

# Current Biology

## The mutation rate and the age of the sex chromosomes in *Silene latifolia*

--Manuscript Draft--

Manuscript Number:	CURRENT-BIOLOGY-D-18-00340R1
Full Title:	The mutation rate and the age of the sex chromosomes in <i>Silene latifolia</i>
Article Type:	Report
Corresponding Author:	Dmitry A Filatov, PhD University of Oxford Oxford, UNITED KINGDOM
First Author:	Marc Krasovec
Order of Authors:	Marc Krasovec
	Michael Chester
	Kate Ridout
	Dmitry A Filatov
Abstract:	<p>Many aspects of sex chromosome evolution are common to both plants and animals [1], but the process of Y chromosome degeneration, where genes on the Y become non-functional over time, may be much slower in plants due to purifying selection against deleterious mutations in the haploid gametophyte [2, 3]. Testing for differences in Y degeneration between the kingdoms has been hindered by the absence of accurate age estimates for plant sex chromosomes. Here we used genome resequencing to estimate the spontaneous mutation rate and the age of the sex chromosomes in white campion (<i>Silene latifolia</i>). Screening of single nucleotide polymorphisms (SNPs) in parents and 10 F1 progeny identified 39 de novo mutations and yielded a rate of <math>7.31 \times 10^{-9}</math> (95% CI: <math>5.20 \times 10^{-9}</math> - <math>8.00 \times 10^{-9}</math>) mutations per site per haploid genome per generation. Applying this mutation rate to the synonymous divergence between homologous X- and Y-linked genes (gametologs) gave age estimates of 11.00 and 6.32 million years, for the old and young strata, respectively. Based on SNP segregation patterns, we inferred which genes were Y-linked and found that at least 47% are already dysfunctional. Applying our new estimates for the age of the sex chromosomes indicates that the rate of Y degeneration in <i>S. latifolia</i> is nearly two-fold slower when compared to animal sex chromosomes of a similar age. Our revised estimates support Y degeneration taking place more slowly in plants, a discrepancy that may be explained by differences in the lifecycles of animals and plants.</p>

# The mutation rate and the age of the sex chromosomes in *Silene latifolia*

Marc Krasovec<sup>1</sup>, Michael Chester<sup>1,2</sup>, Kate Ridout<sup>1,3</sup> and Dmitry A. Filatov<sup>1,4</sup>

<sup>1</sup>Department of Plant Sciences, University of Oxford, Oxford OX1 3RB, United Kingdom

Current address:

<sup>2</sup>Royal Botanic Gardens, Kew, Richmond, Surrey, TW9 3AE, United Kingdom

<sup>3</sup>RDM Nuffield Division of Clinical Laboratory Sciences, John Radcliffe Hospital, Oxford OX3 9DU, United Kingdom

<sup>4</sup>Corresponding author and lead contact: dmitry.filatov@plants.ox.ac.uk

**Keywords:** white campion, sex chromosome, mutation rate, Y degeneration

## SUMMARY

Many aspects of sex chromosome evolution are common to both plants and animals [1], but the process of Y chromosome degeneration, where genes on the Y become non-functional over time, may be much slower in plants due to purifying selection against deleterious mutations in the haploid gametophyte [2, 3]. Testing for differences in Y degeneration between the kingdoms has been hindered by the absence of accurate age estimates for plant sex chromosomes. Here we used genome resequencing to estimate the spontaneous mutation rate and the age of the sex chromosomes in white campion (*Silene latifolia*). Screening of single nucleotide polymorphisms (SNPs) in parents and 10 F<sub>1</sub> progeny identified 39 *de novo* mutations and yielded a rate of  $7.31 \times 10^{-9}$  (95% CI:  $5.20 \times 10^{-9}$  -  $8.00 \times 10^{-9}$ ) mutations per site per haploid genome per generation. Applying this mutation rate to the synonymous divergence between homologous X- and Y-linked genes (gametologs) gave age estimates of 11.00 and 6.32 million years, for the old and young strata, respectively. Based on SNP segregation patterns, we inferred which genes were Y-linked and found that at least 47% are already dysfunctional. Applying our new estimates for the age of the sex chromosomes indicates that the rate of Y degeneration in *S. latifolia* is nearly two-fold slower when compared to animal sex chromosomes of a similar age. Our revised estimates support Y degeneration taking place more slowly in plants, a discrepancy that may be explained by differences in the lifecycles of animals and plants.

## RESULTS and DISCUSSION

*Silene latifolia* has emerged as a useful model species for studies of sex chromosome evolution in plants [1, 4, 5]. The sex chromosomes in *S. latifolia* evolved from a pair homologous autosomes [6] that stopped recombining and gradually diverged into highly distinct X- and Y-chromosomes. Based on a relatively low silent site divergence ( $D_s < 25\%$ , e.g. [7]) between homologous X- and Y-linked genes, it is thought that sex chromosomes in this species are of recent origin, though the actual age is not known because the rate of sequence divergence has not been accurately estimated. To estimate this rate and calculate the age of sex chromosomes we assembled a reference genome for a male *S. latifolia* and then measured the per nucleotide per generation spontaneous mutation rate in this species.

### Assembly and Annotation of a Male *Silene latifolia* Reference Genome

To obtain a *S. latifolia* male genome we conducted genomic sequencing and *de novo* assembly of a male (Sa984) that was used as a paternal parent in a genetic cross. For sequencing, we used a combination of long error-corrected PacBio reads and Illumina sequencing data (118 Gb of data with insert sizes from 150bp to 10kb). The resulting assembly contained 319,549 contigs totalling 1,185.24 Mb in length (N50 = 10,785; see assembly summary statistics, Table S1). Although this corresponds to about 1/3<sup>rd</sup> of the estimated 2.8 Gb of *S. latifolia* genome [8, 9], the assembly contained 198 (79.8%) of the 248 ultraconserved core eukaryotic genes [10], suggesting that the assembled part of the genome is enriched for genic sequences, while the rest is likely to be represented mostly by repetitive sequences. Broadly consistent with this, 61.4% of the male *S. latifolia* genome was previously estimated to comprise repeats, including transposable elements and rDNA [11]. Following repetitive sequence masking and annotation (see methods) the partial assembly was used as a reference sequence throughout the paper.

A fraction of the genomic contigs were classified as X-linked (21.756 Mb in 812 contigs, Table S2), Y-linked (7.78 Mb in 511 contigs, Table S3) and pseudoautosomal (8.92 Mb in 398 contigs, Table S4), based on SNP segregation and a blastn search of known sex-linked genes [7] against the genome. In addition, 73,922 further putative Y-linked sequences with a total length of 112.69 Mb were identified using the difference in the depth of sequencing coverage between the Y-linked genes and the rest of the genome (Table S3 and Figure S1). Altogether, this yielded 151 Mb (75,643 contigs) of sex-linked (including the pseudo-autosomal region) and 1.03 Gb (243,909 contigs) of non-sex-linked sequences (Table S1). The identified sex-linked sequences correspond to ~11% of the genome assembly, which is close to the estimated relative size of the Y chromosome (~10% of the genome) based on cytogenetic analyses [4]. The X-linked, Y-linked and non-sex-linked contigs contained 1247, 1003 and

19195 putative protein-coding genes, respectively. Among these sex-linked genes, 953 had homologs present on both the X- and the Y-linked contigs (with 201 present in the genetic map published previously [7]).

### **The Spontaneous Mutation Rate in *Silene latifolia***

Relatively little is known about spontaneous mutation rates in plants compared to animals, where about a dozen species have already been studied (Table 1). To measure the spontaneous per generation mutation rate in *S. latifolia* we looked for *de novo* mutations that occurred over one generation in a genetic cross between individuals sampled from the wild. For this purpose, we conducted Illumina genomic sequencing of the parents and 10 F<sub>1</sub> progeny (six females and four males; 433 Gb sequence data in total) and compared these genomes to identify single nucleotide polymorphisms (SNPs) that are present in only one F<sub>1</sub> progeny and absent from the parents. Such SNPs, which are private to one of the F<sub>1</sub> individuals, must be either *de novo* mutations or false-positives arising due to sequencing errors.

In total, we analysed 2,666 Mb of callable sites across the 10 F<sub>1</sub> individuals (sites with minimal mapping quality  $\geq 30$  and sequence coverage  $\geq 10$  in both parents and F<sub>1</sub> progeny). After quality filtering (see methods) and verification using high-throughput targeted amplification and sequencing (Illumina TSCA), we confirmed 39 SNPs, private to one of the F<sub>1</sub> individuals, as true *de novo* mutations (Table S5). These *de novo* mutations were verified manually by PCR amplification and Sanger sequencing. The false-negative rate in this analysis was less than 1% (see methods). This gives a mutation rate of  $\mu = 39 / (2 \times 2,666 \times 10^6) = 7.31 \times 10^{-9}$  (CI 95% Poisson distribution:  $5.20 \times 10^{-9} - 8.00 \times 10^{-9}$ ) mutations per site per haploid genome per generation.

