using 1:1:1 orthologs among *S. vulgaris*, *S. conica*, and *S. latifolia* X. Pseudogenes were identified as gametologs with premature stop codons. (**E**) Distribution of Y over X gene expression ratio in four *S. latifolia* male flower buds. Box plot represents median, quartiles, and whiskers. (**F**) Sum of repeat lengths around X-Y gene pairs, from 4000 bp upstream to 4000 bp downstream of the gene. (**G**) Mapping of 24–nucleotide (nt) small RNAs (RPM) on X-Y gene pairs for three *S. latifolia* females and males in flower buds and leaves. The red dot represents the mean. Female X alleles are in red, male X alleles are in green, and Y alleles are in blue. (**H**) Plot of X/Y DNA methylation in CHH context. Female X alleles are in red, male X alleles are in green, and Y alleles are in blue. Average proportion of methylated reads at cytosine positions are shown along sliding windows with a 95% confidence interval. Pairwise synteny plots in (A) were generated with RIdeogram or pafR based on Orthofinder's orthologous relationships. Quantitative plots in other panels were prepared with ggplot2.

nonrecombining pericentromeric regions are typically TE rich because recombination helps purge deleterious TE insertions (*24*); this pericentromeric effect on the *S. latifolia* X chromosome is notably large (fig. S2B). Estimates of the ages of complete LTR retrotransposon insertions show that the insertions in the centers in the pericentromeric regions tend to be oldest and those at the borders the youngest (fig. S12), consistent with TEs accumulating at the margins of these pericentromeric regions and expanding them.

