

 Open access • Posted Content • DOI:10.1101/179044

Maternal X chromosome upregulation in both sexes initiates dosage compensation evolution — [Source link](#)

[Aline Muyle](#), [Nicklaus Zemp](#), [Cécile Fruchard](#), [Radim Cegan](#) ...+7 more authors

Institutions: [Academy of Sciences of the Czech Republic](#)

Published on: 21 Aug 2017 - [bioRxiv](#) (Cold Spring Harbor Laboratory)

Topics: [Dosage compensation](#), [X chromosome](#) and [Silene latifolia](#)

Related papers:

- [Sex chromosome equality in plants.](#)
- [Dosage compensation of the sex chromosomes and autosomes.](#)
- [Sex-Specific Embryonic Gene Expression in Species with Newly Evolved Sex Chromosomes](#)
- [Insights into avian incomplete dosage compensation : Sex-biased gene expression coevolves with sex chromosome degeneration in the common whitethroat](#)
- [Gene dosage compensation and the evolution of sex chromosomes.](#)

Share this paper:    

View more about this paper here: <https://typeset.io/papers/maternal-x-chromosome-upregulation-in-both-sexes-initiates-35s247nujt>



HAL
open science

Genomic imprinting mediates dosage compensation in a young plant XY system.

Aline Muyle, Niklaus Zemp, Cécile Fruchard, Radim Cegan, Jan Vrána, Clothilde Deschamps, Raquel Tavares, Roman Hobza, Franck Picard, Alex Widmer, et al.

► To cite this version:

Aline Muyle, Niklaus Zemp, Cécile Fruchard, Radim Cegan, Jan Vrána, et al.. Genomic imprinting mediates dosage compensation in a young plant XY system.: An article peer-reviewed and recommended by Peer Community In Evolutionary Biology (PCI Evol Biol). 2018. hal-01851376

HAL Id: hal-01851376

<https://hal.archives-ouvertes.fr/hal-01851376>

Preprint submitted on 30 Jul 2018

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

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



Peer Community In Evolutionary Biology

Muyle A, Zemp N, Fruchard C, Cegan R, Vrana J, Deschamps C, Tavares R, Picard F, Hobza R, Widmer A and Marais G. 2018. **Genomic imprinting mediates dosage compensation in a young plant XY system.** bioRxiv. *bioRxiv*, 179044, <https://doi.org/10.1101/179044>

An article reviewed and recommended by *Peer Community In Evolutionary Biology*:

<http://dx.doi.org/10.24072/pci.evolbiol.100044>

15 This preprint has been reviewed and recommended by Peer Community In Evolutionary
16 Biology (<http://dx.doi.org/10.24072/pci.evolbiol.100044>).

17 Sex chromosomes have repeatedly evolved from a pair of autosomes¹. Consequently, X
18 and Y chromosomes initially have similar gene content, but ongoing Y degeneration leads to
19 reduced Y gene expression and eventual Y gene loss. The resulting imbalance in gene expression
20 between Y genes and the rest of the genome is expected to reduce male fitness, especially when
21 protein networks have components from both autosomes and sex chromosomes. A diverse set of
22 dosage compensating mechanisms that alleviates these negative effects has been described in
23 animals²⁻⁴. However, the early steps in the evolution of dosage compensation remain unknown
24 and dosage compensation is poorly understood in plants⁵. Here we show a novel dosage
25 compensation mechanism in the evolutionarily young XY sex determination system of the plant
26 *Silene latifolia*. Genomic imprinting results in higher expression from the maternal X
27 chromosome in both males and females. This compensates for reduced Y expression in males but
28 results in X overexpression in females and may be detrimental. It could represent a transient
29 early stage in the evolution of dosage compensation. Our finding has striking resemblance to the
30 first stage proposed by Ohno for the evolution of X inactivation in mammals.

31 In *Drosophila*, the X chromosome is upregulated specifically in males, resulting in
32 complete dosage compensation through both ancestral expression recovery in males and equal
33 expression between the sexes (hereafter sex equality)⁶. In *Caenorhabditis elegans*, both X
34 chromosomes are downregulated in XX hermaphrodites resulting in sex equality, but only a few
35 genes have their X expression doubled for ancestral expression recovery⁷. In placental mammals,
36 including humans, one X chromosome is randomly inactivated in XX females, resulting in sex
37 equality but without recovering the ancestral expression of sex chromosomes, except for a few

38 dosage-sensitive genes whose X expression was doubled in both sexes⁸⁻¹². In the marsupials, the
39 paternal X chromosome is consistently inactivated in XX females¹³. Differential expression that
40 depends on the parent of origin is known as genomic imprinting¹⁴, and this mechanism also
41 operates in the mouse placenta¹⁵.

42 Despite the plethora of studies on gene expression on sex chromosomes, it is not yet clear
43 if genomic imprinting is commonly involved in the early steps of dosage compensation
44 evolution. In a seminal work, Ohno hypothesized a two-step process for the evolution of dosage
45 compensation¹⁶. In the first step, expression from the X is doubled, thereby mediating the
46 recovery of ancestral expression in XY males. Second, the resulting overexpression in XX
47 females selects for X inactivation. This scenario is consistent with the fact that sexual selection is
48 often stronger on males than on females. Under this scenario, selection on XY males to
49 upregulate their single X chromosome should be stronger than selection on females, leading to
50 overexpression in females until a second correcting mechanism evolves³. However, in order to
51 understand these early steps of dosage compensation evolution, species with young sex
52 chromosomes must be studied.

53 The plant *Silene latifolia* is an ideal model to study early steps of sex chromosome
54 evolution thanks to its pair of X/Y chromosomes that evolved ~4 Mya¹⁷. Dosage compensation is
55 poorly understood in plants⁵. Thus far only sex equality has been studied. Equal expression
56 levels were observed for males and females for some genes despite Y expression degeneration¹⁸⁻
57 ²³. However, the mechanisms through which sex equality is achieved – and whether ancestral
58 expression is recovered in *S. latifolia* males – remain unknown. To address these questions, we
59 have developed an approach relying on (i) the use of an outgroup without sex chromosomes as
60 an ancestral autosomal reference⁵ in order to determine whether X chromosome expression

61 increased or decreased in *S. latifolia*, (ii) the application of methods to study allele-specific
62 expression while correcting for reference mapping bias⁵, and (iii) a statistical framework to
63 quantify dosage compensation⁵.

64 Because only ~25% of the large and highly repetitive *S. latifolia* genome has been
65 assembled so far²³, we used an RNA-seq approach based on the sequencing of a cross (parents
66 and a few offspring of each sex), to infer sex-linked contigs (i.e. contigs located on the non-
67 recombining region of the sex chromosome pair)²⁴. X/Y contigs show both X and Y expression,
68 while X-hemizygous contigs are X-linked contigs without Y allele expression. We made
69 inferences separately for three tissues: flower buds, seedlings and leaves (Supplementary Table
70 S2). Results are consistent across tissues and flower buds and leaves are shown in
71 Supplementary Materials. In seedlings, ~1100 sex-linked contigs were inferred. Among these,
72 15% of contigs with significant expression differences between males and females were removed
73 for further analyses (Supplementary Table S2 and Materials and Methods). These are likely
74 involved in sex-specific functions and are not expected to be dosage compensated²⁵. This was
75 done as a usual procedure for studying dosage compensation, however the resulting trends and
76 significance levels are not affected. About half of the non sex-biased sex-linked contigs could be
77 validated by independent data using three sources: literature, a genetic map and sequence data
78 from Y flow-sorted chromosomes (see Supplementary Table S2 and Materials and Methods). X-
79 hemizygous contigs are more difficult to identify than X/Y contigs using an RNA-seq approach
80 (see Supplementary Text S1). This explains conflicting earlier results on dosage compensation in
81 *S. latifolia*⁵. A study using genomic data (i.e. not affected by the aforementioned ascertainment
82 bias) found sex-equality in approximately half of the studied X-hemizygous genes²³. In our set of
83 X-hemizygous contigs, no evidence for dosage compensation was found (Supplementary Text

84 S1), in agreement with previous work relying on an RNA-seq approach^{18,22}. This could be due to
85 an over-representation of dosage insensitive genes in our set of X-hemizygous contigs
86 (Supplementary Text S1).

87 We estimated paternal and maternal allele expression levels in males and females for sex-
88 linked and autosomal contigs in *S. latifolia* after correcting for reference mapping bias (Materials
89 and Methods). We then compared these allelic expression levels to one or two closely related
90 outgroups without sex chromosomes in order to polarise expression changes in *S. latifolia*. For
91 autosomal contigs, expression levels did not differ between *S. latifolia* and the outgroups (Figure
92 1). This is due to the close relatedness of the outgroups (~5My, Supplementary Figure S1), and
93 validates their use as a reference for ancestral expression levels. We used the ratio of Y over X
94 expression levels in *S. latifolia* males as a proxy for Y degeneration and then grouped contigs on
95 this basis. As expression of the Y allele decreased (paternal allele in blue in Figure 1), expression
96 of the corresponding X allele in males increased (maternal allele in red in Figure 1). This is the
97 first direct evidence for ancestral expression recovery in *S. latifolia*, i.e. ancestral expression
98 levels are reestablished in males despite Y expression degeneration. In females, expression of the
99 maternal X allele also increased with Y degeneration (gray bars in Figure 1), similarly to the
100 maternal X allele in males. The paternal X alleles in females, however, maintained ancestral
101 expression levels, regardless of Y degeneration (black bars in Figure 1). Consequently, sex
102 equality is not achieved in *S. latifolia* due to upregulation of sex-linked genes in females
103 compared to ancestral expression levels. These results were confirmed in two other tissues and
104 when analysing independently validated contigs only (although statistical power is sometimes
105 lacking due to the limited number of validated contigs, Supplementary Figures S2-S7).

106 Upregulation of the maternal X allele both in males and females of *S. latifolia* (Figure 1
107 and Supplementary Figures S2-S7) establishes a role for genomic imprinting in dosage
108 compensation. In order to statistically test this inference at the SNP level, we used a linear
109 regression model with mixed effects (Materials and Methods). Outgroup species were used as a
110 reference and expression levels in *S. latifolia* were then analyzed while accounting for the
111 variability due to contigs and individuals. The joint effect of the parental origin and the
112 degeneration level was estimated, which allowed computing expression differences between
113 maternal and paternal alleles in females for different Y/X degeneration categories (Figure 2).
114 Maternal and paternal alleles of autosomal SNPs were similarly expressed in females, indicating
115 a global absence of genomic imprinting for these SNPs. However, for X/Y SNPs, the difference
116 between the maternal and paternal X in females increased with Y degeneration. These results
117 were confirmed in two other tissues and when analysing independently validated contigs only
118 (although statistical power is sometimes lacking due to the limited number of validated contigs,
119 Supplementary Figures S8-S13).

120 Previous studies that showed sex equality in *S. latifolia* could have been explained by
121 simple buffering mechanisms, where one copy of a gene is expressed at a higher level when
122 haploid than when diploid, due to higher availability of the cell machinery or adjustments in gene
123 expression networks^{23,26,27}. However, the upregulation of the X chromosome we reveal here in *S.*
124 *latifolia* males cannot be explained by buffering mechanisms alone, as the maternal X in females
125 would otherwise not be upregulated. Instead, our findings indicate that a specific dosage
126 compensation mechanism relying on genomic imprinting has evolved in *S. latifolia*. This
127 apparent convergent evolution with marsupials is mediated by different mechanisms (in
128 marsupials the paternal X is inactivated¹³, while in *S. latifolia* the maternal X is upregulated).

129 An exciting challenge ahead will be to understand how upregulation of the maternal X is
130 achieved in *S. latifolia* males and females at the molecular level. Chromosome staining suggests
131 that DNA methylation is involved. Indeed, one arm of one of the two X chromosomes in females
132 was hypomethylated, as well as the same arm of the single X in males²⁸ (Figure 3 and
133 Supplementary Figure S14). Based on our results, we hypothesize that the hypomethylated X
134 chromosome corresponds to the maternal, upregulated X. Unfortunately, parental origin of the X
135 chromosomes was not established in this study²⁸. It would be of interest in the future to study
136 DNA methylation patterns in *S. latifolia* paternal and maternal X chromosomes, along with the
137 homologous pair of autosomes in a closely related species without sex chromosomes. The
138 methylation pattern observed by chromosome staining suggests that dosage compensation in *S.*
139 *latifolia* could be a chromosome arm-wide phenomenon. To test this hypothesis with expression
140 data, positions of genes along the X chromosome remain to be elucidated.