The *S. latifolia* spontaneous mutation rate represents the first such estimate for the Caryophyllales, which comprises over 11,000 species, including carnations, pinks, amaranths, portulacas and vegetables such as beetroot and spinach. Only two other spontaneous mutation rate estimates are available for land plants, *Arabidopsis thaliana* [12] and *Prunus persica* [13], belonging to the Brassicales and Asterales, respectively. The addition of *S. latifolia* to this list reveals a remarkable similarity in mutation rates across three plant orders (Table 1), suggesting that eudicots, and possibly flowering plants generally, have similar per nucleotide per generation mutation rates.

### **Patterns of Mutation in the *Silene latifolia* Genome**

Two of the 39 *de novo* mutations appeared in sex-linked sequences: one T to G substitution on the Y chromosome (scaffold 8750 position 5404) and another T to G substitution on the X chromosome (scaffold 1550 position 27017). The remaining 37 *de novo* mutations are not

sex-linked (Table S5). Given the relative sizes of the X- or Y-linked and autosomal genomic scaffolds, there is no significant difference in mutation rates between the sex chromosomes and the autosomes ( $\chi^2$  test;  $P = 0.479$ ).

We observed that the GC content is not at equilibrium in *S. latifolia*, with an excess of G/C to A/T *de novo* mutations (18 out of 39) compared with mutations in the opposite direction (Figure 1 and Table S5). Three of the G/C to A/T *de novo* mutations appeared at CpG sites. With a total genome size of ~2.8 Gb and assuming a constant nucleotide mutation rate over time, this mutation pressure alone should lead to a depletion of ~4 G or C nucleotides per haploid genome per generation. Importantly, the non-equilibrium state of the GC-content is predicted to affect the overall mutation rate [14]. We estimate that, at the predicted equilibrium GC content ( $GC_{eq} = 24.8\%$ ), the *S. latifolia* mutation rate would be  $\mu_{eq} = 6.78 \times 10^{-9}$ , ~7% lower than the observed rate, a modest effect compared to *A. thaliana* and *P. persica*, where, respectively, 64% and 40% reductions, compared with the predicted equilibrium rates, are estimated [12-14].

The mutation rate in *S. latifolia* fits the general trend of a positive correlation between the mutation rate and genome size reported previously [14, 15] (Pearson correlation test,  $\rho = 0.70$ ,  $P$ -value = 0.0037,  $n = 15$ ; data from Table 1). Because of the negative correlation between effective population size and genome size [16] this trend is thought to result from the so called “drift barrier” hypothesis postulating that most species have the minimum possible mutation rate achievable by natural selection given the population size of the species [17].

### **The Age of the Sex Chromosomes in *Silene latifolia***

The cessation of recombination between the *S. latifolia* X- and Y-chromosomes has evolved in several steps, as indicated by the presence of “evolutionary strata” [6, 7, 18-21]. Using the strata delimited in a previous study (see Figure 1 in [7]) provided us with 86 and 115 genes with known positions in the genetic map, belonging to the older and younger strata, respectively. Based on the *S. latifolia* mutation rate reported above and the mean synonymous divergence between homologous X- and Y-linked genes for each stratum ( $D_s = 0.107$  and 0.062 for old and young strata, respectively), we estimate that the older and the younger strata began to diverge 7.33 (95% CI: 5.21 – 10.02) and 4.20 (95% CI: 2.99 – 5.74) million generations (MG) ago. Assuming that the generation time of *S. latifolia* in nature is 1.5 years [22], the ages of the two strata are 11.00 (95%CI: 7.83 - 15.03) and 6.32 (95%CI: 3.0 - 8.64) MY, respectively.

It is worth noting that our sex chromosome age estimate is conditional on several assumptions. In particular, we assumed that the synonymous substitution rate is constant over time, which may not be correct. Synonymous substitutions are not entirely neutral [23] due to selection on exonic splice site enhancers [24] and codon bias [25]. The efficacy of selection

on silent sites likely varies over time due to its dependence on the effective population size, causing variation in synonymous rate and adding uncertainty to molecular clock-based estimates [26]. Furthermore, we assumed that mutation rate is the same on the autosomes and the sex chromosomes. If this assumption is not correct, then we may have under- or overestimated the age of sex chromosomes. The number of *de novo* mutations detected in our analysis was too small to test whether the mutation rate on the sex chromosomes differs from that on the autosomes. In animals, the larger number of cell divisions in male gametogenesis compared to female gametogenesis results in male-biased mutation rates [27, 28] – the so called “male driven evolution” effect that increases the mutation rate on the Y-chromosome, and reduces the X-chromosome rate relative to the autosomes [29, 30]. However, similar synonymous substitution rates for the X- and Y-linked genes of *S. latifolia* indicate no male-driven evolution in this species [3]. Thus, our genome-wide estimate of the mutation rate in *S. latifolia* should be applicable to the sex chromosomes.

Our study is the first to date the age of sex chromosomes in *S. latifolia* based on the actual mutation rate in the species. Our mutation rate-based estimate is close to the widely used approximation of ~10 MY (Desfeux *et al.* 1996), though a wider range of ages is also mentioned in the literature, e.g. Rautenberg *et al.* [31] argued that the sex chromosomes in this species must have originated as recently as ~5 MY ago, while other papers mentioned 5-10 MY [32] or 10-20 MY [6, 18].

### **The Extent of Y-chromosome Degeneration in *Silene latifolia***

Knowing the age of the *S. latifolia* sex chromosomes allowed us to estimate the rate at which its Y chromosome has degenerated. From the 1003 Y-linked genes, 94 were excluded due to missing data that did not allow us to assess whether the open reading frame is intact. Out of the remaining 909 Y-linked genes analysed, 433 genes (47.6%) were interrupted by a premature stop codon, and 12 genes (1.3%) had lost the start codon when compared to the corresponding X-linked gametolog. For comparison, significantly fewer X-linked genes had premature stop codons (12 out of 945; contingency 2x2,  $\chi^2 = 344.17$ ,  $P < 0.00001$ ). No X-linked genes had a missing start codon. The remaining 51.1% of Y-linked genes had intact structures, suggesting that about half of these genes remain functional. Assuming that the X and Y chromosomes started diverging 11.0 MYA, we estimate that the average rate of gene loss is ~40 per MY (4.3%). Based on 201 genes with known genetic map locations on the X chromosome [7], the degeneration rates for the older and younger strata are 4.64% and 6.92% per MY (or 7% and 10.4% per million generations), respectively.

A higher rate of Y degeneration on the younger *S. latifolia* stratum is consistent with Y degeneration being driven by interference between selection acting at many linked genes [33], which predicts that it will occur fastest immediately after recombination stops, and will become

slower as functional genes reduce in number. A slowing in the rate of degeneration has been documented for mammalian Y [34] and bird W chromosomes [35], and the same may be true in *S. latifolia*. This may explain why so little degeneration was reported for the homomorphic sex chromosomes of two other plants, papaya [36, 37] and persimmon [38]. The non-recombining regions in papaya and persimmon are small (~7Mb and ~1Mb respectively) and may carry too few functional genes to undergo degeneration.

To place our results into the context of other species, we compiled the degeneration rates of Y (or W in birds) chromosomes from other multicellular eukaryotes (Table 2). The rate gene loss varies from 0.5% per MY in human (considering the entire Y chromosome) to 26.6% per MY for the neo-Y chromosome of *Drosophila miranda*. The latter species does not have evolutionary strata due to lack of recombination in *Drosophila* males, so the entire neo-Y stopped recombining once it evolved father to son inheritance about 1.75 MY ago [39]. Although the rate of Y degeneration in *D. miranda* looks exceptionally high, this may be due to very short generation time in this species. Assuming that there are 10 *D. miranda* generations per year reduces its Y degeneration rate to only 2.6% per million generations (MG), bringing it in line with the rates in other animal species (Figure 2). As reported previously [34, 40], the rate of genetic degeneration is negatively correlated with the age of the stratum (Pearson correlation test,  $\rho = -0.85$ ,  $P\text{-value} = 2 \times 10^{-5}$ ,  $n = 16$ , Figure 2, data from Table 2). The correlation is stronger if only animal data are taken into account (Pearson correlation test,  $\rho = -0.94$ ,  $P\text{-value} = 6 \times 10^{-7}$ ,  $n = 14$ , data from Table 2).