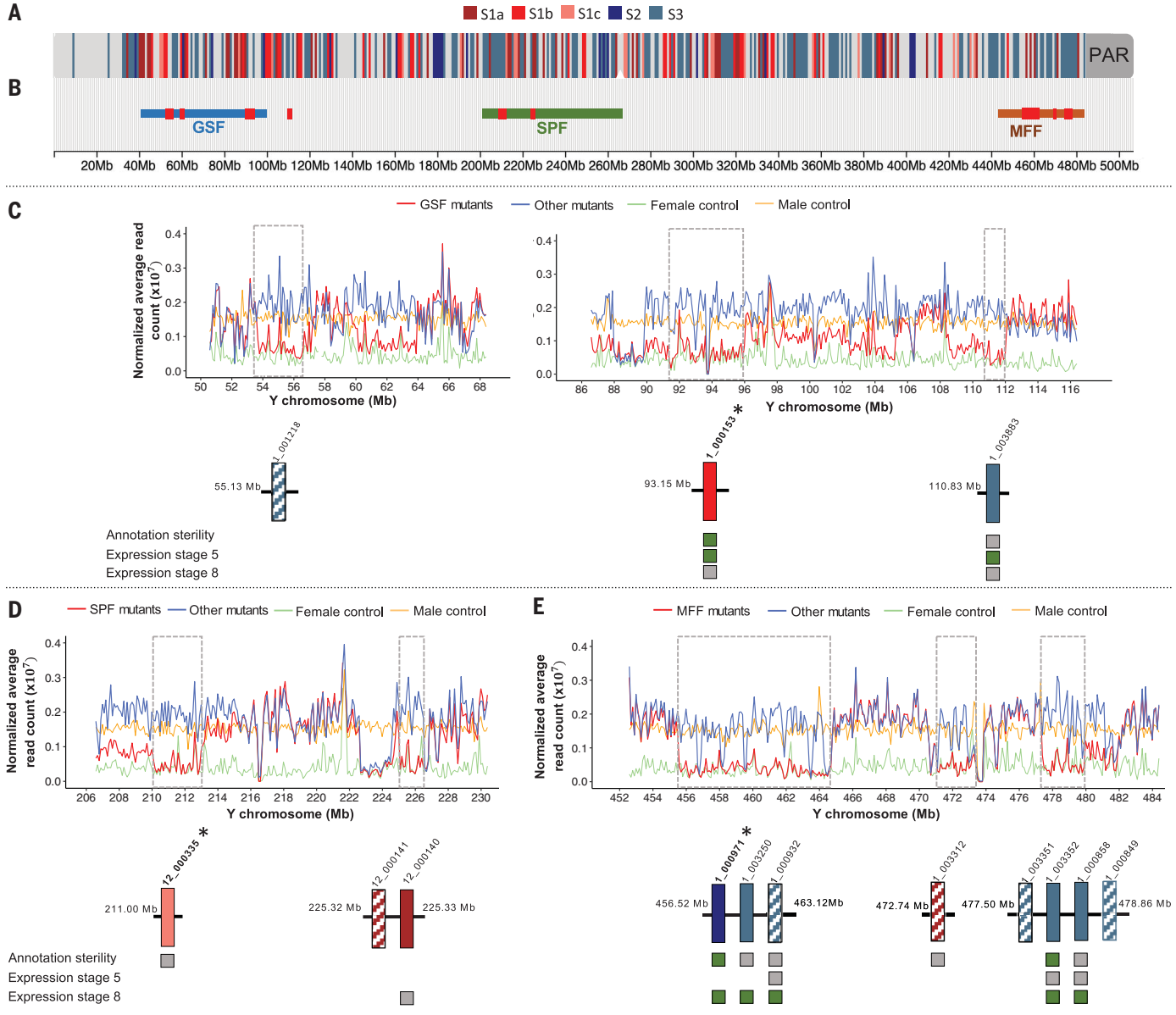

Fig. 3. Identification of deleted genes in the altered sex phenotype mutants. (**A**) The Y chromosome sequence and the evolutionary strata (indicated by 892 Y gametologs). Note that the strata on the Y are mixed up owing to the extensive rearrangements that occurred on this chromosome compared with the X (see text and Fig. 2). The Y is oriented with the PAR on the right side. (**B**) The putative sex-determining regions (MFF, SPF, and GSF) with estimated locations from Bergero *et al*. (*16*). (**C**) Zoom-in on the GSF deletions. Only deletions containing a gene are detailed. Each coverage plot shows the mean normalized read count for GSF mutants, other mutants, control male, and control female. GSF deletions (rectangles) are inferred when the GSF mutant coverage is similar to or lower than that of the control female while the other mutants' coverage was similar to that of the male control. Three genes deleted in all GSF mutants are shown with Y position, stratum assignment [as in (A), hashed fill color indicates that the stratum was inferred from the closest genes due to absence of an X gametolog], and a summary of female sterility annotation and RNA-seq–based expression at stage 5 and 8 in normal male flowers [green, expectations for GSF candidate met (i.e., expressed at stage 5); gray, expectations not met; no square, no data available]. (**D**) Zoom-in on the SPF deletions. Same legend as (C), except for annotation (information on male sterility is indicated). (**E**) Zoom-in on the MFF deletions. Same legend as (D), except for expression (expression for a MFF candidate is expected to be stage 8). The best-supported candidates for sex-determining genes (discussed in the text) are pinpointed by asterisks.

The Y chromosome exhibits signs of considerable degeneration. We found 1541 1:1:1 orthologs in *S. latifolia*, *S. conica*, and *S. vulgaris* on the X chromosome, of which 963 (62%) have no detectable ortholog on the Y chromosome. We also mapped all 3520 genes in our X assembly onto our Y assembly and found homologous genes of about the same size for 1654 (47%), whereas the other 1866 genes (53%) were missing from the Y. A model-based phylogenetic analysis of gene gains and losses also indicated that about 1519 genes (56%) were lost from the Y out of 2694 genes originally present (fig. S13A). Thus, 53 to 62% of the genes appear to have been lost since the Y stopped recombining with the X ~11 Ma, which is higher than previous estimates (*25*). Complete loss of genes is more prominent in older strata, whereas pseudogenes show the opposite trend, consistent with the conclusion that recombination was suppressed more recently in strata S2 and S3, resulting in more recent losses of function (Fig. 2D). We detected a similar proportion of genes with premature stop codons on both sex chromosomes

