Title: The *Silene latifolia* genome and its giant Y chromosome

Authors: Carol Moraga[1,2,3,4]†, Catarina Branco[1,5,6,7]†, Quentin Rougemont[8]†, Paris Veltsos[9], Pavel Jedlička[10], Aline Muyle[11], Melissa Hanique[12], Eric Tannier[1], Xiaodong Liu[13], Eddy Mendoza-Galindo[11], Claire Lemaitre[14], Peter D. Fields[15], Corinne Cruaud[16], Karine Labadie[16], Caroline Belser[16], Jerome Briolay[17], Sylvain Santoni[18], Radim Cegan[10], Raquel Linheiro[1,5,6,7], Ricardo C. Rodríguez de la Vega[8], Gabriele Adam[12], Adil El Filali[1], Vinciane Mossion[19], Adnane Boualem[12], Raquel Tavares[5,6,7], 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], Tatiana Giraud[8], Abdelhafid Bendahmane[12], Alex Di Genova[2,3,4], Amin Madoui[16,32], Roman Hobza[10], Gabriel A. B. Marais[1,5,6,7]*

Affiliations:

1. Laboratoire Biométrie et Biologie Evolutive (LBBE), CNRS / Univ. Lyon 1, Campus de la Doua, 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.Millennium Institute for Integrative Biology iBio, Chile.

5. CIBIO, Centro de Investigação em Biodiversidade e Recursos Genéticos, InBIO, Laboratório Associado, Campus de Vairão, Universidade do Porto, 4485-661 Vairão, Portugal.

6. Departamento de Biologia, Faculdade de Ciências, Universidade do Porto, 4099-002, Porto, Portugal.

7. BIOPOLIS Program in Genomics, Biodiversity and Land Planning, CIBIO, Campus de Vairão, 4485-661 Vairão, Portugal.

8. Université Paris-Saclay, CNRS, AgroParisTech, Laboratoire Ecologie Systématique et Evolution, UMR 8079, Bâtiment 680, 12 route RD128, 91190 Gif-sur-Yvette, France.

9. Department of Biology, Vrije Universiteit Brussel, Pleinlaan 2, 1050 Brussels, Belgium.

10. Department of Plant Developmental Genetics, Institute of Biophysics of the Czech Academy of Sciences, Kralovopolska 135, 61200, Brno, Czech Republic.

11. Centre d'Ecologie Fonctionnelle et Evolutive (CEFE), University of Montpellier, CNRS, EPHE, IRD, Montpellier, France.

12. Institute of Plant Sciences Paris-Saclay (IPS2), Université Paris-Saclay, CNRS, INRAE, Université d'Evry, Gif-sur-Yvette, France.

13. Department of Biology, University of Copenhagen, Ole Maaløes Vej 5, DK-2200 Copenhagen, Denmark.

14. Institut de recherche en informatique et systèmes aléatoires (IRISA), Univ. Rennes, Inria, CNRS, Rennes F-35000, France.

15. Department of Biology, Colorado State University, Fort Collins, CO, USA.

16. Genoscope, Institut François Jacob, CEA, CNRS, Univ Evry, Université Paris-Saclay, Evry, France.

17. Développement de Techniques et Analyse Moléculaire de la Biodiversité (DTAMB), Univ. Lyon 1, Campus de la Doua, Villeurbanne, France.

18. Genomic platform, Amélioration génétique et adaptation des plantes méditerranéennes et tropicales (AGAP), Univ. Montpellier, CIRAD, INRAE, Montpellier, France.

19. Department of Ecology and Genetics, Division of plant ecology and evolution, Uppsala University, Norbyvägen 18D, SE-75236 Uppsala, Sweden.

20. Efor, Grosspeter Tower (Spaces), Grosspeteranlage 29, 4052 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, USA.

23. Santé de la Vigne et Qualité du Vin (SVQV), INRAE, Colmar, France.

24. University of Louisville School of Medicine, Undergraduate Medical Education, 500 South Preston Street, Louisville, KY, 40202, USA.

25. Institute of Experimental Botany of the Czech Academy of Sciences, Olomouc, Czech Republic.

26. Plant Reproduction Lab, Institute of Experimental Botany, Czech Academy of Sciences, Rozvojová 263, Prague, 16502 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, Norbyvägen 18D, SE-75236 Uppsala, Sweden.

30. Department of Biology, Indiana University, Bloomington, IN 47405 USA.

31. Institute of Ecology and Evolution, School of Biological Sciences, University of Edinburgh, Ashworth Laboratories, Charlotte Auerbach Road Edinburgh EH9 3FL, UK.

32. 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.

†These authors contributed equally to this work

*Corresponding author. Email: gmarais@cibio.up.pt

Abstract: In some species, the Y is a tiny chromosome but the dioecious plant *Silene latifolia* has a giant ~550 Mb Y chromosome, which has remained unsequenced so far. Here we used a hybrid approach to obtain a high-quality male *S. latifolia* genome. Using mutants for sexual phenotype, we identified candidate sex-determining genes on the Y. Comparative analysis of the sex chromosomes with outgroups showed the Y is surprisingly rearranged and degenerated for a ~11 MY-old system. Recombination suppression between X and Y extended in a stepwise process, and triggered a massive accumulation of repeats on the Y, as well as in the non-recombining pericentromeric region of the X, leading to giant sex chromosomes.

One-Sentence Summary: This work uncovers the structure, function, and evolution of one of the largest giant Y chromosomes, that of the model plant *Silene latifolia*, which is almost 10 times larger than the human Y, despite similar genome sizes.

Main Text:

Among the multiple paths the evolution of sex chromosomes can take, some lead to giant Y chromosomes (*1,2*). Giant Y chromosomes have been first identified in dioecious plant species (*3*) but also exist in animals (*4*). In the last decade, great advances have been made in studying plant sex chromosomes thanks to genomics and bioinformatics (*5,6*), but no giant plant Y chromosome has been assembled yet. Giant Y chromosomes may result from massive accumulation of transposable elements (TEs) but their chromosome organization, their precise role in sex determination, and their evolution remain poorly known (*3*).

Silene latifolia (Caryophyllaceae) is a dioecious plant that has been studied since Darwin's time (*7*). *S. latifolia* has an XY sex-determination system, which was discovered 100 years ago (*8*). The Y is ~550 Mb, and the X ~400 Mb, of the total haploid genome size of ~2.7 Gb (*9*). Genetic maps show that the X and the Y are largely non-recombining and share only a single pseudo-autosomal region (PAR) (*10*). Recombination has been suppressed progressively forming groups of X/Y gene pairs with differing synonymous divergence, called evolutionary strata (*11,12*). The repeat-richness (*13*) and size of the *S. latifolia* Y have, however, so far prevented its assembly. Mutants with deletions on the Y chromosome and altered sex phenotypes indicate the presence of three sex-determining regions (*14*): one female-suppressing region (carrying a gynoecium-suppressing factor, GSF) and two male-promoting (carrying a stamen-promoting factor - SPF and male-fertiliy factor - MFF) (*15*). A GSF candidate gene has recently been proposed (*16*), but the other sex-determining genes remain unknown.

Here we used an Oxford-Nanopore-Technology (ONT)-based sequencing approach to obtain the *S. latifolia* genome, in order to study the repeat-rich Y chromosome. We also used high-quality genome assemblies of closely related non-dioecious *Silene* species as outgroups to make inferences about the evolution of the *S. latifolia* sex chromosomes. We sequenced mutants with Y deletions for three sexual phenotypes (hermaphrodites and asexuals with early/intermediate and late anther development arrest) in order to pinpoint candidate sex-determining genes/regions. Furthermore, we generated expression data at two critical stages for male and female flower development to help identify sex-determining genes.

