



**HAL**  
open science

## **A derived ZW chromosome system in *Amborella trichopoda*, representing the sister lineage to all other extant flowering plants**

Jos Käfer, Adam Bewick, Amélie Andres-robin, Garance Lapetoule, Alex Harkess, José Caius, Gabriel Marais, Bruno Fogliani, Gildas Gâteblé, Paula Ralph, et al.

### ► To cite this version:

Jos Käfer, Adam Bewick, Amélie Andres-robin, Garance Lapetoule, Alex Harkess, et al.. A derived ZW chromosome system in *Amborella trichopoda*, representing the sister lineage to all other extant flowering plants. *New Phytologist*, 2022, 233, pp.1636-1642. 10.1111/nph.17662 . hal-03370272

**HAL Id: hal-03370272**

**<https://hal.science/hal-03370272v1>**

Submitted on 8 Oct 2021

**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.

1            **A derived ZW chromosome system in *Amborella***  
2            ***trichopoda*, representing the sister lineage to all other**  
3            **extant flowering plants**

4            Jos Käfer<sup>1</sup>, Adam Bewick<sup>2,3</sup>, Amélie Andres-Robin<sup>4</sup>, Garance Lapetoule<sup>1</sup>, Alex Harkess<sup>5,6</sup>,  
5            José Caius<sup>7</sup>, Bruno Fogliani<sup>8,9</sup>, Gildas Gâteblé<sup>8</sup>, Paula Ralph<sup>10</sup>, Claude W. dePamphilis<sup>10</sup>,  
6            Franck Picard<sup>1</sup>, Charlie Scutt<sup>4,\*</sup>, Gabriel AB Marais<sup>1,11,\*</sup>, James Leebens-Mack<sup>2,\*</sup>

7            <sup>1</sup> Université de Lyon; Université Lyon 1; CNRS; UMR 5558; Laboratoire de Biométrie et Biologie Evolutive;  
8            F-69622, Villeurbanne, France; <sup>2</sup> University of Georgia, Department of Plant Biology, Athens, GA 30602-7271,  
9            USA; <sup>3</sup> Bayer Crop Science, Chesterfield, MO 63017, USA; <sup>4</sup> Ecole Normale Supérieure de Lyon, CNRS, Labo-  
10            ratoire Reproduction et Développement des plantes UMR 5667, F-69364 Lyon, France; <sup>5</sup> Department of Crop,  
11            Soil, and Environmental Sciences, Auburn University, Auburn, AL 36849; <sup>6</sup> HudsonAlpha Institute for Biotech-  
12            nology, Huntsville, AL, 35806; <sup>7</sup> Univ. Paris-Saclay, Institute of Plant Sciences, Plateforme transcriptOmique  
13            de l'IP2S (POPS), F-91190 Gif-sur-Yvette, France; <sup>8</sup> Institut Agronomique néo-Calédonien (IAC), BP 73 Port  
14            Laguerre, 98890 Païta, New Caledonia; <sup>9</sup> Institute of Exact and Applied Sciences (ISEA), Université de la  
15            Nouvelle-Calédonie, BP R4, Nouméa Cedex, 98851 New Caledonia; <sup>10</sup> Department of Biology and Huck In-  
16            stitutes of the Life Sciences, The Pennsylvania State University, University Park, PA 16802, USA; <sup>11</sup> LEAF-  
17            Linking Landscape, Environment, Agriculture and Food, Instituto Superior de Agronomia, Universidade de  
18            Lisboa, 1349-017 Lisbon, Portugal

19            \* These authors contributed equally to this work.