141 Our study is the first to establish female upregulation of the X chromosome compared to
142 autosomes, as predicted by Ohno. An earlier report in *Tribolium castaneum* was later shown to
143 be due to biases from inclusion of gonads in whole body extracts⁴. X overexpression in females
144 may be deleterious. Its occurrence suggests that reduced expression of sex-linked genes in males
145 is more deleterious than overexpression in females. This potentially suboptimal situation may be
146 transitory and a consequence of the young age of *S. latifolia* sex chromosomes. Sex equality may
147 evolve at a later stage, following the evolutionary path trajectory originally proposed by Ohno
148 for placental mammals¹⁶.

149 **Methods**

150 **Sequence data and inference of sex-linkage.** RNA-seq data was generated in *S. latifolia* for a
151 cross (parents and progeny) for three tissues (seedlings, leaves and flower buds) and analysed
152 using the SEX-DETECTOR pipeline²⁴. RNA-seq data was also generated for two outgroup species
153 (*S. viscosa* and *S. vulgaris*). Reference mapping bias was corrected using the program GSNAP²⁹.
154 Inferences of sex-linked contigs were validated using three sources of information (literature, a
155 genetic map and flow-sorted Y chromosome sequences). See Supplementary Text S2 for details.
156 **Allelic expression levels.** Contigwise autosomal, X, Y, X+X and X+Y normalised allelic
157 expression levels were computed by summing read numbers for each X-linked or Y-linked alleles
158 for filtered SNPs of the contigs (Supplementary Text S2) for each individual separately and then
159 normalised using the library size and the number of studied sex-linked SNPs in the contig:

$$160 \quad E = r / (n * l) \quad (1)$$

161 With E = normalised expression level for a given individual, r = sum of total read counts, n =
162 number of studied SNPs, l = library size of the individual (number of mapped reads). Allelic
163 expression levels were then averaged among individuals for each contig. In order to make *S.*
164 *latifolia* expression levels comparable to *S. viscosa* and *S. vulgaris*, *S. viscosa* and *S. vulgaris*
165 expression levels were estimated using only the filtered SNP positions used in *S. latifolia*.
166 Normalised expression levels computed as explained in equation (1) in the two outgroups were
167 then averaged together for leaves and flower buds as expression levels are highly correlated (R^2
168 0.7 and 0.5 for flower buds and leaves respectively and p -value $< 2.10^{-6}$ in both cases). Averaging
169 expression levels between the two outgroups allows to get closer to the ancestral autosomal
170 expression level.

171 **Sex-biased expression.** Sex-biased contigs were inferred as in Zemp et al³⁰. See Supplementary
172 Text S2 for more detail.

173 **Expression divergence between *S. latifolia* and the two outgroups at the contig level.** The
174 normalised difference in allelic expression between *S. latifolia* and the two outgroups (hereafter
175 Δ) was computed in order to study how sex chromosome expression levels evolved in *S. latifolia*
176 compared to autosomal expression levels in the two outgroups: Δ is equal to zero if *S. latifolia* and
177 the outgroups have equal expression levels, Δ is positive if *S. latifolia* has higher expression
178 levels compared to the outgroups and Δ is negative otherwise:

$$179 \quad \Delta = (S. \textit{latifolia} \text{ expression level} - \text{outgroup expression level}) / (\text{outgroup expression} \\ 180 \text{ level}) \quad (2)$$

181 Sex-linked contigs were grouped by categories of degeneration level using the average Y
182 over X expression ratio in males. 200 autosomal contigs were randomly selected in order to have
183 similar statistical power among gene categories. Δ values for each allele (maternal and paternal
184 in males and females) and each gene category were compared to zero using a Wilcoxon test. P-
185 values were corrected for multiple testing using a Benjamini and Hochberg correction. The
186 estimated median Δ , confidence intervals and adjusted p-values were then used to plot Figure 1
187 and Supplementary Figures S2 to S7.

188 **Expression differences between maternal and paternal alleles at the SNP level.** Maternal and
189 paternal alleles expression were compared in *S. latifolia* for autosomal and sex-linked SNPs. In
190 order to deal with the difference in numbers of autosomal versus sex-linked contigs
191 (Supplementary Table S2), 200 autosomal contigs were randomly selected in order to keep
192 comparable powers of detection. Allelic expression levels in *S. latifolia* for each individual at
193 every SNP position were analysed using a linear regression model with mixed effects with the R
194 package lme4. We assumed a normal distribution of the read count data after log transformation.
195 In order to account for inter-individual and inter-contig variability, a random “individual” and a

196 random “contig” effect were included in the model. The aim of this modeling framework was to
197 estimate the joint effect of the chromosomal origin of alleles (paternal or maternal in males or
198 females) and the status of the gene (autosomal or sex-linked with various levels of Y
199 degeneration defined by the average Y over X expression ratio in males). Two fixed effects with
200 interaction were therefore considered in the model, see equation (3). In order to estimate the
201 changes in sex-linked gene expression levels since the evolution of sex chromosomes, we used
202 the average of the two outgroup expression levels as a reference (offset) for every SNP position,
203 divided by two in order to be comparable to *S. latifolia* allelic expression levels.

$$204 \quad \log(\text{Expression}+1) \sim \text{Chromosome} * \text{Degeneration} + (1|\text{individual}) + (1|\text{Contig}), \text{offset} = \\ 205 \quad \log(\text{outgroup average expression}/2 + 1) \quad (3)$$

206 All effects of the model (fixed or random) were proved highly significant (p-values <
207 $2.2 \cdot 10^{-16}$) using comparison of the fit of model (3) to simpler nested models (removing one effect
208 at a time in model (3)). In order to statistically test whether there was a difference between the
209 effects of paternal and maternal alleles in females in different degeneration categories we used
210 the contrasts provided by the lmerTest package in R. This strategy provided estimates,
211 confidence intervals and p-values of the difference between the two effects of paternal and
212 maternal origin in females in interaction with degeneration levels, while normalising by the
213 expression of the two outgroups. Moreover, the presence of random effects allows to account for
214 inter-individual and inter-contig variability. Finally, p-values were corrected for multiple testing
215 using a Benjamini and Hochberg correction. These values were used to plot Figure 2 and
216 Supplementary Figures S8 to S13.

217 **Data Availability.** The new sequence data presented here can be downloaded from the European
218 Nucleotide Archive (ENA) under accession number PRJEB24933.

219 **References**

1. Bachtrog, D. Y-chromosome evolution: emerging insights into processes of Y-chromosome degeneration. *Nat. Rev. Genet.* **14**, 113–124 (2013).
2. Charlesworth, B. The evolution of chromosomal sex determination and dosage compensation. *Curr. Biol. CB* **6**, 149–162 (1996).
3. Mank, J. E. Sex chromosome dosage compensation: definitely not for everyone. *Trends Genet. TIG* **29**, 677–683 (2013).
4. Gu, L. & Walters, J. Sex chromosome dosage compensation evolution in animals: a beautiful theory, undermined by facts and bedeviled by details. *Genome Biol. Evol.* (2017).
5. Muyle, A., Shearn, R. & Marais, G. A. The Evolution of Sex Chromosomes and Dosage Compensation in Plants. *Genome Biol. Evol.* **9**, 627–645 (2017).
6. Kuroda, M. I., Hilfiker, A. & Lucchesi, J. C. Dosage Compensation in *Drosophila*-a Model for the Coordinate Regulation of Transcription. *Genetics* **204**, 435–450 (2016).
7. Albritton, S. E. *et al.* Sex-biased gene expression and evolution of the x chromosome in nematodes. *Genetics* **197**, 865–883 (2014).
8. Ercan, S. Mechanisms of x chromosome dosage compensation. *J. Genomics* **3**, 1–19 (2015).
9. Pessia, E., Engelstädter, J. & Marais, G. A. B. The evolution of X chromosome inactivation in mammals: the demise of Ohno's hypothesis? *Cell. Mol. Life Sci. CMLS* **71**, 1383–1394 (2014).
10. Disteche, C. M. Dosage compensation of the sex chromosomes. *Annu. Rev. Genet.* **46**, 537–560 (2012).

11. Julien, P. *et al.* Mechanisms and evolutionary patterns of mammalian and avian dosage compensation. *PLoS Biol.* **10**, e1001328 (2012).
12. Pessia, E., Makino, T., Bailly-Bechet, M., McLysaght, A. & Marais, G. A. B. 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).
13. Whitworth, D. J. & Pask, A. J. The X factor: X chromosome dosage compensation in the evolutionarily divergent monotremes and marsupials. *Semin. Cell Dev. Biol.* **56**, 117–121 (2016).
14. Patten, M. M. *et al.* The evolution of genomic imprinting: theories, predictions and empirical tests. *Heredity* **113**, 119–128 (2014).
15. Wagschal, A. & Feil, R. Genomic imprinting in the placenta. *Cytogenet. Genome Res.* **113**, 90–98 (2006).
16. Ohno, S. Sex chromosomes and sex-linked genes. *Springer-Verl. Berl. Heidelberg. N. Y.* (1967).
17. Rautenberg, A., Hathaway, L., Oxelman, B. & Prentice, H. C. Geographic and phylogenetic patterns in *Silene* section *Melandrium* (Caryophyllaceae) as inferred from chloroplast and nuclear DNA sequences. *Mol. Phylogenet. Evol.* **57**, 978–991 (2010).
18. Chibalina, M. V. & Filatov, D. A. Plant Y chromosome degeneration is retarded by haploid purifying selection. *Curr. Biol. CB* **21**, 1475–1479 (2011).
19. Bergero, R. & Charlesworth, D. Preservation of the Y transcriptome in a 10-million-year-old plant sex chromosome system. *Curr. Biol. CB* **21**, 1470–1474 (2011).
20. Muyle, A. *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).
21. Kazama, Y. *et al.* *SIWUS1*; an X-linked gene having no homologous Y-linked copy in *Silene latifolia*. *G3 Bethesda Md* **2**, 1269–1278 (2012).
22. Bergero, R., Qiu, S. & Charlesworth, D. Gene Loss from a Plant Sex Chromosome System. *Curr. Biol. CB* (2015). doi:10.1016/j.cub.2015.03.015

23. Papadopulos, A. S. T., Chester, M., Ridout, K. & Filatov, D. A. Rapid Y degeneration and dosage compensation in plant sex chromosomes. *Proc. Natl. Acad. Sci. U. S. A.* **112**, 13021–13026 (2015).
24. Muyle, A. *et al.* SEX-DETECTOR: A Probabilistic Approach to Study Sex Chromosomes in Non-Model Organisms. *Genome Biol. Evol.* **8**, 2530–2543 (2016).
25. Ellegren, H. & Parsch, J. The evolution of sex-biased genes and sex-biased gene expression. *Nat. Rev. Genet.* **8**, 689–698 (2007).
26. Malone, J. H. *et al.* Mediation of *Drosophila* autosomal dosage effects and compensation by network interactions. *Genome Biol.* **13**, r28 (2012).
27. Veitia, R. A., Veyrunes, F., Bottani, S. & Birchler, J. A. X chromosome inactivation and active X upregulation in therian mammals: facts, questions, and hypotheses. *J. Mol. Cell Biol.* **7**, 2–11 (2015).
28. Siroky, J., Castiglione, M. R. & Vyskot, B. DNA methylation patterns of *Melandrium album* chromosomes. *Chromosome Res. Int. J. Mol. Supramol. Evol. Asp. Chromosome Biol.* **6**, 441–446 (1998).
29. Wu, T. D. & Nacu, S. Fast and SNP-tolerant detection of complex variants and splicing in short reads. *Bioinforma. Oxf. Engl.* **26**, 873–881 (2010).
30. Zemp, N. *et al.* Evolution of sex-biased gene expression in a dioecious plant. *Nat. Plants* **2**, 16168 (2016).

220 **Supplementary Materials:**

221 Supplementary Information includes Supplementary Texts S1-S2, Supplementary Figures S1-
222 S14 and Supplementary Tables S1-S3.