The structure and gene content of the sex chromosomes

To assemble the complex *S. latifolia* genome, we produced an inbred population from 17 generations of brother-sister crosses to reduce heterozygosity (see Material and Methods for the details of the sequencing, assembly and annotation methodologies). We selected one male and one female individual from this inbred population for sequencing. The male was sequenced using ONT at 100X coverage to ensure 50X coverage for both sex chromosomes (Table S1). Assembly was performed using an initial hybrid long-read approach including Flye (Figure S1). The ONT contigs were polished using 200X of short Illumina reads and then scaffolded using Bionano optical mapping data followed by Omni-C contact data (see Table 1 for genome statistics and Table S2 for more details). The polished consensus sequence reached a phred quality score of ~30 (1/1000 base error). All scaffolds were anchored onto our genetic map, except scaffolds 1, 12, 13, 16 and 41 (Figure S2). These unplaced scaffolds were classified as likely fragments of the Y chromosome based on female/male sequence depths and mapping of previously well-characterized sex-linked genes (see below). They were assembled into a single sequence. Re-mapping the Omni-C data onto

the chromosome sequences obtained in this manner further improved the scaffolding of the X and Y chromosomes. The final chromosome sequences show high consistency with the contact data (Figure S3). During the assembly process, the two haplotypes of each sequenced region were collapsed into a single sequence. We present this a haploid version of the *S. latifolia* genome with a single pseudomolecule for each chromosome, including separate assemblies of X and the Y chromosomes, except for the PARs on the X and Y, which were also collapsed (the resulting single PAR sequence is assembled on the X chromosome only (see below). The sizes of the genome and sex chromosomes obtained closely match the previous C-value-based size estimates (Table 1). Gene annotation identified 35,436 protein-coding genes. BUSCO scores were high at different steps of the assembly/annotation process and the final score was 86.5% (Tables 1 and S2). Repeat analysis revealed that 79.24% of the *S. latifolia* male genome consists of repeats (Table S3). Annotated repeats represented 61.18% of the *S. latifolia* genome with a very high abundance of two LTR retrotransposons *Copia* (20.73%) and *Gypsy* (33.69%). Telomere-associated repeats were found at the expected locations for all chromosome sequences, which constitute telomere-to-telomere assemblies (Figure 1A). Genes are concentrated in the chromosomal arms and are sparse in the pericentromeric regions, and vice versa for repeats, as in many eukaryotic genomes (*17*).

Our assemblies of the X and the Y chromosomes are of high quality. In systems in which the X and Y are differentiated (with SNPs and indels) and using stringent mapping parameters, female/male sequence-depth ratio of ~1, ~2 and ~0 for autosomes, X, and Y chromosomes, respectively, are expected, and were observed in our data (Figure 1A). A smaller set of experimentally validated sex-linked genes (compiled in Muyle et al. 2018) also mapped as expected to their previously assigned X or Y position (Figures 2 and 3). The X chromosome sequence obtained is 346 Mb long. The distribution of genes and repeats (in particular centromere- and telomere associated repeats) is also as expected for a typical metacentric chromosome (Figure 1A). Sex-specific recombination data and other data identified the pseudo-autosomal region (Figure 1B). The PAR is a small gene-rich region (25 Mb with 1,286 genes). The Y chromosome assembly is 497 Mb long (522 Mb if the PAR is included) and includes some of the largest scaffolds (1 - 364.5 Mb, 12 - 45.7 Mb, 13 - 41.9 Mb).

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 found three evolutionary strata, and extensive rearrangements on the X, and especially the Y chromosome. Our analyses of rearrangements and estimates of synonymous site divergence (d_S) between the X and Y copies used a set of 355 gametologs (= X/Y gene pairs) with 1:1 orthology with the outgroup species *Silene conica* and *Silene vulgaris*. Change-point analysis identified four regions in the non-recombining region with different pairwise mean X-Y d_S values in adjacent X chromosome regions, based on the gene order in the assembly (Figures 2 and S4A). We defined three evolutionary strata, S1, S2 and S3, based on the different d_S levels (Figures 2A and S4B). The oldest stratum, region S1, is split in two by S3. S3 has a lower d_S mean than its two flanking regions, which does not fit the expected pattern of decreasing divergence with proximity to the PAR; it has d_S value similar to that of the S2 region but much higher synteny, and it was not formed via the same type of rearrangements (see below). S1 and S3 are small regions located within the first 27 Mb of the X chromosome q arm. S2, on the other hand, is very large and includes most of the X chromosome. Using a

molecular clock approach, we found that strata S2 and S3 evolved 4.9 [3.8, 6] and 4.4 [3.3, 5.6] MY ago respectively and stratum S1 10.8 [9.7, 11.9] MY ago, i.e. when the *S. latifolia* sex chromosomes originated (*18*).

Comparisons of gene order between *S. latifolia* and two non-dioecious close relatives, *S. conica* and *S. vulgaris* revealed large syntenic blocks with some rearrangements (Figure S5). The *S. latifolia* Y chromosome stands out as highly rearranged compared with either the X or both outgroups, with the notable exception of the S3 stratum, which includes in particular a syntenic block of 4 Mb (Figure S6A and S6C). The *S. latifolia* X shows homology with four *S. vulgaris* scaffolds (1, 3, 6 and 16; Figures S5) and with chromosome 5 of *S. conica* (its closest relative), and smaller parts of chromosomes 1, 2 and 6 (Figures S5 and S6B). Reconstruction of the rearrangements between the X, the Y and the outgroups (Figures 2B, S6C-E and S7) showed that stratum S1 may have resulted from two inversions, one on the X encompassing S1a to S1b and one on the Y including S1c, that occurred early in the evolution of the sex chromosomes.

Stratum S3 had a lower d_S than stratum S1 and was the only region showing extended synteny between the X and Y chromosomes, which suggests that S3 has recently translocated into the middle of the oldest stratum S1 (Figure S6C). However, comparisons between *S. latifolia* X and the outgroups suggest that S3 was ancestrally located within the S1 stratum at the very same place (Figures 2 and S6D-E). To reconcile these findings, we propose that S3 (initially within the S1 stratum) was lost from the Y and later regained by a recent duplicative-translocation from the X (Figure 2B). Stratum S2 is likely slightly older than S3 as it is more rearranged, and probably arose through a different mechanism (Figure 2C). Reconstruction of the rearrangements between the X, the Y and the outgroups (Figures 2 and S8) could not associate S2 formation to a single event. We found several inversions, some of them pericentric, as previously speculated (*19*).

The Y and the X chromosomes are giant sized in *S. latifolia* and we infer that this is due to TE accumulation in both sex chromosomes. Our repeat analysis indeed revealed massive TE accumulation on the Y, but also, to a lesser extent, on the X (Tables 1 and S3, Figure 1). This explains why the X and Y, are, respectively, 4 and 5.5 times larger than a typical *S. conica* chromosome such as chromosome 5, with which most of the orthologs with the sex-linked genes are from. The *S. latifolia* autosomes average size is twice that of their *S. conica* homologs, supporting the view that TEs on sex chromosomes constitute a reservoir spreading genome-wide (*20*). In Eukaryotes, the non-recombining pericentromeric regions are TE-rich as recombination helps purge deleterious TE insertions (*21*) and this pericentromeric effect on the X chromosome is strikingly large (Figure S2B).

The Y chromosome exhibits signs of considerable degeneration. Out of 1,541 1:1 orthologs in *S. latifolia* X, *S. conica* and *S. vulgaris*, 963 had no detected ortholog on the Y chromosome. A model-based phylogenetic analysis of gene gain and loss confirmed this observation. As many as 58% of the genes on the Y appeared to have been lost since it stopped recombining with the X about ~11 MY ago. In addition, we detected more genes with premature stop codons on the Y compared to the X, except for stratum 3 (Figure 2E), suggesting the X might also be degenerating in this stratum. Among the gametologs with apparently functional X and Y copies, 77% of those with significant differences in rates of non-synonymous versus synonymous (d_N/d_S) changes between X and Y had the Y copies with higher d_N/d_S values, indicating relaxed selection (Figure 2D). Another form of degeneration is when Y-linked genes have lower expression than their X counterparts, which

has been reported in *S. latifolia* (*22-24*); this may be explained by epigenetic modifications (Figure 2F-I), as Y genes had more TEs in their vicinity and bear hallmarks of silencing, i.e., a higher number of 24 nucleotide small RNA mapping and higher levels of DNA methylation, especially around the promoter in the CHH context, compared to X genes (other contexts are shown in Figure S9). Using RAD-seq data, we found that the Y chromosome exhibits considerably lower genetic diversity *as c*ompared to the X chromosome and autosomes (Figure S10), in agreement with this chromosome undergoing genetic degeneration .

The sex-determining genes on the Y chromosome

We identified candidate sex-determining genes by sequencing at low coverage 18 mutants for sexual phenotypes (Table S4 and Material and Methods). Hermaphrodite mutants have deletions in the GSF region, asexual mutants in which anther development was stopped at early or intermediate stages in flower development have a deletions in the SPF region, and asexual mutants with pollen defects (late events) have deletions in the MFF region. SPF and MFF mutants were phenotypically very diverse, suggesting they have several different deleted genes (*25,26*).

Figure 3A shows the mapping of the mutant reads onto the reference Y chromosome along with reads from normal (U17) males and females, which indicate the expected coverage for presence (male) or absence (female) along the Y chromosome. Note that female coverage is generally low but not always null depending on how many X reads can mismap on the Y chromosome. A region is inferred as deleted in a mutant if the coverage is similar (or lower) to the normal female and different from the normal male. While the data are noisy, regions of very low coverage in all mutants of a given category (but not other categories) are still easily identified and are limited in number (3 for GSF; 2 for SPF; 3 for MFF categories). They were located as expected on the Y based on a map of the Y built using genetic markers to genotype the mutants and locate their deletions (*26*). In particular, MFF deletions cluster near the PAR and Y4; GSF deletions are close to Slss, DD44, Cyp and Y6a; SPF deletions are close to Y6b.

