

1 **Maternal X chromosome upregulation in both sexes initiates dosage**  
2 **compensation evolution**

3  
4 **Authors:** Aline Muyle<sup>1\*</sup>, Nicklaus Zemp<sup>2</sup>, Cécile Fruchard<sup>1</sup>, Radim Cegan<sup>3</sup>, Jan Vrana<sup>4</sup>,  
5 Clothilde Deschamps<sup>5</sup>, Raquel Tavares<sup>1</sup>, Franck Picard<sup>1</sup>, Roman Hobza<sup>3,4</sup>, Alex Widmer<sup>2</sup>, Gabriel  
6 AB Marais<sup>1</sup>

7 **Affiliations:**

8 <sup>1</sup>Laboratoire “Biométrie et Biologie Evolutive”, CNRS / Université Lyon 1, France.

9 <sup>2</sup>Institute of Integrative Biology, ETH Zurich, Switzerland.

10 <sup>3</sup>Institute of Biophysics, Academy of Sciences of the Czech Republic, Brno.

11 <sup>4</sup>Center for Biotechnological and Agricultural Research in Hana, Olomouc, Czech Republic

12 <sup>5</sup>Center for Bioinformatics in Rhône-Alpes (PRABI), Lyon, France

13 \*Correspondence to: [aline.muyle@univ-lyon1.fr](mailto:aline.muyle@univ-lyon1.fr)

14

15 **Abstract:** When sex chromosomes evolve, Y genes become less expressed than their X  
16 counterpart. This degeneration is compensated through various mechanisms in different animal  
17 species, that reestablish ancestral expression levels in males or balance males and females. This  
18 phenomenon, called dosage compensation, has been observed for some genes in the plant *Silene*  
19 *latifolia*. However, the mechanism involved remains unknown. Using an outgroup without sex  
20 chromosomes as a reference, we show that the maternal X chromosome is hyperexpressed in  
21 both sexes in *S. latifolia*. This compensates for the lower Y expression in males. However, the  
22 paternal X chromosome in females maintained its ancestral expression levels, causing global  
23 hyperexpression of sex chromosomes in females, which is likely to be suboptimal. Because *S.*  
24 *latifolia* sex chromosomes have evolved fairly recently, our findings provide insights into the  
25 first steps of dosage compensation evolution, in addition to revealing a link between dosage  
26 compensation and imprinting.

27

28 **One Sentence Summary:** In *Silene latifolia*, Y chromosome expression degeneration is  
29 compensated by hyperexpression of the maternal X chromosome in both males and females.

30

31 **Main Text**

32 Sex chromosomes have repeatedly evolved from autosomes in different taxa (1). Y  
33 chromosomes initially have similar gene content compared to the X, but with time Y genes  
34 become less expressed and eventually get lost. This causes reduced dosage in XY males  
35 compared to XX females, with potential negative effects on fitness. In some species, a  
36 mechanism called dosage compensation evolved that reestablishes ancestral expression levels in  
37 males. In some cases only, this is accompanied by dosage balance, i.e. balanced expression  
38 between males and females (2–4). For example, in *Drosophila* the X chromosome is  
39 hyperexpressed specifically in males, resulting in complete dosage compensation along with  
40 dosage balance (5). In placental mammals, on the other hand, one X is inactivated in females,  
41 resulting in dosage balance between males and females without dosage compensation, except for  
42 a few genes which expression was doubled (6–10). The situation is similar in *Caenorhabditis*  
43 *elegans*, where both X chromosomes are downregulated in hermaphrodites and only a few genes  
44 have their expression doubled for complete dosage compensation (11). It is currently unknown  
45 how these mechanisms evolved, that is to say whether expression was doubled for a few genes  
46 first or whether X chromosome inactivation happened first. To answer this question, species with  
47 younger sex chromosomes must be studied.

48 The plant *Silene latifolia* is an ideal model to study early steps of sex chromosome  
49 evolution thanks to its X/Y pair that evolved ~5 Mya (12). In this species, only dosage balance  
50 has been studied so far. Equal expression levels were observed for males and females for some  
51 genes in spite of Y expression degeneration (13–18). However, the mechanism through which  
52 dosage balance is achieved and whether dosage compensation evolved in *S. latifolia* is unknown.  
53 In order to elucidate these questions, an outgroup without the sex chromosome pair must be used  
54 as an ancestral autosomal reference (19). This allows to determine whether X chromosome  
55 expression increased or decreased in *S. latifolia* compared to the outgroup. Also, there is a  
56 technical bias that was not appropriately corrected in previous analyses of dosage compensation

57 in *S. latifolia* (19). When RNA-seq reads are mapped on a transcriptome reference, there is a bias  
58 for reads to map better when they carry the same allele as the reference. This influences  
59 downstream expression level estimates. Here, we present an analysis of dosage compensation in  
60 *S. latifolia* using an outgroup species as a reference and correcting for reference mapping bias.  
61 We also develop a new statistical approach, as required to fully characterize dosage  
62 compensation (19).

63 Because only a third of the large and highly repetitive *S. latifolia* genome has so far been  
64 sequenced (18), we used an RNA-seq approach based on the sequencing of a cross (parents and  
65 progeny of each sex), to infer sex-linked contigs (i.e. contigs located on the non-recombining  
66 region of the sex chromosome pair) (20). X/Y contigs and X-hemizygous contigs (X-linked  
67 contigs without Y allele expression) were inferred separately for three tissues: flower buds,  
68 seedlings and leaves (Supplementary Table S2). About half of the inferred sex-linked contigs  
69 could be validated by independent data (Supplementary Table S2 and Materials and Methods).  
70 Contigs with sex-biased expression are not expected to be dosage compensated as they are likely  
71 to be involved in sex-specific functions. They were therefore removed for further analyses  
72 (Supplementary Table S2 and Materials and Methods). X-hemizygous contigs have previously  
73 been characterized as partially dosage balanced (17, 18) and are discussed in Supplementary Text  
74 S1.

75 Paternal and maternal allele expression levels in males and females were estimated for  
76 sex-linked and autosomal contigs in *S. latifolia* after correcting for reference mapping bias  
77 (Materials and Methods). Then, the corresponding allelic expression levels in one or two closely  
78 related outgroups without sex chromosomes was used as a reference to orientate expression  
79 changes in *S. latifolia*. For autosomal contigs, expression levels remained statistically similar  
80 between *S. latifolia* and the outgroups (Figure 1). This is due to the close relatedness of *S.*  
81 *latifolia* and the outgroups, and validates the use of the outgroups as a reference for ancestral  
82 expression levels. We used the ratio of Y over X expression levels in males as a proxy for Y  
83 degeneration in order to categorize X/Y contigs. As the Y allele degenerates (paternal allele in  
84 blue in Figure 1), the expression of the X allele in males increases (maternal allele in red in  
85 Figure 1). This is the first evidence of dosage compensation in *S. latifolia*, i.e. ancestral  
86 expression levels are reestablished in males in spite of Y expression degeneration. In females, the  
87 maternal X allele expression increases with Y degeneration (gray bars in Figure 1), in a way

88 similar to the maternal X allele in males. The paternal X allele in females, however, maintained  
89 its ancestral expression level regardless of Y degeneration (black bars in Figure 1). Hence,  
90 dosage balance between males and females is not achieved in *S. latifolia* due to a global  
91 hyperexpression of sex-linked contigs in females compared to ancestral expression levels. All  
92 results were confirmed on two other tissues as well as on validated contigs only (Supplementary  
93 Figures S1-S6).

94 The maternal X allele is hyperexpressed both in males and females (Figure 1 and  
95 Supplementary Figures S1-S6), suggesting a link between dosage compensation and imprinting  
96 (the differential expression of alleles depending on the parent of origin). In order to statistically  
97 test this observation, we used a linear regression model with mixed effects (Materials and  
98 Methods). The outgroups were used as a reference and expression levels in *S. latifolia* were then  
99 analyzed while taking into account the variability due to contigs and individuals. The interaction  
100 between the effect of the parental origin and the degeneration level was estimated, which allowed  
101 to plot the difference in expression between the maternal and paternal alleles in females for  
102 different degeneration categories (Figure 2). In autosomal contigs, the maternal and paternal  
103 alleles are expressed in a similar way in females, showing a global absence of imprinting for  
104 these contigs. However, for X/Y contigs, the difference between the maternal and paternal X in  
105 females increases with Y degeneration, demonstrating the existence of a link between the  
106 evolution of dosage compensation and imprinting in *S. latifolia*. Results were confirmed on two  
107 other tissues as well as on validated contigs only (Supplementary Figures S7-S12).

108 Our findings provide the first characterization of a dosage compensation mechanism  
109 where sex chromosomes are hyperexpressed in females compared to ancestral expression levels  
110 (4). This suggests that underexpression of sex-linked genes in males is more deleterious than  
111 hyperexpression in females. The young age of *S. latifolia* sex chromosomes could explain this  
112 suboptimal pattern, and dosage balance between sexes might later evolve in this species. The  
113 results imply that hyperexpression of sex-linked genes in males and females evolves prior to X-  
114 inactivation or X downregulation, as hypothesized by Ohno for placental mammals (21).

115 Previous studies that showed dosage balance between sexes in *S. latifolia* were criticized  
116 as being a possible result of buffering mechanisms, where one copy of a gene is expressed at a  
117 higher level when haploid than when diploid due to higher availability of the cell machinery (18,

118 22, 23). However, the observed hyperexpression of the X chromosome in *S. latifolia* males cannot  
119 be explained by buffering mechanisms alone, otherwise the maternal X in females wouldn't be  
120 hyperexpressed. Rather, our findings provide evidence for a strong link between dosage  
121 compensation evolution and imprinting. This link has previously been observed in marsupials,  
122 where the paternal X chromosome is systematically the one to be inactivated (24). This is also  
123 the case in mice placenta (25).

124 It would be interesting to understand how hyperexpression of the maternal X is achieved  
125 in *S. latifolia* males and females at the molecular level. Chromosome staining suggests that DNA  
126 methylation could be involved. Indeed, one arm of one of the two X chromosomes in females  
127 was shown to be hypomethylated, as well as the same arm of the single X in males (26). Given  
128 our results, it is very likely that this hypomethylated X chromosome is the maternal  
129 hyperexpressed X. Unfortunately, parental origin of the X chromosomes was unknown in this  
130 study (26). It would be extremely insightful in future to study DNA methylation patterns in *S.*  
131 *latifolia* paternal and maternal X chromosomes, along with a homologous pair in a closely  
132 related species without sex chromosomes.

133

134

## 135 **References and Notes**

1. D. Bachtrog, Y-chromosome evolution: emerging insights into processes of Y-chromosome degeneration. *Nat. Rev. Genet.* **14**, 113–124 (2013).
2. B. Charlesworth, The evolution of chromosomal sex determination and dosage compensation. *Curr. Biol. CB.* **6**, 149–162 (1996).
3. J. E. Mank, Sex chromosome dosage compensation: definitely not for everyone. *Trends Genet. TIG.* **29**, 677–683 (2013).
4. L. Gu, J. Walters, Sex chromosome dosage compensation evolution in animals: a beautiful theory, undermined by facts and bedeviled by details. *Genome Biol. Evol.* (2017).
5. M. I. Kuroda, A. Hilfiker, J. C. Lucchesi, Dosage Compensation in *Drosophila*-a Model for the Coordinate Regulation of Transcription. *Genetics.* **204**, 435–450 (2016).