223 **Acknowledgments**

224 This project was supported through an ANR grant ANR-14-CE19-0021-01 to G.A.B.M, SNSF
225 grant 160123 to A.W and Czech Science Agency grant number 16-08698S to R.H. We thank
226 Brandon Gaut, Tatiana Giraud and Judith Mank for comments on the manuscript.

227 **Author contributions**

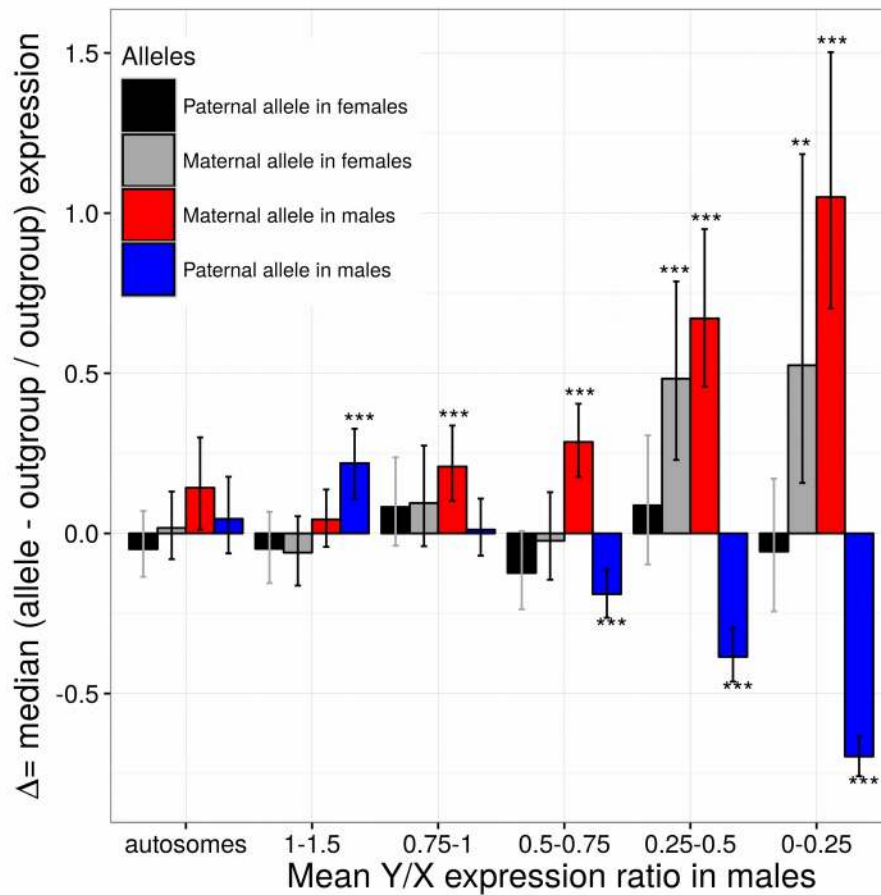
228 Aline Muyle, Niklaus Zemp, Alex Widmer and Gabriel Marais conceived the study and
229 experimental design. Niklaus Zemp and Alex Widmer prepared and sequenced the plant material.
230 Aline Muyle ran SEX-DETECTOR on the RNA-seq datasets for the three tissues, analysed the data,
231 prepared Tables and Figures and wrote the Supplementary Material with inputs from other
232 authors. Niklaus Zemp generated the X chromosome genetic map (with help from Aline Muyle
233 for the mapping and genotyping part). Radim Cegan, Jan Vrana and Roman Hobza did the Y
234 chromosome flow cytometry sorting and sequencing. Clothilde Deschamps did the first assembly
235 of the sorted Y chromosome and improved it with RNA-seq data with the help of Cecile
236 Fruchard. Aline Muyle did the blasts to validate the inferences of SEX-DETECTOR. Raquel Tavares
237 did the GO term analysis. Aline Muyle and Frank Picard did the statistical analyses of the data.
238 Gabriel Marais and Aline Muyle wrote the main text of the manuscript with inputs from other
239 authors.

240 **Author information**

241 The authors declare no competing interests.

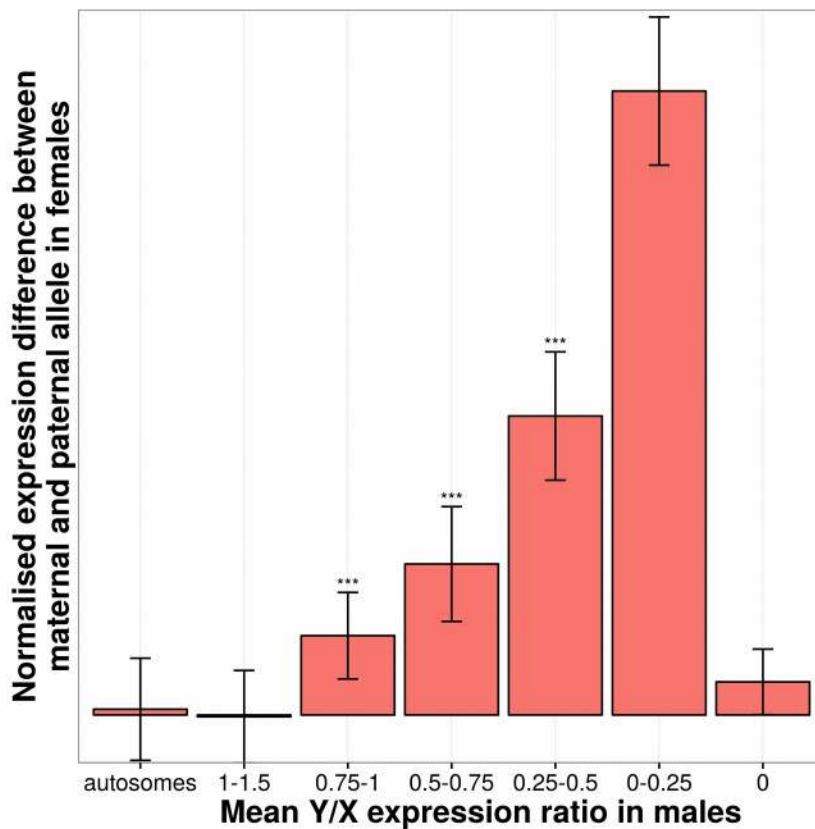
242 Correspondence and requests for materials should be addressed to Aline Muyle (email:
243 muyle.aline@gmail.fr).

244 **Figures**



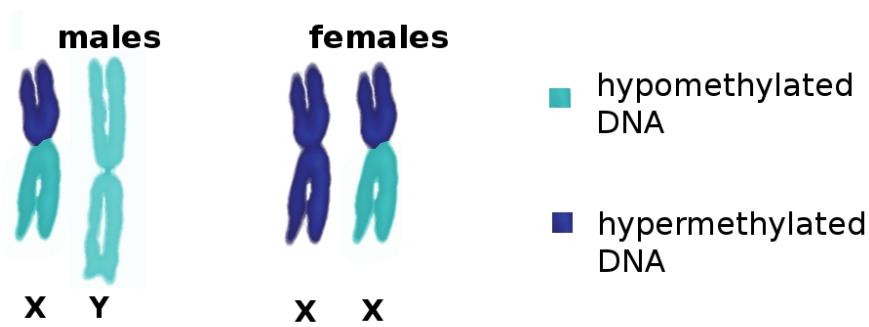
245 **Figure 1:** Normalised difference (hereafter Δ) in allelic expression levels between *S. latifolia* and
 246 the outgroup without sex chromosomes *S. vulgaris*, in autosomal and sex-linked contigs
 247 for the seedling tissue. If Δ is lower, higher or equal to zero, then expression in *S.*
 248 *latifolia* is respectively lower, higher or equal to the outgroup (See Materials and
 249 Methods for details). For all contig categories, Δ was compared to zero using a Wilcoxon
 250 test. The median Δ , confidence intervals and p-values adjusted for multiple testing using
 251 a Benjamini and Hochberg correction are shown (***: p-value < 0.001; **: p-value <
 252 0.01, *: p-value < 0.05). Allelic expression at SNP positions was averaged for each contig
 253 separately and the Y/X ratio was used as a proxy for Y degeneration to group contigs.

254 Contigs with sex-biased expression were removed, as well as contigs with Y/X
255 expression ratios above 1.5. Sample sizes for the different contig categories are:
256 autosomal: 200; 1-1.5:148; 0.75-1:139; 0.5-0.75:160; 0.25-0.5:114; 0-0.25:79 (we
257 randomly selected 200 autosomal contigs to ensure similar statistical power among gene
258 categories).



259 **Figure 2:** Normalised expression difference between maternal and paternal alleles in *S. latifolia*
260 females in autosomal and sex-linked SNPs in the seedling tissue. The Y axis unit is the

261 normalised allelic read count difference in log scale. A linear regression model with
262 mixed effects was used to estimate the normalised difference between the effect of
263 paternal and maternal origin of alleles in interaction with the contig status (autosomal or
264 sex-linked with various levels of Y degeneration), while accounting for inter-contig and
265 inter-individual variability (see Materials and Methods for details). The analysis is SNP-
266 wise and reveals consistent patterns across SNPs. See Fig. 1 legend for sample sizes for
267 the different contig categories and statistical significance symbols.



268 **Figure 3:** Illustration of DNA methylation staining results in *S. latifolia* from Siroky et al.²⁸. See
269 Supplementary Figure S14 for the original Figure. One arm of one of the two X chromosomes in
270 females was hypomethylated, as well as the same arm of the single X in males.

271 **Supplementary Materials**

272 **Supplementary Text S1: Dosage compensation in X-hemizygous** 273 **genes**

274 The first papers on dosage compensation in *S. latifolia* were contradictory because they
275 focused on different gene sets. Muyle et al.¹ focused on X/Y gene pairs while other papers
276 focused on X-hemizygous genes^{2,3}. However, the X-hemizygous gene sets returned by the RNA-
277 seq approach used in those papers is less reliable than the X/Y gene sets⁴. A gene might be
278 inferred as X-hemizygous simply because the – still functional – Y copy is not expressed in the
279 tissue sampled for RNA-seq. In *S. latifolia*, X-hemizygous genes tend to be less expressed than
280 X/Y genes and are less likely to be detected by segregation analysis as efficient SNP calling
281 requires a certain read depth, see⁴. Moreover, X-hemizygous genes are inferred from X
282 polymorphisms while X/Y genes can be detected both with X and X/Y polymorphisms, which
283 are more numerous. Another inherent bias to X-hemizygous contig inference comes from the
284 assembly step. If the X and the Y copy are too divergent to be assembled together, the X contig
285 will be wrongly inferred as X-hemizygous because Y alleles will be absent from the contig (this
286 bias was at least partly corrected in the analyses presented here, see Material and Method section
287 5.1). The inferences of X-hemizygous genes using the RNA-seq approach (including SEX-
288 DETector) imply a higher rate of both false positives and false negatives than those for X/Y gene
289 pairs. In Papadopoulos et al.⁵, 25% of the X/Y chromosomes were sequenced using a genomic
290 approach. A much higher fraction of X-hemizygous genes was found than in previous RNA-seq
291 papers^{2,3}. Papadopoulos et al.⁵ did find evidence for dosage compensation in approximately half

292 of X-hemizygous genes (see their figure 3D). Due to limitations of the RNA-seq approach in
293 inferring X-hemizygous genes, results on X-hemizygous contigs are analysed separately here.

294 Poor dosage compensation of X-hemizygous contigs compared to X/Y contigs with high
295 Y degeneration was observed across all tissues (Supplementary Figures 2 to 7). Also, the parental
296 origin of the X chromosome has limited to no effect on female X expression levels for X-
297 hemizygous contigs, unlike X/Y contigs (Supplementary Figures 8 to 13). A reason that could
298 explain such a different pattern for X-hemizygous genes compared to X/Y genes is the possible
299 dosage insensitivity of X-hemizygous genes. X-hemizygous genes could have lost their Y copy
300 because dosage was not important for them and selection neither slowed down the loss of the Y
301 copy nor selected for dosage compensation when degeneration inevitably occurred⁶. A well
302 described characteristic of dosage sensitive genes is that they tend to code proteins involved in
303 large complexes⁷. Gene Ontology was studied using the Blast2GO PRO version 2.7.2³⁰ as in⁸.
304 Using the GO-term analysis, our set of X-hemizygous contigs were found to be significantly
305 depleted in ribosomal protein coding genes compared to autosomal genes (p-value $1.3 \cdot 10^{-4}$),
306 which is consistent with the global dosage insensitivity of X-hemizygous genes in *S. latifolia*.
307 This depletion in large protein complexes was not found when comparing X/Y genes to
308 autosomal genes.