(fig. S13B), which suggests that the X is also losing genes, as expected given its lack of recombination in stratum S2 (*26*). Among the gametologs with apparently functional X and Y copies, 77% of those with significantly different rates of nonsynonymous versus synonymous substitutions (d_N versus d_S) in the X and Y lineages had higher values in Y lineages, indicating less effective selection (fig. S13C). Restriction site–associated DNA sequencing (RAD-seq) estimates within two populations showed that the Y chromosome has considerably lower nucleotide diversity than the X or the autosomes (fig. S14), as predicted for a chromosome with lower effective population size and undergoing genetic hitchhiking processes (*1*, *27*). Degenerated Y-linked genes have lower expression compared with their X counterparts (Fig. 2E), as already reported in *S. latifolia* (*28*–*30*). This may be explained by epigenetic modifications because Y genes tend to have more TEs nearby than other genes (Fig. 2F) and show two hallmarks of silencing—a higher number of mapped 24-nucleotide small RNAs (Fig. 2G) and higher DNA methylation levels, especially at proximal promoters [200 base pairs (bp) upstream of the transcription start site (TSS)] and gene bodies in the CHH context (where H is any base except G, see Fig. 2H; other methylation contexts are shown in fig. S15).

The sex-determining genes on the Y chromosome

Y chromosome deletions display three sex phenotype categories (*15*, *16*, *31*, *32*): hermaphrodite mutants, asexual mutants in which anther development stops early or at intermediate stages in flower development, and males with pollen defects (late effects). Genetic markers roughly located these deletions in three Y chromosome sex-determining regions (*15*, *16*): a single female-suppressing region [carrying a gynoecium-suppressing factor (GSF)] and two male-promoting regions [one carrying a stamen-promoting factor (SPF) and the other a male fertility factor (MFF) affecting pollen production]. The SPF and MFF mutants are phenotypically quite diverse, and these regions could include several genes affecting those phenotypes (*16*, *33*).

We identified candidate sex-determining genes by low-coverage sequencing of 18 well-characterized sex phenotype mutants with Y deletions (table S5). We mapped the mutants' reads onto our reference Y chromosome along with reads from a phenotypically normal control individual of each sex from the same U17 population as our reference (fig. S16). Although female coverage on the Y is generally low, it is not always zero because X reads may mis-map to the Y chromosome. We therefore classified a region as deleted in a mutant if its coverage was similar to the control female value or less. Coverage similar to the control male indicates absence of Y deletions. Although the data are noisy, they clearly identify deletions specific to each mutant category: four for the GSF mutant category, two for SPF, and three for MFF (Fig. 3B), almost all in the three sex-determining regions previously defined (*16*) (fig. S3).