6. E. Pessia, J. Engelstädter, G. A. B. Marais, The evolution of X chromosome inactivation in mammals: the demise of Ohno's hypothesis? *Cell. Mol. Life Sci. CMLS.* **71**, 1383–1394 (2014).
7. C. M. Disteche, Dosage compensation of the sex chromosomes. *Annu. Rev. Genet.* **46**, 537–560 (2012).
8. P. Julien *et al.*, Mechanisms and evolutionary patterns of mammalian and avian dosage compensation. *PLoS Biol.* **10**, e1001328 (2012).
9. E. Pessia, T. Makino, M. Bailly-Bechet, A. McLysaght, G. A. B. Marais, Mammalian X chromosome inactivation evolved as a dosage-compensation mechanism for dosage-sensitive genes on the X chromosome. *Proc. Natl. Acad. Sci. U. S. A.* **109**, 5346–5351 (2012).
10. S. Ercan, Mechanisms of x chromosome dosage compensation. *J. Genomics.* **3**, 1–19 (2015).
11. S. E. Albritton *et al.*, Sex-biased gene expression and evolution of the x chromosome in nematodes. *Genetics.* **197**, 865–883 (2014).
12. A. Rautenberg, L. Hathaway, B. Oxelman, H. C. Prentice, Geographic and phylogenetic patterns in *Silene* section *Melandrium* (Caryophyllaceae) as inferred from chloroplast and nuclear DNA sequences. *Mol. Phylogenet. Evol.* **57**, 978–991 (2010).
13. M. V. Chibalina, D. A. Filatov, Plant Y chromosome degeneration is retarded by haploid purifying selection. *Curr. Biol. CB.* **21**, 1475–1479 (2011).
14. R. Bergero, D. Charlesworth, Preservation of the Y transcriptome in a 10-million-year-old plant sex chromosome system. *Curr. Biol. CB.* **21**, 1470–1474 (2011).
15. A. Muyle *et al.*, Rapid de novo evolution of X chromosome dosage compensation in *Silene latifolia*, a plant with young sex chromosomes. *PLoS Biol.* **10**, e1001308 (2012).

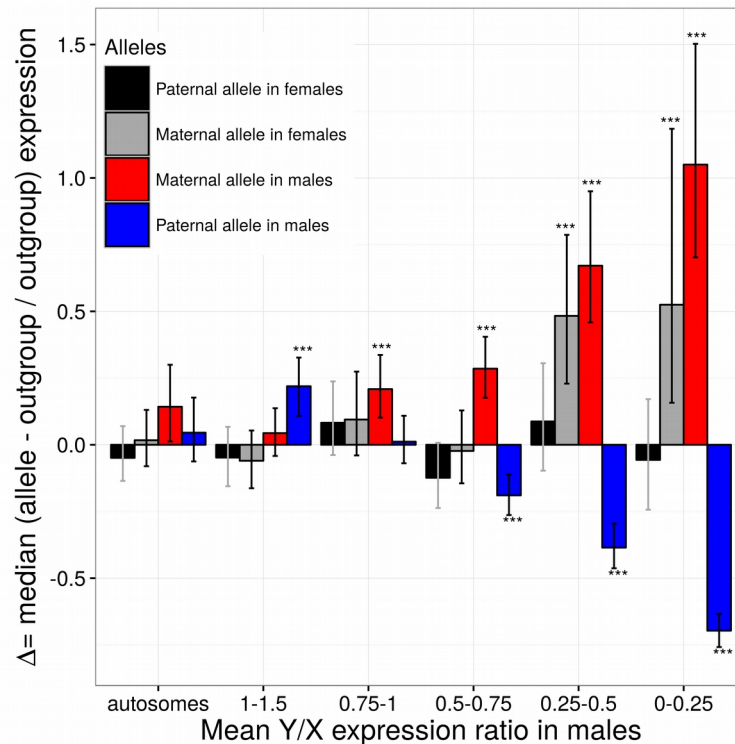
16. Y. Kazama *et al.*, SIWUS1; an X-linked gene having no homologous Y-linked copy in *Silene latifolia*. *G3 Bethesda Md.* **2**, 1269–1278 (2012).
17. R. Bergero, S. Qiu, D. Charlesworth, Gene Loss from a Plant Sex Chromosome System. *Curr. Biol. CB* (2015), doi:10.1016/j.cub.2015.03.015.
18. A. S. T. Papadopoulos, M. Chester, K. Ridout, D. A. Filatov, Rapid Y degeneration and dosage compensation in plant sex chromosomes. *Proc. Natl. Acad. Sci. U. S. A.* **112**, 13021–13026 (2015).
19. A. Muyle, R. Shearn, G. A. Marais, The Evolution of Sex Chromosomes and Dosage Compensation in Plants. *Genome Biol. Evol.* **9**, 627–645 (2017).
20. A. Muyle *et al.*, SEX-DETECTOR: A Probabilistic Approach to Study Sex Chromosomes in Non-Model Organisms. *Genome Biol. Evol.* **8**, 2530–2543 (2016).
21. S. Ohno, Sex chromosomes and sex-linked genes. *Springer-Verl. Berl. Heidelb. N. Y.* (1967).
22. J. H. Malone *et al.*, Mediation of *Drosophila* autosomal dosage effects and compensation by network interactions. *Genome Biol.* **13**, r28 (2012).
23. R. A. Veitia, F. Veyrunes, S. Bottani, J. A. Birchler, X chromosome inactivation and active X upregulation in therian mammals: facts, questions, and hypotheses. *J. Mol. Cell Biol.* **7**, 2–11 (2015).
24. D. J. Whitworth, A. J. Pask, The X factor: X chromosome dosage compensation in the evolutionarily divergent monotremes and marsupials. *Semin. Cell Dev. Biol.* **56**, 117–121 (2016).
25. A. Wagschal, R. Feil, Genomic imprinting in the placenta. *Cytogenet. Genome Res.* **113**, 90–98 (2006).

26. J. Siroky, M. R. Castiglione, B. Vyskot, DNA methylation patterns of *Melandrium album* chromosomes. *Chromosome Res. Int. J. Mol. Supramol. Evol. Asp. Chromosome Biol.* **6**, 441–446 (1998).

136

137

## 138 Figures

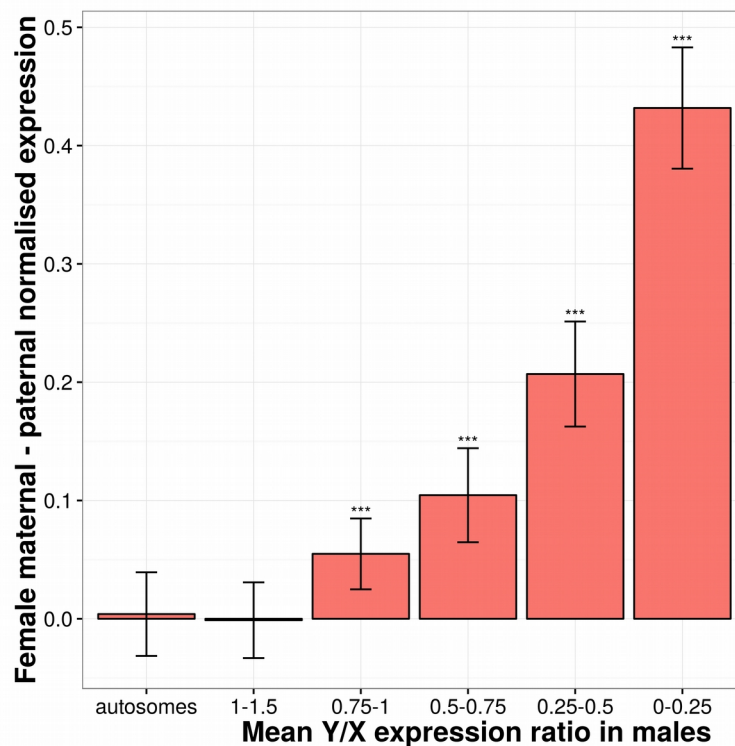


139

140 **Fig. 1:** Normalised difference in allelic expression levels between *S. latifolia* and the outgroup  
141 without sex chromosomes *S. vulgaris* (hereafter  $\Delta$ ), in autosomal and sex-linked contigs  
142 for the seedling tissue. Maternal and paternal allelic read numbers were summed at SNP  
143 positions and normalised for each individual separately, then averaged among individuals  
144 for each contig.  $\Delta$  was computed as follows:  $\Delta = (\text{allelic expression in } S. \textit{latifolia} - \text{allelic}$   
145  $\text{expression in the outgroup}) / \text{allelic expression in the outgroup}$ . If  $\Delta$  is lower, higher or  
146 equal to zero, then expression in *S. latifolia* is respectively lower, higher or equal to the  
147 outgroup. For all contig categories,  $\Delta$  was compared to zero using a Wilcoxon test. The  
148 median  $\Delta$ , confidence intervals and p-values adjusted for multiple testing using a



149 Benjamini and Hochberg correction are shown (\*\*\*: p-value < 0.001; \*\*: p-value < 0.01,  
150 \*: p-value < 0.05). The Y/X ratio was computed in *S. latifolia* males and averaged among  
151 individuals to use as a proxy for Y degeneration. Contigs with sex-biased expression were  
152 removed, as well as contigs with Y/X expression ratios above 1.5. Sample sizes for the  
153 different contig categories are: autosomal:200; 1-1.5:148; 0.75-1:139; 0.5-0.75:160; 0.25-  
154 0.5:114; 0-0.25:79 (note that 200 autosomal contigs were randomly selected in order to  
155 have similar statistical power among gene categories). In the absence of dosage  
156 compensation, the single X in males should be expressed at levels similar to the outgroup  
157 that does not have sex chromosomes, in other words, without dosage compensation  $\Delta$   
158 should be close to zero for the maternal allele in males (red bars). Results show that the  
159 maternal allele is hyper-expressed in *S. latifolia* when the Y chromosome is degenerated,  
160 both in males and females.  
161



162  
163 **Fig. 2:** Normalised expression difference between the maternal and paternal allele in *S. latifolia*  
164 females in autosomal and sex-linked contigs for the seedling tissue. The Y axis unit is the  
165 normalised allelic read count difference in log scale. A linear regression model with

166 mixed effects was used to study allelic expression in *S. latifolia* for every SNP position.  
167 In order to measure the changes in *S. latifolia* expression due to sex chromosomes  
168 evolution, the outgroup *S. vulgaris* that does not have sex chromosomes was used as a  
169 reference in the model (see Materials and Methods for details). The framework provided  
170 estimates for the normalised difference between the effect of paternal and maternal origin  
171 of alleles in interaction with the contig status (autosomal or sex-linked with various levels  
172 of Y degeneration), while accounting for inter-contig and inter-individual variability. See  
173 Fig. 1 legend for sample sizes for the different contig categories and statistical  
174 significance symbols. Results show that Y degeneration is linked to a significant  
175 expression difference between the paternal and maternal alleles in females, which is not  
176 observed in autosomal and non-degenerated sex-linked contigs.