309 **Supplementary Text S2:**

310 **1) Plant material and sequencing**

311 **1.1) RNA-seq Illumina data**

312 RNA-seq data from previous studies were used (the GEO database GEO Series
313 GSE35563, European Nucleotide Archive PRJEB14171), it included flower buds and leaf tissues
314 from individuals of a cross in *S. latifolia* as well as individuals in *S. vulgaris*. In addition to these
315 preexisting data, RNA-seq reads were generated in a comparable way for seedlings of a
316 controlled cross using the same parents in *S. latifolia*, four males and four females were sampled
317 (Seed_lati_female_1, Seed_lati_female_2, Seed_lati_female_3, Seed_lati_female_4,
318 Seed_lati_male_1, Seed_lati_male_2, Seed_lati_male_3 and Seed_lati_male_4). Seedlings were
319 also sequenced for *S. vulgaris* (Seed_vulg_herm_1, Seed_vulg_herm_2, Seed_vulg_herm_3 and
320 Seed_vulg_herm_4). Seedlings were grown in a temperature controlled climate chamber in
321 Eschikon (Switzerland) using the same conditions as in⁸. The *S. latifolia* and *S. vulgaris*
322 seedlings were collected without roots at the four-leaf stage. The sexing of the *S. latifolia*
323 seedlings was done using Y specific markers⁹ that were amplified with the direct PCAR
324 KAPA3G Plant PCR Kit (however male number 3 was later shown to be a female). High quality
325 RNA (RIN > 8.5) was extracted using the total RNA mini kit from Geneaid. Twelve RNA-seq
326 libraries were produced using the Truseq kit v2 from Illumina. Libraries were tagged individually
327 and sequenced in two Illumina HiSeq 2000 channels at the D-BSSE (ETH Zürich, Switzerland)
328 using 100 bp paired-end read protocol.

329 *S. viscosa* seeds we received from botanical gardens or collected in the wild by Bohuslav
330 Janousek and grown under controlled conditions in a greenhouse in Eschikon (Switzerland) and
331 Lyon (France). Similarly to⁸, flower buds after removing the calyx and leaves were collected.
332 Total RNA were extracted through the Spectrum Plant Total RNA kit (Sigma, Inc., USA)
333 following the manufacturer's protocol and treated with a DNase. Libraries were prepared with
334 the TruSeq RNA sample Preparation v2 kit (Illumina Inc., USA). Each 2 nM cDNA library was

335 sequenced using a paired-end protocol on a HiSeq2000 sequencer. Demultiplexing was
336 performed using CASAVA 1.8.1 (Illumina) to produce paired sequence files containing reads for
337 each sample in Illumina FASTQ format. RNA extraction and sequencing were done by the
338 sequencing platform in the AGAP laboratory, Montpellier, France (<http://umr-agap.cirad.fr/>).

339 A female individual from an interspecific *S. latifolia* cross (C1_37) was back crossed
340 with a male from an 11 generation inbred line (U10_49). The offspring (hereafter called BC1
341 individuals) were grown under controlled conditions in a greenhouse in Eschikon (Switzerland).
342 High quality RNA from flower buds as described in¹⁰ was extracted from 48 BC1 individuals (35
343 females and 13 males). 48 RNA-seq libraries were produced using the Truseq kit v2 from
344 Illumina with a median insert size of about 200 bp. Individuals were tagged separately and
345 sequenced in four Illumina HiSeq 2000 channels at the D-BSSE (ETH Zürich, Switzerland) using
346 100bp paired-end read protocol. The parents used for this back cross had previously been
347 sequenced in a similar way^{1,8}.

348 **1.2) DNA-seq data from filtered Y chromosome**

349 Y chromosome DNA was isolated using flow cytometry. The samples for flow cytometric
350 experiments were prepared from root tips according to¹¹ with modifications. Seeds of *S. latifolia*
351 were germinated in a petri dish immersed in water at 25°C for 2 days until optimal length of
352 roots was achieved (1 cm). The root cells were synchronized by treatment with 2mM
353 hydroxyurea at 25°C for 18h. Accumulation of metaphases was achieved using 2.5µM oryzalin.
354 Approximately 200 root tips were necessary to prepare 1ml of sample. The chromosomes were
355 released from the root tips by mechanical homogenization using a Polytron PT1200 homogenizer
356 (Kinematica AG, Littau, Switzerland) at 18,000rpm for 13 s. The crude suspension was filtered
357 and stained with DAPI (2µg/ml). All flow cytometric experiments were performed on FACSAria

358 II SORP flow cytometer (BD Biosciences, San José, Calif., USA). Isolated Y chromosomes were
359 sequenced with 2x100bp PE Illumina HiSeq.

360 **1.3) RNA-seq PacBio data**

361 Plants from an 11 generation inbred line were grown under controlled conditions in a
362 greenhouse in Eschikon (Switzerland). One male (U11_02) was randomly selected. High quality
363 RNA (RIN > 7.5) were extracted using the total RNA mini kit of Geneaid from very small flower
364 buds, small and large flower buds, flowers before anthesis without calyces, rosette leaves,
365 seedlings (4 leaves stage) and pollen. RNA of the different tissues was equally pooled and cDNA
366 was produced using the Clontech SMARTer Kit. The cDNA pool was then normalized using a
367 duplex specific endonuclease of the Evrogen TRIMMER kit. Two ranges were selected (1- 1.3
368 kb and 1.2 -2 kb) using the Pippin Prep (Sage Science). Two SMRTbell libraries were prepared
369 using the C2 Pacific Biosciences (PacBio) chemistry and sequenced with two SMRT Cells runs
370 on a PacBio RS II at the Functional Genomic Center Zurich (FGCZ).

371 **1.4) RNA-seq 454 data**

372 Previously generated 454 data was used^{8,12}.

373 **2) Reference transcriptome assembly**

374 The same reference transcriptome as in Muyle et al.¹² and Zemp et al.⁸ was used.

375 **3) Inference of sex-linked contigs**

376 Autosomal and sex-linked contigs were inferred as in Muyle et al.¹² and Zemp et al.⁸.
377 Illumina reads from the individuals of the cross were mapped onto the assembly using BWA¹³
378 version 0.6.2 with the following parameters: bwa aln -n 5 and bwa sampe. The libraries were
379 then merged using SAMTOOLS version 0.1.18¹⁴. The obtained alignments were locally

380 realigned using GATK IndelRealigner¹⁵ and were analysed using reads2snps¹⁶ version 3.0 with
381 the following parameters: -fis 0 -model M2 -output_genotype best -multi_alleles acc
382 -min_coverage 3 -par false. This allowed to genotype individuals at each loci while allowing for
383 biases in allele expression, and without cleaning for paralogous SNPs. Indeed, X/Y SNPs tend to
384 be filtered out by paraclean, a program which removes paralogous positions¹⁷. A second run of
385 genotyping was done with paraclean in order to later remove paralogous SNPs from autosomal
386 contigs only. SEX-DETECTOR¹² was then used to infer contig segregation types after estimation of
387 parameters using an SEM algorithm. Contig posterior segregation type probabilities were filtered
388 to be higher than 0.8. Because the parents were not sequenced for the leaf and seedling datasets,
389 SEX-DETECTOR was run using the flower bud data for the parents.

390 **4) Reference mapping bias correction**

391 In order to avoid biases towards the reference allele in expression level estimates, a
392 second mapping was done using the program GSNAP¹⁸ with SNP tolerant mapping option. A
393 GSNAP SNP file was generated by home-made perl scripts using the SEX-DETECTOR SNP detail
394 output file. Shortly, for each polymorphic position of all contigs, the most probable posterior
395 SNP type was used to extract the possible alleles and write them to the GSNAP SNP file. This
396 way, reference mapping bias was corrected for both sex-linked and autosomal contigs. Only
397 uniquely mapped and concordant paired reads were kept after this. See Supplementary Table S1
398 for percentage of mapped reads. SEX-DETECTOR was run a second time on this new mapping and
399 the new inferences were used afterwards for all analyses (see Supplementary Table S2 for
400 inference results).

401 **5) Validation of sex-linked contigs**

402 **5.1) Detection of false X-hemizygous contigs**

403 Erroneous inference of X-hemizygous contig can be due to a true X/Y gene which X and
404 Y copies were assembled into different contigs. In order to detect such cases, X-hemizygous
405 contigs were blasted¹⁹ with parameter -e 1E-5 against RNA-seq contigs that have male-limited
406 expression (see section 7 below for how male-limited contigs were inferred). These cases were
407 removed from the analyses presented here.

408 **5.2) Validation using data from literature**

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

411 **5.3) Validation using a genetic map**

412 A genetic map was built and contigs from the X linkage group were used to validate
413 SEX-DETECTOR inferences. RNA-seq reads from the flower bud *S. latifolia* full-sib cross
414 (hereafter CP) and backcross (hereafter BC1) were mapped against the reference transcriptome
415 using BWA¹³ with a maximum number of mismatch equal to 5. Libraries were merged and
416 realigned using GATK¹⁵ and SNPs were analysed using reads2snps¹⁶. Using a customized perl
417 script, SNP genotypes from the parents and the offspring as well as the associated posterior
418 probabilities were extracted from the reads2snps output file. Only SNPs with a reads2snps
419 posterior genotyping probability higher than 0.8 were kept for further analyses. Then, only
420 informative SNPs were kept: both parents had to be homozygous and different between father
421 LEUK144-3 and mother U10_37 in a first generation backcross population design (BC1) and at
422 least one allele had to be different between mother C1_37 and father U10_49 in the cross-
423 pollinator (CP). Filtered SNPs were then converted into a JoinMap format using a customized R
424 script. If more than one informative SNP per contig was present, the SNP was used with less
425 segregation distortion and less missing values. This led to 8,023 BC1 and 16,243 CP markers.

426 Loci with more than 10 % missing values were excluded, resulting in 7,951 BC1 and 15,118 CP
427 markers. Linkage groups were identified using the default setting of JoinMap 4.1²⁰. Robustness
428 of the assignment of the linkage groups was tested using LepMap²¹. Blasting the contigs against
429 known sex-linked genes allowed the identification of the X chromosome linkage group. Contigs
430 could not be ordered along the linkage groups due to the too limited number of individuals that
431 prevented the convergence of contig order. However, contigs were reliably attributed to linkage
432 groups.

433 **5.4) Validation using isolated Y chromosome DNA-seq data**

434 Filtered Y chromosome DNA-seq reads were filtered for quality and Illumina adapters
435 were removed using the ea-utils FASTQ processing utilities²². The optimal kmer value for
436 assembly was searched using KmerGenie²³. Filtered reads were assembled using soapdenovo2²⁴
437 with kmer=49, as suggested by KmerGenie. The obtained assembly was highly fragmented,
438 therefore RNA-seq data was used to join, order and orient the genomic fragments with
439 L_RNA_scaffolder²⁵. The following RNA-seq reads were used (see section 1): one sample of
440 male flower buds sequenced by 454, 6 samples of male flower buds sequenced by Illumina
441 paired-end, 4 samples of male leaves sequenced by Illumina paired-end and one sample of male
442 pooled tissues sequenced by PacBio. The genomic assembly was successively scaffolded with
443 L_RNA_scaffolder using RNA-seq samples one after the other, first 454 samples then Illumina
444 and finally PacBio. The obtained contigs were filtered to be longer than 200pb.

445 **5.5) Set of validated sex-linked and autosomal contigs**

446 The three sources of data (litterature, genetic map and filtered Y sequence data) were
447 compared to SEX-DETECTOR inferred sex-linked RNA-seq contigs using BLAST¹⁹ with parameter
448 -e 1E-5. Blasts were filtered for having a percentage of identity over 90%, an alignment length

449 over 100bp and were manually checked. If a sex-linked RNA-seq contig blasted against a
450 sequence from one of the three data sources (literature, X genetic map or filtered Y DNA-seq) it
451 was then considered as validated. See Supplementary Table S2 for numbers of validated sex-
452 linked contigs.