The rate of degeneration in each of the *S. latifolia* strata is about 60% of that in the animal species with strata of a comparable age (Figure 2), which is consistent with earlier claims that Y degeneration in plants is slower due to stronger purifying selection at the haploid stage of plant life cycle – in the gametophyte and gamete [2, 3]. Alternatively, a slower rate of Y degeneration in *Silene* could be due to the lack of male driven evolution [3], resulting in fewer mutations per generation compared to mammalian Y-chromosomes. However, if this factor were an important contributor, then female-specific W-chromosomes in birds would be expected to degenerate at an even slower rate than the *S. latifolia* Y-chromosome, which is not the case (Figure 2).

## Conclusions

We measured the mutation rate in white campion and found it to be remarkably similar to two other estimates for plants [12, 13]. The similarity in mutation rates across three disparate eudicot lineages suggests a conservation that may extend across eudicots and perhaps even more broadly. Our estimated spontaneous mutation rate in *S. latifolia* allowed us to measure the age of sex chromosomes in that species, which confirms their relatively recent origin of ~11 million years. Although approximately half of the Y-linked genes in white campion are now

interrupted by premature stop codons, by applying our divergence time estimations we find that the Y-chromosome is degenerating considerably slower compared to animals.

## ACKNOWLEDGEMENTS

This work was supported by a grant from the BBSRC (Grant BB/P009808/1 to DAF). We thank Bruno Nevado for helpful comments and the staff at the Wellcome Trust Centre (Oxford) for high throughput sequencing and initial data processing.

## AUTHOR CONTRIBUTIONS

DAF conceived the project, designed the work, generated most of the data and drafted the text of the manuscript. MK did the analyses of mutation rate and the rate of genetic degeneration, MC was involved in verification of *de novo* mutations and KR assembled and annotated the draft genome for a *S. latifolia* male plant used as a reference in this study. All the authors contributed to writing or editing the manuscript.

## DECLARATION OF INTERESTS

The authors declare no competing interests.

## REFERENCES

1. Charlesworth, D. (2016). Plant Sex Chromosomes. *Annu Rev Plant Biol* 67, 397-420.
2. Bergero, R., and Charlesworth, D. (2011). Preservation of the Y transcriptome in a 10-million-year-old plant sex chromosome system. *Curr Biol* 21, 1470-1474.
3. Chibalina, M.V., and Filatov, D.A. (2011). Plant Y chromosome degeneration is retarded by haploid purifying selection. *Curr Biol* 21, 1475-1479.
4. Armstrong, S.J., and Filatov, D.A. (2008). A cytogenetic view of sex chromosome evolution in plants. *Cytogenet Genome Res* 120, 241-246.
5. Bernasconi, G., Antonovics, J., Biere, A., Charlesworth, D., Delph, L.F., Filatov, D., Giraud, T., Hood, M.E., Marais, G.A., McCauley, D., et al. (2009). *Silene* as a model system in ecology and evolution. *Heredity (Edinb)* 103, 5-14.
6. Filatov, D.A. (2005). Evolutionary history of *Silene latifolia* sex chromosomes revealed by genetic mapping of four genes. *Genetics* 170, 975-979.
7. Papadopoulos, A.S., Chester, M., Ridout, K., and Filatov, D.A. (2015). Rapid Y degeneration and dosage compensation in plant sex chromosomes. *Proc Natl Acad Sci U S A* 112, 13021-13026.
8. Zonneveld, B.J.M., Leitch, I.J., and Bennett, M.D. (2005). First nuclear DNA amounts in more than 300 angiosperms. *Ann Bot-London* 96, 229-244.
9. Siroky, J., Lysak, M.A., Dolezel, J., Kejnovsky, E., and Vyskot, B. (2001). Heterogeneity of rDNA distribution and genome size in *Silene* spp. *Chromosome Res* 9, 387-393.
10. Parra, G., Bradnam, K., and Korf, I. (2007). CEGMA: a pipeline to accurately annotate core genes in eukaryotic genomes. *Bioinformatics* 23, 1061-1067.
11. Macas, J., Kejnovsky, E., Neumann, P., Novak, P., Koblikova, A., and Vyskot, B. (2011). Next generation sequencing-based analysis of repetitive DNA in the model dioecious plant *Silene latifolia*. *PLoS One* 6, e27335.



12. Ossowski, S., Schneeberger, K., Lucas-Lledó, J.I., Warthmann, N., Clark, R.M., Shaw, R.G., Weigel, D., and Lynch, M. (2010). The rate and molecular spectrum of spontaneous mutations in *Arabidopsis thaliana*. *Science* 327, 92-94.
13. Xie, Z., Wang, L., Wang, L., Wang, Z., Lu, Z., Tian, D., Yang, S., and Hurst, L.D. (2016). Mutation rate analysis via parent-progeny sequencing of the perennial peach. I. A low rate in woody perennials and a higher mutagenicity in hybrids. *Proc Biol Sci* 283.
14. Krasovec, M., Eyre-Walker, A., Sanchez-Ferandin, S., and Piganeau, G. (2017). Spontaneous mutation rate in the smallest photosynthetic eukaryotes. *Mol Biol Evol* 34, 1770-1779.
15. Smeds, L., Qvarnstrom, A., and Ellegren, H. (2016). Direct estimate of the rate of germline mutation in a bird. *Genome Res* 26, 1211-1218.
16. Lynch, M., and Conery, J.S. (2003). The origins of genome complexity. *Science* 302, 1401-1404.
17. Lynch, M., Ackerman, M.S., Gout, J.F., Long, H., Sung, W., Thomas, W.K., and Foster, P.L. (2016). Genetic drift, selection and the evolution of the mutation rate. *Nat Rev Genet* 17, 704-714.
18. Bergero, R., Forrest, A., Kamau, E., and Charlesworth, D. (2007). Evolutionary strata on the X chromosomes of the dioecious plant *Silene latifolia*: evidence from new sex-linked genes. *Genetics* 175, 1945-1954.
19. Pandey, R.S., and Azad, R.K. (2016). Deciphering evolutionary strata on plant sex chromosomes and fungal mating-type chromosomes through compositional segmentation. *Plant Mol Biol* 90, 359-373.
20. Atanassov, I., Delichere, C., Filatov, D.A., Charlesworth, D., Negrutiu, I., and Moneger, F. (2001). Analysis and evolution of two functional Y-linked loci in a plant sex chromosome system. *Molecular Biology and Evolution* 18, 2162-2168.
21. Bergero, R., Qiu, S., Forrest, A., Borthwick, H., and Charlesworth, D. (2013). Expansion of the pseudo-autosomal region and ongoing recombination suppression in the *Silene latifolia* sex chromosomes. *Genetics* 194, 673-686.
22. Richards, C.M., Emery, S.N., and McCauley, D.E. (2003). Genetic and demographic dynamics of small populations of *Silene latifolia*. *Heredity (Edinb)* 90, 181-186.
23. Chamary, J.V., Parmley, J.L., and Hurst, L.D. (2006). Hearing silence: non-neutral evolution at synonymous sites in mammals. *Nat Rev Genet* 7, 98-108.
24. Caceres, E.F., and Hurst, L.D. (2013). The evolution, impact and properties of exonic splice enhancers. *Genome Biol* 14, R143.
25. Qiu, S., Bergero, R., Zeng, K., and Charlesworth, D. (2011). Patterns of codon usage bias in *Silene latifolia*. *Mol Biol Evol* 28, 771-780.
26. Lanfear, R., Kokko, H., and Eyre-Walker, A. (2014). Population size and the rate of evolution. *Trends Ecol Evol* 29, 33-41.
27. Segurel, L., Wyman, M.J., and Przeworski, M. (2014). Determinants of mutation rate variation in the human germline. *Annu Rev Genomics Hum Genet* 15, 47-70.
28. Venn, O., Turner, I., Mathieson, I., de Groot, N., Bontrop, R., and McVean, G. (2014). Strong male bias drives germline mutation in chimpanzees. *Science* 344, 1272-1275.
29. Malcom, C.M., Wyckoff, G.J., and Lahn, B.T. (2003). Genic mutation rates in mammals: local similarity, chromosomal heterogeneity, and X-versus-autosome disparity. *Mol Biol Evol* 20, 1633-1641.
30. Axelsson, E., Smith, N.G., Sundstrom, H., Berlin, S., and Ellegren, H. (2004). Male-biased mutation rate and divergence in autosomal, Z-linked and W-linked introns of chicken and Turkey. *Mol Biol Evol* 21, 1538-1547.
31. Rautenberg, A., Hathaway, L., Oxelman, B., and Prentice, H.C. (2010). Geographic and phylogenetic patterns in *Silene* section *Melandrium* (Caryophyllaceae) as inferred from chloroplast and nuclear DNA sequences. *Mol Phylogenet Evol* 57, 978-991.
32. Nicolas, M., Marais, G., Hykelova, V., Janousek, B., Laporte, V., Vyskot, B., Mouchiroud, D., Negrutiu, I., Charlesworth, D., and Moneger, F. (2005). A gradual process of recombination restriction in the evolutionary history of the sex chromosomes in dioecious plants. *PLoS Biol* 3, e4.