177

## Supplementary Materials

### 178 **Supplementary Text 1: Dosage compensation in X-hemizygous genes**

179

180 X-hemizygous contigs are inferred by SEX-DETECTOR from X polymorphisms that  
181 segregate in the absence of Y allele expression in the RNA-seq data. For that reason, X-  
182 hemizygous contigs are harder to detect compared to X/Y contigs due to lower polymorphism  
183 (X/Y contigs can be detected both with X and X/Y polymorphisms). Another inherent bias to X-  
184 hemizygous contig inference comes from the assembly step. If the X and the Y copy are too  
185 divergent to be assembled together, the X contig will be detected as X-hemizygous by SEX-  
186 DETECTOR. This bias was at least partly corrected in the analyses presented here (see Material and  
187 Method section 4). Also, previous work showed that only some X-hemizygous genes are dosage  
188 compensated in *Silene latifolia* (1–5). For all the above reasons, results on X-hemizygous contigs  
189 are analysed separately here. Poor dosage compensation of X-hemizygous contigs compared to  
190 X/Y contigs with high Y degeneration was observed across all tissues (Supplementary Figures 1  
191 to 6). Also, the parental origin of the X chromosome has limited to no effect on female X  
192 expression levels for X-hemizygous contigs, unlike X/Y contigs (Supplementary Figures 7 to  
193 12). A reason that could explain such a different pattern for X-hemizygous genes compared to  
194 X/Y genes is the possible dosage insensitivity of X-hemizygous genes. X-hemizygous genes  
195 could have lost their Y copy because dosage was not important for them and selection neither  
196 slowed down the loss of the Y copy nor selected for dosage compensation when degeneration  
197 inevitably occurred (5). A well described characteristic of dosage sensitive genes is that they tend  
198 to code proteins involved in large complexes (6). Using the GO-term analysis, X-hemizygous  
199 contigs were found to be significantly depleted in ribosomal protein coding genes, which is  
200 consistent with the global dosage insensitivity of X-hemizygous genes in *S. latifolia*.

201

202

203

## 204 **Materials and Methods:**

### 205 **1) RNA-seq data**

#### 206 **1.1) Plant material and sequencing**

207 RNA-seq data were generated for a cross in the dioecious plant *S. latifolia* (which has sex  
208 chromosomes), a cross in the gynodioecious plant *S. vulgaris* (which does not have sex  
209 chromosomes) and for individuals of the hermaphrodite plant *S. viscosa* (which does not have  
210 sex chromosomes either). Plants were grown in the greenhouse from seeds collected in the wild.  
211 The following individuals and tissues were sequenced (see Supplementary Table 1 for library  
212 sizes):

213 • For flower buds for *S. latifolia*, both parents were sequenced (a male from a wild  
214 population, Leuk144-3\_father, and a female from a ten-generation inbred line, U10\_37\_mother),  
215 along with their progeny (C1\_01\_male, C1\_3\_male, C1\_04\_male, C1\_05\_male, C1\_26\_female,  
216 C1\_27\_female, C1\_29\_female, C1\_34\_female). For flower buds for *S. vulgaris*, the  
217 hermaphrodite father came from a wild population (Guarda\_1), the female mother from another  
218 wild population (Seebach\_2) and their hermaphrodite (V1\_1, V1\_2, V1\_4) and female (V1\_5,  
219 V1\_8, V1\_9) progeny were sequenced. For flower buds for *S. viscosa*, three individuals were  
220 sequenced (Svi\_BUL, Svi\_KIR, SVi\_RUS).

221 • For leaves for *S. latifolia* the same cross progeny individuals were sequenced  
222 (C1\_01\_male, C1\_3\_male, C1\_04\_male, C1\_05\_male, C1\_26\_female, C1\_27\_female,  
223 C1\_29\_female, C1\_34\_female). For leaves for *S. vulgaris* four hermaphrodites were sequenced  
224 (V2\_10\_L, V2\_11\_L, V2\_12\_L, V2\_13\_L). For leaves for *S. viscosa* two individuals were  
225 sequenced (Svi\_CZE, SVi\_SWE).

226 • For seedlings for *S. latifolia* different cross progeny individuals were sequenced  
227 (Seed\_lati\_female\_1, Seed\_lati\_female\_2, Seed\_lati\_female\_3, Seed\_lati\_female\_4,  
228 Seed\_lati\_male\_1, Seed\_lati\_male\_2, Seed\_lati\_male\_3, Seed\_lati\_male\_4). For seedlings for *S.*  
229 *vulgaris* four hermaphrodites were sequenced (Seed\_vulg\_herm\_1, Seed\_vulg\_herm\_2,  
230 Seed\_vulg\_herm\_3, Seed\_vulg\_herm\_4).

231 For the flower bud and leaf datasets, individuals were grown in a temperature-controlled  
232 greenhouse. The QiagenRNeasy Mini Plant extraction kit was used to extract total RNA two

233 times separately from four flower buds at developmental stages B1–B2 after removing the calyx.  
234 Samples were treated additionally with QiagenDNase. RNA quality was assessed with an Aligent  
235 Bioanalyzer (RIN.9) and quantity with an Invitrogen Qubit. An intron-spanning PCR product  
236 was checked on an agarose gel to exclude the possibility of genomic DNA contamination. Then,  
237 the two extractions of the same individual were pooled. Individuals were tagged and pooled for  
238 sequencing. Samples were sequenced by FASTERIS SA on an Illumina HiSeq2000 following an  
239 Illumina paired-end protocol (fragment lengths 150–250bp, 100 bp sequenced from each end).  
240 Seedlings were grown in a temperature controlled climate chamber in Eschikon (Switzerland)  
241 using the same conditions as in (7). The *S. latifolia* and *S. vulgaris* seedlings were collected  
242 without roots at the four leaf stage. The sexing of the *S. latifolia* seedlings was done using Y  
243 specific markers (8) that were amplified with the direct PCAR KAPA3G Plant PCR Kit (note  
244 male number 3 was later shown to be a female). High quality RNA (RIN > 7.5) was extracted  
245 using the total RNA mini kit of Geneaid. Twelve RNA-seq libraries were produced using the  
246 Truseq kit v2 from Illumina. Libraries were tagged individually and sequenced in two Illumina  
247 Hiseq 2000 channels at the D-BSSE (ETH Zürich, Switzerland) using 100 bp paired-end read  
248 protocol.

249 A normalised 454 library was generated for each sex of *S. latifolia* using bud extracts  
250 from 4 different developmental stages that were pooled.

251 Plants from an 11 generation inbred line were grown under controlled conditions in a  
252 greenhouse in Eschikon (Switzerland). One female (U11\_01) and one male (U11\_02) were  
253 randomly selected. From each plant, high quality RNA (RIN > 7.5) were extracted using the  
254 total RNA mini kit of Geneaid from very small flower buds, small and large flower buds, flowers  
255 before anthesis without calyces, rosette leaves and seedlings (4 leaves stage). Additionally,  
256 pollinated flowers (1 h after pollination) and developing fruits (5 day after pollination) were  
257 extracted from the female and pollen from the male. RNA of the different tissues was equally  
258 pooled for each sex and cDNA was produced using the Clontech SMARTer Kit. The two cDNA  
259 pools were then normalized using a duplex specific endonuclease of the Evrogen TRIMMER kit.  
260 For each sex two ranges were selected (1- 1.3 kb and 1.2 -2 kb) using the Pippin Prep (Sage  
261 Science). Four SMRTbell libraries were prepared using the C2 Pacific Biosciences (PacBio)  
262 chemistry and sequenced on a PacBio RS II at the Functional Genomic Center Zurich (FGCZ).

## 263 **1.2) Reference transcriptome**

264 A reference transcriptome was built from the *Silene latifolia* flower bud data and used for  
265 all tissue types and species afterwards. Adaptors, low quality and identical reads were removed.  
266 The transcriptome was then assembled using TRINITY (9) on the combined 10 individuals  
267 described previously as well as 6 individuals from (10) and the normalized 454 sequencing that  
268 was transformed to illumina using 454-to-illumina-transformed-reads. Isoforms were then  
269 collapsed using /trinity-plugins/rsem-1.2.0/rsem-prepare-reference. PolyA tails, bacterial RNAs  
270 and ribosomal RNAs were removed using ribopicker. ORFs were predicted with trinity  
271 transcripts\_to\_best\_scoring\_ORFs.pl. In order to increase the probability of X and Y sequences  
272 to be assembled in the same contig, ORFs were further assembled using home made perl scripts  
273 to run CAP3 (11) Version Date 10/15/07 inside of TRINITY components with parameter -p 70.

## 274 **2) Inference of sex-linked contigs**

275 Illumina reads from the 10 individuals of the cross were mapped onto the assembly using  
276 BWA (12) version 0.6.2 with the following parameters: bwa aln -n 5 and bwa sampe. The  
277 libraries were then merged using SAMTOOLS version 0.1.18 (13). The obtained alignments  
278 were locally realigned using GATK IndelRealigner (14) and were analysed using reads2snps (15)  
279 version 3.0 with the following parameters: -fis 0 -model M2 -output\_genotype best -multi\_alleles  
280 acc -min\_coverage 3 -par false. This allowed to genotype individuals at each loci while allowing  
281 for biases in allele expression, and without cleaning for paralogous SNPs. Indeed, X/Y SNPs  
282 tend to be filtered out by paraclean, a program which removes paralogous positions (16). SEX-  
283 DETector (17) was then used to infer contig segregation types after estimation of parameters  
284 using an SEM algorithm. Contig posterior segregation type probabilities were filtered to be  
285 higher than 0.8. Because the parents were not sequenced for the leaf and seedling datasets, SEX-  
286 DETector was run using the flower bud data for the parents.

## 287 **3) Reference mapping bias correction**

288 In order to avoid biases towards the reference allele in expression level estimates, a  
289 second mapping was done using the program GSNAP (18) with SNP tolerant mapping option. A  
290 GSNAP SNP file was generated by home-made perl scripts using the SEX-DETECTOR SNP detail  
291 output file. Shortly, for each polymorphic position of all contigs, the most probable posterior  
292 SNP type was used to extract the possible alleles and write them to the GSNAP SNP file. This

293 way, reference mapping bias was corrected for both sex-linked and autosomal contigs. Only  
294 uniquely mapped and concordant paired reads were kept after this. See Supplementary Table S1  
295 for percentage of mapped reads. SEX-DETECTOR was run a second time on this new mapping and  
296 the new inferences were used afterwards for all analyses (see Supplementary Table S2 for  
297 inference results).

#### 298 **4) Validation of sex-linked contigs**

##### 299 **3.1) Detection of false X-hemizygous contigs**

300 Erroneous inference of X-hemizygous contig can be due to a true X/Y gene which X and  
301 Y copies were assembled into different contigs. In order to detect such cases, X-hemizygous  
302 contigs were blasted (19) with parameter -e 1E-5 against RNA-seq contigs that have male-  
303 limited expression (see section 5 below for how male-limited contigs were inferred). These cases  
304 were removed from the analyses presented here.

##### 305 **3.2) Validation using data from literature**

306 A few sex-linked and autosomal genes in *S. latifolia* have already been described in the  
307 literature (see Supplementary Table S3).

##### 308 **3.3) Validation using a genetic map**

309 A genetic map was built and contigs from the X linkage group were used to validate  
310 SEX-DETECTOR inferences.

311 **Plant material and genotyping:** A female individual from an interspecific *S. latifolia* cross  
312 (C1\_37) was back crossed with a male from an 11 generation inbred line (U10\_49). The  
313 offspring (hereafter called BC1 individuals) were grown under controlled conditions in a  
314 greenhouse in Eschikon (Switzerland). High quality RNA from flower buds as described in (7)  
315 was extracted from 48 BC1 individuals (35 females and 13 males). 48 RNA-seq libraries were  
316 produced using the Truseq kit v2 from Illumina with a median insert size of about 200 bp.  
317 Individuals were tagged separately and sequenced in four Illumina HiSeq 2000 channels at the D-  
318 BSSE (ETH Zürich, Switzerland) using 100bp paired-end read protocol. The parents used for  
319 this back cross had previously been sequenced in a similar way (10, 20).