20            **Authors for correspondence:**

21            Jos Käfer, Tel.: +33-472447932, Email: jos.kafer@univ-lyon1.fr;

22            James Leebens-Mack, Tel.: +1-706-583-5573, Email: jleebensmack@uga.edu

23            • Main text total number of words (including figure legends, excluding summary, references, acknowledge-  
24            ments and supporting information): 2924

25            • Number of words in the Introduction: 429

26            • Number of words in the Materials and methods: 467

27            • Number of words in the Results: 1247

28            • Number of words in the Discussion: 530

29            • Number of figures in the main text: 2

30            • Supporting information documents: 4

## 29 Abstract

- 30 • The genetic basis and evolution of sex determination in dioecious plants is emerging as an active area  
31 of research with exciting advances in genome sequencing and analysis technologies. As the sole species  
32 within the sister lineage to all other extant flowering plants, *Amborella trichopoda* is an important model  
33 for understanding the evolution and development of flowers. Plants typically produce only male or female  
34 flowers, but sex determination mechanisms are unknown for the species.
- 35 • Sequence data derived from plants of natural origin and an F1 mapping population were used to identify  
36 sex-linked genes and the non-recombining region.
- 37 • *A. trichopoda* has a ZW sex determination system. Analysis of genes in a 4 Mb non-recombining sex-  
38 determination region reveals recent divergence of Z and W gametologs, and few Z- and W-specific genes.
- 39 • The sex chromosomes of *A. trichopoda* evolved less than 16.5 My ago, long after the divergence of the  
40 extant angiosperms.

## 41 Introduction

42 Dioecy (separate male and female plants) is one of the many sexual systems exhibited by flowering plant species.  
43 Dioecy has evolved repeatedly in angiosperms and *c.* 6% of all species (i.e. more than 15,000 species) are dioe-  
44 cious (Renner, 2014). Though detailed knowledge of the molecular mechanisms of sex determination in dioecious  
45 plants is rather scarce, purely environmental sex determination seems to be rare (Renner, 2014; Charlesworth,  
46 2016), and most dioecious species are therefore thought to have genetic sex determination. However, only *c.* 50  
47 species with homomorphic or heteromorphic sex chromosomes have been reported, either using light microscopy  
48 or classical genetic inheritance studies, and more recently with genomic approaches (Ming et al., 2011; Muyle  
49 et al., 2017; Renner and Müller, 2021), and the exact sex-determining genes have been identified in even fewer  
50 species (Akagi et al., 2014; Harkess et al., 2017; Akagi et al., 2019; Harkess et al., 2020; Müller et al., 2020).  
51 Because dioecy evolved many times independently, different dioecious clades could have different mechanisms  
52 of sex determination, so these need to be identified in each clade separately. Very often, the sex chromosomes  
53 are only mildly diverged and homomorphic (with some notable exceptions; cf Renner and Müller, 2021).

54 *Amborella trichopoda* is an understory shrub endemic to New Caledonia and the sole extant species on the  
55 sister lineage to the clade including all other angiosperms, which form the most diverse division of plants. It  
56 seems clearly established that the most recent common ancestor of all living angiosperms had bisexual flowers,  
57 and that the transition to dioecy in the *A. trichopoda* lineage occurred after it diverged from the last common  
58 ancestor of all other flowering plants (Endress and Doyle, 2009; Sauquet et al., 2017). Despite the independent  
59 origin of dioecy within the *A. trichopoda* lineage, unraveling the mechanism of sex determination in this species  
60 could help to shed light on the evolution and development of unisexual flowers in other angiosperms and the  
61 evolution of sex chromosomes more generally.

62 *A. trichopoda* plants grown from seeds collected in the wild exhibit a sex ratio of nearly 50:50 (Anger et al.,  
63 2017) making genetic sex determination in the species plausible. However, there are also reports of individual  
64 plants exhibiting shifts sex over time (Buzgo et al., 2004; Anger et al., 2017). In fact, the reference genome  
65 sequence (Amborella Genome Project, 2013) was derived from a genotype (Santa Cruz Arboretum accession  
66 no. 75.3) that was first documented as male but cuttings produced female flowers (Buzgo et al., 2004). Here we

67 test for a genetic basis for sex determination in *A. trichopoda*, and go on to characterize the sex determination  
68 region in the *Amborella* genome.