33. Bachtrog, D. (2013). Y-chromosome evolution: emerging insights into processes of Y-chromosome degeneration. *Nat Rev Genet* 14, 113-124.
34. Bellott, D.W., Hughes, J.F., Skaletsky, H., Brown, L.G., Pyntikova, T., Cho, T.J., Koutseva, N., Zaghlul, S., Graves, T., Rock, S., et al. (2014). Mammalian Y chromosomes retain widely expressed dosage-sensitive regulators. *Nature* 508, 494.
35. Zhou, Q., Zhang, J., Bachtrog, D., An, N., Huang, Q., Jarvis, E.D., Gilbert, M.T.P., and Zhang, G. (2014). Complex evolutionary trajectories of sex chromosomes across bird taxa. *Science* 346, 1332.
36. Wang, J., Na, J.K., Yu, Q., Gschwend, A.R., Han, J., Zeng, F., Aryal, R., VanBuren, R., Murray, J.E., Zhang, W., et al. (2012). Sequencing papaya X and Y chromosomes reveals molecular basis of incipient sex chromosome evolution. *Proc Natl Acad Sci U S A* 109, 13710-13715.
37. Liu, Z., Moore, P.H., Ma, H., Ackerman, C.M., Ragiba, M., Yu, Q., Pearl, H.M., Kim, M.S., Charlton, J.W., Stiles, J.I., et al. (2004). A primitive Y chromosome in papaya marks incipient sex chromosome evolution. *Nature* 427, 348-352.
38. Akagi, T., Henry, I.M., Tao, R., and Comai, L. (2014). A Y-chromosome-encoded small RNA acts as a sex determinant in persimmons. *Science* 346, 646-650.
39. Bachtrog, D., Hom, E., Wong, K.M., Maside, X., and de Jong, P. (2008). Genomic degradation of a young Y chromosome in *Drosophila miranda*. *Genome Biol* 9, R30.
40. Hughes, J.F., Skaletsky, H., Brown, L.G., Pyntikova, T., Graves, T., Fulton, R.S., Dugan, S., Ding, Y., Buhay, C.J., Kremitzki, C., et al. (2012). Strict evolutionary conservation followed rapid gene loss on human and rhesus Y chromosomes. *Nature* 483, 82-86.
41. Besenbacher, S., Liu, S., Izarzugaza, J.M., Grove, J., Belling, K., Bork-Jensen, J., Huang, S., Als, T.D., Li, S., Yadav, R., et al. (2015). Novel variation and de novo mutation rates in population-wide de novo assembled Danish trios. *Nat Commun* 6, 5969.
42. Uchimura, A., Higuchi, M., Minakuchi, Y., Ohno, M., Toyoda, A., Fujiyama, A., Miura, I., Wakana, S., Nishino, J., and Yagi, T. (2015). Germline mutation rates and the long-term phenotypic effects of mutation accumulation in wild-type laboratory mice and mutator mice. *Genome Res* 25, 1125-1134.
43. Liu, H., Jia, Y., Sun, X., Tian, D., Hurst, L.D., and Yang, S. (2017). Direct determination of the mutation rate in the bumblebee reveals evidence for weak recombination-associated mutation and an approximate rate constancy in insects. *Mol Biol Evol* 34, 119-130.
44. Keightley, P.D., Pinharanda, A., Ness, R.W., Simpson, F., Dasmahapatra, K.K., Mallet, J., Davey, J.W., and Jiggins, C.D. (2015). Estimation of the spontaneous mutation rate in *Heliconius melpomene*. *Mol Biol Evol* 32, 239-243.
45. Schrider, D.R., Houle, D., Lynch, M., and Hahn, M.W. (2013). Rates and genomic consequences of spontaneous mutational events in *Drosophila melanogaster*. *Genetics* 194, 937-954.
46. Oppold, A.-M., and Pfenninger, M. (2017). Direct estimation of the spontaneous mutation rate by short-term mutation accumulation lines in *Chironomus riparius*. *Evolution Letters* 1, 86-92.
47. Flynn, J.M., Chain, F.J., Schoen, D.J., and Cristescu, M.E. (2017). Spontaneous mutation accumulation in *Daphnia pulex* in selection-free vs. competitive environments. *Mol Biol Evol* 34, 160-173.
48. Weller, A.M., Rodelsperger, C., Eberhardt, G., Molnar, R.I., and Sommer, R.J. (2014). Opposing forces of A/T-biased mutations and G/C-biased gene conversions shape the genome of the nematode *Pristionchus pacificus*. *Genetics* 196, 1145-1152.
49. Denver, D.R., Wilhelm, L.J., Howe, D.K., Gafner, K., Dolan, P.C., and Baer, C.F. (2012). Variation in base-substitution mutation in experimental and natural lineages of *Caenorhabditis* nematodes. *Genome Biol Evol* 4, 513-522.
50. Bachtrog, D. (2008). The temporal dynamics of processes underlying Y chromosome degeneration. *Genetics* 179, 1513-1525.

51. BirdLife\_International (2016). The IUCN Red List of Threatened Species.
52. Bolger, A.M., Lohse, M., and Usadel, B. (2014). Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* 30, 2114-2120.
53. Morgulis, A., Gertz, E.M., Schaffer, A.A., and Agarwala, R. (2006). A fast and symmetric DUST implementation to mask low-complexity DNA sequences. *J Comput Biol* 13, 1028-1040.
54. Luo, R., Liu, B., Xie, Y., Li, Z., Huang, W., Yuan, J., He, G., Chen, Y., Pan, Q., Liu, Y., et al. (2012). SOAPdenovo2: an empirically improved memory-efficient short-read de novo assembler. *Gigascience* 1, 18.
55. Hunt, M., Kikuchi, T., Sanders, M., Newbold, C., Berriman, M., and Otto, T.D. (2013). REAPR: a universal tool for genome assembly evaluation. *Genome Biol* 14, R47.
56. Song, L., Florea, L., and Langmead, B. (2014). Lighter: fast and memory-efficient sequencing error correction without counting. *Genome Biology* 15, 509.
57. Au, K.F., Underwood, J.G., Lee, L., and Wong, W.H. (2012). Improving PacBio long read accuracy by short read alignment. *PLoS One* 7, e46679.
58. English, A.C., Richards, S., Han, Y., Wang, M., Vee, V., Qu, J., Qin, X., Muzny, D.M., Reid, J.G., Worley, K.C., et al. (2012). Mind the gap: upgrading genomes with Pacific Biosciences RS long-read sequencing technology. *PLoS One* 7, e47768.
59. Boetzer, M., Henkel, C.V., Jansen, H.J., Butler, D., and Pirovano, W. (2011). Scaffolding pre-assembled contigs using SSPACE. *Bioinformatics* 27, 578-579.
60. Xue, W., Li, J.T., Zhu, Y.P., Hou, G.Y., Kong, X.F., Kuang, Y.Y., and Sun, X.W. (2013). L\_RNA\_scaffolder: scaffolding genomes with transcripts. *BMC Genomics* 14, 604.
61. Benson, G. (1999). Tandem repeats finder: a program to analyze DNA sequences. *Nucleic Acids Res* 27, 573-580.
62. Holt, C., and Yandell, M. (2011). MAKER2: an annotation pipeline and genome-database management tool for second-generation genome projects. *BMC Bioinformatics* 12, 491.
63. Korf, I. (2004). Gene finding in novel genomes. *BMC Bioinformatics* 5, 59.
64. Stanke, M., Schoffmann, O., Morgenstern, B., and Waack, S. (2006). Gene prediction in eukaryotes with a generalized hidden Markov model that uses hints from external sources. *BMC Bioinformatics* 7, 62.
65. Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G., Durbin, R., and Genome Project Data Processing, S. (2009). The sequence alignment/map format and SAMtools. *Bioinformatics* 25, 2078-2079.
66. Muyle, A., Kafer, J., Zemp, N., Mousset, S., Picard, F., and Marais, G.A. (2016). SEX-DETECTOR: a probabilistic approach to study sex chromosomes in non-model organisms. *Genome Biol Evol* 8, 2530-2543.
67. McKenna, A., Hanna, M., Banks, E., Sivachenko, A., Cibulskis, K., Kernysky, A., Garimella, K., Altshuler, D., Gabriel, S., Daly, M., et al. (2010). The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Res* 20, 1297-1303.
68. Cingolani, P., Platts, A., Wang le, L., Coon, M., Nguyen, T., Wang, L., Land, S.J., Lu, X., and Ruden, D.M. (2012). A program for annotating and predicting the effects of single nucleotide polymorphisms, SnpEff: SNPs in the genome of *Drosophila melanogaster* strain w1118; iso-2; iso-3. *Fly (Austin)* 6, 80-92.
69. Camacho, C., Coulouris, G., Avagyan, V., Ma, N., Papadopoulos, J., Bealer, K., and Madden, T.L. (2009). BLAST+: architecture and applications. *BMC Bioinformatics* 10, 421.
70. Langmead, B., and Salzberg, S.L. (2012). Fast gapped-read alignment with Bowtie 2. *Nat Methods* 9, 357-359.
71. Keightley, P.D., Ness, R.W., Halligan, D.L., and Haddrill, P.R. (2014). Estimation of the spontaneous mutation rate per nucleotide site in a *Drosophila melanogaster* full-sib family. *Genetics* 196, 313-320.
72. Eyre-Walker, A. (2006). The genomic rate of adaptive evolution. *Trends Ecol Evol* 21, 569-575.