320 **Linkage group identification:** RNA-seq reads were mapped against the *S. latifolia* flower buds  
321 reference transcriptome using BWA (12) with a maximum number of mismatch equal to 5.

322 Libraries were merged and realigned using GATK (14) and SNPs were analysed using  
323 reads2snps (15). Using a customized perl script, SNP genotypes from the parents and the  
324 offspring as well as the associated posterior probabilities were extracted from the reads2snps  
325 output file. Informative SNPs according to CP and BC1 design and a posterior odd of 0.8 were  
326 then converted into a JoinMap format using a customized R script. If more than one informative  
327 SNP per contig was present, the SNP was used with less segregation distortion and less missing  
328 values. This led to 8,023 BC1 and 16,243 CP markers. Loci with more than 10 % missing values  
329 were excluded, which left 7,951 BC1 and 15,118 CP markers. Linkage groups were identified  
330 using the default setting of JoinMap 4.1 (21). Robustness of the assignment of the linkage groups  
331 was tested using LepMap (22). Blasting the contigs against known sex-linked genes allowed the  
332 identification of the X chromosome linkage group. Contigs could not be ordered along the  
333 linkage groups due to the too limited number of individuals that prevented the convergence of  
334 contig order. However, contigs were reliably attributed to linkage groups.

#### 335 **3.4) Validation using isolated Y chromosome DNA-seq data**

336 Y chromosome DNA was isolated using flow cytometry. The samples for flow cytometric  
337 experiments were prepared from root tips according to (23) with modifications. Seeds of *S.*  
338 *latifolia* were germinated in a petri dish immersed in water at 25°C for 2 days until optimal  
339 length of roots was achieved (1 cm). The root cells were synchronized by treatment with 2mM  
340 hydroxyurea at 25°C for 18h. Accumulation of metaphases was achieved using 2.5µM oryzalin.  
341 Approximately 200 root tips were necessary to prepare 1ml of sample. The chromosomes were  
342 released from the root tips by mechanical homogenization using a Polytron PT1200 homogenizer  
343 (Kinematica AG, Littau, Switzerland) at 18,000rpm for 13 s. The crude suspension was filtered  
344 and stained with DAPI (2µg/ml). All flow cytometric experiments were performed on FACSARIA  
345 II SORP flow cytometer (BD Biosciences, San José, Calif., USA). Isolated Y chromosomes were  
346 sequenced with 2x100bp PE Illumina HiSeq.

347 Reads were filtered for quality and Illumina adapters were removed using the ea-  
348 utils FASTQ processing utilities (24). The optimal kmer value for assembly was searched using  
349 KmerGenie (25). Filtered reads were assembled using soapdenovo2 (26) with kmer=49, as  
350 suggested by KmerGenie. The obtained assembly was highly fragmented, therefore RNA-seq  
351 data was used to join, order and orient the genomic fragments with L\_RNA\_scaffolder (27). The  
352 following RNA-seq reads were used (see section 1.1): one sample of male flower buds



353 sequenced by 454, 6 samples of male flower buds sequenced by Illumina paired-end, 4 samples  
354 of male leaves sequenced by Illumina paired-end and one sample of male pooled tissues  
355 sequenced by PacBio. The genomic assembly was successively scaffolded with  
356 L\_RNA\_scaffolder using RNA-seq samples one after the other, first 454 samples then Illumina  
357 and finally PacBio. The obtained contigs were filtered to be longer than 200pb.

### 358 **3.5) Set of validated sex-linked and autosomal contigs**

359 The three sources of data (litterature, genetic map and filtered Y sequence data) were  
360 compared to SEX-DETECTOR inferred sex-linked RNA-seq contigs using BLAST (19) with  
361 parameter -e 1E-5. Blasts were filtered for having a percentage of identity over 90%, an  
362 alignment length over 100bp and were manually checked. If a sex-linked RNA-seq contig  
363 blasted against a sequence from one of the three data sources (litterature, X genetic map or  
364 filtered Y DNA-seq) it was then considered as validated. See Supplementary Table S2 for  
365 numbers of validated sex-linked contigs.

### 366 **4) Expression level estimates**

#### 367 **4.1) whole contig expression levels**

368 Whole contig mean expression levels were obtained for each individual using GATK  
369 DepthOfCoverage (14) as the sum of every position coverage, divided by the length of the  
370 contig. Normalised expression levels, in RPKM (28), were then computed for each individual by  
371 dividing by the value by the library size of the individual (total number of mapped reads),  
372 accounting for different depths of coverage among individuals. Whole contig mean male and  
373 female expression levels were then computed by averaging male and female individuals for each  
374 contig.

#### 375 **4.2) Allelic expression levels**

376 In order to study separately X and Y allele expression levels in males and females,  
377 expression levels were studied at the SNP level. In *S. latifolia*, for each sex-linked contig  
378 expression levels were estimated using read counts from both X/Y and X-hemizygous  
379 informative SNPs. SNPs were attributed to an X/Y or X hemizygous segregation type if the  
380 according posterior probability was higher than 0.5. SNPs are considered informative if the  
381 father is heterozygous and has a genotype that is different from the mother (otherwise it is not

382 possible to tell apart the X from the Y allele and therefore it is not possible to compute X and Y  
383 expression separately). X/Y SNPs for which at least one female had over two percent of her  
384 reads belonging to the Y allele were removed as unlikely to be true X/Y SNPs.

385 Contigwise X, Y, X+X and X+Y normalised expression levels in RPKM (28) were  
386 computed by summing read numbers for each X-linked or Y-linked alleles for all SNPs of the  
387 contig and each individual separately and then normalised using the library size and the number  
388 of studied sex-linked SNPs in the contig:

$$389 \quad E = r / (n * l) \quad (1)$$

390 With E = normalised expression level for a given individual, r = sum of total read counts,  
391 n = number of studied SNPs, l = library size of the individual (number of mapped reads).

392 For contigs that only have X/X SNPs (SNPs for which the father's X is different to both  
393 Xs from the mother), Y expression level is only computed from the father as all males are  
394 homozygous in the progeny. Such contigs were therefore removed when having under 3 X/X  
395 SNPs to avoid approximations on the contig mean Y/X expression level (39 contigs removed in  
396 the flower buds dataset, 44 in the leaves dataset and 40 in the seedlings dataset).

397 In order to make *S. latifolia* expression levels comparable to *S. viscosa* and *S. vulgaris*  
398 for sex-linked contigs, *S. viscosa* and *S. vulgaris* expression levels were estimated using only the  
399 positions used in *S. latifolia* (informative X/Y or X-hemizygous SNPs). The read count of every  
400 position in every contig and for every *S. viscosa* and *S. vulgaris* individual was given by GATK  
401 DepthOfCoverage (14). Only positions corresponding to informative X/Y or X-hemizygous  
402 SNPs in *S. latifolia* were used to compute the expression level for each contig and each  
403 individual as explained in equation (1).

404 Contigwise *S. latifolia* X, Y, X+X, X+Y allelic expression levels were then averaged  
405 among individuals. Autosomal normalised expression levels in the two outgroups (*S. vulgaris*  
406 and *S. viscosa*) were averaged together.

#### 407 **5) Identification of sex-biased contigs**

408 The analysis was done separately for the three tissues (flower buds, seedling and rosette  
409 leaves) as in (20). Contigs were filtered to be sufficiently expressed in at least half of the female  
410 and/or male libraries using a cutoff of 1 count per million reads. Contigs with sex-biased

411 expression were identified using a common dispersion in edgeR (29) including the different  
412 replicates for the contigs inferred to sex-linked and X-hemizygous separately. See  
413 Supplementary Table S2 for numbers of sex-biased contigs removed in order to study dosage  
414 compensation.

415 Male-limited expressed contigs were identified by calculating the mean expression values  
416 (FPKM) in both sexes and selecting those which were exclusively expressed in males.

#### 417 **6) Analysis of expression divergence between *S. latifolia* and the two outgroups**

418 The normalised difference in allelic expression between *S. latifolia* and the two outgroups  
419 (hereafter  $\Delta$ ) was computed in order to study how sex chromosome expression levels evolved in  
420 *S. latifolia* compared to autosomal expression levels in the two outgroups:  $\Delta$  is equal to zero if *S.*  
421 *latifolia* and the outgroups have equal expression levels,  $\Delta$  is positive if *S. latifolia* has higher  
422 expression levels compared to the outgroups and  $\Delta$  is negative otherwise (2).

423 
$$\Delta = (S. latifolia \text{ expression level} - \text{outgroup expression level}) / (\text{outgroup expression}$$
  
424 level) (2)

425 Sex-linked contigs were grouped by categories of degeneration level using the average Y  
426 over X expression ratio in males. 200 autosomal contigs were randomly selected in order to have  
427 similar statistical power among gene categories.  $\Delta$  values for each allele (maternal and paternal  
428 in males and females) and each gene category were compared to zero using a Wilcoxon test. P-  
429 values were corrected for multiple testing using a Benjamini and Hochberg correction. The  
430 estimated median  $\Delta$ , confidence intervals and adjusted p-values were then used to plot Figure 1  
431 and Supplementary Figures S1 to S6.

#### 432 **7) Analysis of expression level difference between maternal and paternal alleles**

433 Maternal and paternal alleles expression were compared in *S. latifolia* for autosomal and  
434 sex-linked contigs. In order to deal with the difference in numbers of autosomal versus sex-  
435 linked contigs (Supplementary Table S2), 200 autosomal contigs were randomly selected in order  
436 to keep comparable powers of detection. Allelic expression levels in *S. latifolia* for each  
437 individual at every SNP position of each contig were analysed using a linear regression model  
438 with mixed effects (with the R package lme4, 30). We assumed a normal distribution of the read  
439 count data after log transformation. In order to account for inter-individual and inter-contig

440 variability, a random “individual” and a random “contig” effect were included in the model. The  
441 aim of this modeling framework was to estimate the joint effect of the chromosomal origin of  
442 alleles (paternal or maternal in males or females) and the status of the gene (autosomal or sex-  
443 linked with various levels of Y degeneration defined by the average Y over X expression ratio in  
444 males). Two fixed effects with interaction were therefore considered in the model (3). In order to  
445 estimate the changes in sex-linked gene expression levels since the evolution of sex  
446 chromosomes, we used the average of the two outgroup expression levels as a reference (offset)  
447 for every SNP position, divided by two in order to be comparable to *S. latifolia* allelic expression  
448 levels.

449  $\log(\text{Expression}+1) \sim \text{Chromosome} * \text{Degeneration} + (1|\text{individual}) + (1|\text{Contig}), \text{offset} =$   
450  $\log(\text{outgroup average expression}/2 + 1)$  (3)

451 All effects of the model (fixed or random) were proved highly significant (p-values <  
452  $2.2 \cdot 10^{-16}$ ) using comparison of the fit of model (3) to simpler nested models (removing one effect  
453 at a time in model (3)). In order to statistically test whether there was a difference between the  
454 effects of paternal and maternal alleles in females in different degeneration categories we used  
455 the contrasts provided by the lmerTest package in R (30). This strategy provided estimates,  
456 confidence intervals and p-values of the difference between the two effects of paternal and  
457 maternal origin in females in interaction with degeneration levels, while normalising by the  
458 expression of the two outgroups. Moreover, the presence of random effects allows to account for  
459 inter-individual and inter-contig variability. Finally, p-values were corrected for multiple testing  
460 using a Benjamini and Hochberg correction. These values were used to plot Figure 2 and  
461 Supplementary Figures S7 to S12.