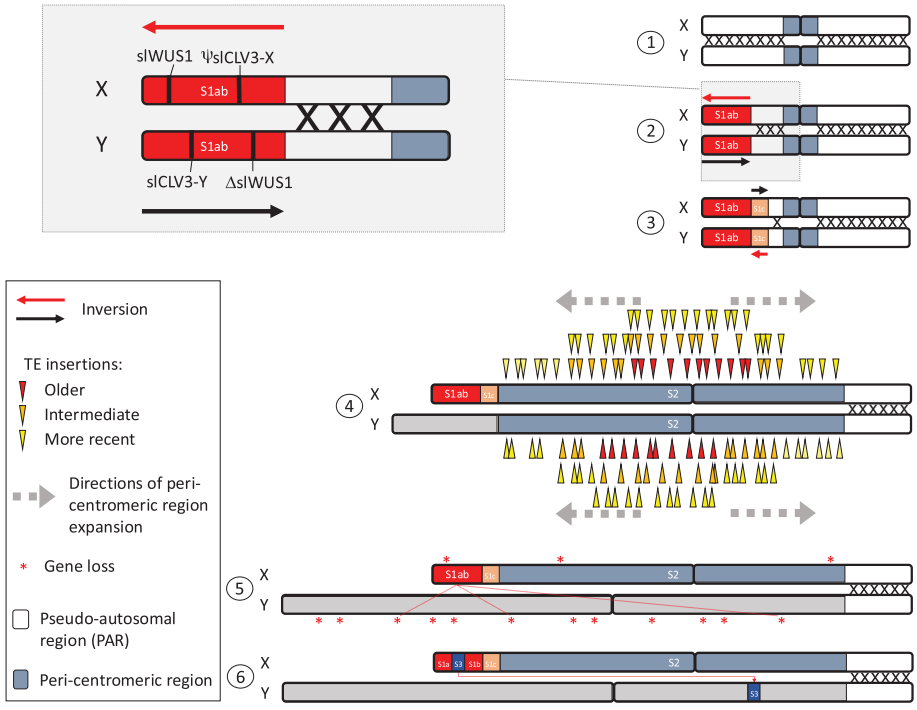

Fig. 4. Scenario for the evolution of sex chromosomes in *S. latifolia*. (1) Pair of autosomes, ancestors of the sex chromosomes; (2) ~11 Ma: inversion on the X generating the stratum S1ab, indicated by a red arrow; (3) ~11 Ma: inversion on the Y generating the stratum S1c, indicated by a red arrow [the inset at the top left represents a zoom-in on the S1 stratum, showing the candidate sex-determining genes *Clavata3* (GSF) and *Wuschel1*. *slCLV3* (GSF); and *slWUS1* is in stratum S1ab defined by the inversion on the X]; (4) ~5 Ma: pericentromeric region extension by TE accumulation forming stratum S2, TE insertions being depicted by arrowheads [MFF candidates were found in this stratum (not shown)]; (5) gene losses on the Y chromosome, those from future stratum S3 (located in S1) are highlighted by red dashed lines connecting X and Y genes; and (6) ~4.5 Ma: duplication and translocation of a fragment from the X to the Y, possibly for compensating gene losses from (5), formation of S3, X-Y d_S values become 0 in the X translocated region.

Figure 3 shows the genes located within the deletions, those with functional annotations of sterility terms, and their expression during early (stage 5) or late (stage 8) flower development (table S6). The GSF deletions include the Y *Clavata3* gene (*slCLV3-Y* with gene ID scaffold1_000153), confirming this recently proposed GSF candidate (*17*, *18*). The functional role of *slCLV3* in male and female organ development is not well understood, but differences in the balance of gene expression in the Clavata-Wuschel pathway between males and females were proposed to explain carpel formation versus inhibition in female and male flowers in *S. latifolia*, respectively (*17*, *18*). Indeed, *Clavata3*, known as a carpel inhibitor in *Arabidopsis thaliana* and other plants, has a functional copy on the Y and a gametolog pseudogene on the X [*ψslCLV3-X* (*17*, *18*)]. *Wuschel1*, known as a carpel promoter in *A. thaliana* and other plants, is present on the X (*slWUS1-X*) and deleted from the Y [*ΔslWUS1-Y* (*18*, *34*)]. Our results independently confirm the role of this gene pair in determining female fertility or sterility in *S. latifolia* (Fig. 3).

We found several MFF candidate genes, including *slCyp704B1-Y* (gene ID scaffold1_000971), which is homologous to the *A. thaliana Cyp704B1* gene, encoding a cytochrome P450 protein crucial for pollen maturation. *slCyp704B1-Y* is expressed in the tapetum and is involved in the synthesis of sporopollenin (a pollen cell wall component). Its inactivation causes male sterility in *A. thaliana*. *slCyp704B1-Y* is expressed during late development (stage 8) of *S. latifolia* male flowers only based on RNA sequencing (RNA-seq) and quantitative polymerase chain reaction (qPCR) (fig. S17), consistent with an MFF function. Another MFF candidate is *slTHI1-Y* (gene ID scaffold1_003352), which is homologous to the *A. thaliana THI1* gene, encoding a papain-like cysteine protease that is also expressed in the tapetum, is important for pollen

maturation (through involvement in proteolysis and tapetal cell degeneration), and is also annotated as a male fertility gene.

We identified one SPF candidate based on deletion mutants (gene ID scaffold12_000335, homolog to the *A. thaliana Scarecrow-like 4* and *7* transcription factors *SCL4/7*). The *slSCL4-Y* gene is pseudogenized by a premature stop codon but is expressed in flower buds at developmental stage 5, when sex is determined (fig. S19). Because deletion of this pseudogene leads to the loss of the male function in *S. latifolia* flowers, we hypothesize that its expression as a long noncoding RNA (lncRNA) is important for sex determination. However, the mechanism through which it may regulate sex determination remains unclear (supplementary text, section S2) (*19*).