## Figure Legends

### **Figure 1. The distribution of 39 *de novo* mutations identified in ten F<sub>1</sub> progeny**

*R*-values in the top right corner are the estimated rates of different mutation types in the spectrum of mutations. See also Table S5.

### **Figure 2. Correlation between the rate of the Y (or W) degeneration and the age of the**

**stratum.** Source data are shown in Table 2. Blue: *Homo sapiens*, green: *S. latifolia*, orange: five bird species, purple: *Drosophila miranda* neo-Y. Multiple points are shown for single species that have strata of different ages. Green bars are the confidence intervals for the age of the two strata in *S. latifolia*. The degeneration rate on the Y axis is the percentage of genes lost per million generations.

**Table 1. Genome Size (G) and Mutation Rates per Nucleotide ( $\mu$ ) in Multicellular Eukaryotes**

	<b>Species</b>	<b>G (Mb)</b>	<b><math>\mu</math> (<math>\times 10^{-9}</math>)</b>	<b>Reference</b>
<b>Plants</b>				
Angiosperms	<i>Silene latifolia</i>	2641	7.31	This study
	<i>Arabidopsis thaliana</i>	134	7.00	[12]
	<i>Prunus persica</i>	212	7.77	[13]
<b>Metazoans</b>				
Mammals	<i>Homo sapiens</i>	3309	12.9	[41]
	<i>Mus musculus</i>	2671	5.40	[42]
Birds	<i>Ficedula albicollis</i>	1118	4.60	[15]
	<i>Bombus terrestris</i>	250	3.60	[43]
Insects	<i>Heliconius melpomene</i>	275	2.90	[44]
	<i>Apis mellifera</i>	236	3.40	[43]
	<i>Drosophila melanogaster</i>	148	5.49	[45]
	<i>Chironomus riparius</i>	155	2.10	[46]
	<i>Daphnia pulex</i>	197	2.30	[47]
Crustaceans	<i>Pristionchus pacificus</i>	133	2.00	[48]
	<i>Caenorhabditis briggsae</i>	108	1.34	[49]
	<i>Caenorhabditis elegans</i>	100	1.48	[49]

**Table 2. Degeneration Rates of the Y (or W) Chromosomes in Plant and Animals**

Species	Age (MY)	GT (Y)	Age (MG)	% Degen.	Sex system	Ref.
<i>Silene latifolia</i>	11.00	1.50	7.33	51.04	XY	This study
	6.30	1.50	4.20	43.73		
<i>Drosophila miranda</i>	1.75	0.10	17.50	46.55	XY	[39]
<i>Homo sapiens</i>	150	10.00	15.00	83.33	XY	[50]
	100	15.00	6.67	71.43		
	50	20.00	2.50	41.67		
White-throated tinamou ( <i>Tinamus guttatus</i> )	50.00	6.80	7.35	89.50	WZ	[35]
	29.00	6.80	4.26	76.85		
Brown mesite ( <i>Mesitornis unicolor</i> )	69.00	4.20	16.43	91.08	WZ	
	48.00	4.20	11.43	66.72		
White-tailed tropicbird ( <i>Phaethon lepturus</i> )	96.00	11.10	8.65	96.00	WZ	
	70.00	11.10	6.31	83.30		
Crested ibis ( <i>Nipponia nippon</i> )	69.00	10.10	6.83	85.56	WZ	
	53.00	10.10	5.25	91.69		
Barn owl ( <i>Tyto alba</i> )	69.00	6.20	11.13	93.84	WZ	
	49.00	6.20	7.90	86.24		

MY: million years, MG: million generations, GT: generation time in years, % Degen: total percentage of gene losses. Species known to have strata of different ages are shown with multiple rows. To convert the age in time to the age in generations, we assumed 10 generations per year for *Drosophila miranda* and 10 to 20 years per generation for the human strata, provided that generation time was likely shorter for ancestral lineage of mammals when older strata evolved (the oldest human stratum excluded due to uncertainty with generation time back then). For birds, we used generation times from the International Union for Conservation of Nature web site [51].

## STAR METHODS

### CONTACT FOR REAGENT AND RESOURCE SHARING

Further information regarding the manuscript and requests for reagents may be directed to, and will be fulfilled by the lead contact, Dmitry A. Filatov (Dmitry.Filatov@plants.ox.ac.uk).

### EXPERIMENTAL MODEL AND SUBJECT DETAILS

Twelve individuals of the plant *Silene latifolia* (white campion) were used for this study: one male parent (individual number Sa984), one female parent (individual number Sa985) and 10 F<sub>1</sub> progeny (4 males, 6 females) from the Sa984 x Sa985 cross. All individuals were grown in the glasshouse at the University of Oxford (Department of Plant Sciences).

### METHOD DETAILS

#### ***Genomic sequencing data***

To measure mutation rates in *S. latifolia* we generated and analysed genome sequence data for the parents and 10 F<sub>1</sub> offspring (four males and six females) of a genetic cross between female (Sa985) and male (Sa984) plants grown from seed collected in 2005 in Austria and England, respectively. Both the parents and the F<sub>1</sub> progeny of this cross were grown in the glasshouse in Oxford. DNA was extracted from fresh plant leaves of both parents and F<sub>1</sub> progeny with the Plant DNeasy Mini kit (Qiagen) and sequenced on an Illumina HiSeq2000 platform with 100 bp paired-end reads. Furthermore, to improve the genomic assembly we generated mate pair Illumina reads with ~2 kb, ~5 kb and ~10 kb insert sizes as well as PacBio sequence data for the male parent of the cross.

#### ***Genomic assembly and annotation of the male reference genome***

A partial genome sequence is available only for a female *S. latifolia* [7]. To generate a sequence for a male *S. latifolia* we used a combination of paired end and mate pair Illumina reads with a range of insert sizes (200bp to 10kb) generated for the paternal individual of the genetic cross (Sa984). The reads were trimmed and adaptors were removed using Trimmomatic v.0.30 with default parameters [52]. Duplicate read pairs were collapsed using fastx-collapser from the Fastx-Toolkit ([http://hannonlab.cshl.edu/fastx\\_toolkit/](http://hannonlab.cshl.edu/fastx_toolkit/)). Low complexity regions were masked using DUST [53] with default parameters. Filtered paired-end and mate-paired reads were assembled using SOAPdenovo2 [54] with k-mer values between 35 and 55 (odd values only). The best assembly was chosen using REAPR [55]. GapCloser [54] was run on the best assembly to correct false joins and fill gaps.

To assemble longer genomic scaffolds we also used Pacific Biosciences (PacBio) reads following error correction with CLC genomics workbench genome finishing module

version 1.7 and Bowtie2 version 2.1.0 [70] in combination with LSC version 0.3.1 [57]. Error-corrected PacBio reads were then used to extend scaffolds, fill gaps, and join scaffolds using PBJelly2 [58]. Additional scaffolding was carried out using SSPACE [59] and L\_RNA\_Scaffolder [60].

Repeat masking was carried out on the assembled genome using a combination of three programs, Tandem Repeats Finder (TRF; [61]) for masking tandem repeats, RepeatModeler (<http://www.repeatmasker.org>) for predicting *Silene*-specific transposable elements, and RepeatMasker (<http://www.repeatmasker.org>), which was used to identify any additional tandem repeats and transposable elements using a repeat library for the Caryophyllaceae. All annotations were carried out on the masked genome. A combination of CEGMA (Core Eukaryotic Genes Mapping Approach; [10]) proteins and assembled *Silene* transcripts from previous work [3, 7] were passed to the gene prediction program MAKER2 [62], following the GMOD (Generic Model Organism Database; <http://gmod.org/wiki/MAKER>) guidelines. The predicted gene set was then filtered to preserve only complete coding genes (*i.e.* genes beginning with Methionine and ending with a stop codon), and used to train SNAP (Semi-HMM-based Nucleic Acid Parser; [63]) and AUGUSTUS [64], the *ab initio* gene predictors. MAKER2 was then run again, combining the gene set from the previous run with annotated Caryophyllaceae proteins (NCBI GeneBank), and SNAP and AUGUSTUS predictions to produce the final gene set.