## 462 **8) GO term analysis**

463 All contigs of the reference transcriptome were blasted against the Uniprot database using blastx  
464 with an expectation value cutoff of 0.001, and an upper limit of hit results per sequence of 20.  
465 Significant blastx hits were obtained for 35,763 contigs. Gene Ontology terms were associated  
466 with the contigs using the Blast2GO PRO version V 2.7.2 (31) and the Data Base version  
467 b2g\_sep14, with 41,136 GO terms and 4,097 Enzymes available. Blast hits were related to the  
468 functional GO annotation by using the “Mapping” function of Blast2GO. The complete  
469 annotation of the sequences was performed using Annotation, ANNEX (increasing annotation

470 step), GO-SLIM (using the goslim\_plant.obo option to reduce the GO vocabulary) and InterPro  
471 (by choosing the whole set of InterPro annotation applications available: BlastProDom,  
472 FPrintScan, HMMPiR, HMMPfam, HMMSmart, HMMTigr, HMMPanther, ProfileScan,  
473 HAMAP, PatternScan, SuperFamily, Gene3D, Phobius, Coils, SignalIPHMM and TMHMM).  
474 Annotation using the default parameters (E-Value-Hit-Filtrer 1.0E-6; Annotation CutOff 55; GO  
475 Weight 5; Hsp-Hit Coverage CutOff 0) yielded a total of 27,152 annotated contigs (after merging  
476 the InterPro results with the GO annotation). Tests for enrichment were performed on all GO  
477 terms associated with the sequences, using a two-tailed Fisher's exact test with correction for  
478 multiple tests by using a false discovery rate of 0.05. Double IDs (both on the test and on the  
479 reference set) were removed.

480

481 **Supplementary material references**

1. R. Bergero, S. Qiu, D. Charlesworth, Gene Loss from a Plant Sex Chromosome System. *Curr. Biol. CB* (2015), doi:10.1016/j.cub.2015.03.015.
2. A. S. T. Papadopulos, M. Chester, K. Ridout, D. A. Filatov, Rapid Y degeneration and dosage compensation in plant sex chromosomes. *Proc. Natl. Acad. Sci. U. S. A.* **112**, 13021–13026 (2015).
3. M. V. Chibalina, D. A. Filatov, Plant Y chromosome degeneration is retarded by haploid purifying selection. *Curr. Biol. CB.* **21**, 1475–1479 (2011).
4. M. Toups, P. Veltsos, J. R. Pannell, Plant sex chromosomes: lost genes with little compensation. *Curr. Biol. CB.* **25**, R427-430 (2015).
5. A. Muyle, R. Shearn, G. A. Marais, The Evolution of Sex Chromosomes and Dosage Compensation in Plants. *Genome Biol. Evol.* **9**, 627–645 (2017).
6. E. Pessia, T. Makino, M. Bailly-Bechet, A. McLysaght, G. A. B. Marais, Mammalian X chromosome inactivation evolved as a dosage-compensation mechanism for dosage-sensitive genes on the X chromosome. *Proc. Natl. Acad. Sci. U. S. A.* **109**, 5346–5351 (2012).

7. N. Zemp, A. Minder, A. Widmer, Identification of internal reference genes for gene expression normalization between the two sexes in dioecious white Campion. *PLoS One*. **9**, e92893 (2014).
8. R. Hobza, A. Widmer, Efficient molecular sexing in dioecious *Silene latifolia* and *S. dioica* and paternity analysis in F(1) hybrids. *Mol. Ecol. Resour.* **8**, 1274–1276 (2008).
9. B. J. Haas *et al.*, De novo transcript sequence reconstruction from RNA-seq using the Trinity platform for reference generation and analysis. *Nat. Protoc.* **8**, 1494–1512 (2013).
10. A. Muyle *et al.*, Rapid de novo evolution of X chromosome dosage compensation in *Silene latifolia*, a plant with young sex chromosomes. *PLoS Biol.* **10**, e1001308 (2012).
11. X. Huang, A. Madan, CAP3: A DNA Sequence Assembly Program. *Genome Res.* **9**, 868–877 (1999).
12. H. Li, R. Durbin, Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinforma. Oxf. Engl.* **25**, 1754–1760 (2009).
13. H. Li *et al.*, The Sequence Alignment/Map format and SAMtools. *Bioinforma. Oxf. Engl.* **25**, 2078–2079 (2009).
14. M. A. DePristo *et al.*, A framework for variation discovery and genotyping using next-generation DNA sequencing data. *Nat. Genet.* **43**, 491–498 (2011).
15. G. Tsagkogeorga, V. Cahais, N. Galtier, The population genomics of a fast evolver: high levels of diversity, functional constraint, and molecular adaptation in the tunicate *Ciona intestinalis*. *Genome Biol. Evol.* **4**, 740–749 (2012).
16. P. Gayral *et al.*, Reference-free population genomics from next-generation transcriptome data and the vertebrate-invertebrate gap. *PLoS Genet.* **9**, e1003457 (2013).
17. A. Muyle *et al.*, SEX-DETECTOR: A Probabilistic Approach to Study Sex Chromosomes in Non-Model Organisms. *Genome Biol. Evol.* **8**, 2530–2543 (2016).

18. T. D. Wu, S. Nacu, Fast and SNP-tolerant detection of complex variants and splicing in short reads. *Bioinforma. Oxf. Engl.* **26**, 873–881 (2010).
19. S. F. Altschul, W. Gish, W. Miller, E. W. Myers, D. J. Lipman, Basic local alignment search tool. *J. Mol. Biol.* **215**, 403–410 (1990).
20. N. Zemp *et al.*, Evolution of sex-biased gene expression in a dioecious plant. *Nat. Plants.* **2**, 16168 (2016).
21. J. W. VAN Ooijen, Multipoint maximum likelihood mapping in a full-sib family of an outbreeding species. *Genet. Res.* **93**, 343–349 (2011).
22. P. Rastas, L. Paulin, I. Hanski, R. Lehtonen, P. Auvinen, Lep-MAP: fast and accurate linkage map construction for large SNP datasets. *Bioinforma. Oxf. Engl.* **29**, 3128–3134 (2013).
23. J. Vrána, H. Simková, M. Kubaláková, J. Cíhalíková, J. Doležel, Flow cytometric chromosome sorting in plants: the next generation. *Methods San Diego Calif.* **57**, 331–337 (2012).
24. E. Aronesty, *ea-utils*: “Command-line tools for processing biological sequencing data” (2011; <http://code.google.com/p/ea-utils>).
25. R. Chikhi, P. Medvedev, Informed and automated k-mer size selection for genome assembly. *Bioinforma. Oxf. Engl.* **30**, 31–37 (2014).
26. R. Luo *et al.*, SOAPdenovo2: an empirically improved memory-efficient short-read de novo assembler. *GigaScience.* **1**, 18 (2012).
27. W. Xue *et al.*, L\_RNA\_scaffolder: scaffolding genomes with transcripts. *BMC Genomics.* **14**, 604 (2013).
28. A. Oshlack, M. D. Robinson, M. D. Young, From RNA-seq reads to differential expression results. *Genome Biol.* **11**, 220 (2010).

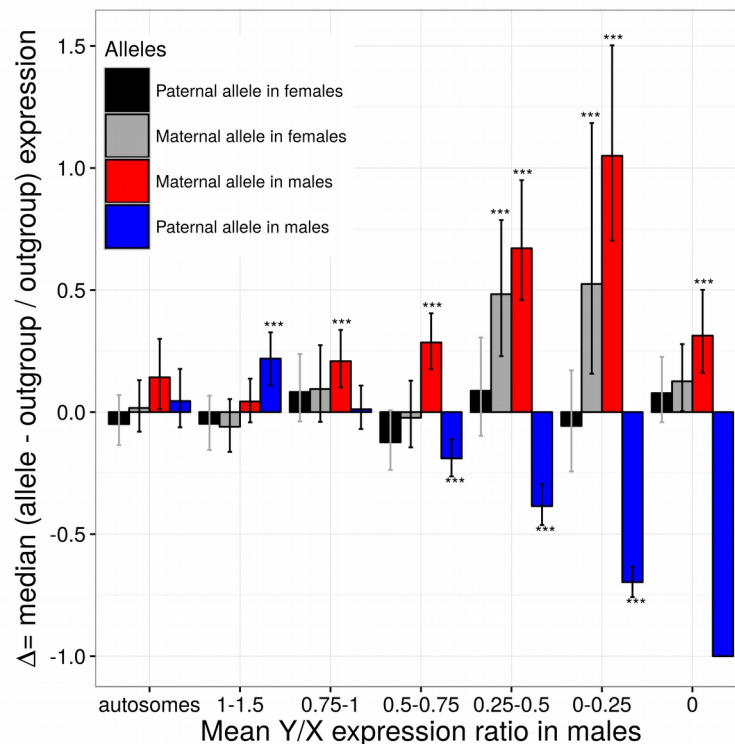
29. M. D. Robinson, D. J. McCarthy, G. K. Smyth, edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinforma. Oxf. Engl.* **26**, 139–140 (2010).
30. R Development Core Team, R: A language and environment for statistical computing, reference index version 2.15.0. *R Found. Stat. Comput. Vienna Austria* (2012).
31. A. Conesa *et al.*, Blast2GO: a universal tool for annotation, visualization and analysis in functional genomics research. *Bioinforma. Oxf. Engl.* **21**, 3674–3676 (2005).