## 69 Materials and methods

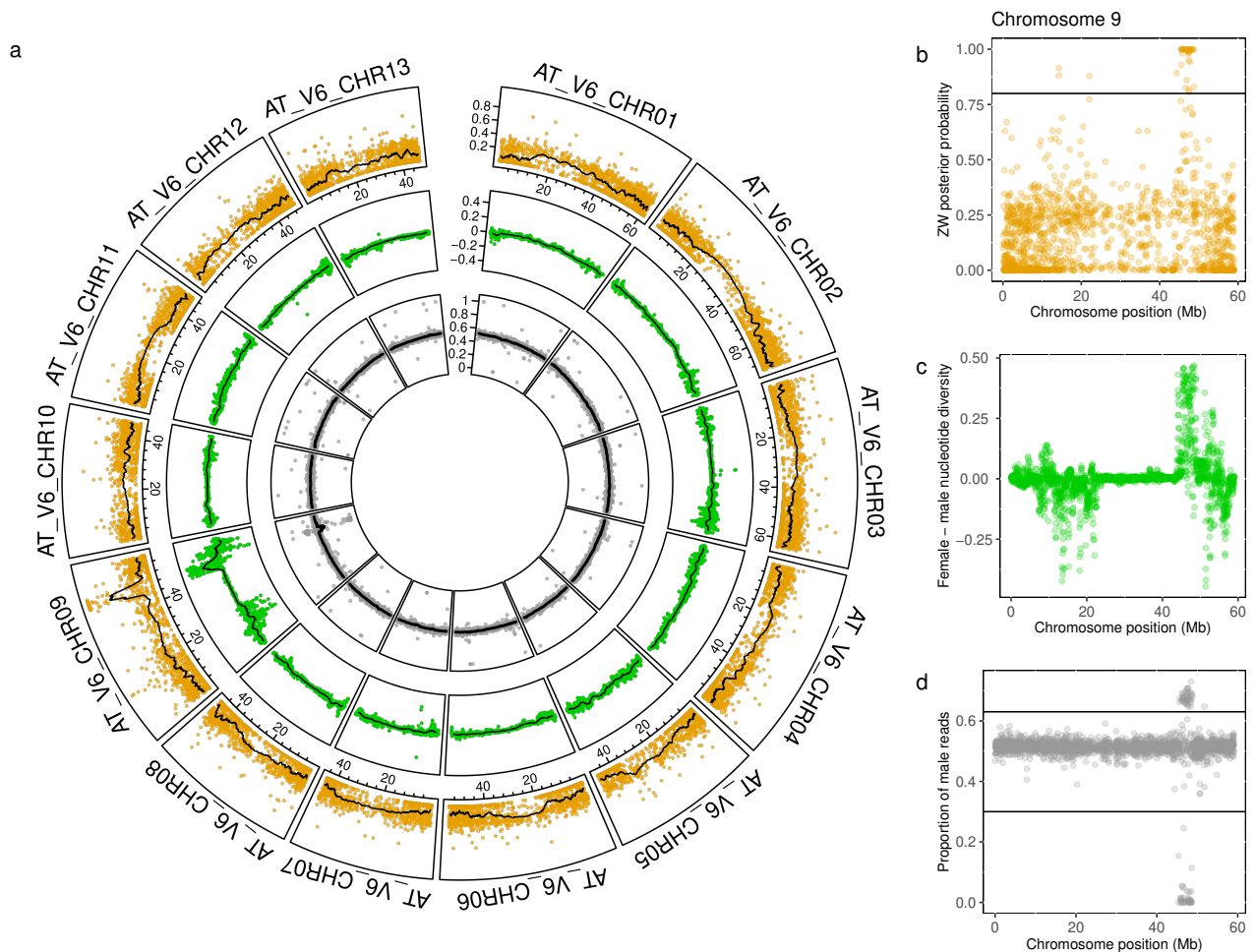
70 For the initial inference of sex-linkage, we obtained RNA-seq data of 20 individuals (10 females, 10 males;  
71 see Supporting Information Table S1 and Supporting Information, Materials and Methods S1) grown from  
72 seeds collected in the wild (hereafter referred to as the “wild population”. Sequence reads were aligned with  
73 STAR (Dobin et al., 2013) to the draft chromosomal assembly of the *A. trichopoda* genome (V6.1, available on  
74 CoGe, genome ID 50948), consisting of 13 pseudomolecules containing  $\sim 95\%$  of the genomic scaffolds. Single  
75 Nucleotide Polymorphisms (SNPs) were inferred with reads2snp (Gayral et al., 2013) and analyzed with SDpop,  
76 a probabilistic framework based on the modeling of allele and genotype frequencies in populations (Käfer et al.,  
77 2021). These RNAseq data from flower buds (between stages 5 and 6 according to Buzgo et al. 2004) were also  
78 used to infer sex-biased expression with edgeR (Robinson et al., 2010).