### ***Identification of sex-linked genomic contigs in the male reference genome***

We identified putative sex-linked sequences with three different approaches. First, a coverage analysis was used to identify Y-linked sequences in the male reference genome using Illumina sequences of five males (male parent and four F<sub>1</sub> progeny) and seven females (female parent and six F<sub>1</sub> progeny). All alignments were built with Bowtie2 using standard parameters [56] and treated with Samtools [65]. Two criteria were used to define putative Y-linked contigs: (i) they were not mapped in females, and (ii) had half the mean mapping coverage in males. Second, we used blastn [69] of the known sex-linked genes based on previous studies [3, 7] to identify both Y- and X-linked contigs. A contig was considered sex-linked if only Y (or X) genes mapped to it. If X and Y homologs mapped to the same contig, Y or X-linkage was assigned based on which of the two homologs had the highest similarity. If, after applying these criteria, contigs appeared to be both Y and X-linked, they were assigned to the pseudo-autosomal region (PAR). Third, we ran SEX-DETECTOR, which uses SNP segregation to identify sex-linked sequences [66]. Input data for SEX-DETECTOR comprised the male reference genome and the genomic sequence reads from the two parents and the ten F<sub>1</sub> progeny. The genotype file (alr format) which is required for SEX-DETECTOR was created by Reads2snp [66] (parameters: -min 3 -bqt 20 -rqt 10). SEX-DETECTOR was run assuming an XY sex system with



the 'cross' pipeline. All contigs not identified as sex-linked by the methods described above were classed as autosomal sequences.

### ***Identification and verification of de novo mutations***

SNPs that are present only in a single F<sub>1</sub> progeny, but not in the parents (singleton SNPs) must either be errors or *de novo* mutations. To minimise the number of PCR and sequencing errors due to library preparation process we prepared and sequenced at least three separate libraries per individual. Only those singleton SNPs that were found in more than one library were kept for further analyses. As our aim was to estimate the age of sex chromosomes rather than to comprehensively characterise the patterns of *de novo* mutations, we therefore focused on single nucleotide mutations and ignored more complex types of mutations (insertions, deletions). The Illumina reads of each of the parents and F<sub>1</sub> progeny were mapped to the *S. latifolia* male genome with Bowtie2 [56] and the resulting SAM files were processed with Samtools [65]. We performed the SNP calling with HaplotypeCaller from GATK [67] in the recommended pipeline (i.e. MarkDuplicates, IndelRealigner, RealignerTargetCreator). To distinguish between *de novo* mutations and standing SNPs and limit the number of false positives, we filtered out the SNPs if:

- (i) the alternative allele was present in another F<sub>1</sub> genome
- (ii) the alternative allele was present in the parents
- (iii) the coverage was below 10 (both in the F<sub>1</sub> and parent genomes)
- (iv) the mapping quality was below 30 (MQ value in the vcf file)
- (v) the alternative allele was supported by less than one third of the total number of reads mapping to that position.

To check the accuracy of this SNP filtering procedure, we used a high-throughput targeted amplification and sequencing (Illumina TSCA), followed by manual verification. Sequencing-based verification was done for each of the 39 candidate *de novo* SNPs by PCR and Sanger re-sequencing in both parent and F<sub>1</sub> progeny. Once the set of *de novo* mutations was identified, we used SnpEff [68] to infer the possible effect of each mutation on the phenotype.

Following the method from Keightley and co-workers [44, 71], we estimated the false negative rate in the experiment by artificially introducing 1,000 mutations into raw sequence reads using the empirical distributions of numbers of non-reference bases at heterozygous sites in the offspring. Modified reads were extracted from the SAM files and re-mapped against the reference genome. We then repeated the procedure of SNP calling and filtering with the same pipeline as described above. The proportion of artificial mutations at callable sites that we failed to detect was <1%, implying that the false negative rate was low.

### **Mutation rate analysis**

The per-nucleotide mutation rate was calculated as the number of confirmed *de novo* mutations (*i.e.* 39) divided by the number of callable sites in each of the F<sub>1</sub> individuals (Table S1). Callable sites are only those positions where the reference base is known, the coverage is  $\geq 10$  and mapping quality (MQ)  $> 30$ . We calculated the 95% confidence interval of the mutation rate with a Poisson distribution with 39 events, assuming that mutations are rare events. To place our estimate of the mutation rate in the context of other species, we used mutation rates that have been published for multicellular eukaryotes (15 species, Table 1). The correlation between  $\mu$  and the genome size was measured using the Pearson correlation test. The spectrum of mutations ( $R_1$  [GC->AT],  $R_2$  [AT->GC],  $R_3$  [GC->GC] and  $R_4$  [AT->AT]) was analysed based on the observed number of mutations of each type. The strength of mutational GC bias in *S. latifolia* was calculated from the ratio  $R_1/R_2$ , while the equilibrium GC content was calculated using the equation  $GC_{eq} = R_2/(R_1+R_2)$ . The distribution of mutations was compared to the random expectation with a Chi-square test, with a  $H_0$  hypothesis that mutations occur randomly and independently. To establish whether a *de novo* mutation was X-linked, Y-linked or autosomal, we analysed the segregation of SNPs adjacent to the *de novo* mutation. Exclusive inheritance of SNPs from fathers to sons or to daughters indicates Y-linkage or X-linkage, respectively.

### **The age of the sex chromosomes**

Assuming that synonymous mutations are neutral, the time of divergence (in generations,  $t_g$ ) between homologous Y- and X-linked genes can be calculated from equation  $t_g = d_s/2\mu$ , where  $d_s$  is the synonymous sequence divergence and  $\mu$  is the per nucleotide mutation rate [72]. The age of the sex chromosomes (in years,  $t_y$ ) was obtained from  $t_g$  assuming 1.5 years per generation in *S. latifolia* [22]. Estimates of  $d_s$  for hundreds of *S. latifolia* X- and Y-linked gametologs are available from a previous study [7]. To estimate the age of the evolutionary strata in *S. latifolia* we used a subset of sex-linked genes (201 homologous X- and Y-linked gene pairs, Table S2) for which the positions of X-linked genes in the genetic map are known [7]. The ages were calculated for each gene, at each position in the genetic map and for each of the two strata. The older and the younger strata are defined as the regions from 0 to 46 cM and from 46 to 63 cM of the genetic map, respectively [7]. The older and the younger strata contain 86 and 115 genetically mapped genes, respectively. The rest of the X-chromosome (63 to 114 cM; 107 mapped genes [7]) represents the pseudoautosomal region (PAR) that pairs and recombines in male meioses.

## **QUANTIFICATION AND STATISTICAL ANALYSIS**

The summary statistics for the genome assembly (Table S1) were obtained with a custom perl script `fnaSummary.pl`. The proportion of the genes included in the genome assembly was assessed with CEGMA [10] pipeline that was run in a virtual machine ([http://korflab.ucdavis.edu/datasets/cegma/cegma\\_vm.html](http://korflab.ucdavis.edu/datasets/cegma/cegma_vm.html)). The distribution of read coverage across genomic contigs (Figure S1) was obtained with `samtools idxstats` command [65] to calculate the RPKM values for each contig for all males and all females. The randomness in the distribution of *de novo* mutations across the genome was tested with a chi-square test in R v.3.4.4 (<http://www.R-project.org>). The correlation tests using the data listed in Table 1 and 2 were performed with Pearson linear model implemented in R. Details are provided in the Methods section above. Both figure 1 and figure 2 were plotted in R.

#### **DATA AND SOFTWARE AVAILABILITY**

High throughput sequencing data generated in this paper is available from NCBI under bioproject number PRJNA448569. The assembly of the male *S. latifolia* genome is available from NCBI with genome accession number: QBIE000000000.