One of the three GSF deletions includes *Clavata3* (*scaffold_1_000153*), the recently proposed GSF candidate gene (*16*). Differences in the balance of the Clavata-Wuschel pathway in males and females has been proposed to explain carpel formation/inhibition in female and male flowers as *Clavata3* (a carpel inhibitor) has a functional copy on the Y and a pseudogene on the X, while *Wuschel* (a carpel promoter) is present on the X and deleted from the Y (*27*). Consistent with this, both *Clavata3* and *Wuschel* are in stratum 1, as expected if both changed during the first step of sex-chromosome evolution, i.e., they support the model involving male-sterility and female-suppressing mutations (*28*). Figure 3B focuses on genes located in the deletions, combining information on coverage, gene expression during early or late flower development (Table S5) and functional annotation of sterility terms. We also considered genes that are absent in at least one mutant of a category, possibly explaining the observed phenotypic variability observed among the mutants. We found several MFF candidate genes. A notable MFF candidate is the gene *scaffold_1_000971* that encodes a cytochrome P450 protein and is homologous to the *Arabidopsis thaliana* Cyp704B1 gene, which is crucial for pollen maturation, is expressed in the tapetum, and involved in the sporopollenin synthesis (a pollen cell wall component). Its inactivation causes

male-sterility in *A. thaliana*. This gene is expressed in stage 8 but not stage 5 in *S. latifolia* male flower development consistent with a MFF gene. Another MFF candidate gene is *scaffold_1_003352* that encodes for a papain-like cysteine protease also expressed in tapetum and important for pollen maturation, through involvement in proteolysis and tapetal cell degeneration). It is also annotated as a male-sterility gene. No clear SPF candidate was found and this region is not shown in Figure 3B. *scaffold_1_000971* is probably located in stratum 1 while *scaffold_1_003352* is located in stratum 2. Our best MFF candidate is thus located in S1 as the best GSF one, which suggests that S1 might have formed when successive closely linked female-suppressing/male-enahncing mutations appeared during the evolution of dioecy (*28*).

Conclusions

We produced high-quality *S. latifolia* sex chromosome assemblies, which provided insights about their structure, function, and evolution. We found that the evolution of the *S. latifolia* sex chromosomes started ~11 MY ago with the differentiation of a small region including the GSF and MFF sex-determining genes. This first stratum probably formed by two paracentric inversions on both the X and the Y. More recently, another stratum including the centromere, S2, was formed ~5 MY ago. This generated a very large non-recombining region on the Y, in which a massive accumulation of TEs occurred, and from which they dispersed throughout the chromosome. These repeats and the lack of recombination probably allowed chromosomal rearrangements to occur on the Y. The absence of recombination also led to genetic degeneration, with as much as 58% of the genes on the X being lost from the Y. Interestingly, similar changes also affected the X chromosome, as its non-recombining pericentromeric region also expanded to a giant size via massive TE accumulation, which might have been driven by a reduced effective population size of the X chromosome (*29,30*).

References and Notes

1. Charlesworth, D., Charlesworth, B., & Marais, G. (2005). Steps in the evolution of heteromorphic sex chromosomes. *Heredity*, 95(2), 118–128. https://doi.org/10.1038/sj.hdy.6800697
2. Bachtrog, D., Mank, J. E., Peichel, C. L., Kirkpatrick, M., Otto, S. P., Ashman, T. L., Hahn, M. W., Kitano, J., Mayrose, I., Ming, R., Perrin, N., Ross, L., Valenzuela, N., Vamosi, J. C., & Tree of Sex Consortium (2014). Sex determination: why so many ways of doing it? *PLoS biology*, 12(7), e1001899. https://doi.org/10.1371/journal.pbio.1001899
3. Reviewed in Ming, R., Bendahmane, A., & Renner, S. S. (2011). Sex chromosomes in land plants. *Annual review of plant biology*, 62, 485–514. https://doi.org/10.1146/annurev-arplant-042110-103914
4. Mahajan, S., Wei, K. H., Nalley, M. J., Gibilisco, L., & Bachtrog, D. (2018). De novo assembly of a young Drosophila Y chromosome using single-molecule sequencing and chromatin conformation capture. *PLoS biology*, 16(7), e2006348. https://doi.org/10.1371/journal.pbio.2006348
5. Reviewed by Muyle, A., Shearn, R., & Marais, G. A. (2017). The Evolution of Sex Chromosomes and Dosage Compensation in Plants. *Genome biology and evolution*, 9(3), 627–645. https://doi.org/10.1093/gbe/evw282
6. Reviewed by Renner, S. S., & Müller, N. A. (2021). Plant sex chromosomes defy evolutionary models of expanding recombination suppression and genetic degeneration. *Nature plants*, 7(4), 392–402. https://doi.org/10.1038/s41477-021-00884-3
7. Bernasconi, G., Antonovics, J., Biere, A., Charlesworth, D., Delph, L. F., Filatov, D., Giraud, T., Hood, M. E., Marais, G. A., McCauley, D., Pannell, J. R., Shykoff, J. A., Vyskot, B., Wolfe, L. M., & Widmer, A. (2009). Silene as a model system in ecology and evolution. *Heredity*, 103(1), 5–14. https://doi.org/10.1038/hdy.2009.34
8. Blackburn, K. B. (1923). Sex chromosomes in plants. *Nature*, 112(2819), 687-688. https://doi.org/10.1038/112687c0
9. Matsunaga, S., Hizume, M., Kawano, S., & Kuroiwa, T. (1994). Cytological analyses in Melandrium album: genome size, chromosome size and fluorescence in situ hybridization. *Cytologia*, 59(1), 135-141. https://doi.org/10.1508/cytologia.59.135
10. Qiu, S., Bergero, R., Guirao-Rico, S., Campos, J. L., Cezard, T., Gharbi, K., & Charlesworth, D. (2016). RAD mapping reveals an evolving, polymorphic and fuzzy boundary of a plant pseudoautosomal region. *Molecular ecology*, 25(1), 414–430. https://doi.org/10.1111/mec.13297
11. Nicolas, M., Marais, G., Hykelova, V., Janousek, B., Laporte, V., Vyskot, B., Mouchiroud, D., Negrutiu, I., Charlesworth, D., & Monéger, F. (2005). A gradual process of recombination restriction in the evolutionary history of the sex chromosomes in dioecious plants. *PLoS biology*, 3(1), e4. https://doi.org/10.1371/journal.pbio.0030004
12. Filatov D. A. (2005). Evolutionary history of Silene latifolia sex chromosomes revealed by genetic mapping of four genes. *Genetics*, 170(2), 975–979. https://doi.org/10.1534/genetics.104.037069

13. Hobza, R., Cegan, R., Jesionek, W., Kejnovsky, E., Vyskot, B., & Kubat, Z. (2017). Impact of Repetitive Elements on the Y Chromosome Formation in Plants. *Genes*, 8(11), 302. https://doi.org/10.3390/genes8110302

14. Westergaard, M. (1958). The mechanism of sex determination in dioecious flowering plants. Advances in genetics, 9, 217-281. https://doi.org/10.1016/S0065-2660(08)60163-7

15. The presence of these three factors was confirmed by Bergero, R., Charlesworth, D., Filatov, D. A., & Moore, R. C. (2008). Defining regions and rearrangements of the *Silene latifolia* Y chromosome. *Genetics*, 178(4), 2045–2053. https://doi.org/10.1534/genetics.107.084566

16. Kazama, Y., Kitoh, M., Kobayashi, T., Ishii, K., Krasovec, M., Yasui, Y., Abe, T., Kawano, S., & Filatov, D. A. (2022). A CLAVATA3-like Gene Acts as a Gynoecium Suppression Function in White Campion. *Molecular biology and evolution*, 39(10), msac195. https://doi.org/10.1093/molbev/msac195

17. Gaut, B. S., Wright, S. I., Rizzon, C., Dvorak, J., & Anderson, L. K. (2007). Recombination: an underappreciated factor in the evolution of plant genomes. *Nature reviews Genetics*, 8(1), 77–84. https://doi.org/10.1038/nrg1970

18. Krasovec, M., Chester, M., Ridout, K., & Filatov, D. A. (2018). The Mutation Rate and the Age of the Sex Chromosomes in Silene latifolia. *Current biology*, 28(11), 1832–1838.e4. https://doi.org/10.1016/j.cub.2018.04.069

19. Bačovský, V., Čegan, R., Šimoníková, D., Hřibová, E., & Hobza, R. (2020). The Formation of Sex Chromosomes in *Silene latifolia* and *S. dioica* Was Accompanied by Multiple Chromosomal Rearrangements. *Frontiers in plant science*, 11, 205. https://doi.org/10.3389/fpls.2020.00205

20. Duhamel, M., Hood, M. E., Rodríguez de la Vega, R. C., Giraud, T. (2023). Dynamics of transposable element accumulation in the non-recombining regions of mating-type chromosomes in anther-smut fungi* co-senior authors. *Nature Communications*, in press.