482

483



## 484 Supplementary Figures

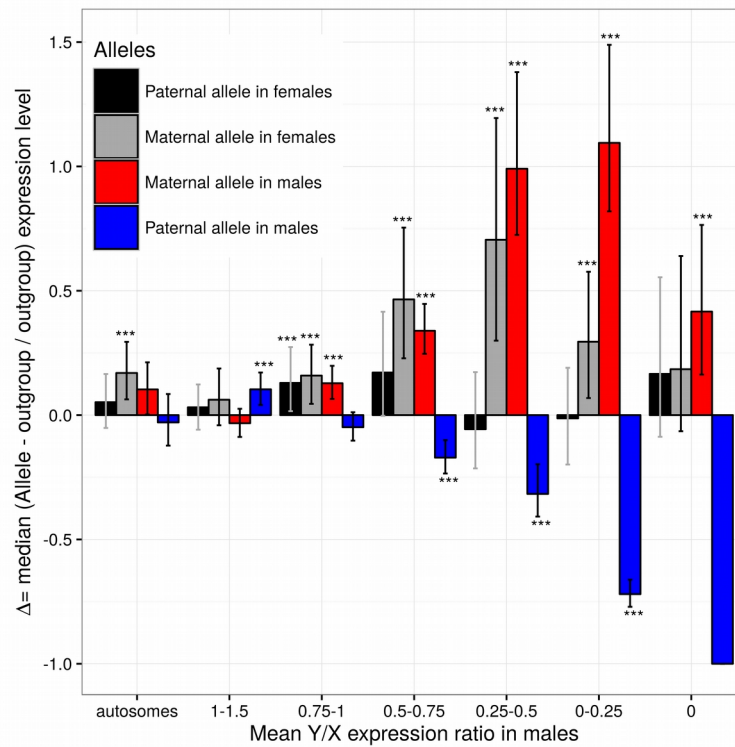


485

486 **Supplementary Figure S1:** Normalised difference in allelic expression levels between *S.*  
487 *latifolia* and the two outgroups without sex chromosomes *S. vulgaris* and *S. viscosa* (hereafter  
488  $\Delta$ ), in autosomal and sex-linked contigs for the **seedling** tissue. Maternal and paternal allelic read  
489 numbers were summed at SNP positions and normalised for each individual separately, then  
490 averaged among individuals for each contig.  $\Delta$  was computed as follows:  $\Delta = (\text{allelic expression}$   
491  $\text{in } S. \textit{latifolia} - \text{allelic expression in the outgroup}) / \text{allelic expression in the outgroup}$ . If  $\Delta$  is  
492 lower, higher or equal to zero, then expression in *S. latifolia* is respectively lower, higher or  
493 equal to the outgroup. For all contig categories,  $\Delta$  was compared to zero using a Wilcoxon test.  
494 The median  $\Delta$ , confidence intervals and p-values adjusted for multiple testing using a Benjamini  
495 and Hochberg correction are shown (\*\*\*: p-value < 0.001; \*\*: p-value < 0.01, \*: p-value < 0.05).  
496 The Y/X ratio was computed in *S. latifolia* males and averaged among individuals to use as a  
497 proxy for Y degeneration. Contigs with sex-biased expression were removed, as well as contigs  
498 with Y/X expression ratios above 1.5. Sample sizes for the different contig categories are:  
499 autosomal:200; 1-1.5:148; 0.75-1:139; 0.5-0.75:160; 0.25-0.5:114; 0-0.25:79; 0:205 (note that  
500 200 autosomal contigs were randomly selected in order to have similar statistical power among

501 gene categories). In the absence of dosage compensation, the single X in males should be  
502 expressed at levels similar to the outgroup that does not have sex chromosomes, in other words,  
503 without dosage compensation  $\Delta$  should be close to zero for the maternal allele in males (red  
504 bars). Results show that the maternal allele is hyper-expressed in *S. latifolia* when the Y  
505 chromosome is degenerated, both in males and females.

506



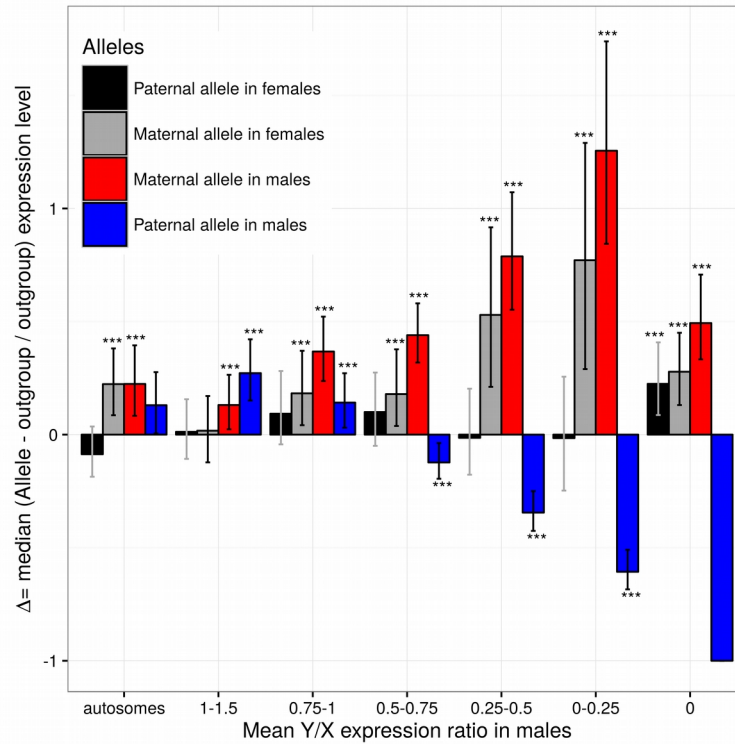
507

508 **Supplementary Figure S2:** Normalised difference in allelic expression levels between *S.*  
509 *latifolia* and the two outgroups without sex chromosomes *S. vulgaris* and *S. viscosa* ( $\Delta$ ), in  
510 autosomal and sex-linked contigs for the **flower bud** tissue. Same legend as Supplementary  
511 Figure 1 except for sample sizes for the different contig categories: autosomal:200; 1-1.5:95;  
512 0.75-1:195; 0.5-0.75:203; 0.25-0.5:176; 0-0.25:116; 0:103.

513

514

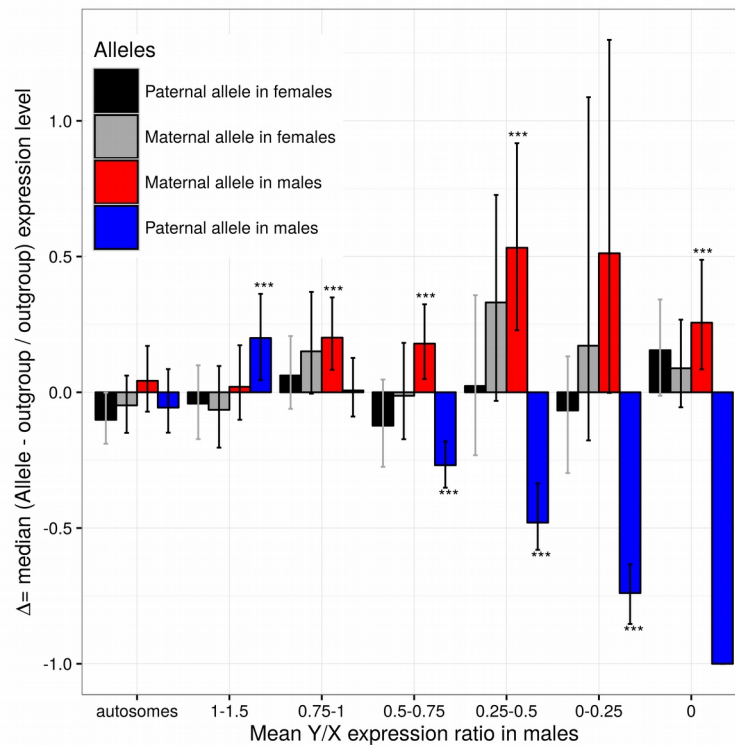
515



516 **Supplementary Figure S3:** Normalised difference in allelic expression levels between *S.*  
517 *latifolia* and the two outgroups without sex chromosomes *S. vulgaris* and *S. viscosa* ( $\Delta$ ), in  
518 autosomal and sex-linked contigs for the **leaf** tissue. Same legend as Supplementary Figure 1  
519 except for sample sizes for the different contig categories: autosomal:200; 1-1.5:159; 0.75-1:132;  
520 0.5-0.75:147; 0.25-0.5:126; 0-0.25:71; 0:275.

521

522



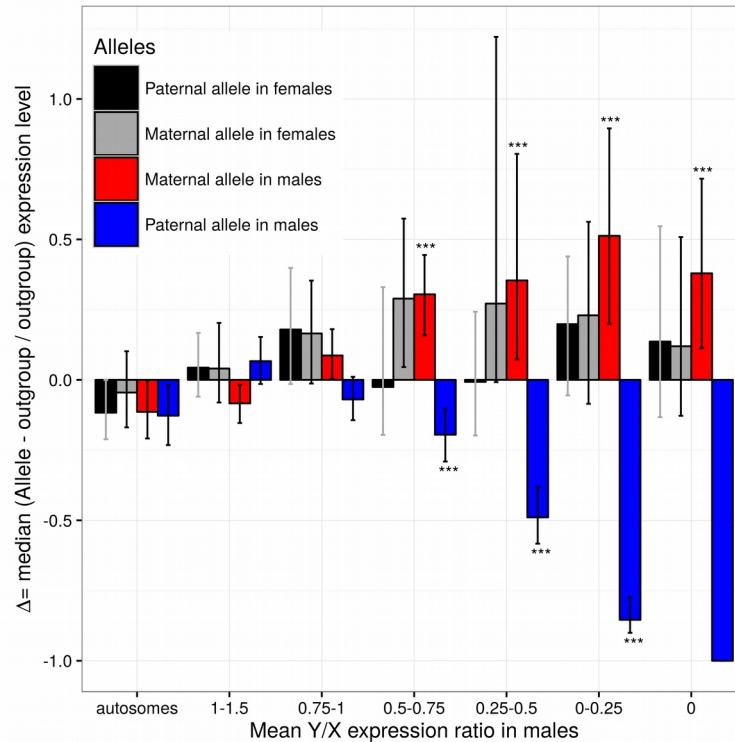
523

524 **Supplementary Figure S4:** Normalised difference in allelic expression levels between *S.*  
525 *latifolia* and the two outgroups without sex chromosomes *S. vulgaris* and *S. viscosa* ( $\Delta$ ), in  
526 autosomal and sex-linked contigs that were **validated** (see Materials and Methods), for the  
527 **seedling** tissue. Same legend as Supplementary Figure 1 except for sample sizes for the different  
528 contig categories: autosomal:77; 1-1.5:71; 0.75-1:82; 0.5-0.75:91; 0.25-0.5:44; 0-0.25:29; 0:89.

529

530

531

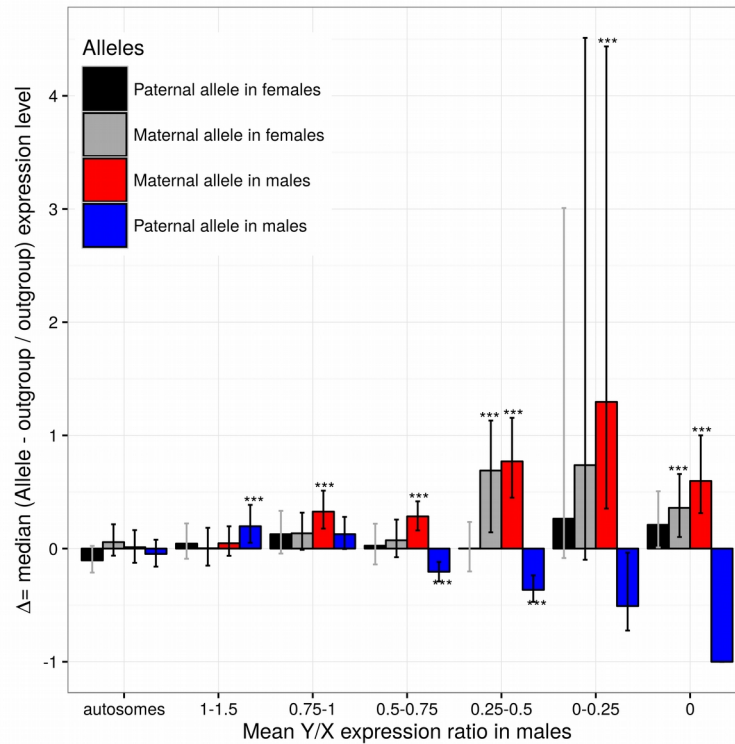


532

533 **Supplementary Figure S5:** Normalised difference in allelic expression levels between *S.*  
534 *latifolia* and the two outgroups without sex chromosomes *S. vulgaris* and *S. viscosa* ( $\Delta$ ), in  
535 autosomal and sex-linked contigs that were **validated** (see Materials and Methods), for the  
536 **flower bud** tissue. Same legend as Supplementary Figure 1 except for sample sizes for the  
537 different contig categories: autosomal:74; 1-1.5:86; 0.75-1:91; 0.5-0.75:67; 0.25-0.5:45; 0-  
538 0.25:31; 0:55.