453 **6) Expression level estimates**

454 **6.1) whole contig expression levels**

455 Whole contig mean expression levels were obtained for each individual using GATK
456 DepthOfCoverage¹⁵ as the sum of every position coverage, divided by the length of the contig.
457 Normalised expression levels, in RPKM²⁶, were then computed for each individual by dividing
458 by the value by the library size of the individual (total number of mapped reads), accounting for
459 different depths of coverage among individuals. Whole contig mean male and female expression
460 levels were then computed by averaging male and female individuals for each contig.

461 **6.2) Allelic expression levels filtering**

462 In order to study separately X and Y allele expression levels in males and females,
463 expression levels were studied at the SNP level. In *S. latifolia*, for each sex-linked contig
464 expression levels were estimated using read counts from both X/Y and X-hemizygous
465 informative SNPs. SNPs were attributed to an X/Y or X hemizygous segregation type if the
466 according posterior probability was higher than 0.5. SNPs are considered informative if the
467 father is heterozygous and has a genotype that is different from the mother (otherwise it is not
468 possible to tell apart the X from the Y allele and therefore it is not possible to compute X and Y
469 expression separately). X/Y SNPs for which at least one female had over two percent of her

470 reads belonging to the Y allele were removed as unlikely to be true X/Y SNPs. Informative
471 autosomal SNPs from autosomal contigs were used in a similar way.

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

477 In order to make *S. latifolia* expression levels comparable to *S. viscosa* and *S. vulgaris*
478 for sex-linked contigs, *S. viscosa* and *S. vulgaris* expression levels were estimated using only the
479 positions used in *S. latifolia* (informative X/Y or X-hemizygous SNPs). The read count of every
480 position in every contig and for every *S. viscosa* and *S. vulgaris* individual was given by GATK
481 DepthOfCoverage¹⁵. Only positions corresponding to informative autosomal, X/Y or X-
482 hemizygous SNPs in *S. latifolia* were used to compute the expression level for each contig and
483 each individual as explained in equation (1).

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

487 **7) Identification of contigs with sex-biased expression**

488 The analysis was done separately for the three tissues (flower buds, seedling and rosette
489 leaves) as in Zemp et al.⁸ using the R package edgeR²⁷. See Supplementary Table S2 for number
490 of sex-biased contigs removed in order to study dosage compensation. Male-limited expressed

491 contigs were identified by calculating the mean expression values (FPKM) in both sexes and
492 selecting those which were exclusively expressed in males.

493 **Supplementary References**

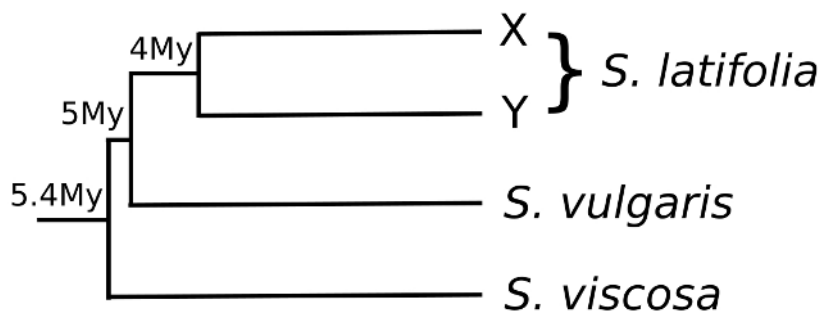
1. Muyle, A. *et al.* Rapid de novo evolution of X chromosome dosage compensation in *Silene latifolia*, a plant with young sex chromosomes. *PLoS Biol.* **10**, e1001308 (2012).
2. Chibalina, M. V. & Filatov, D. A. Plant Y chromosome degeneration is retarded by haploid purifying selection. *Curr. Biol. CB* **21**, 1475–1479 (2011).
3. Bergero, R., Qiu, S. & Charlesworth, D. Gene Loss from a Plant Sex Chromosome System. *Curr. Biol. CB* (2015). doi:10.1016/j.cub.2015.03.015
4. Blavet, N. *et al.* Identifying new sex-linked genes through BAC sequencing in the dioecious plant *Silene latifolia*. *BMC Genomics* (2015).
5. Papadopulos, A. S. T., Chester, M., Ridout, K. & Filatov, D. A. Rapid Y degeneration and dosage compensation in plant sex chromosomes. *Proc. Natl. Acad. Sci. U. S. A.* **112**, 13021–13026 (2015).
6. Muyle, A., Shearn, R. & Marais, G. A. The Evolution of Sex Chromosomes and Dosage Compensation in Plants. *Genome Biol. Evol.* **9**, 627–645 (2017).
7. Pessia, E., Makino, T., Bailly-Bechet, M., McLysaght, A. & Marais, G. A. B. 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).
8. Zemp, N. *et al.* Evolution of sex-biased gene expression in a dioecious plant. *Nat. Plants* **2**, 16168 (2016).
9. Hobza, R. & Widmer, A. Efficient molecular sexing in dioecious *Silene latifolia* and *S. dioica* and paternity analysis in F(1) hybrids. *Mol. Ecol. Resour.* **8**, 1274–1276 (2008).

10. Zemp, N., Minder, A. & Widmer, A. Identification of internal reference genes for gene expression normalization between the two sexes in dioecious white Campion. *PLoS One* **9**, e92893 (2014).
11. Vrána, J., Simková, H., Kubaláková, M., Cíhalíková, J. & Doležel, J. Flow cytometric chromosome sorting in plants: the next generation. *Methods San Diego Calif* **57**, 331–337 (2012).
12. Muyle, A. *et al.* SEX-DETECTOR: A Probabilistic Approach to Study Sex Chromosomes in Non-Model Organisms. *Genome Biol. Evol.* **8**, 2530–2543 (2016).
13. Li, H. & Durbin, R. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinforma. Oxf. Engl.* **25**, 1754–1760 (2009).
14. Li, H. *et al.* The Sequence Alignment/Map format and SAMtools. *Bioinforma. Oxf. Engl.* **25**, 2078–2079 (2009).
15. DePristo, M. A. *et al.* A framework for variation discovery and genotyping using next-generation DNA sequencing data. *Nat. Genet.* **43**, 491–498 (2011).
16. Tsagkogeorga, G., Cahais, V. & Galtier, N. 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).
17. Gayral, P. *et al.* Reference-free population genomics from next-generation transcriptome data and the vertebrate-invertebrate gap. *PLoS Genet.* **9**, e1003457 (2013).
18. Wu, T. D. & Nacu, S. Fast and SNP-tolerant detection of complex variants and splicing in short reads. *Bioinforma. Oxf. Engl.* **26**, 873–881 (2010).
19. Altschul, S. F., Gish, W., Miller, W., Myers, E. W. & Lipman, D. J. Basic local alignment search tool. *J. Mol. Biol.* **215**, 403–410 (1990).
20. VAN Ooijen, J. W. Multipoint maximum likelihood mapping in a full-sib family of an outbreeding species. *Genet. Res.* **93**, 343–349 (2011).

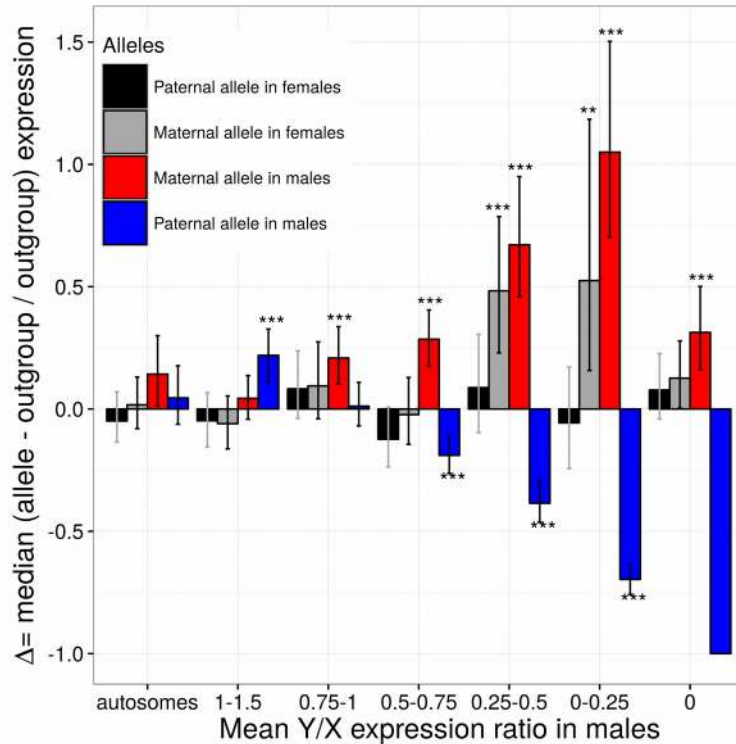
21. Rastas, P., Paulin, L., Hanski, I., Lehtonen, R. & Auvinen, P. Lep-MAP: fast and accurate linkage map construction for large SNP datasets. *Bioinforma. Oxf. Engl.* **29**, 3128–3134 (2013).
22. Aronesty, E. *ea-utils: 'Command-line tools for processing biological sequencing data'*. (2011).
23. Chikhi, R. & Medvedev, P. Informed and automated k-mer size selection for genome assembly. *Bioinforma. Oxf. Engl.* **30**, 31–37 (2014).
24. Luo, R. *et al.* SOAPdenovo2: an empirically improved memory-efficient short-read de novo assembler. *GigaScience* **1**, 18 (2012).
25. Xue, W. *et al.* L_RNA_scaffolder: scaffolding genomes with transcripts. *BMC Genomics* **14**, 604 (2013).
26. Oshlack, A., Robinson, M. D. & Young, M. D. From RNA-seq reads to differential expression results. *Genome Biol.* **11**, 220 (2010).
27. Robinson, M. D., McCarthy, D. J. & Smyth, G. K. edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinforma. Oxf. Engl.* **26**, 139–140 (2010).
28. R Development Core Team. R: A language and environment for statistical computing, reference index version 2.15.0. *R Foundation for Statistical Computing, Vienna, Austria* (2012).
29. Benjamini, Y. & Hochberg, Y. Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing. *J. R. Stat. Soc. Ser. B Methodol.* **57**, 289–300 (1995).
30. Conesa, A. *et al.* Blast2GO: a universal tool for annotation, visualization and analysis in functional genomics research. *Bioinforma. Oxf. Engl.* **21**, 3674–3676 (2005).
31. Rautenberg, A., Hathaway, L., Oxelman, B. & Prentice, H. C. Geographic and phylogenetic patterns in *Silene* section *Melandrium* (Caryophyllaceae) as inferred from chloroplast and nuclear DNA sequences. *Mol. Phylogenet. Evol.* **57**, 978–991 (2010).

32. Jenkins, C. & Keller, S. R. A phylogenetic comparative study of preadaptation for invasiveness in the genus *Silene* (Caryophyllaceae). *Biol. Invasions* **13**, 1471–1486 (2010).
33. Rautenberg, A., Sloan, D. B., Aldén, V. & Oxelman, B. Phylogenetic Relationships of *Silene multinervia* and *Silene* Section *Conoimorpha* (Caryophyllaceae). *Syst. Bot.* **37**, 226–237 (2012).
34. Siroky, J., Castiglione, M. R. & Vyskot, B. DNA methylation patterns of *Melandrium album* chromosomes. *Chromosome Res. Int. J. Mol. Supramol. Evol. Asp. Chromosome Biol.* **6**, 441–446 (1998).