21. Charlesworth, B., Sniegowski, P., & Stephan, W. (1994). The evolutionary dynamics of repetitive DNA in eukaryotes. *Nature*, 371(6494), 215–220. https://doi.org/10.1038/371215a0

22. Chibalina, M. V., & Filatov, D. A. (2011). Plant Y chromosome degeneration is retarded by haploid purifying selection. *Current biology*, 21(17), 1475–1479. https://doi.org/10.1016/j.cub.2011.07.045

23. Bergero, R., & Charlesworth, D. (2011). Preservation of the Y transcriptome in a 10-million-year-old plant sex chromosome system. *Current biology*, 21(17), 1470–1474. https://doi.org/10.1016/j.cub.2011.07.032

24. Muyle, A., Zemp, N., Deschamps, C., Mousset, S., Widmer, A., & Marais, G. A. (2012). Rapid de novo evolution of X chromosome dosage compensation in Silene latifolia, a plant with young sex chromosomes. *PLoS biology*, 10(4), e1001308. https://doi.org/10.1371/journal.pbio.1001308

25. Zluvova, J., Georgiev, S., Janousek, B., Charlesworth, D., Vyskot, B., & Negrutiu, I. (2007). Early events in the evolution of the Silene latifolia Y chromosome: male specialization and recombination arrest. *Genetics*, 177(1), 375–386. https://doi.org/10.1534/genetics.107.071175

26. Bergero, R., Charlesworth, D., Filatov, D. A., & Moore, R. C. (2008). Defining regions and rearrangements of the Silene latifolia Y chromosome. *Genetics*, 178(4), 2045–2053. https://doi.org/10.1534/genetics.107.084566

27. Kazama, Y., Nishihara, K., Bergero, R., Fujiwara, M. T., Abe, T., Charlesworth, D., & Kawano, S. (2012). SlWUS1; an X-linked gene having no homologous Y-linked copy in *Silene latifolia*. *G3* (Bethesda, Md.), 2(10), 1269–1278. https://doi.org/10.1534/g3.112.003749

28. Charlesworth, B., & Charlesworth, D. (1978). A model for the evolution of dioecy and gynodioecy. *The American Naturalist*, 112(988), 975-997.

29. Wang, J., Na, J. K., Yu, Q., Gschwend, A. R., Han, J., Zeng, F., Aryal, R., VanBuren, R., Murray, J. E., Zhang, W., Navajas-Pérez, R., Feltus, F. A., Lemke, C., Tong, E. J., Chen, C., Wai, C. M., Singh, R., Wang, M. L., Min, X. J., Alam, M., … Ming, R. (2012). Sequencing papaya X and Yh chromosomes reveals molecular basis of incipient sex chromosome evolution. *Proceedings of the National Academy of Sciences of the United States of America*, 109(34), 13710–13715. https://doi.org/10.1073/pnas.1207833109

30. Bellott, D. W., Skaletsky, H., Cho, T. J., Brown, L., Locke, D., Chen, N., Galkina, S., Pyntikova, T., Koutseva, N., Graves, T., Kremitzki, C., Warren, W. C., Clark, A. G., Gaginskaya, E., Wilson, R. K., & Page, D. C. (2017). Avian W and mammalian Y chromosomes convergently retained dosage-sensitive regulators. *Nature genetics*, 49(3), 387–394. https://doi.org/10.1038/ng.3778

31. Lebel-Hardenack, S., Hauser, E., Law, T. F., Schmid, J., & Grant, S. R. (2002). Mapping of sex determination loci on the white campion (*Silene latifolia*) Y chromosome using amplified fragment length polymorphism. *Genetics*, 160(2), 717–725. https://doi.org/10.1093/genetics/160.2.717

32. Zemp, N., Tavares, R., Muyle, A., Charlesworth, D., Marais, G. A., & Widmer, A. (2016). Evolution of sex-biased gene expression in a dioecious plant. *Nature plants*, 2(11), 16168. https://doi.org/10.1038/nplants.2016.168

33. Alberti, A., Poulain, J., Engelen, S., Labadie, K., Romac, S., Ferrera, I., Albini, G., Aury, J. M., Belser, C., Bertrand, A., Cruaud, C., Da Silva, C., Dossat, C., Gavory, F., Gas, S., Guy, J., Haquelle, M., Jacoby, E., Jaillon, O., Lemainque, A., … Wincker, P. (2017). Viral to metazoan marine plankton nucleotide sequences from the Tara Oceans expedition. *Scientific data*, 4, 170093. https://doi.org/10.1038/sdata.2017.93

34. Engelen S, Aury JM. fastxtend. https://www.genoscope.cns.fr/fastxtend

35. Li, R., Li, Y., Kristiansen, K., & Wang, J. (2008). SOAP: short oligonucleotide alignment program. *Bioinformatics* (Oxford, England), 24(5), 713–714. https://doi.org/10.1093/bioinformatics/btn025

36. Bolger, A. M., Lohse, M., & Usadel, B. (2014). Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* (Oxford, England), 30(15), 2114–2120. https://doi.org/10.1093/bioinformatics/btu170

37. Papadopulos, A. S., Chester, M., Ridout, K., & Filatov, D. A. (2015). Rapid Y degeneration and dosage compensation in plant sex chromosomes. *Proceedings of the National Academy of Sciences of the United States of America*, 112(42), 13021–13026. https://doi.org/10.1073/pnas.1508454112

38. Langmead, B., & Salzberg, S. L. (2012). Fast gapped-read alignment with Bowtie 2. *Nature methods*, 9(4), 357–359. https://doi.org/10.1038/nmeth.1923

39. Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G., Durbin, R., & 1000 Genome Project Data Processing Subgroup (2009). The Sequence Alignment/Map format and SAMtools. *Bioinformatics* (Oxford, England), 25(16), 2078–2079. https://doi.org/10.1093/bioinformatics/btp352

40. Koboldt, D. C., Chen, K., Wylie, T., Larson, D. E., McLellan, M. D., Mardis, E. R., Weinstock, G. M., Wilson, R. K., & Ding, L. (2009). VarScan: variant detection in massively parallel sequencing of individual and pooled samples. *Bioinformatics* (Oxford, England), 25(17), 2283–2285. https://doi.org/10.1093/bioinformatics/btp373

41. Rastas, P., Paulin, L., Hanski, I., Lehtonen, R., & Auvinen, P. (2013). Lep-MAP: fast and accurate linkage map construction for large SNP datasets. *Bioinformatics* (Oxford, England), 29(24), 3128–3134. https://doi.org/10.1093/bioinformatics/btt563

42. Bergero, R., Qiu, S., Forrest, A., Borthwick, H., & Charlesworth, D. (2013). Expansion of the pseudo-autosomal region and ongoing recombination suppression in the Silene latifolia sex chromosomes. *Genetics*, 194(3), 673–686. https://doi.org/10.1534/genetics.113.150755

43. Kolmogorov, M., Yuan, J., Lin, Y., & Pevzner, P. A. (2019). Assembly of long, error-prone reads using repeat graphs. *Nature biotechnology*, 37(5), 540–546. https://doi.org/10.1038/s41587-019-0072-8

44. Walker, B. J., Abeel, T., Shea, T., Priest, M., Abouelliel, A., Sakthikumar, S., Cuomo, C. A., Zeng, Q., Wortman, J., Young, S. K., & Earl, A. M. (2014). Pilon: an integrated tool for comprehensive microbial variant detection and genome assembly improvement. *PloS one*, 9(11), e112963. https://doi.org/10.1371/journal.pone.0112963

45. Vaser, R., Sović, I., Nagarajan, N., & Šikić, M. (2017). Fast and accurate de novo genome assembly from long uncorrected reads. *Genome research*, 27(5), 737–746. https://doi.org/10.1101/gr.214270.116

46. Warren, R. L., Coombe, L., Mohamadi, H., Zhang, J., Jaquish, B., Isabel, N., Jones, S. J. M., Bousquet, J., Bohlmann, J., & Birol, I. (2019). ntEdit: scalable genome sequence polishing. *Bioinformatics* (Oxford, England), 35(21), 4430–4432. https://doi.org/10.1093/bioinformatics/btz400

47. Manni, M., Berkeley, M. R., Seppey, M., Simão, F. A., & Zdobnov, E. M. (2021). BUSCO Update: Novel and Streamlined Workflows along with Broader and Deeper Phylogenetic Coverage for Scoring of Eukaryotic, Prokaryotic, and Viral Genomes. *Molecular biology and evolution*, 38(10), 4647–4654. https://doi.org/10.1093/molbev/msab199