79 We furthermore used shotgun DNA seq data from an F1 mapping population: parents along with 16 female  
80 and 17 male F1 offspring; these data will be referred to as the “F1 mapping population”, see Supporting  
81 Information Table S2. These data were also aligned to the V6.1 draft chromosomal assembly (using BWA; Li  
82 and Durbin 2009). Genotypes were inferred with freebayes (Garrison and Marth, 2012). Putative W- and Z-  
83 specific regions in the assembly were identified using coverage differences in genomic windows of 20 Kpb between  
84 females and males. These were used to construct masks of the W- and Z-specific sequences for remapping the  
85 sequencing data on the W- and Z-masked genomes (see Supporting Information, Materials and Methods S1).

86 Using the mappings of both datasets on the masked genome assemblies, we used SDpop and SEX-DETECTOR++  
87 (Muyle et al., 2016; Badouin et al., 2020) to infer the Z- and W-haplotypes of the detected gametologous genes.  
88 Synonymous divergence ( $dS$ ) between these haplotypes was calculated with codeML (Yang, 2007). Because  
89 of high sampling variance in gene-wise divergence values (Takahata and Nei, 1985), average  $dS$  values per  
90 genes using both datasets were calculated. To estimate the time since recombination suppression, we used the  
91 maximum average  $dS$  value in two ways. First, based on population genetics results, the divergence between  
92 two sequences that do not recombine is  $2\mu t + \pi$  (Nei and Li, 1979; Takahata and Nei, 1985), where  $\pi$  is the  
93 nucleotide diversity,  $t$  the time since separation and  $\mu$  the per-base mutation rate, which is again linked to  $\pi$   
94 and the effective population size  $N_e$  by  $\pi = 4N_e\mu$ . We used the estimates of population genetic parameters and  
95 generation time obtained by the Amborella Genome Project (2013). Second, we compared the obtained  $dS$  value  
96 to the estimated divergence of orthologous genes between *Amborella* and *Nuphar* by the One Thousand Plant  
97 Transcriptomes Initiative (2019), and scaled this difference using estimates of the crown age of the angiosperms  
98 (Salomo et al., 2017; Sauquet and Magallon, 2018).