Based on the location of these genes on the X chromosome, both *slWUS1* and *slCLV3* are in stratum S1 (in S1a and S1b, respectively). The *slCyp704B1-Y* MFF candidate has no X gametolog, but its orthologs are located in the *S. conica* chromosome 1 and *S. vulgaris* scaffold 1 in syntenic blocks homologous to the middle of the *S. latifolia* X chromosome stratum S2 (around positions 140 to 150 Mb). *slTHI1* is in stratum S3. Therefore, both MFF candidates appear to be a late addition to the nonrecombining region of the Y chromosome. An inversion on the X coincides with S1ab, and an inversion on the Y coincides with S1c, which suggests that these inversions suppressed recombination between the sex chromosomes, thereby forming strata S1ab and S1c (Fig. 4, stratum S1 panel). GSF potential sex-determining genes are thus concentrated in the small, oldest stratum S1. This is consistent with models for the evolution of dioecy in plants, with a yet-unknown initial male-sterility mutation and subsequent partially female-suppressing mutations creating selection for suppressed recombination (*35*). This model predicts that multiple linked mutations will create complete maleness. The candidate GSF genes that we detect in *S. latifolia*, together with the paracentric inversions forming S1ab, are consistent with this model. The MFF region, on the other hand, evolved well after dioecy was established and is associated with stratum S2. MFF genes may simply be male-function genes lost from the X because there is no selection to keep them on that chromosome (*26*), or they may have sexually antagonistic effects that led to polymorphisms favoring stratum S2 formation (*1*).

Conclusions

Our high-quality *S. latifolia* sex chromosome assemblies provide insights about their structure, function, and evolution (Fig. 4). These sex chromosomes originated ~11 Ma with the differentiation of a small region carrying the GSF sex-determining genes. A first stratum was probably formed by one paracentric inversion on each of the X and the Y. More recently, another stratum, S2, encompassing the centromere and including MFF, formed ~5 Ma, probably in the context of a general expansion of the pericentromeric regions in the *S. latifolia* genome, whose cause remains unclear. This generated a very large, nonrecombining, and TE-rich region on the Y. This is similar to what was observed in *Rumex hastatulus*, in which the Y chromosome is massively rearranged and repeat rich despite its recent origin (<10 Ma) (*36*). The absence of recombination also led to genetic degeneration, with half of the ancestral genes being lost from the Y, despite the expression of Y-linked genes in the haploid stages in plants (*37*, *38*). The X chromosome's pericentromeric region is also extremely large, perhaps because of increased TE accumulation under a reduced effective population size and absence of recombination of this stratum S2 on the X chromosome relative to autosomes (*39*, *40*).