48. Roach, M. J., Schmidt, S. A., & Borneman, A. R. (2018). Purge Haplotigs: allelic contig reassignment for third-gen diploid genome assemblies. *BMC bioinformatics*, 19(1), 460. https://doi.org/10.1186/s12859-018-2485-7

49. Istace, B., Belser, C., & Aury, J. M. (2020). BiSCoT: improving large eukaryotic genome assemblies with optical maps. *PeerJ*, 8, e10150. https://doi.org/10.7717/peerj.10150

50. Aury, J. M., & Istace, B. (2021). Hapo-G, haplotype-aware polishing of genome assemblies with accurate reads. *NAR genomics and bioinformatics*, 3(2), lqab034. https://doi.org/10.1093/nargab/lqab034

51. Ghurye, J., Pop, M., Koren, S., Bickhart, D., & Chin, C. S. (2017). Scaffolding of long read assemblies using long range contact information. *BMC genomics*, 18(1), 527. https://doi.org/10.1186/s12864-017-3879-z

52. Tang, H., Zhang, X., Miao, C., Zhang, J., Ming, R., Schnable, J. C., Schnable, P. S., Lyons, E., & Lu, J. (2015). ALLMAPS: robust scaffold ordering based on multiple maps. Genome biology, 16(1), 3. https://doi.org/10.1186/s13059-014-0573-1

53. Brůna, T., Hoff, K. J., Lomsadze, A., Stanke, M., & Borodovsky, M. (2021). BRAKER2: automatic eukaryotic genome annotation with GeneMark-EP+ and AUGUSTUS supported by a protein database. *NAR genomics and bioinformatics*, 3(1), lqaa108. https://doi.org/10.1093/nargab/lqaa108

54. Stiehler, F., Steinborn, M., Scholz, S., Dey, D., Weber, A. P. M., & Denton, A. K. (2021). Helixer: cross-species gene annotation of large eukaryotic genomes using deep learning. *Bioinformatics* (Oxford, England), 36(22-23), 5291–5298. https://doi.org/10.1093/bioinformatics/btaa1044

55. Tarailo-Graovac, M., & Chen, N. (2009). Using RepeatMasker to identify repetitive elements in genomic sequences. *Current protocols in bioinformatics*, Chapter 4, 4.10.1–4.10.14. https://doi.org/10.1002/0471250953.bi0410s25

56. Flynn, J. M., Hubley, R., Goubert, C., Rosen, J., Clark, A. G., Feschotte, C., & Smit, A. F. (2020). RepeatModeler2 for automated genomic discovery of transposable element families. *Proceedings of the National Academy of Sciences of the United States of America*, 117(17), 9451–9457. https://doi.org/10.1073/pnas.1921046117

57. Chen, S. (2023). Ultrafast one-pass FASTQ data preprocessing, quality control, and deduplication using fastp. iMeta, e107. https://doi.org/10.1002/imt2.107

58. Wu, T. D., Reeder, J., Lawrence, M., Becker, G., & Brauer, M. J. (2016). GMAP and GSNAP for Genomic Sequence Alignment: Enhancements to Speed, Accuracy, and Functionality. *Methods in molecular biology* (Clifton, N.J.), 1418, 283–334. https://doi.org/10.1007/978-1-4939-3578-9_15

59. Stanke, M., Keller, O., Gunduz, I., Hayes, A., Waack, S., & Morgenstern, B. (2006). AUGUSTUS: ab initio prediction of alternative transcripts. *Nucleic acids research*, 34(Web Server issue), W435–W439. https://doi.org/10.1093/nar/gkl200

60. Buchfink, B., Xie, C., & Huson, D. H. (2015). Fast and sensitive protein alignment using DIAMOND. *Nature methods*, 12(1), 59–60. https://doi.org/10.1038/nmeth.3176

61. Törönen, P., Medlar, A., & Holm, L. (2018). PANNZER2: a rapid functional annotation web server. *Nucleic acids research*, 46(W1), W84–W88. https://doi.org/10.1093/nar/gky350

62. 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., Jiang, N., Hirsch, C. N., & Hufford, M. B. (2019). Benchmarking transposable element annotation methods for creation of a streamlined, comprehensive pipeline. *Genome biology*, 20(1), 275. https://doi.org/10.1186/s13059-019-1905-y

63. Cermak, T., Kubat, Z., Hobza, R., Koblizkova, A., Widmer, A., Macas, J., Vyskot, B., & Kejnovsky, E. (2008). Survey of repetitive sequences in Silene latifolia with respect to their

distribution on sex chromosomes. *Chromosome research*, 16(7), 961–976. https://doi.org/10.1007/s10577-008-1254-2

64. Macas, J., Kejnovský, E., Neumann, P., Novák, P., Koblížková, A., & Vyskot, B. (2011). Next generation sequencing-based analysis of repetitive DNA in the model dioecious [corrected] plant *Silene latifolia*. *PloS one*, 6(11), e27335. https://doi.org/10.1371/journal.pone.0027335

65. Fields, P. D., Weber, M. M., Waneka, G., Broz, A. K., & Sloan, D. B. (2023). Chromosome-level genome assembly for the angiosperm *Silene conica*. *bioRxiv*, 2023-09. https://doi.org/10.1101/2023.09.05.556365

66. Emms, D. M., & Kelly, S. (2019). OrthoFinder: phylogenetic orthology inference for comparative genomics. *Genome biology*, 20(1), 238. https://doi.org/10.1186/s13059-019-1832-y

67. Camacho, C., Coulouris, G., Avagyan, V., Ma, N., Papadopoulos, J., Bealer, K., & Madden, T. L. (2009). BLAST+: architecture and applications. *BMC bioinformatics*, 10, 421. https://doi.org/10.1186/1471-2105-10-421

68. Yang Z. (2007). PAML 4: phylogenetic analysis by maximum likelihood. *Molecular biology and evolution*, 24(8), 1586–1591. https://doi.org/10.1093/molbev/msm088

69. Abascal, F., Zardoya, R., & Telford, M. J. (2010). TranslatorX: multiple alignment of nucleotide sequences guided by amino acid translations. *Nucleic acids research*, 38(Web Server issue), W7–W13. https://doi.org/10.1093/nar/gkq291

70. Edgar R. C. (2004). MUSCLE: multiple sequence alignment with high accuracy and high throughput. Nucleic acids research, 32(5), 1792–1797. https://doi.org/10.1093/nar/gkh340

71. Lindeløv, J. K. (2020, January 5). mcp: An R Package for Regression With Multiple Change Points. https://doi.org/10.31219/osf.io/fzqxv

72. Richards, C. M., Emery, S. N., & McCauley, D. E. (2003). Genetic and demographic dynamics of small populations of Silene latifolia. *Heredity*, 90(2), 181–186. https://doi.org/10.1038/sj.hdy.6800214

73. Ranwez, V., Douzery, E. J. P., Cambon, C., Chantret, N., & Delsuc, F. (2018). MACSE v2: Toolkit for the Alignment of Coding Sequences Accounting for Frameshifts and Stop Codons. *Molecular biology and evolution*, 35(10), 2582–2584. https://doi.org/10.1093/molbev/msy159

74. Following Marais, G. A., Nicolas, M., Bergero, R., Chambrier, P., Kejnovsky, E., Monéger, F., Hobza, R., Widmer, A., & Charlesworth, D. (2008). Evidence for degeneration of the Y chromosome in the dioecious plant *Silene latifolia*. *Current biology*, 18(7), 545–549. https://doi.org/10.1016/j.cub.2008.03.023

75. Álvarez-Carretero, S., Kapli, P., & Yang, Z. (2023). Beginner's Guide on the Use of PAML to Detect Positive Selection. *Molecular biology and evolution*, 40(4), msad041. https://doi.org/10.1093/molbev/msad041

76. Mendes, F. K., Vanderpool, D., Fulton, B., & Hahn, M. W. (2021). CAFE 5 models variation in evolutionary rates among gene families. *Bioinformatics* (Oxford, England), 36(22-23), 5516–5518. https://doi.org/10.1093/bioinformatics/btaa1022

77. Kozlov, A. M., Darriba, D., Flouri, T., Morel, B., & Stamatakis, A. (2019). RAxML-NG: a fast, scalable and user-friendly tool for maximum likelihood phylogenetic inference. *Bioinformatics* (Oxford, England), 35(21), 4453–4455. https://doi.org/10.1093/bioinformatics/btz305

78. Kumar, S., Suleski, M., Craig, J. M., Kasprowicz, A. E., Sanderford, M., Li, M., Stecher, G., & Hedges, S. B. (2022). TimeTree 5: An Expanded Resource for Species Divergence Times. *Molecular biology and evolution*, 39(8), msac174. Advance online publication. https://doi.org/10.1093/molbev/msac174