## 99 Results

100 We applied SDpop (Käfer et al., 2021) to the wild population RNA-seq data, and found Z/W segregation  
101 of sex-linked SNPs. SDpop inferred 49 ZW gametologs and nine putatively Z-hemizygous genes (Supporting  
102 Information, Results S1 and Supporting Information Table S3), which represents 0.28% of all transcribed genes.  
103 Of the ZW gametologs, 33 (67%) were located in a single region of *c.* 4 Mb on chromosome 9 of the reference  
104 genome assembly (Fig. 1). Twelve of the genes inferred as ZW occurred on scaffolds that were not assigned



**Fig. 1** Sex-linkage inferred from sequencing data in the whole *Amborella trichopoda* genome (a) and on chromosome 9 (b, c, d). Orange (panel a, outer graphs; panel b): SDpop posterior scores for the ZW segregation type per gene. The black line in (b) indicates the posterior score threshold (0.8) for classification as sex-linked. Green: (panel a, middle graphs; panel c): difference in average nucleotide diversity per gene between females and males. Grey (panel a, inner graphs; panel d): proportion of reads from males for each gene. The black lines in (d) indicate the thresholds (0.30 and 0.63) beyond which the genes were considered putatively sex-linked. In (a), the black lines represent the average in a sliding window of 50 genes.

105 to 13 chromosomal pseudomolecules, suggesting that they may reside on alternative haplotypes for the non-  
 106 recombining region of chromosome 9 in the primary chromosomal assembly.

107 Using the 33 genotyped and phenotyped progeny and their parents from the F1 mapping population, we  
 108 first assessed whether 671 female-specific SNP variants identified in the wild population transcripts were only  
 109 present in the seed parent and female progeny. Indeed, males in the mapping population were found to have no  
 110 more than five of the putatively female-specific variants, whereas females contained at least 465 of the female-  
 111 specific SNP variants identified in the wild-collected population (Supporting Information Table S2). Thus, the  
 112 divergence between Z and W genes is conserved among *A. trichopoda* populations.

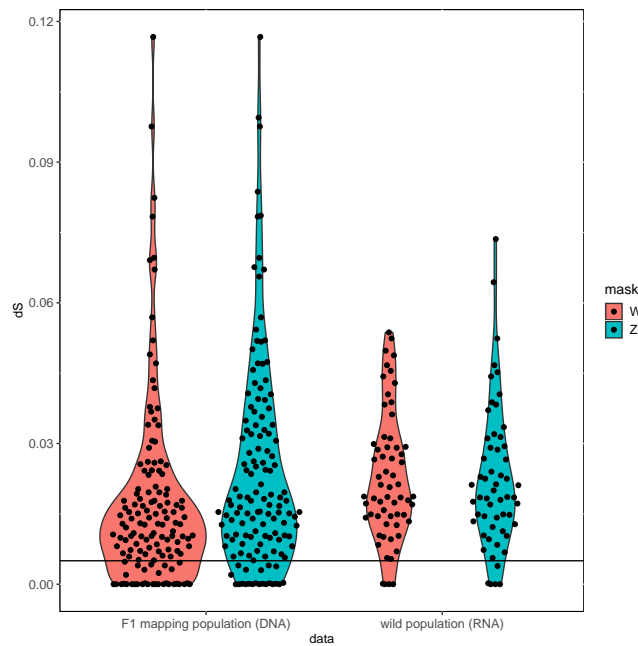
113 We then used the mapping population to estimate the frequency of the sites that were heterozygous in  
 114 females and homozygous in males, and found a clear excess of female-heterozygous SNPs in the sex-determining  
 115 region identified by SDpop (Fig. 1). Furthermore, the male-versus-female coverage ratio showed a remarkable  
 116 pattern in this region, with some genes having equal coverage for both sexes, others no male coverage, and

117 yet others for which two thirds of the reads were from males. As the existence of sex chromosomes (or any  
118 structural heterozygosity) was not considered in the haploid assembly process, the assembled sex determination  
119 region placed on chromosome 9 is a chimeric mix of fragments derived from the W and Z chromosomes likely  
120 together with regions of consensus sequence for the W and Z haplotypes. Accordingly, divergent regions of the  
121 W and Z haplotypes not included in the chromosomal pseudomolecules were assembled in unassigned scaffolds.  
122 For the inference of divergence and sex-biased gene expression, we thus obtained two masks for the assembly  
123 based on the coverage, one discarding W-specific regions, and the other Z-specific regions of the chromosome 9  
124 assembly and scaffolds that were not included in the chromosomal pseudomolecules (see Supporting Information,  
125 Results S1; Supporting Information Fig. S1 and Supporting Information Fig. S2; and Data Files S1 and S2).