494 **Supplementary Figures**

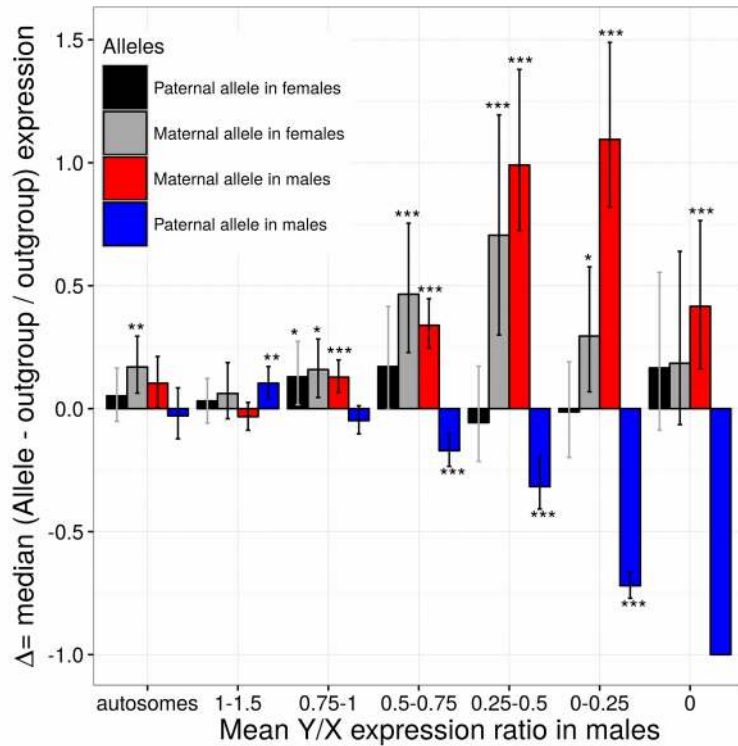


495 **Supplementary Figure S1:** Relatedness among the three studied species, extracted from³¹ ages
 496 at the nodes are shown in million years (My). The exact relationship among species is poorly
 497 resolved^{31–33}. In some phylogenies *S. viscosa* is closest to *S. latifolia*, whereas in others *S.*
 498 *vulgaris* is closest as shown here, and in others both species are equally diverged to *S. latifolia*.

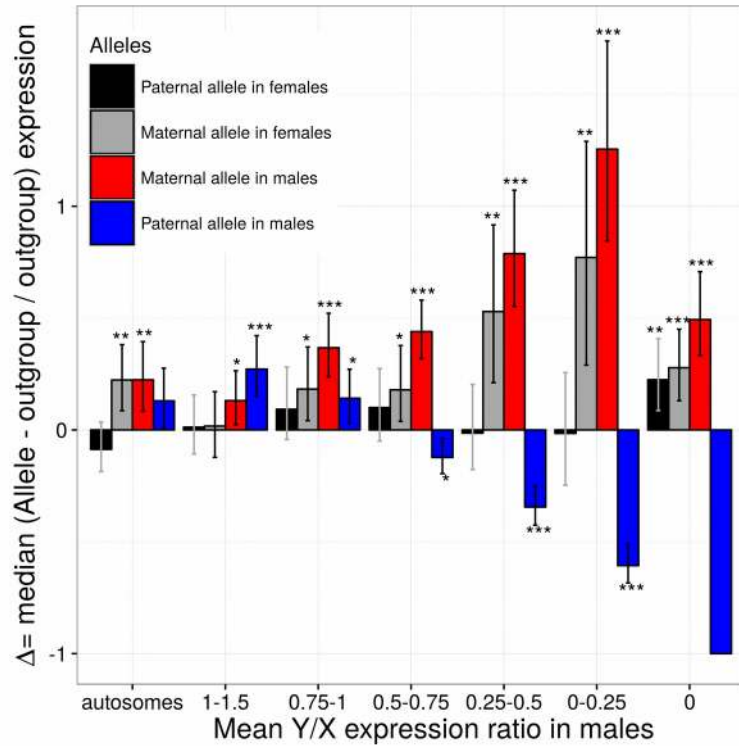


499 **Supplementary Figure S2:** Normalised difference in allelic expression levels between *S.*
500 *latifolia* and the two outgroups without sex chromosomes *S. vulgaris* and *S. viscosa* (hereafter
501 Δ), in autosomal and sex-linked contigs for the **seedling** tissue. Maternal and paternal allelic read
502 numbers were summed at SNP positions and normalised for each individual separately, then
503 averaged among individuals for each contig. Δ was computed as follows: $\Delta = (\text{allelic expression}$
504 $\text{in } S. \textit{latifolia} - \text{allelic expression in the outgroup}) / \text{allelic expression in the outgroup}$. If Δ is
505 lower, higher or equal to zero, then expression in *S. latifolia* is respectively lower, higher or
506 equal to the outgroup. For all contig categories, Δ was compared to zero using a Wilcoxon test.
507 The median Δ , confidence intervals and p-values adjusted for multiple testing using a Benjamini
508 and Hochberg correction are shown (***: p-value < 0.001; **: p-value < 0.01, *: p-value < 0.05).
509 The Y/X ratio was computed in *S. latifolia* males and averaged among individuals to use as a

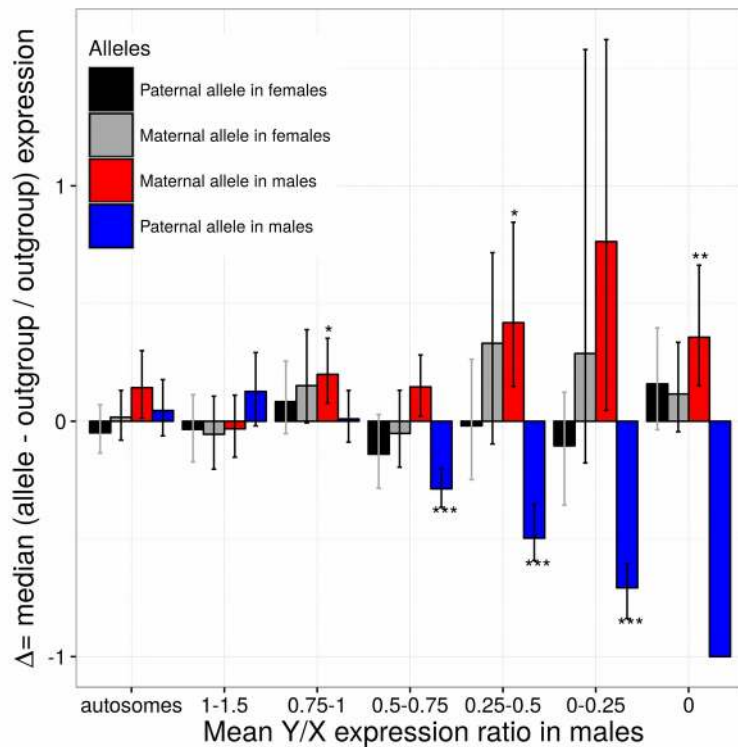
510 proxy for Y degeneration. X-hemizygous contigs have a Y/X ratio equal to zero. Contigs with
511 sex-biased expression were removed, as well as contigs with Y/X expression ratios above 1.5.
512 Sample sizes for the different contig categories are: autosomal:200; 1-1.5:148; 0.75-1:139; 0.5-
513 0.75:160; 0.25-0.5:114; 0-0.25:79; 0:205 (note that 200 autosomal contigs were randomly
514 selected in order to have similar statistical power among gene categories). In the absence of
515 dosage compensation, the single X in males should be expressed at levels similar to the outgroup
516 that does not have sex chromosomes, in other words, without dosage compensation Δ should be
517 close to zero for the maternal allele in males (red bars). Results show that the maternal allele is
518 hyper-expressed in *S. latifolia* when the Y chromosome is degenerated, both in males and
519 females.



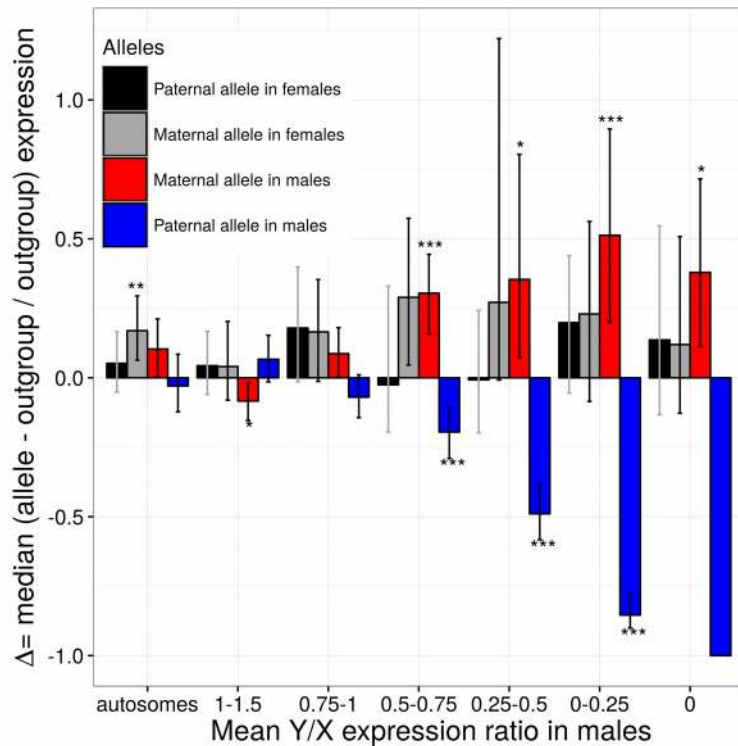
520 **Supplementary Figure S3:** Normalised difference in allelic expression levels between *S.*
 521 *latifolia* and the two outgroups without sex chromosomes *S. vulgaris* and *S. viscosa* (Δ), in
 522 autosomal and sex-linked contigs for the **flower bud** tissue. Same legend as Supplementary
 523 Figure S2 except for sample sizes for the different contig categories: autosomal:200; 1-1.5:95;
 524 0.75-1:195; 0.5-0.75:203; 0.25-0.5:176; 0-0.25:116; 0:103.



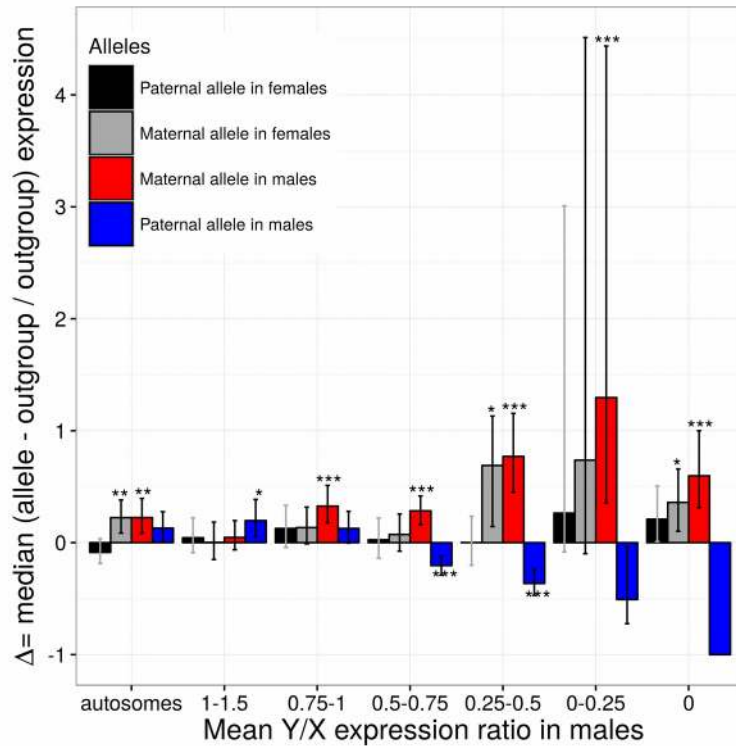
525 **Supplementary Figure S4:** Normalised difference in allelic expression levels between *S.*
 526 *latifolia* and the two outgroups without sex chromosomes *S. vulgaris* and *S. viscosa* (Δ), in
 527 autosomal and sex-linked contigs for the **leaf** tissue. Same legend as Supplementary Figure S2
 528 except for sample sizes for the different contig categories: autosomal:200; 1-1.5:159; 0.75-1:132;
 529 0.5-0.75:147; 0.25-0.5:126; 0-0.25:71; 0:275.