REFERENCES AND NOTES

1. D. Charlesworth, B. Charlesworth, G. Marais, *Heredity* **95**, 118–128 (2005).
2. D. Bachtrog *et al*., *PLOS Biol.* **12**, e1001899 (2014).
3. R. Ming, A. Bendahmane, S. S. Renner, *Annu. Rev. Plant Biol.* **62**, 485–514 (2011).
4. L. A. M. Rosolen, M. R. Vicari, M. C. Almeida, *Cytogenet. Genome Res.* **156**, 215–222 (2018).
5. A. Muyle, R. Shearn, G. A. Marais, *Genome Biol. Evol.* **9**, 627–645 (2017).
6. S. S. Renner, N. A. Müller, *Nat. Plants* **7**, 392–402 (2021).
7. G. Bernasconi *et al*., *Heredity* **103**, 5–14 (2009).
8. K. B. Blackburn, *Nature* **112**, 687–688 (1923).
9. S. Matsunaga, M. Hizume, S. Kawano, T. Kuroiwa, *Cytologia* **59**, 135–141 (1994).
10. S. Qiu *et al*., *Mol. Ecol.* **25**, 414–430 (2016).
11. M. Nicolas *et al*., *PLOS Biol.* **3**, e4 (2004).
12. D. A. Filatov, *Genetics* **170**, 975–979 (2005).
13. J. Yue *et al*., *Curr. Biol.* **33**, 2504–2514.e3 (2023).
14. R. Hobza *et al*., *Genes* **8**, 302 (2017).
15. M. Westergaard, *Adv. Genet.* **9**, 217–281 (1958).
16. R. Bergero, D. Charlesworth, D. A. Filatov, R. C. Moore, *Genetics* **178**, 2045–2053 (2008).
17. Y. Kazama *et al*., *Mol. Biol. Evol.* **39**, msac195 (2022).
18. Y. Kazama, T. Kobayashi, D. A. Filatov, *BioEssays* **45**, 2300111 (2023).
19. See the supplementary materials.
20. A. Muyle *et al*., *Genome Biol. Evol.* **8**, 2530–2543 (2016).
21. A. Muyle *et al*., *Nat. Plants* **4**, 677–680 (2018).
22. V. Bačovský, R. Čegan, D. Šimoníková, E. Hřibová, R. Hobza, *Front. Plant Sci.* **11**, 205 (2020).
23. M. Krasovec, M. Chester, K. Ridout, D. A. Filatov, *Curr. Biol.* **28**, 1832–1838.e4 (2018).
24. B. Charlesworth, P. Sniegowski, W. Stephan, *Nature* **371**, 215–220 (1994).
25. A. S. Papadopulos, M. Chester, K. Ridout, D. A. Filatov, *Proc. Natl. Acad. Sci. U.S.A.* **112**, 13021–13026 (2015).
26. A. Mrnjavac, K. A. Khudiakova, N. H. Barton, B. Vicoso, *Evol. Lett.* **7**, 4–12 (2023).
27. I. Gordo, B. Charlesworth, *Curr. Biol.* **11**, R684–R686 (2001).
28. M. V. Chibalina, D. A. Filatov, *Curr. Biol.* **21**, 1475–1479 (2011).
29. R. Bergero, D. Charlesworth, *Curr. Biol.* **21**, 1470–1474 (2011).
30. A. Muyle *et al*., *PLOS Biol.* **10**, e1001308 (2012).
31. A. Lardon, S. Georgiev, A. Aghmir, G. Le Merrer, I. Negrutiu, *Genetics* **151**, 1173–1185 (1999).
32. I. Farbos *et al*., *Genetics* **151**, 1187–1196 (1999).
33. J. Zluvova *et al*., *Genetics* **177**, 375–386 (2007).
34. Y. Kazama *et al*., *G3* **2**, 1269–1278 (2012).
35. B. Charlesworth, D. Charlesworth, *Am. Nat.* **112**, 975–997 (1978).
36. B. Sacchi *et al*., *Mol. Biol. Evol.* **41**, msae074 (2024).
37. G. Sandler, F. E. G. Beaudry, S. C. H. Barrett, S. I. Wright, *Evol. Lett.* **2**, 368–377 (2018).
38. J. E. Mank, *Phil. Trans. R. Soc. B* **377**, 20210218 (2022).
39. J. Wang *et al*., *Proc. Natl. Acad. Sci. U.S.A.* **109**, 13710–13715 (2012).
40. D. W. Bellott *et al*., *Nature* **466**, 612–616 (2010).
41. C. Moraga, A. Di Genova, Silene-genome/genome-paper: V1.0, version 1.0, Zenodo (2024); https://doi.org/10.5281/zenodo.14434538.