126 Realigning the wild population RNAseq data on these masked assemblies allowed detection of slightly more  
127 sex-linked genes using SDpop (Supporting Information Table S3). As these RNAseq data only contain expressed  
128 genes (*c.* 50% of all genes), we used the DNA-reseq data from the mapping population to expand our inference  
129 of putatively sex-linked genes. First, we analysed the data with SEX-DETECTOR (Muyle et al., 2016; Badouin  
130 et al., 2020) that assigns genes as sex-linked or not based on the transmission of alleles from parents to progeny.  
131 When the non-recombining region is small, as is the case here, many genes are in linkage disequilibrium with this  
132 region due to the limited number of crossovers in one generation, and SEX-DETECTOR overestimates the size of  
133 the non-recombining sex-determination region (Supporting Information Fig. S3). We thus used the divergence  
134 between the predicted Z and W haplotypes as a second criterion for inference of sex-linkage: genes with  $dS$   
135 values higher than 0.005 (i.e., 2.5 times higher than the genome-wide heterozygosity) were considered as potential  
136 ZW gametologs. Second, we used the coverage data from the mapping-population sequencing: we considered  
137 as potentially sex-linked the genes that Fig. 1d and Supporting Information Fig. S2). Using these criteria,  
138 we identified 211 potentially sex-linked genes, of which 140 are on chromosome 9 and 69 on the unassigned  
139 scaffolds (Supplementary Data File 3). Among these, stringent analyses of the wild-population and F1 mapping  
140 populations data implicated four gene models as W-specific (AmTr\_v6.0\_c9.20670.1, AmTr\_v6.0\_c9.20870.1,  
141 and AmTr\_v6.0\_s3897.A.10.1) and three as Z-specific (AmTr\_v6.0\_c9.20390.1, AmTr\_v6.0\_c9.20460.1, and  
142 AmTr\_v6.0\_s2752.A.10.1; Data File S3). These assignments are conservative and it is likely that more W-  
143 and Z-specific genes may be identified in the future as high-fidelity long read data are generated and used to  
144 construct a fully phased diploid assembly.

145 We calculated the synonymous divergence ( $dS$ ) using the mapping on the two masked assemblies, and using  
146 the RNAseq as well as the DNA-reseq data (Fig. 2). As the sampling variance of these values is predictably  
147 high, we used the average of the two sequence datasets as our best estimate for the  $dS$  value of each gene  
148 (cf Takahata and Nei, 1985). The maximum  $dS$  value is between 0.05 and 0.06, depending on the masking  
149 of the reference. We calculated the time since recombination suppression based on these values in two ways,  
150 either using a population genetics-based (i.e., comparing gametolog divergence to the average level of nucleotide  
151 polymorphism) or phylogenetically based (i.e. comparing gametolog divergence to divergence estimates with  
152 other species; see Supporting Information, Results S1). The phylogeny-based estimate yields a wider range of  
153 divergence-time estimates (3.2 to 16.5 million years ago) than the population genetics-based estimate (7.2 to 8.7  
154 million years ago), but both estimates rely heavily on their respective underlying assumptions (e.g. molecular  
155 clock, generation time, past population size) which cannot be directly inferred nor tested.

156 Using our RNAseq data (from the wild population), we quantified sex-related expression biases. Whereas the  
157 number of female-biased genes is approximately equal to the number of male-biased genes genome-wide, male-



**Fig. 2** Distribution of synonymous divergence ( $dS$ ) between the Z and W copies of gametologous genes of *Amborella trichopoda*. For each dataset (mapping population DNA-reseq or wild population RNAseq), Z and W haplotypes were obtained after mapping on the “W masked” (red) and “Z masked” (blue) genomes. The areas of the “violins” are scaled according to the number of gametolog pairs (points). The horizontal black line indicates the threshold (0.005) above which genes were considered as sex-linked in the mapping population data.