## SUPPLEMENTAL ITEM TITLES

Table S2. The list of X-linked genomic scaffolds (Excel file), related to text section “Assembly and Annotation of a Male *Silene latifolia* Reference Genome” and Table S1

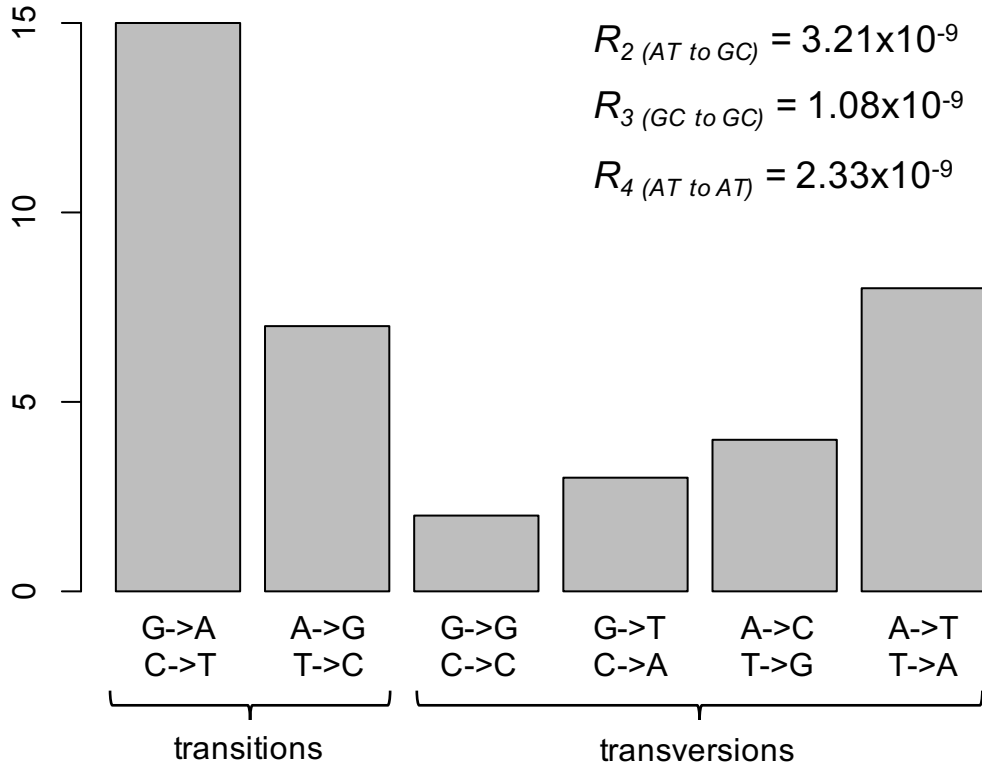
Table S3. The list of Y-linked genomic scaffolds (Excel file), related to text section “Assembly and Annotation of a Male *Silene latifolia* Reference Genome” and Table S1

Table S4. The list of pseudoautosomal genomic scaffolds (Excel file), related to text section “Assembly and Annotation of a Male *Silene latifolia* Reference Genome” and Table S1

## KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Critical Commercial Assays		
Illumina high throughput sequencing	WTCHG genomic facility, Oxford	Paired end genomic sequencing
Illumina TSCA genotyping	Illumina	FC-130-1001
QIAGEN DNEASY Plant Kit	QIAGEN	cat # 69104
Deposited Data		
WGS data	This paper	NCBI Bioproject: PRJNA448569
Whole genome assembly for <i>Silene latifolia</i> male	This paper	NCBI; genome accession number: QBIE00000000
Experimental Models: Organisms/Strains		
<i>S. latifolia</i> male plant	[7]	Sa984
<i>S. latifolia</i> female plant	[7]	Sa985
<i>S. latifolia</i> F1 progeny of the cross Sa985 x Sa984	This paper	N/A
Software and Algorithms		
Trimmomatic v.0.30	[51]	<a href="http://www.usadellab.org/cms/?page=trimmomatic">http://www.usadellab.org/cms/?page=trimmomatic</a>
fastx-collapser	Fastx-toolkit	<a href="http://hannonlab.cshl.edu/fastx_toolkit/">http://hannonlab.cshl.edu/fastx_toolkit/</a>
DUST	[52]	<a href="https://www.ncbi.nlm.nih.gov/IEB/ToolBox/CPP_DOC/lxr/source/src/app/dustmasker/">https://www.ncbi.nlm.nih.gov/IEB/ToolBox/CPP_DOC/lxr/source/src/app/dustmasker/</a>
SOAPdenovo2	[53]	<a href="http://soapdenovo2.sourceforge.net/">http://soapdenovo2.sourceforge.net/</a>
REAPR	[54]	<a href="http://www.sanger.ac.uk/science/tools/reapr">http://www.sanger.ac.uk/science/tools/reapr</a>
CLC genomics workbench genome finishing module version 1.7	QIAGEN	<a href="https://www.qiagenbioinformatics.com/plugins/clc-genome-finishing-module/">https://www.qiagenbioinformatics.com/plugins/clc-genome-finishing-module/</a>
Bowtie2 v. 2.1.0	[55]	<a href="http://bowtie-bio.sourceforge.net/bowtie2/">http://bowtie-bio.sourceforge.net/bowtie2/</a>
LSC v. 0.3.1	[56]	<a href="https://www.healthcare.uiowa.edu/labs/au/LSC/">https://www.healthcare.uiowa.edu/labs/au/LSC/</a>
PBJelly2	[57]	<a href="https://sourceforge.net/projects/pb-jelly/">https://sourceforge.net/projects/pb-jelly/</a>
SSPACE	[58]	<a href="https://www.baseclear.com/bioinformatics-tools/">https://www.baseclear.com/bioinformatics-tools/</a>
L_RNA_Scaffolder	[59]	<a href="http://www.fishbrowser.org/software/L_RNA_scaffolder/">http://www.fishbrowser.org/software/L_RNA_scaffolder/</a>
Tandem Repeats Finder	[60]	<a href="https://tandem.bu.edu/trf/trf.html">https://tandem.bu.edu/trf/trf.html</a>
RepeatModeler	N/A	<a href="http://www.repeatmasker.org">http://www.repeatmasker.org</a>
RepeatMasker	N/A	<a href="http://www.repeatmasker.org">http://www.repeatmasker.org</a>
MAKER2	[61]	<a href="http://www.yandell-lab.org/software/maker.html">http://www.yandell-lab.org/software/maker.html</a>
Semi-HMM-based Nucleic Acid Parser	[62]	<a href="http://korflab.ucdavis.edu/software.html">http://korflab.ucdavis.edu/software.html</a>
AUGUSTUS	[63]	<a href="http://augustus.gobics.de/">http://augustus.gobics.de/</a>
Samtools	[64]	<a href="http://samtools.sourceforge.net/">http://samtools.sourceforge.net/</a>
SEX-DETECTOR	[65]	<a href="http://lbbe.univ-lyon1.fr/-SEX-DETECTOR-.html?lang=fr">http://lbbe.univ-lyon1.fr/-SEX-DETECTOR-.html?lang=fr</a>
HaplotypeCaller	[66]	<a href="https://software.broadinstitute.org/gatk/">https://software.broadinstitute.org/gatk/</a>
SnEff	[67]	<a href="http://snpeff.sourceforge.net/">http://snpeff.sourceforge.net/</a>
blastn	[68]	<a href="https://blast.ncbi.nlm.nih.gov/Blast.cgi">https://blast.ncbi.nlm.nih.gov/Blast.cgi</a>
CEGMA VM	[10]	<a href="http://korflab.ucdavis.edu/datasets/cegma/cegma_vm.html">http://korflab.ucdavis.edu/datasets/cegma/cegma_vm.html</a>