79. Li H. (2018). Minimap2: pairwise alignment for nucleotide sequences. *Bioinformatics* (Oxford, England), 34(18), 3094–3100. https://doi.org/10.1093/bioinformatics/bty191

80. Lovell, J. T., Sreedasyam, A., Schranz, M. E., Wilson, M., Carlson, J. W., Harkess, A., Emms, D., Goodstein, D. M., & Schmutz, J. (2022). GENESPACE tracks regions of interest and gene copy number variation across multiple genomes. *eLife*, 11, e78526. https://doi.org/10.7554/eLife.78526

81. Larget, B., Kadane, J. B., & Simon, D. L. (2005). A Bayesian approach to the estimation of ancestral genome arrangements. *Molecular phylogenetics and evolution*, 36(2), 214–223. https://doi.org/10.1016/j.ympev.2005.03.026

82. Miklós, I., & Tannier, E. (2010). Bayesian sampling of genomic rearrangement scenarios via double cut and join. *Bioinformatics* (Oxford, England), 26(24), 3012–3019. https://doi.org/10.1093/bioinformatics/btq574

83. Liu, X., Glémin, S., & Karrenberg, S. (2020). Evolution of putative barrier loci at an intermediate stage of speciation with gene flow in campions (*Silene*). *Molecular ecology*, 29(18), 3511–3525. https://doi.org/10.1111/mec.15571

84. Rochette, N. C., Rivera-Colón, A. G., & Catchen, J. M. (2019). Stacks 2: Analytical methods for paired-end sequencing improve RADseq-based population genomics. *Molecular ecology*, 28(21), 4737–4754. https://doi.org/10.1111/mec.15253

85. Li, H. (2013). Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. *arXiv* preprint arXiv:1303.3997. https://doi.org/10.48550/arXiv.1303.3997

86. Garrison, E., & Marth, G. (2012). Haplotype-based variant detection from short-read sequencing. *arXiv* preprint arXiv:1207.3907. https://doi.org/10.48550/arXiv.1207.3907

87. O'Leary, S. J., Puritz, J. B., Willis, S. C., Hollenbeck, C. M., & Portnoy, D. S. (2018). These aren't the loci you'e looking for: Principles of effective SNP filtering for molecular ecologists. *Molecular ecology*, 10.1111/mec.14792. Advance online publication. https://doi.org/10.1111/mec.14792

88. Korunes, K. L., & Samuk, K. (2021). pixy: Unbiased estimation of nucleotide diversity and divergence in the presence of missing data. *Molecular ecology resources*, 21(4), 1359–1368. https://doi.org/10.1111/1755-0998.13326

89. Andrews, S. (2010). FastQC: A Quality Control Tool for High Throughput Sequence Data. Available online at: http://www.bioinformatics.babraham.ac.uk/projects/fastqc/

90. Martin, M. (2011). Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet journal*, 17(1), 10-12. https://doi.org/10.14806/ej.17.1.200

91. Shahid, S., & Axtell, M. J. (2014). Identification and annotation of small RNA genes using ShortStack. *Methods* (San Diego, Calif.), 67(1), 20–27. https://doi.org/10.1016/j.ymeth.2013.10.004

92. Johnson, N. R., Yeoh, J. M., Coruh, C., & Axtell, M. J. (2016). Improved Placement of Multi-mapping Small RNAs. *G3* (Bethesda, Md.), 6(7), 2103–2111. https://doi.org/10.1534/g3.116.030452

93. Quinlan, A. R., & Hall, I. M. (2010). BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics* (Oxford, England), 26(6), 841–842. https://doi.org/10.1093/bioinformatics/btq033

94. Krueger, F. (2015). Trim Galore!: A wrapper around Cutadapt and FastQC to consistently apply adapter and quality trimming to FastQ files, with extra functionality for RRBS data. Babraham Institute.

95. Krueger, F., & Andrews, S. R. (2011). Bismark: a flexible aligner and methylation caller for Bisulfite-Seq applications. *Bioinformatics* (Oxford, England), 27(11), 1571–1572. https://doi.org/10.1093/bioinformatics/btr167

96. Bray, N. L., Pimentel, H., Melsted, P., & Pachter, L. (2016). Near-optimal probabilistic RNA-seq quantification. *Nature biotechnology*, 34(5), 525–527. https://doi.org/10.1038/nbt.3519

97. Musich, R., Cadle-Davidson, L., & Osier, M. V. (2021). Comparison of Short-Read Sequence Aligners Indicates Strengths and Weaknesses for Biologists to Consider. *Frontiers in plant science*, 12, 657240. https://doi.org/10.3389/fpls.2021.657240

98. Zluvova, J., Nicolas, M., Berger, A., Negrutiu, I., & Monéger, F. (2006). Premature arrest of the male flower meristem precedes sexual dimorphism in the dioecious plant Silene latifolia. *Proceedings of the National Academy of Sciences of the United States of America*, 103(49), 18854–18859. https://doi.org/10.1073/pnas.0606622103

99. Dobin, A., Davis, C. A., Schlesinger, F., Drenkow, J., Zaleski, C., Jha, S., Batut, P., Chaisson, M., & Gingeras, T. R. (2013). STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* (Oxford, England), 29(1), 15–21. https://doi.org/10.1093/bioinformatics/bts635

100. Liao, Y., Smyth, G. K., & Shi, W. (2014). featureCounts: an efficient general purpose program for assigning sequence reads to genomic features. *Bioinformatics* (Oxford, England), 30(7), 923–930. https://doi.org/10.1093/bioinformatics/btt656

101. Love, M. I., Huber, W., & Anders, S. (2014). Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome biology*, 15(12), 550. https://doi.org/10.1186/s13059-014-0550-8

102. Wu, T., Hu, E., Xu, S., Chen, M., Guo, P., Dai, Z., Feng, T., Zhou, L., Tang, W., Zhan, L., Fu, X., Liu, S., Bo, X., & Yu, G. (2021). clusterProfiler 4.0: A universal enrichment tool for interpreting omics data. *Innovation* (Cambridge (Mass.)), 2(3), 100141. https://doi.org/10.1016/j.xinn.2021.100141

Acknowledgments: We thank Elise Lacroix technician at Univ. Lyon 1 glasshouse for her help for growing the *S. latifolia* U17 plants.

Funding:

Agence Nationale de la Recherche (ANR) grant ANR-20-CE20-0015-01 (GABM, ABe).

Fédération de Recherche "Biodiversité, Eau, Environnement, Ville & Santé" (FR BIOEENVIS) of University of Lyon 1 grant (GABM, JB)

The Czech Science Foundation grants 21-00580S and 22-00364S (PJ)

Author contributions:

Conceptualization: GABM, RH, AMa, ADG, ABe, TG, DC, AMu

Formal analysis: CM, CB, QR, PV, PJ, AMu, MH, ET, XL, EMG, CL, PDF, CBe, RCe, RRV, GA, AEF, VM, AV, SK, ADG, AMa

Methodology: CM, CB, RL, ADG, GABM

Investigation: CM, CB, QR, PV, PJ, AMu, MH, ET, XL, EMG, CL, PDF, CC, KL, CBe, JB, SS, RCe, RL, RRV, GA, ABo, RT, DBS, SK, LFD, DC, TG, ABe, ADG, AMa, RH, GABM

Resources: RH, ADG, LFD, DBS, MSO, DRT, OAJA, MK, HS, HSi, PC, ZT, HT, LW, SD, BS, DP AMu

Software: CM, CB, ET, RL, RRV, AV, ADG

Validation: AC, RCo, CF

Visualization: CM, CB, QR, PV, AMu, ET, RL, RRV, DP, SK, AMa,

Funding acquisition: GABM, ABe

Project administration: GABM

Supervision: GABM, RH, AMa, ADG, ABe, TG, LFD, SK, DBS, RCo, RT, ABo, RRV, RL, AMu

Writing – original draft: GABM, AMa, ABe, SK, DBS, SS, JB, PDF, XL, ET, AMa, QR, CB, CM

Writing – review & editing: GABM, TG, DC, LFD, SK, ABo, RRV, AMu, PV

Competing interests: Authors declare that they have no competing interests

Data and materials availability: All data will be available on NCBI and codes on github.