539

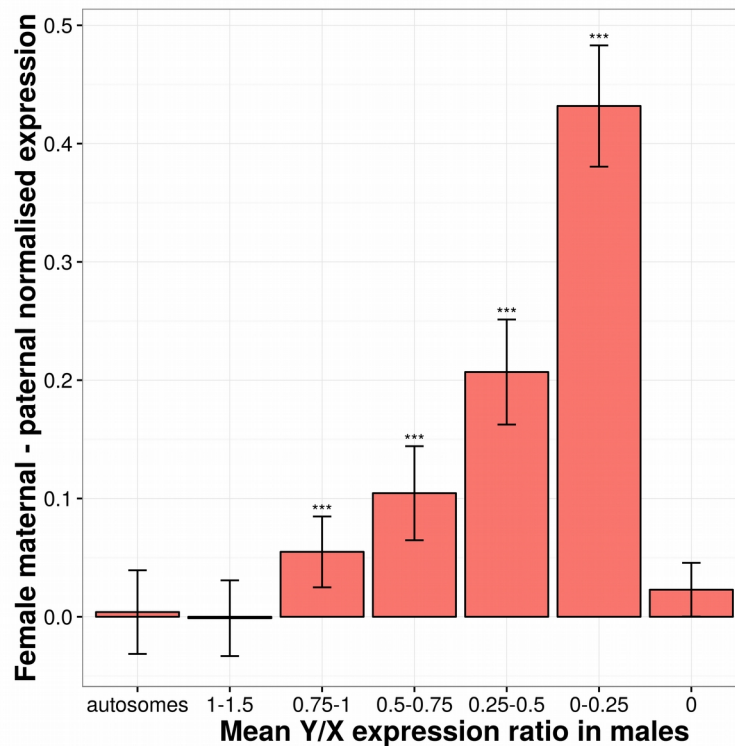
540



541

542 **Supplementary Figure S6:** Normalised difference in allelic expression levels between *S.*  
543 *latifolia* and the two outgroups without sex chromosomes *S. vulgaris* and *S. viscosa* ( $\Delta$ ), in  
544 autosomal and sex-linked contigs that were **validated** (see Materials and Methods), for the **leaf**  
545 tissue. Same legend as Supplementary Figure 1 except for sample sizes for the different contig  
546 categories: autosomal:79; 1-1.5:84; 0.75-1:74; 0.5-0.75:77; 0.25-0.5:52; 0-0.25:19; 0:119.

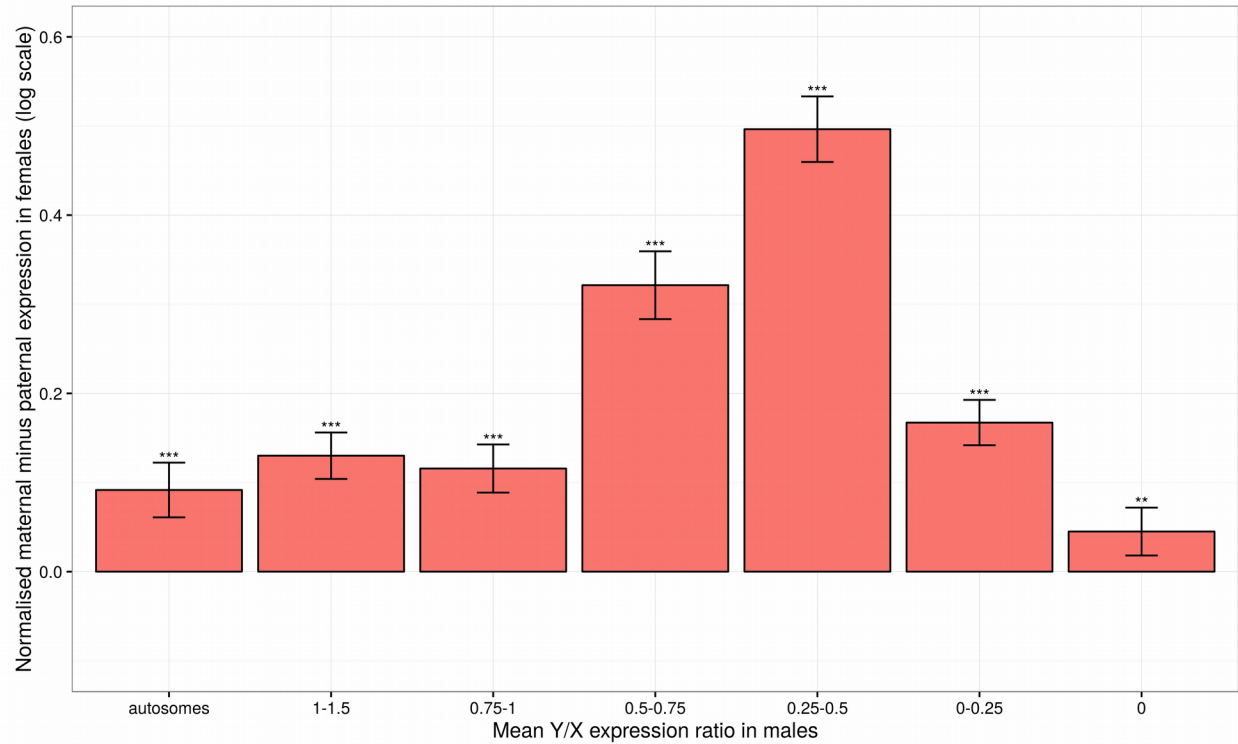
547



548

549 **Supplementary Figure S7:** Normalised expression difference between the maternal and paternal  
550 allele in *S. latifolia* females in autosomal and sex-linked contigs for the **seedling** tissue. The Y  
551 axis unit is the normalised allelic read count difference in log scale. A linear regression model  
552 with mixed effects was used to study allelic expression in *S. latifolia* for every SNP position. In  
553 order to measure the changes in *S. latifolia* expression due to sex chromosomes evolution, the  
554 outgroup *S. vulgaris* that does not have sex chromosomes was used as a reference in the model  
555 (see Materials and Methods for details). The framework provided estimates for the normalised  
556 difference between the effect of paternal and maternal origin of alleles in interaction with the  
557 contig status (autosomal or sex-linked with various levels of Y degeneration), while accounting  
558 for inter-contig and inter-individual variability. See Supplementary Figure S1 legend for sample  
559 sizes for the different contig categories and statistical significance symbols. Results show that Y  
560 degeneration is linked to a significant expression difference between the paternal and maternal  
561 alleles in females, which is not observed in autosomal and non-degenerated sex-linked contigs.

562



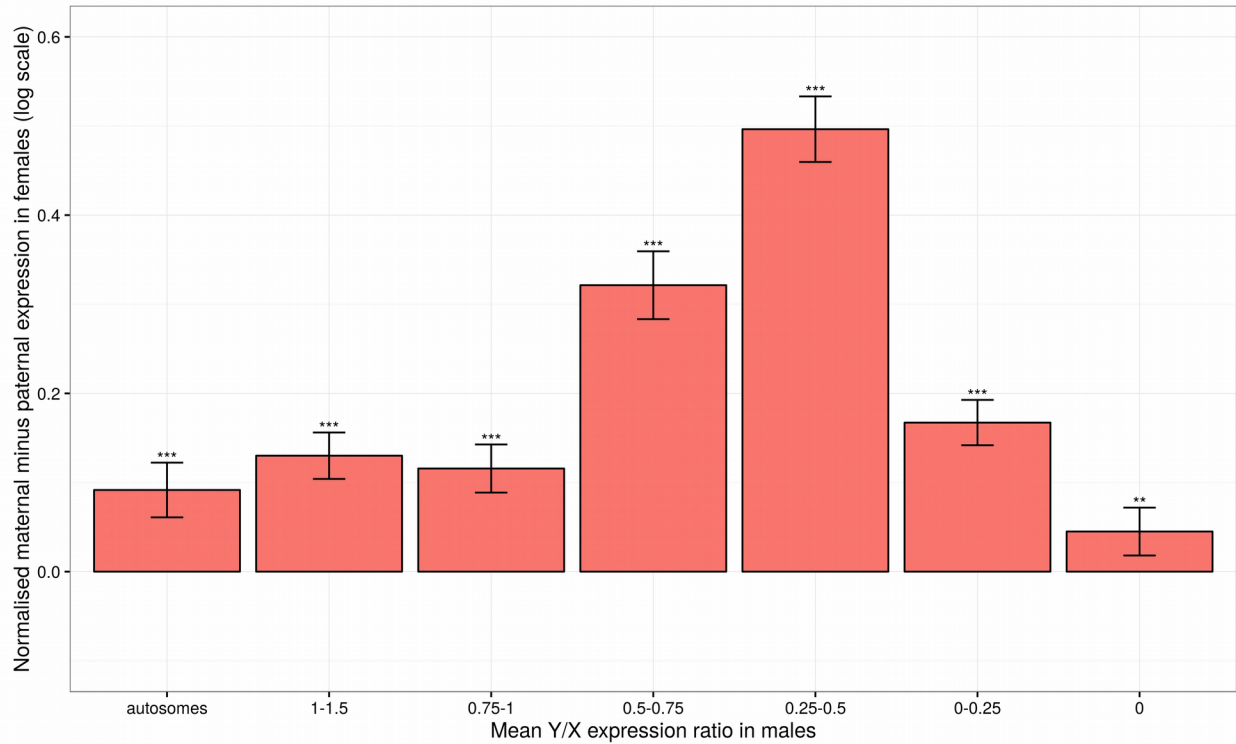
563

564 **Supplementary Figure S8:** Normalised expression difference between the maternal and paternal  
565 allele in *S. latifolia* females in autosomal and sex-linked contigs for the **flower bud** tissue. See  
566 supplementary Figure 7 for legend and Supplementary Figure S2 for sample sizes for the  
567 different contig categories.