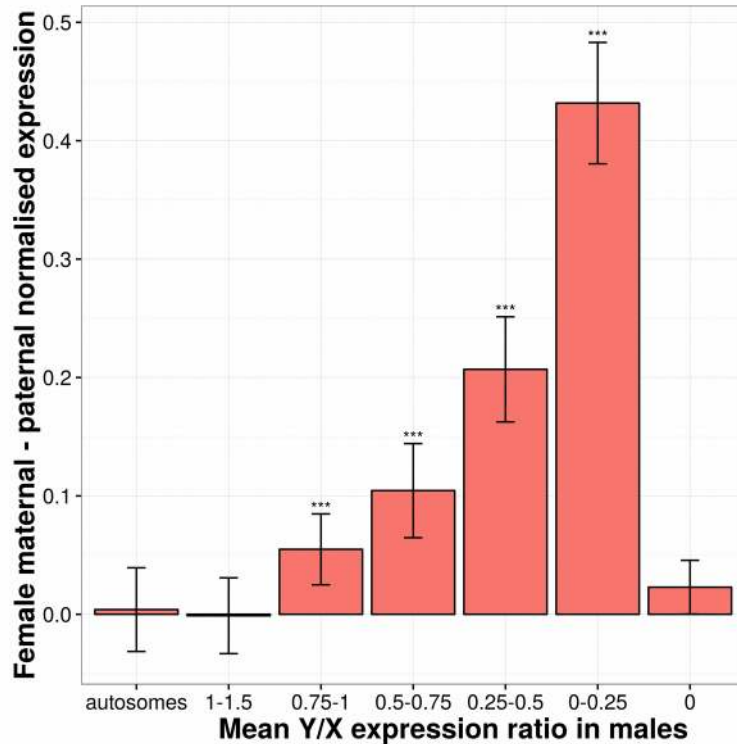
530 **Supplementary Figure S5:** Normalised difference in allelic expression levels between *S.*
 531 *latifolia* and the two outgroups without sex chromosomes *S. vulgaris* and *S. viscosa* (Δ), in
 532 autosomal and sex-linked contigs that were **validated** (see Materials and Methods), for the
 533 **seedling** tissue. Same legend as Supplementary Figure S2 except for sample sizes for the
 534 different contig categories: autosomal:77; 1-1.5:71; 0.75-1:82; 0.5-0.75:91; 0.25-0.5:44; 0-
 535 0.25:29; 0:89.



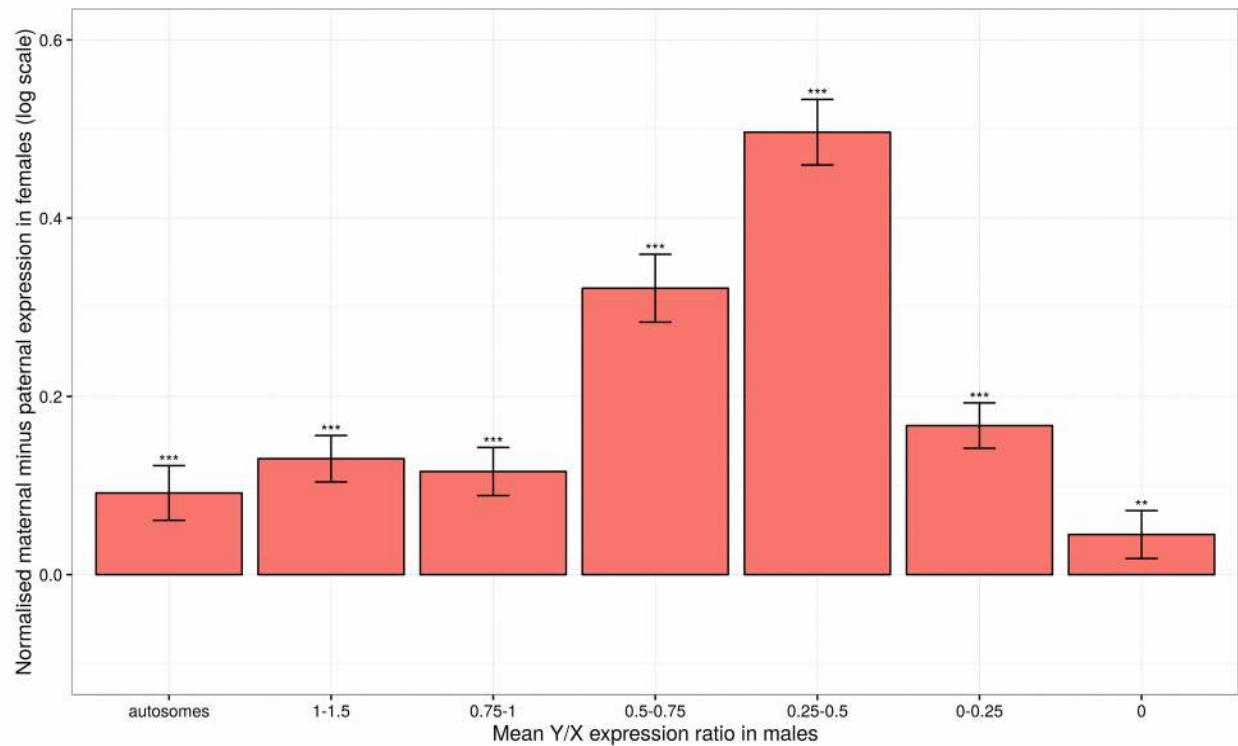
536 **Supplementary Figure S6:** Normalised difference in allelic expression levels between *S.*
 537 *latifolia* and the two outgroups without sex chromosomes *S. vulgaris* and *S. viscosa* (Δ), in
 538 autosomal and sex-linked contigs that were **validated** (see Materials and Methods), for the
 539 **flower bud** tissue. Same legend as Supplementary Figure S2 except for sample sizes for the
 540 different contig categories: autosomal:74; 1-1.5:86; 0.75-1:91; 0.5-0.75:67; 0.25-0.5:45; 0-
 541 0.25:31; 0:55.



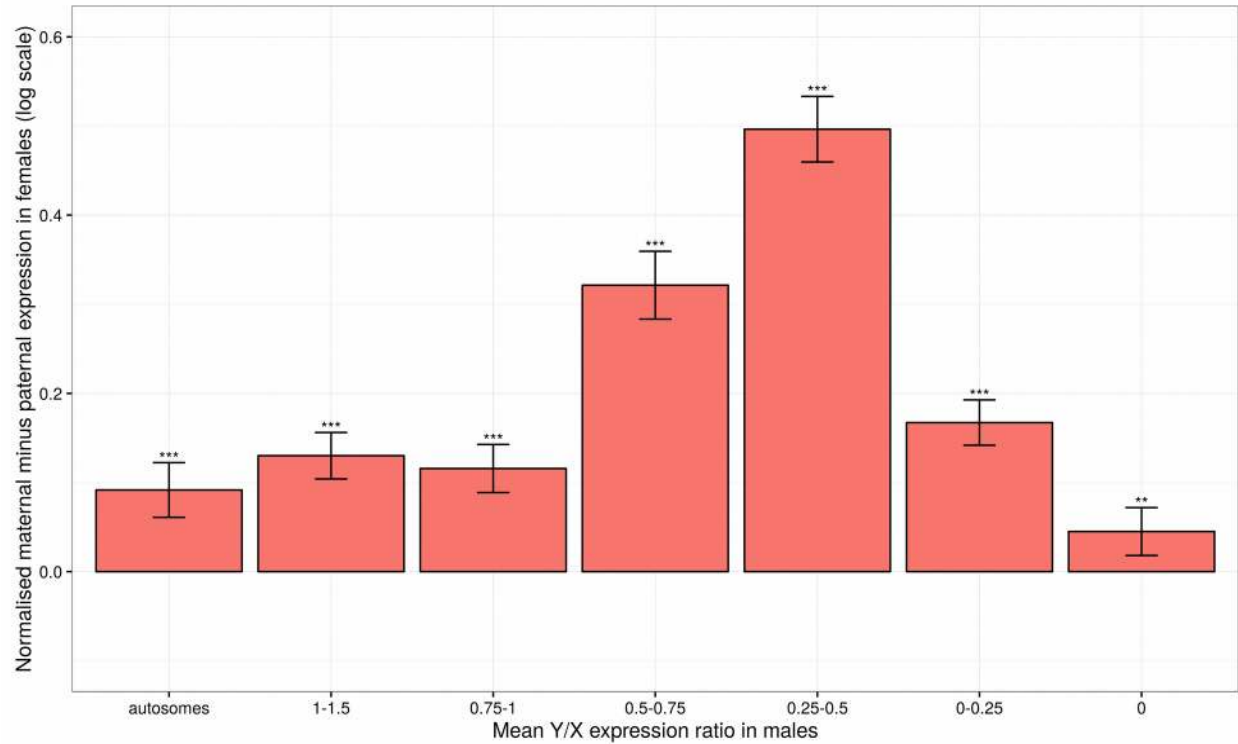
542 **Supplementary Figure S7:** Normalised difference in allelic expression levels between *S.*
 543 *latifolia* and the two outgroups without sex chromosomes *S. vulgaris* and *S. viscosa* (Δ), in
 544 autosomal and sex-linked contigs that were **validated** (see Materials and Methods), for the **leaf**
 545 tissue. Same legend as Supplementary Figure S2 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.



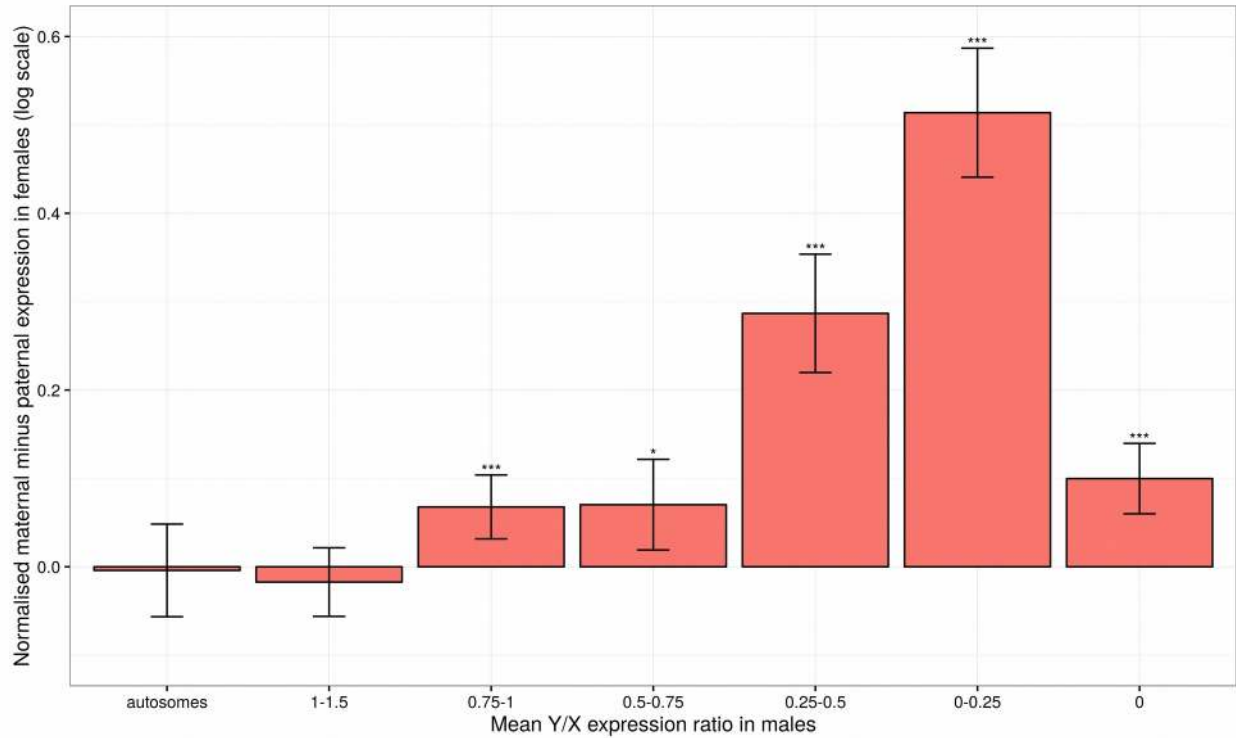
547 **Supplementary Figure S8:** Normalised expression difference between the maternal and paternal
548 allele in *S. latifolia* females in autosomal and sex-linked contigs for the **seedling** tissue. The Y
549 axis unit is the normalised allelic read count difference in log scale. A linear regression model
550 with mixed effects was used to study allelic expression in *S. latifolia* for every SNP position. In
551 order to measure the changes in *S. latifolia* expression due to sex chromosomes evolution, the
552 outgroup *S. vulgaris* that does not have sex chromosomes was used as a reference in the model
553 (see Materials and Methods for details). The framework provided estimates for the normalised
554 difference between the effect of paternal and maternal origin of alleles in interaction with the
555 contig status (autosomal or sex-linked with various levels of Y degeneration), while accounting
556 for inter-contig and inter-individual variability. See Supplementary Figure S2 legend for sample
557 sizes for the different contig categories and statistical significance symbols. Results show that Y
558 degeneration is linked to a significant expression difference between the paternal and maternal
559 alleles in females, which is not observed in autosomal and non-degenerated sex-linked contigs.