Figure 1  
Number of  
mutations



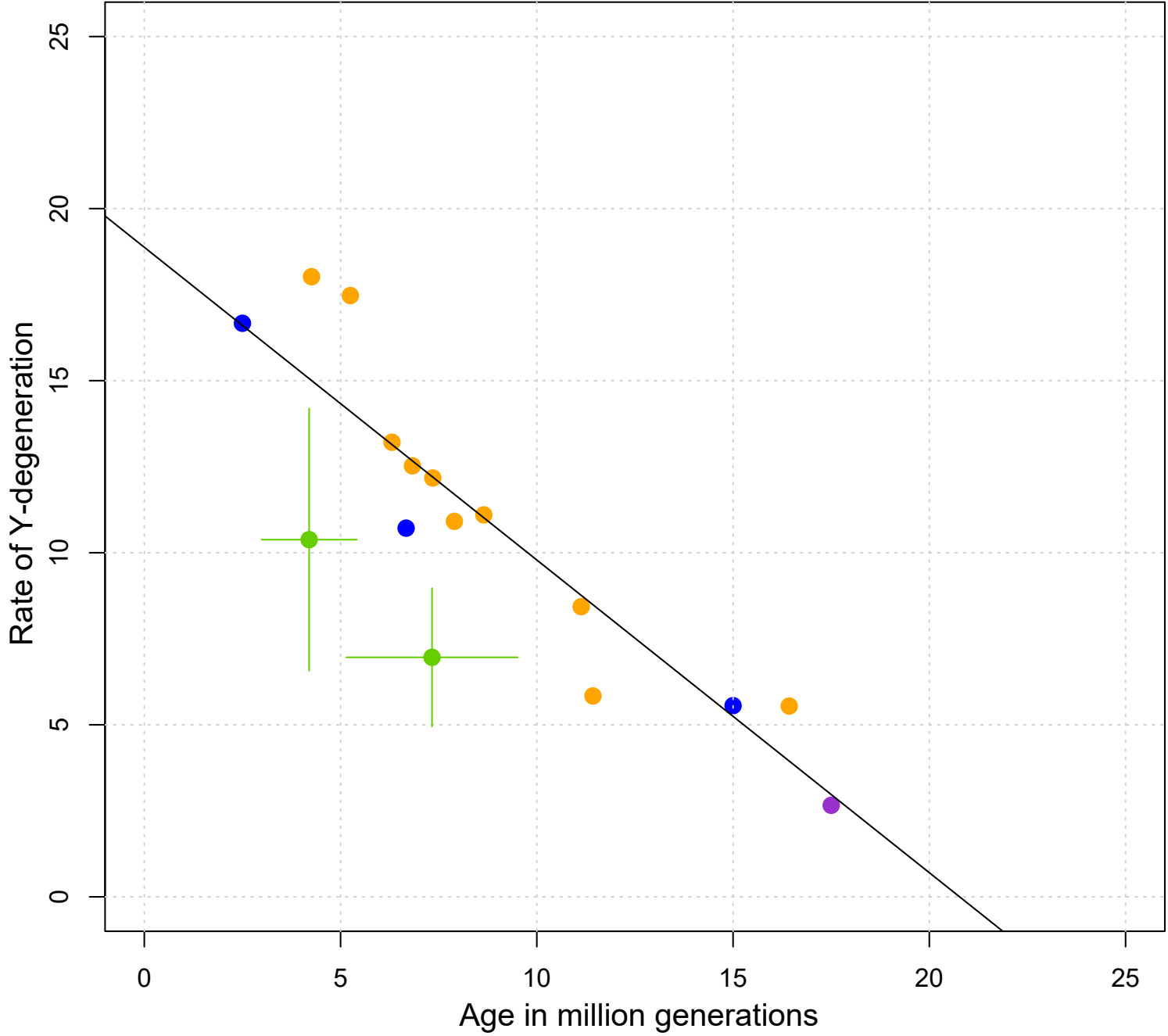
$$R_1 (GC \text{ to } AT) = 9.71 \times 10^{-9}$$

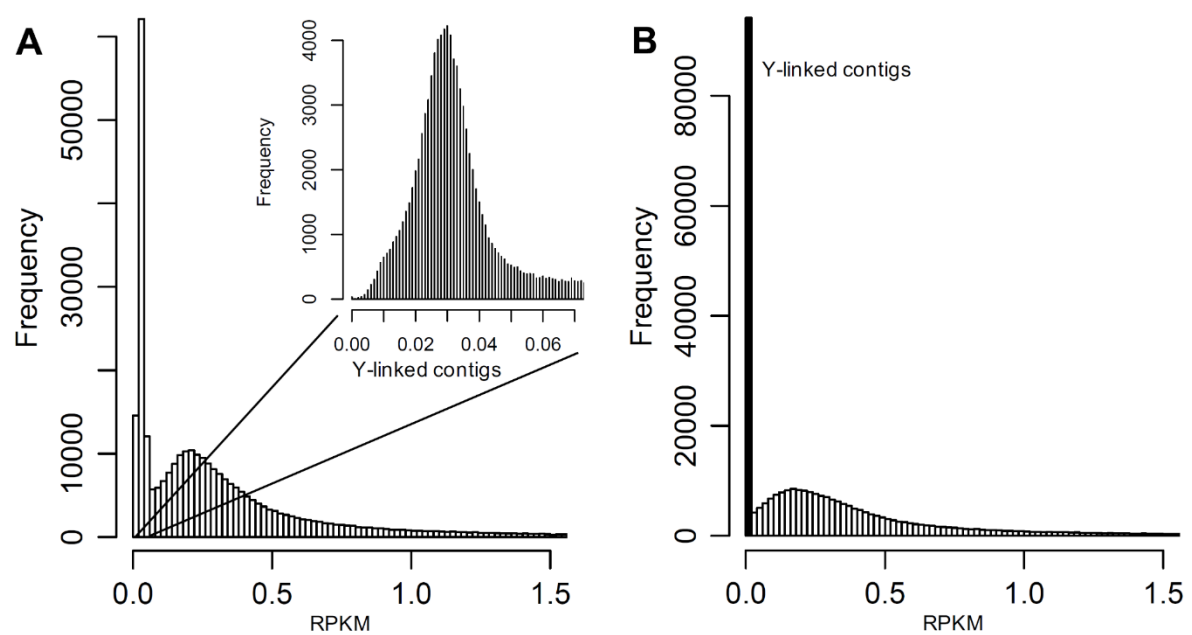
$$R_2 (AT \text{ to } GC) = 3.21 \times 10^{-9}$$

$$R_3 (GC \text{ to } GC) = 1.08 \times 10^{-9}$$

$$R_4 (AT \text{ to } AT) = 2.33 \times 10^{-9}$$

Figure 2





**Figure S1. The distribution of normalised sequence coverage (RPKM) in males (A) and females (B), related to the coverage-based identification of Y-linked contigs in the section “Assembly and Annotation of a Male *Silene latifolia* Reference Genome”**



Total contigs	319,549
Shortest contig (bp)	200
Longest contig (bp)	371,899
Contigs > 10,000 (bp)	23,010
Total length (bp)	1,185,243,700
N50 (bp)	10,785
Contigs > N50	21,428
X linked contigs	812
Total X linked length (bp)	21,756,364
Y linked contigs	74,433
Total Y linked length (bp)	120,392,727
PAR contigs	398
Total PAR length (bp)	8,919,216

(PAR = pseudoautosomal region)

**Table S1. Summary statistics for the *Silene latifolia* Sa984 male reference genome assembly, related to text section “Assembly and Annotation of a Male *Silene latifolia* Reference Genome”**

Scf	Pos	Ref	Alt	Effect	Linkage
898	17180	C	T	Intergenic	Aut
916	57274	G	A	Intergenic	Aut
1017	4319	C	T	Intergenic	Aut
1095	1271	T	G	Intergenic	Aut
1115	33150	T	A	Intergenic	Aut
1127	18681	A	T	Intergenic	Aut
1550	27017	T	G	Intergenic	X
4146	5394	G	C	Intergenic	Aut
5846	25355	C	T	Intergenic	Aut
6552	37709	A	T	Intergenic	Aut
8527	4934	G	A	Intergenic	Aut
8750	5404	T	G	Intergenic	Y
9085	16868	G	A	Intergenic	Aut
11338	13849	G	A	Intergenic	Aut
14109	231	G	A	Intergenic	Aut
15753	13834	G	C	Intergenic	Aut
16012	372	T	A	Intergenic	Aut
16417	10953	G	A	Intergenic	Aut
18411	1127	G	A	Intergenic	Aut
18895	358	C	A	Intergenic	Aut
22236	1353	T	A	Intergenic	Aut
23385	16463	G	A	Intergenic	Aut
24571	612	T	C	Intergenic	Aut
27165	6781	C	T	Intergenic	Aut
32129	72	T	A	Intergenic	Aut
32282	7907	A	G	Intergenic	Aut
34014	1156	C	T	Intergenic	Aut
47558	1790	A	G	Intergenic	Aut
55489	1976	A	G	Intergenic	Aut
57259	2693	C	T	Intergenic	Aut
65577	3800	A	G	Intergenic	Aut
104988	507	A	C	Intergenic	Aut
129750	410	G	T	Intergenic	Aut
132723	262	T	C	Intergenic	Aut
229154	704	T	A	Intergenic	Aut
236317	218	T	C	Intergenic	Aut
291949	457	A	T	Intergenic	Aut
299105	56	G	T	Intergenic	Aut
314884	143	C	T	Intergenic	Aut

Scf: scaffold number, Pos: position, Ref: reference allele, Alt: predicted mutation. The effect column gives results from SnpEff. The sequence column gives the location of the mutation deduced by SNP segregation around the mutation (Aut for autosomal, X or Y for sex linked).

**Table S5. List of *de novo* mutations identified in the study, related to Figure 1**



[Click here to access/download](#)

**Supplemental Videos and Spreadsheets**  
**Suppl\_Table\_S2.xlsx**





[Click here to access/download](#)

**Supplemental Videos and Spreadsheets**  
**Suppl\_Table\_S3.xlsx**





[Click here to access/download](#)

**Supplemental Videos and Spreadsheets**  
**Suppl\_Table\_S4.xlsx**