Supplementary Materials

Materials and Methods

Figs. S1 to S11

Tables S1 to S5

References (*31–102*)

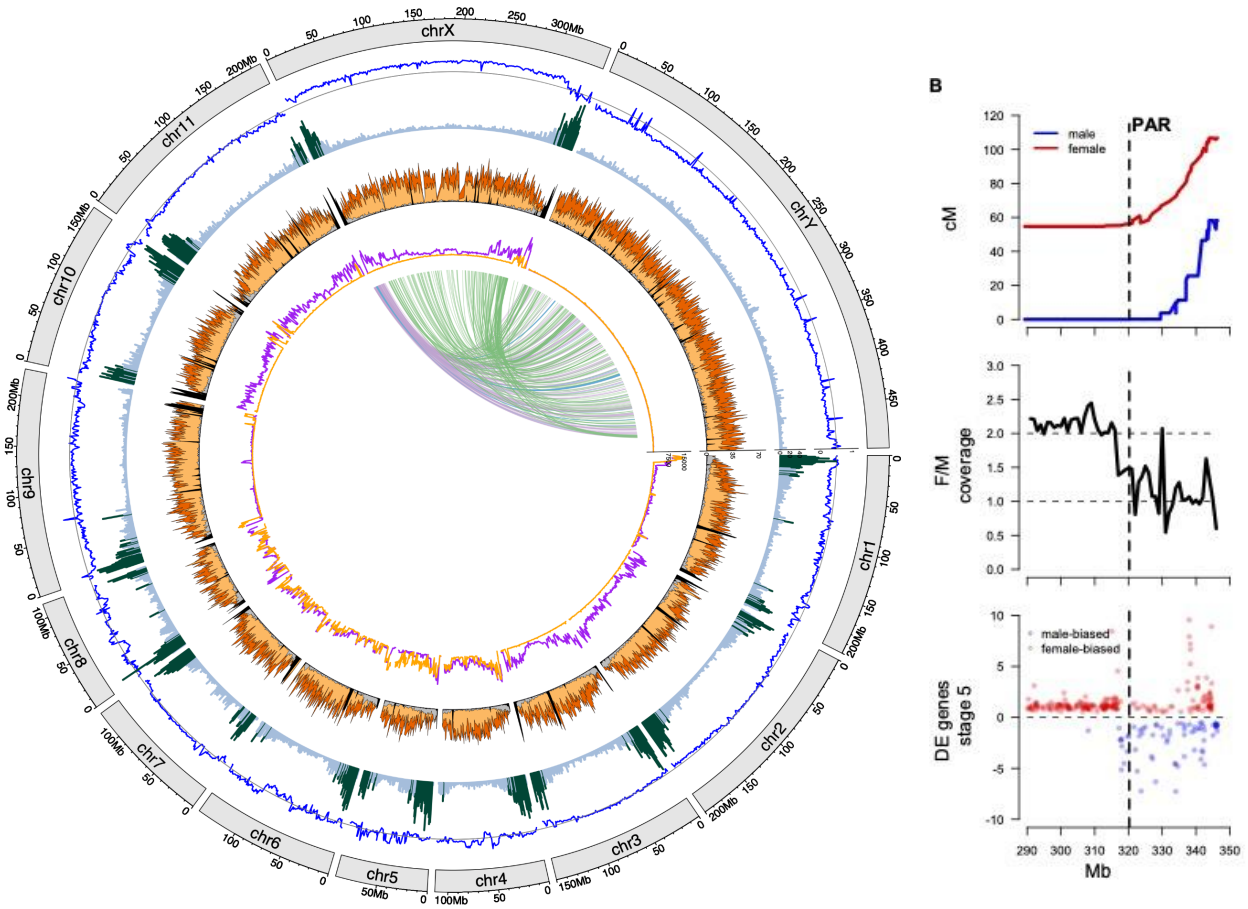
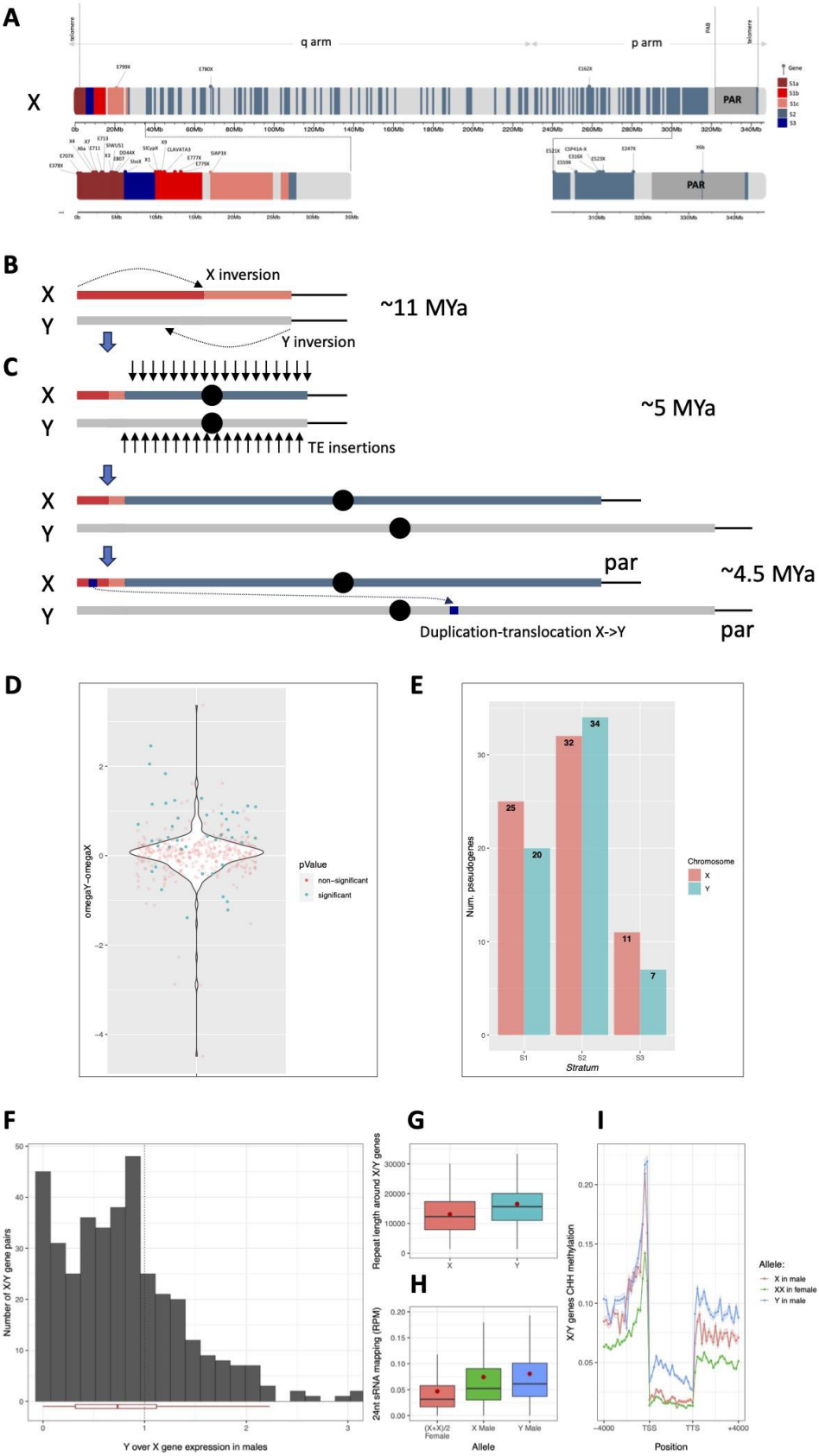


Fig. 1. View of the assembly of the *S. latifolia* male genome assembly. (A) Circos plot of the *S. latifolia* male genome**.** Circles included in the Circos plot are from outer to inner circles: 1. The coverage Female/Male ratio distribution. 2. Gene density highlights the most populated regions (dark green); they are defined as those with a density larger than the average plus 1 sd. 3. Family repeats density distribution showing satellite elements (black), LTR elements: Ty3/Gypsy (orange), Ty1/Copia (yellow), LINE (violet), and Helitron (grey). 4. SNP Female and male distribution (purple for female and orange for male). This analysis is consistent with the sequenced male being highly homozygous, although some chromosomes show heterozygosity. It is also consistent with the female being a sister of the male. 5. Rearrangements between X and Y chromosomes based on the gametologs. (B) Zoom in on the X chromosome showing recombination in males (blue) and females (red) defining the pseudo-autosomal boundary at 321 Mb, female/male sequence coverage ratio, and significant differential expression between male and female flowers (stage 5); the full analysis of differential gene expression is shown in Figure S11. All panels have data summarised in 1Mb windows.