158 biased genes seem to be over-represented in the non-recombining region (Supporting Information Table S4;  
 159 Supporting Information Fig. S4). Of the sex-linked female-biased genes, only two almost completely lacked  
 160 male expression (AmTr\_v6.0\_c9.19430.1 and AmTr\_v6.0\_s1359.A.20.1; Supporting Information, Results S1).  
 161 These genes could have a role in female sex determination, but their function is currently not known.

162 The mapping population data showed another intriguing pattern on chromosome 9: male offspring were more  
 163 often heterozygous than female offspring in a region from *c.* 10 to *c.* 20 Mb (Fig. 1). Upon closer inspection, it  
 164 appears that this occurs particularly at SNP positions for which the mother, but not the father, was heterozygous  
 165 (Supporting Information Fig. S5). This pattern is not typical of pseudoautosomal transmission in regions closely  
 166 linked to the non-recombining region, nor can it be explained by allele loss (e.g. through lethality) or by increased  
 167 heterozygosity in the mother (Supporting Information, Results S1; Supporting Information Fig. S6). An obligate  
 168 recombination event between the left and the right arms of chromosome 9 during meiosis in the mother would  
 169 be able to cause such a pattern, although this also would imply that the less frequent (minor) alleles all occur on  
 170 the same copy of chromosome 9. Although a weak deviation from purely autosomal allele-genotype equilibria  
 171 was also detected in the wild population (Fig. 1), more data (e.g. using different parents for an F1) are needed  
 172 before a precise mechanistic hypothesis for this pattern can be formulated and tested.

## 173 Discussion

174 We here show that *A. trichopoda* has sex chromosomes of the ZW type, with female heterogamy. Based on the  
 175 current haploid genome assembly, the non-recombining region is estimated to comprise only *c.* 4 Mb long of  
 176 the 59 Mb sex chromosome. The maximum synonymous divergence of *c.* 5 to 6% indicates that recombination

177 suppression occurred rather recently, at least much more recently than the divergence of *Amborella* and other  
178 living angiosperms. Although it might be possible that the first genes for which recombination was suppressed  
179 are not expressed or have been lost, the fact that Z and W haplotype sequences were collapsed in the current  
180 assembly and that we were not able to unambiguously identify more than a handful of Z- or W- specific genes  
181 also suggests that the *A. trichopoda* sex chromosomes are rather young.

182 Sex change is well documented in *A. trichopoda*, in wild as well as in cultivated plants (Buzgo et al., 2004;  
183 Grosse-Veldmann et al., 2011; Anger et al., 2017). Our results show that the individual used for reference genome  
184 sequencing (Amborella Genome Project, 2013) is genetically female with a Z and W gametologs, although it had  
185 produced male flowers in the past. Sex changes are not uncommon in dioecious angiosperms, even in those with  
186 clearly heteromorphic sex chromosomes (Ainsworth, 2000). The identification of the non-recombining region  
187 should allow researchers to determine the genetic sex for comparison with the phenotypic sex, and help gain  
188 insight in the genes involved in sex determination.

189 Most plant sex chromosomes described so far are of the XY type (Ming et al., 2011; Renner and Müller,  
190 2021), and this observation is considered by some as support for the hypothesis that dioecy evolves from gyn-  
191 odioecy through the occurrence of a recessive female-determining gene and a dominant male-determining gene  
192 (Charlesworth and Charlesworth, 1978). However, few sex determination mechanisms have been described at  
193 the gene level (cf Renner and Müller, 2021), and there seem to be many different developmental and evolutionary  
194 pathways to the production of unisexual flowers (Diggle et al., 2011; Golenberg and West, 2013).