ACKNOWLEDGMENTS

We thank E. Lacroix, technician at Université Claude Bernard Lyon 1 glasshouse, for her help for growing the *S. latifolia* U17 plants; C. Scutt for discussion about floral development; and three anonymous referees for suggestions and comments that improved this manuscript. All the computationally demanding bioinformatics were performed using the computing facilities of the CC LBBE/PRABI of Université Claude Bernard Lyon 1; the supercomputing infrastructure of the High-Performance Computing UOH laboratory (FIC 40059065-0) of Universidad de O'Higgins, Rancagua; the supercomputing infrastructure of the NLHPC (CCSS210001), by the e-INFRA CZ project (ID: 90254), supported by the Ministry of Education, Youth and Sports of the Czech Republic; and by the National Academic Infrastructure for Supercomputing in Sweden (NAISS, projects 2023/23-315 and 2023/22-242), funded by the Swedish Research Council through grant agreement no. 2022-06725. **Funding:** This study received support from Agence Nationale de la Recherche (ANR) grants ANR-20-CE20-0015-01 (G.A.B.M. and A.Be.), ANR-20-CE02-0015 (G.A.B.M.), and ANR-22-CE02-0024 (A.M.); the Fédération de Recherche "Biodiversité, Eau, Environnement, Ville & Santé" (FR BIOEENVIS) of University of Lyon 1 grant (G.A.B.M. and J.B.); the Czech Science Foundation grants 21-00580S and 22-00364S (P.J.); the European Research Council (ERC) EvolSexChrom (832352) grant (T.G.); the ERC NectarGland (101095736) grant (A.Be.); the CNRS Biology starting grant (A.M.); the ANID Chile, Grants Fondecyt Regular 1221029, and SIA grant SA77210017 (A.D.G.); the US National Science Foundation grant MCB-2048407 (D.B.S.); the US National Institutes of Health grant R35GM148134 (D.B.S.); and the US National Science Foundation grant DEB-1353970 (L.F.D.).
Author contributions: Conceptualization: G.A.B.M., R.H., M.-A.M., A.D.G., A.Be., T.G., D.C., A.M.; Formal analysis: C.M., C.Br., Q.R., P.V., P.J., A.M., M.H., R.C.R.d.l.V., E.T., X.L., E.M.-G., C.L., P.D.F., C.Be., R.Ce., G.A., A.E.F., V.M., A.V., S.K., A.D.G., M.-A.M.; Methodology: C.M., C.Br., R.L., A.D.G., G.A.B.M.; Investigation: C.M., C.Br., Q.R., P.V., P.J., A.M., M.H., R.C.R.d.l.V., E.T., X.L., E.M.-G., C.L., P.D.F., C.C., K.L., C.Be., J.B., S.S., R.Ce., R.L., G.A., A.Bo., R.T., D.B.S., S.K., L.F.D., D.C., T.G., A.Be., A.D.G., M.-A.M., R.H., G.A.B.M.; Resources: R.H., A.D.G., L.F.D., D.B.S., M.S.O., D.R.T., O.A.J.A., M.K., H.Šť., H.Ši., P.C., Z.T., H.T., L.W., S.D., B.S., D.P., A.M.; Software: C.M., C.Br., E.T., R.L., R.C.R.d.l.V., A.V., A.D.G.; Validation: A.C., R.Co., C.F.; Visualization: C.M., C.Br., Q.R., P.V., A.M., R.C.R.d.l.V., E.T., R.L., D.P., S.K., M.-A.M.; Funding acquisition: G.A.B.M., A.Be., J.B., P.J., T.G., A.M., A.D.G., D.B.S., L.F.D.; Project administration: G.A.B.M.; Supervision: G.A.B.M., R.H., M.-A.M., A.D.G., A.Be., T.G., L.F.D., S.K., D.B.S., R.Co., R.T., A.Bo., R.C.R.d.l.V., R.L., A.M.; Writing – original draft: G.A.B.M., M.-A.M., A.Be., S.K., D.B.S., S.S., J.B., P.D.F., X.L., E.T., A.M., Q.R., C.Br., C.M.; Writing – review & editing: G.A.B.M., T.G., D.C., L.F.D., S.K., R.C.R.d.l.V., C.Br., E.M.-G., A.Bo., A.M., P.V., C.M., Q.R., P.J., R.Ce., R.H., and all authors. **Competing interests:** The authors declare that they have no competing interests. **Data and materials availability:** Sequencing data (long reads, short reads, and Omni-C datasets), genome assembly, and annotation are available under the project PRJNA1132743 on the National Center for Biotechnology Information (NCBI). All analyses and pipelines to generate figures are available on GitHub (https://github.com/Silene-genome/genome-paper) and Zenodo (*41*). **License information:** Copyright © 2025 the authors, some rights reserved; exclusive licensee American Association for the Advancement of Science. No claim to original US government works. https://www.science.org/about/science-licenses-journal-article-reuse

SUPPLEMENTARY MATERIALS

science.org/doi/10.1126/science.adj7430
Materials and Methods
Supplementary Text
Figs. S1 to S20
Tables S1 to S7
References (*42*–*132*)
MDAR Reproducibility Checklist

Submitted 18 September 2023; resubmitted 22 April 2024
Accepted 18 December 2024
10.1126/science.adj7430