568

569

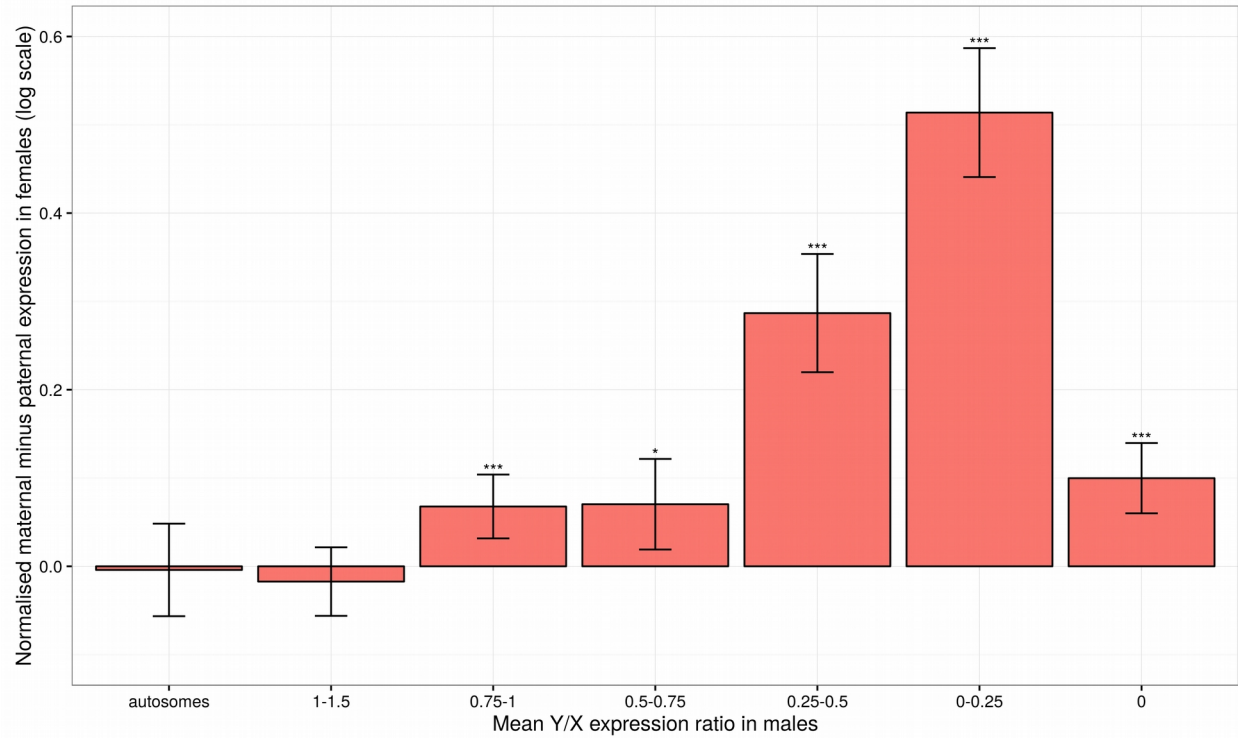




570

571 **Supplementary Figure S9:** Normalised expression difference between the maternal and paternal  
572 allele in *S. latifolia* females in autosomal and sex-linked contigs for the **leaf** tissue. See  
573 supplementary Figure 7 for legend and Supplementary Figure S3 for sample sizes for the  
574 different contig categories.

575



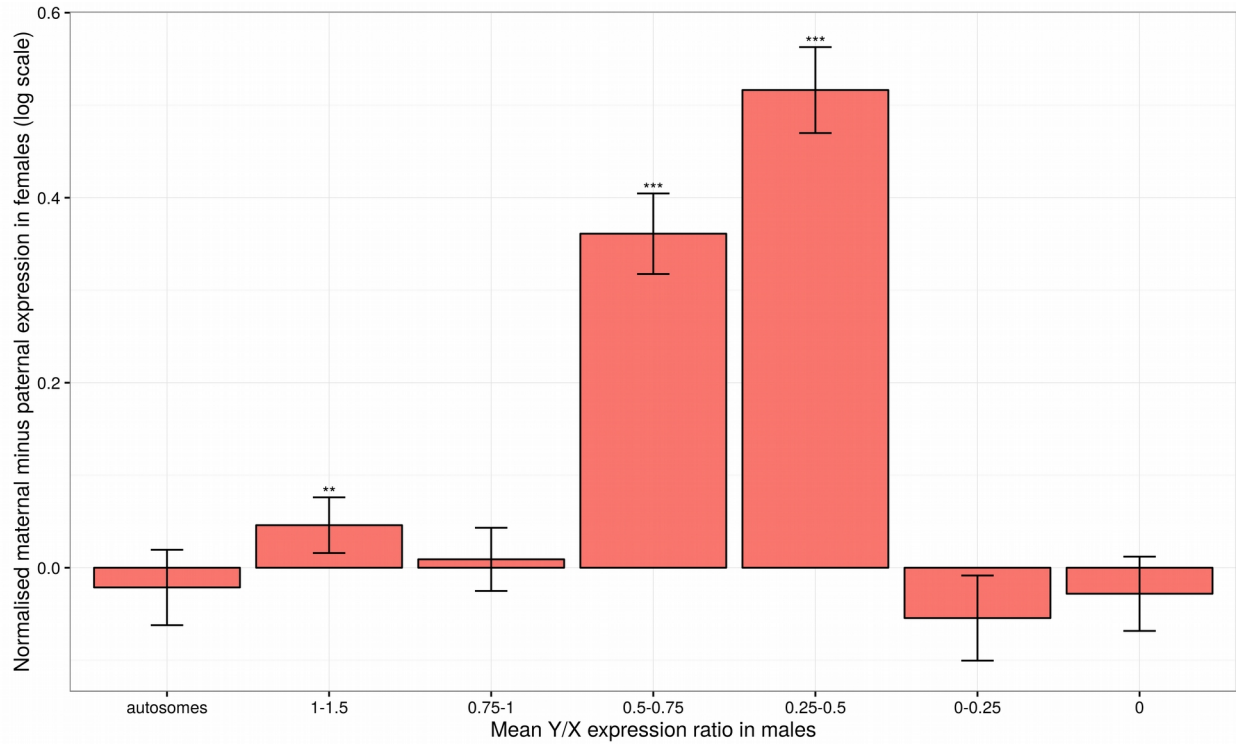
576

577 **Supplementary Figure S10:** Normalised expression difference between the maternal and  
578 paternal allele in *S. latifolia* females in autosomal and sex-linked **validated** contigs for the  
579 **seedling** tissue. See supplementary Figure 7 for legend and Supplementary Figure S4 for sample  
580 sizes for the different contig categories.

581

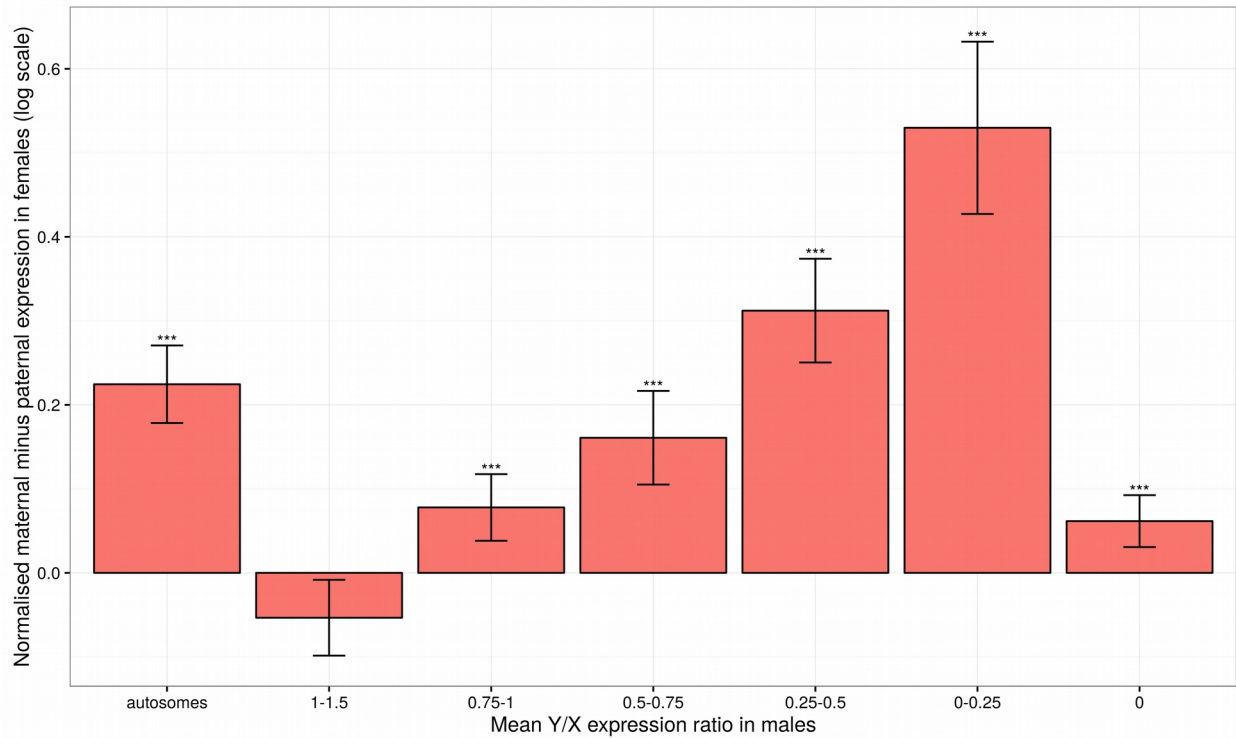
582

583



584 **Supplementary Figure S11:** Normalised expression difference between the maternal and  
585 paternal allele in *S. latifolia* females in autosomal and sex-linked **validated** contigs for the  
586 **flower bud** tissue. See supplementary Figure 7 for legend and Supplementary Figure S5 for  
587 sample sizes for the different contig categories.

588



589

590 **Supplementary Figure S12:** Normalised expression difference between the maternal and  
591 paternal allele in *S. latifolia* females in autosomal and sex-linked **validated** contigs for the **leaf**  
592 tissue. See supplementary Figure 7 for legend and Supplementary Figure S6 for sample sizes for  
593 the different contig categories.

594

595

## 596 **Supplementary Tables**

597 **Supplementary Table S1:** library sizes (number of reads) of each individual and mapping  
598 statistics.

599 **Supplementary Table S2:** Number of contigs after SEX-DETECTOR inferences, removal of sex-  
600 bias and selection of validated contigs in the three tissues.

	Tissue type		
	flower buds	leaves	seedlings
<b>number of ORFs</b>	46178		
<b>Unassigned</b>	33172	33564	33781
<b>Autosomal</b>	11662	11558	11292
<b>X/Y</b>	1140	772	844
<b>X-hemizygous</b>	204	284	261
<b>X/Y non sex-biased</b>	901	733	732
<b>X-hemizygous non sex-biased</b>	103	275	205
<b>X/Y non sex-biased validated</b>	339	345	365
<b>X-hemizygous non sex-biased validated</b>	55	119	89
<b>Autosomal validated</b>	74	79	77

601

602 **Supplementary Table S3:** list of known sex-linked genes in *S. latifolia* and associated literature  
603 references.

604

605

## 606 **Author contributions**

607 Aline Muyle, Niklaus Zemp, Alex Widmer and Gabriel Marais conceived the study and  
608 experimental design. Niklaus Zemp and Alex Widmer prepared and sequenced the RNA-seq  
609 datasets and assembled the reference transcriptome. Aline Muyle ran SEX-DETECTOR on the  
610 RNA-seq datasets for the three tissues (including mapping and genotyping steps), produced  
611 allelic expression levels, analysed the data, prepared Tables and Figures and wrote the  
612 Supplementary Material with inputs from other authors. Niklaus Zemp generated the X  
613 chromosome genetic map (with help from Aline Muyle for the mapping and genotyping part).  
614 Radim Cegan, Jan Vrana and Roman Hobza did the Y chromosome flow cytometry sorting and  
615 sequencing. Clothilde Deschamps did the first assembly of the sorted Y chromosome and

616 improved it with RNA-seq data with the help of Cecile Fruchard. Aline Muyle did the blasts to  
617 validate the inferences of SEX-DETECTOR. Raquel Tavares did the GO term analysis. Aline Muyle  
618 and Frank Picard did the statistical analyses of the data. Gabriel Marais and Aline Muyle wrote  
619 the main text of the manuscript with inputs from other authors.  
620