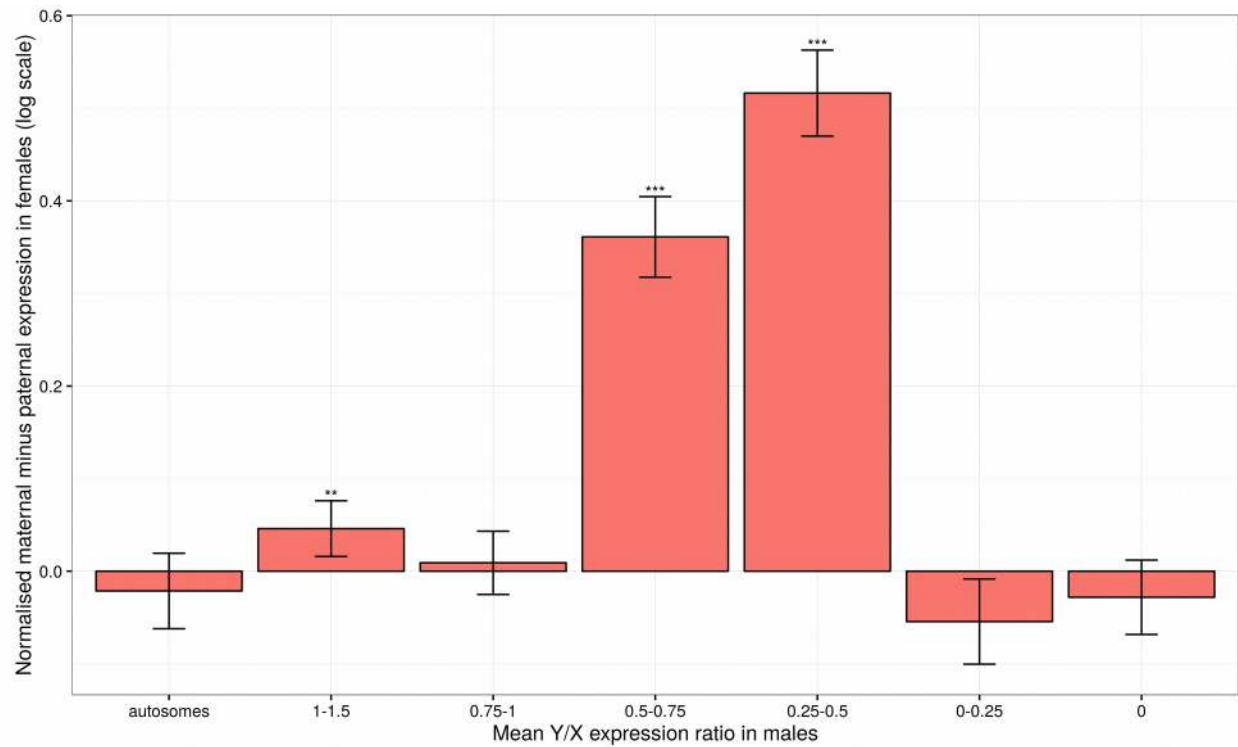
560 **Supplementary Figure S9:** Normalised expression difference between the maternal and paternal
 561 allele in *S. latifolia* females in autosomal and sex-linked contigs for the **flower bud** tissue. See
 562 supplementary Figure S8 for legend and Supplementary Figure S3 for sample sizes for the
 563 different contig categories.



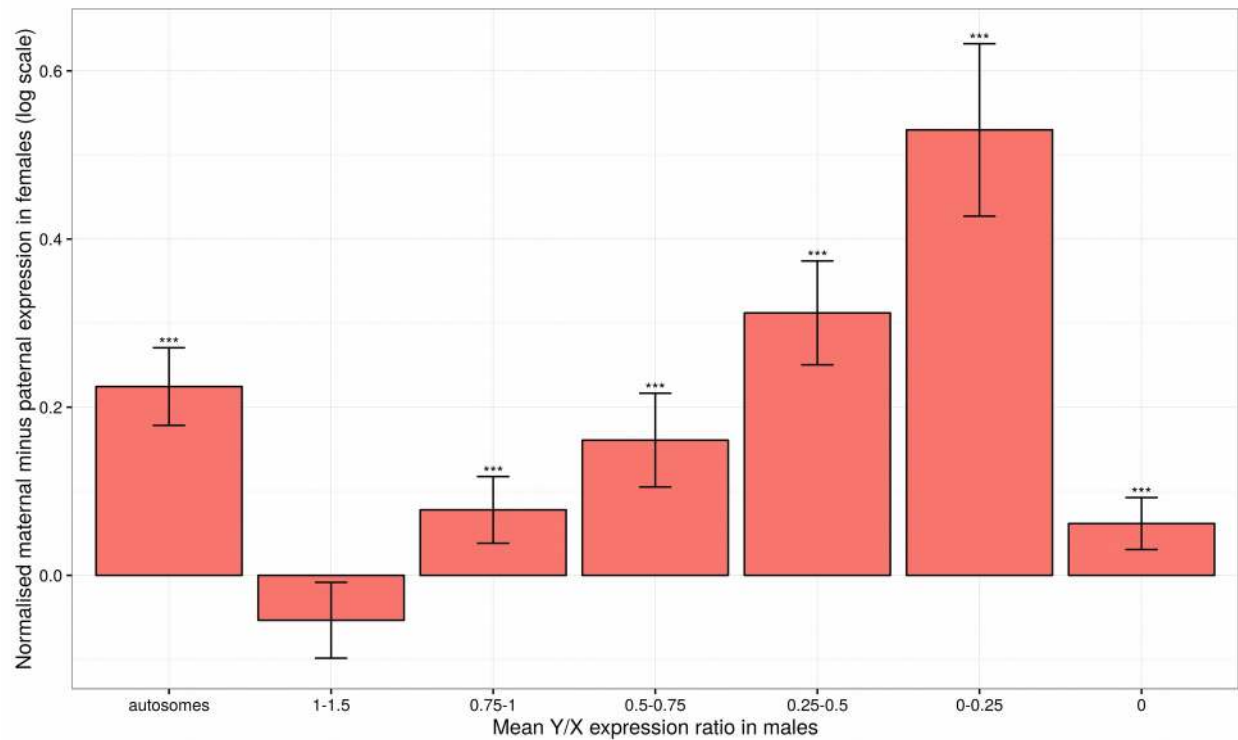
564 **Supplementary Figure S10:** Normalised expression difference between the maternal and
565 paternal allele in *S. latifolia* females in autosomal and sex-linked contigs for the **leaf** tissue. See
566 supplementary Figure S8 for legend and Supplementary Figure S4 for sample sizes for the
567 different contig categories.



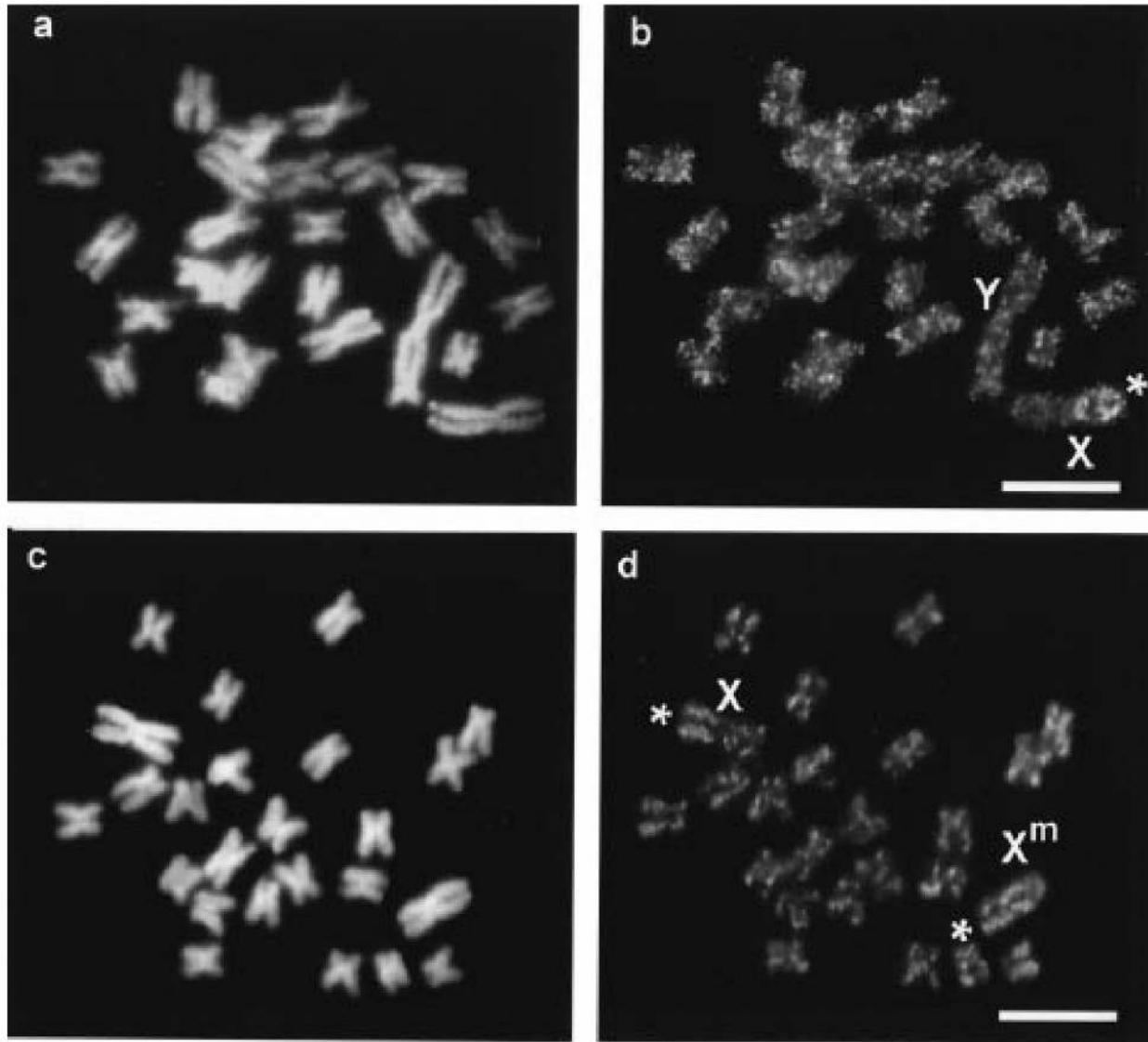
568 **Supplementary Figure S11:** Normalised expression difference between the maternal and
 569 paternal allele in *S. latifolia* females in autosomal and sex-linked **validated** contigs for the
 570 **seedling** tissue. See supplementary Figure S8 for legend and Supplementary Figure S5 for
 571 sample sizes for the different contig categories.



572 **Supplementary Figure S12:** Normalised expression difference between the maternal and
 573 paternal allele in *S. latifolia* females in autosomal and sex-linked **validated** contigs for the
 574 **flower bud** tissue. See supplementary Figure S8 for legend and Supplementary Figure S6 for
 575 sample sizes for the different contig categories.



576 **Supplementary Figure S13:** Normalised expression difference between the maternal and
 577 paternal allele in *S. latifolia* females in autosomal and sex-linked **validated** contigs for the **leaf**
 578 tissue. See supplementary Figure S8 for legend and Supplementary Figure S7 for sample sizes
 579 for the different contig categories.



580 **Supplementary Figure S14:** Original DNA methylation staining results from Siroky et al
 581 1998³⁴. **(a)** Male metaphase chromosomes stained with PI. **(b)** FITC-anti-5-mC signals on the
 582 same chromosomes. The hypomethylated shorter X arm is marked by an asterisk; The X and Y
 583 chromosomes are indicated. **(c)** Female metaphase chromosomes stained with PI. **(d)** FITC-anti-
 584 5-mC signals on the same chromosomes. Shorter arms of the Xs are indicated by asterisks. The
 585 hypermethylated X chromosome is marked as X^m. Bars = 5μm.

586 **Supplementary Tables**

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

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

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

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