1

Fig. 2. The evolution of the sex chromosomes. (A) The X chromosome sequence. The 417 gametologs genes are placed and colored according to their evolutionary stratum (S1a, S1b, S1c, S2 and S3). Zoom in of the extremities of the X chromosome are also shown. Genes previously characterized in the literature are identified by a ball-headed pin colored according to their stratum and located over the X chromosome sequence. (B) Schematic view of the first step of the evolution of the sex chromosomes with a zoom in on the first 35 Mb of the p arm of the X chromosome. The formation of stratum 1 (from S1a to S1c) is shown. Strata are not represented on the Y. (C) Global view of the other steps of the evolution of the sex chromosomes (full chromosomes are shown). (D) Violin plot of dN/dS differences among X and Y copies, gametologs pairs with significant differences (p-value < 0.05) are shown in blue. (E) Genes with premature stop codons on the sex chromosomes shown by strata. (F) Distribution of Y over X gene expression ratio in four *S. latifolia* males in flower buds. The boxplots below the distribution represent the median, the first and third quartiles (hinges) and 1.5 times the distance between the first and third quartiles (whiskers). (G) Sum of repeat lengths around X/Y gene pairs, from 4000 bp upstream to 4000 bp downstream of the gene. If repeats went beyond the borders (+/- 4000bp), their length was not included. (H) Mapping of 24nt small RNA (in reads per million mapped reads, abbreviated RPM) on X/Y gene pairs for three *S. latifolia* females and three males in flower buds and leaves. The boxplots represent the median, the first and third quartiles (hinges) and 1.5 times the distance between the first and third quartiles (whiskers). The red dot stands for the mean. Both X alleles in females are represented in red and their sRNA mapping was divided by two for stochiometric comparison to the X allele in males (represented in green) and the Y allele in males (represented in blue). (I) Plot of X/Y genes DNA methylation in CHH context. Both X alleles in the female are represented in red, the X allele in the male in green and the Y allele in the male in blue. All X/Y genes were combined to plot the average proportion of methylated reads at cytosine positions along sliding windows; the 95% confidence interval is represented as a small ribbon around the curve. Twenty windows of 200 bp were studied upstream of the transcription start site (TSS) and twenty windows of 200 bp were studied downstream of the transcription termination site (TTS). The gene body (from TSS to TTS) was divided into twenty windows of equal size.

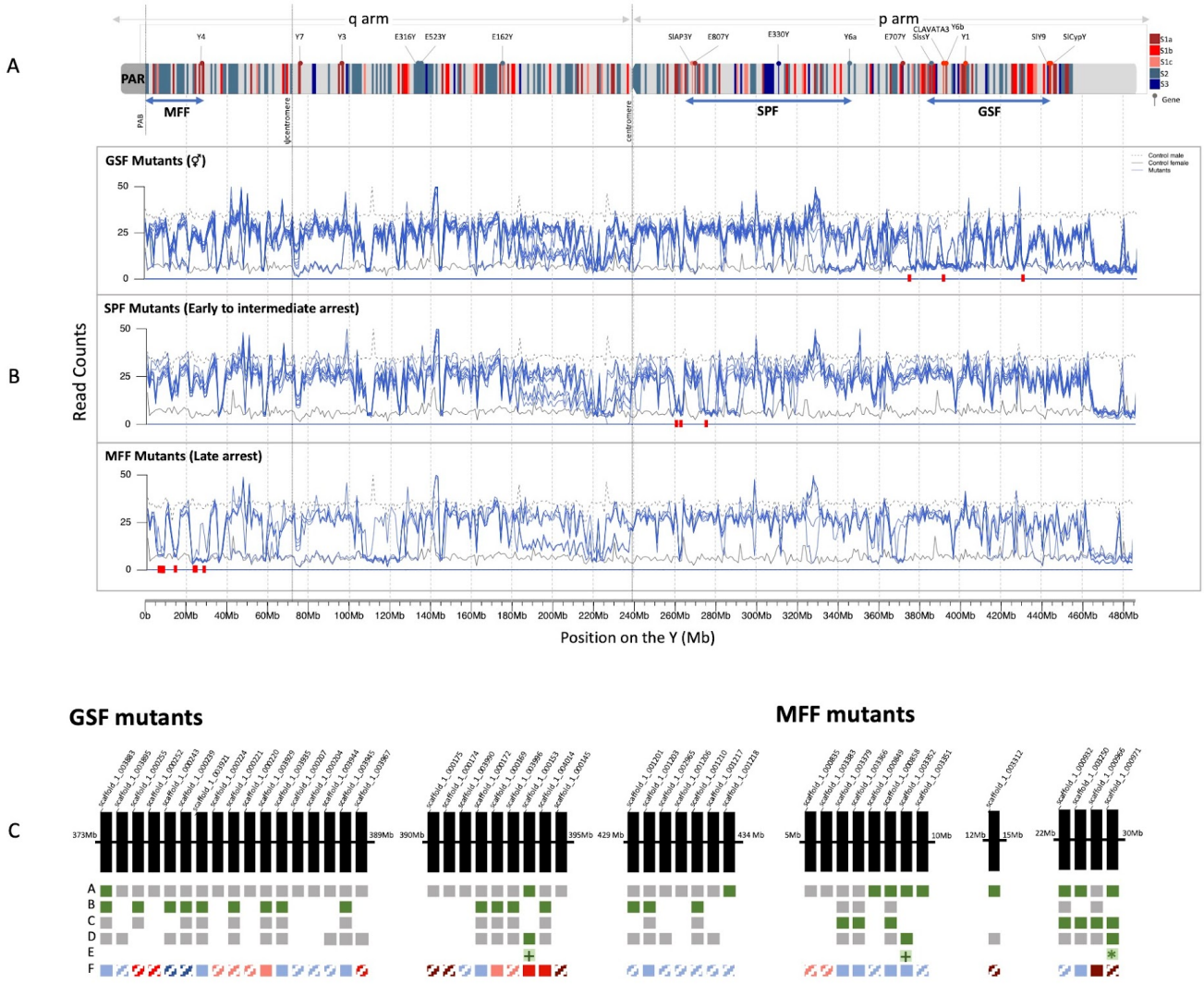


Fig. 3. Y deletion mutants analysis. (A) The Y chromosome sequence and its putative sex-determining regions (MFF, SPF, and GSF) with estimated location from Bergero et al. (2008). The 417 gametologs genes are placed and colored according to their evolutionary stratum (S1a, S1b, S1c, S2 and S3). Additionally, genes previously characterized in the literature are identified by a ball-headed pin colored according to their stratum and located over the Y chromosome sequence. (B) Each plot shows the read count of all individuals (blue) grouped by mutant's phenotypic category (from top to bottom: GSF mutants, SPF mutants, and MFF mutants) after mapping their reads onto the reference Y chromosome. The read count of the female and male control are also present in the plot (solid and dashed black line, respectively) and coincide with the maximum and the minimum (male and female values, respectively) read count of the mutants. Red rectangles depict genes that are deleted in all mutants of a given phenotypic category and present in all the remaining mutants (i.e., phenotype-specific deleted genes): 3 in GSF mutants at 370–440Mb, 3 in SPF mutants at 260Mb–280Mb and 8 in MFF mutants at 5Mb–30Mb. The maximum number of reads was set to 50. Window size=1Mb. (C) Phenotype-specific deleted genes and their neighbor genes with a relevant presence/absence pattern within the low covered regions in the GSF and MFF mutants' phenotypic category: row A concerns the

presence/absence of the gene among mutants, in which green is for phenotype-specific deleted genes and gray is for genes deleted in at least one mutant of the category of interest and present in all the remaining mutants; row B and C concern the expression of each gene in a normal male at stage 5 and 8 of development, respectively: green is for genes expressed in the stage expected for sex determining genes (stage 5 for GSF genes or stage 8 for MFF genes), gray is for genes that do not follow the expected pattern of expression in sex determining genes, and blank means the gene is not expressed in any of the stages; row D concerns the functional annotation of each gene: green is for genes annotated with a sex determining function, gray is for genes annotated with no clear sex determining function, and blank is for genes without available functional annotation; row E regards they possible role as a sex determining gene, in which we highlight with asterisk (*) a very good candidate to sex-determining gene and with a plus signal (+) good candidates to sex-determining genes; row F indicates the stratum of each gene: brown for S1a, red for S1b, pink for S1c, light blue for S3 (for the genes with striped squares the stratum was inferred from the closest genes).

Table 1. Statistics for the genome and the sex chromosomes. All assembly metrics were computed by the total assembly and sex chromosomes.

	Whole genome	X chromosome	Y chromosome
Total assembled size (bp)	2,739,705,433	346,484,273	497,814,031
Number of contigs	1553	55	29
Number of scaffolds	929	21	4
N50 (bp)	100,527,128	50,520,191	364,495,762
N90 (bp)	24,182,774	21,428,974	364,495,762
Largest length size (bp)	364,495,762	133,616,578	364,495,762
Anchored of the total sequences (anchoring rate)	2,586,997,243 (95%)		
Annotated protein-coding genes	35,436	3,520	2,344
Mean gene length (bp)	4,155	4,332	4,045
BUSCO score	86.5%		
Identified repeats	79.24%	77.75%	81.38%
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%