195 Our inference that the ZW sex determination system is a recently derived trait in *A. trichopoda* underscores  
196 the fact that the species is not ancestral nor “basal”. The *Amborella* lineage and the ancestral lineage of  
197 all other flowering plants diverged from their most recent common ancestor (MRCA) in the early Cretaceous  
198 or even late Jurassic (Magallón et al., 2015; Silvestro et al., 2021). Thus, the biology of extant *Amborella*  
199 populations has been molded by more than 130 million years of evolution, since the early diversification of the  
200 extant flowering plants. However, *A. trichopoda* has retained some characteristics that are hypothesized to be  
201 shared with the MCRA of living angiosperms (e.g. free floral organs, including an undifferentiated perianth of  
202 tepals, and carpels that contain an apical, secretion-filled canal for pollen-tube growth; Endress and Igersheim  
203 2000; Sauquet et al. 2017), and no genome duplication has occurred in the *Amborella* lineage since diverged  
204 from the other angiosperms (Amborella Genome Project, 2013). Thus, further study of the genes involved in  
205 sex determination in *A. trichopoda* might shed light on the regulatory pathways that underlie development of  
206 unisexual flowers in other angiosperms.

## 207 **Acknowledgements**

208 We thank the Environmental Service of the Northern Province of New Caledonia for permission for seed col-  
209 lection and the staff of the IAC for horticultural assistance, John Bowers and the Amborella Genome Project  
210 contributors for work on the publicly available chromosomal assembly of the *Amborella trichopoda* genome used  
211 in this study, Nicolas Lartillot and Elise Lucotte for discussions, as well as Susanne Renner and two anonymous  
212 reviewers for comments on the manuscript. *Amborella* genome mapping and analyses were supported by U.S.  
213 NSF Plant Genome Research Program (grant 0922742 to CWdP, JLM and co-PIs). This work was also sup-  
214 ported by funding from the Agence Nationale de la Recherche (grant number ANR-14-CE19-0021-01 to GABM  
215 and CS).



## 216 Author contributions

217 Conceptualization: GABM, CS, JK, JLM, AB Methodology: JK, FP, GABM, JLM, AB Software: JK, FP,  
218 Formal analysis: JK, AB, GL, AB, JB, AH Investigation: JK, AB, AAR, GL, JC, CS, GABM, JLM Resources:  
219 BF, GG, PR, CWdP Writing - Original Draft: JK, GABM, JLM, CS Writing - Review & Editing: all authors,  
220 Visualization: JK, AB, Supervision: GABM, JLM, CS, JK Project administration: GABM, JLM, CS, Funding  
221 acquisition: GABM, CS, JLM, CWdP

## 222 Data availability

223 The data that support the findings of this study are openly available in SRA, bioprojects PRJNA748676 and  
224 PRJNA212863. We used the annotated V6.1 draft genome of *Amborella trichopoda* available on CoGe, genome  
225 ID 50948 ([genomevolution.org/coge/GenomeInfo.pl?gid=50948](http://genomevolution.org/coge/GenomeInfo.pl?gid=50948)).

## 226 References

- 227 **Ainsworth C. 2000.** Boys and girls come out to play: The molecular biology of dioecious plants. *Ann. Bot.*  
228 **86(2):** 211–221.
- 229 **Akagi T, Henry IM, Tao R, and Comai L. 2014.** Plant genetics. A Y-chromosome-encoded small RNA  
230 acts as a sex determinant in persimmons. *Science* **346(6209):** 646–650.
- 231 **Akagi T, Pilkington SM, Varkonyi-Gasic E, Henry IM, Sugano SS, Sonoda M, Firl A, McNeilage**  
232 **MA, Douglas MJ, Wang T, et al. 2019.** Two Y-chromosome-encoded genes determine sex in kiwifruit.  
233 *Nat. Plants* **5(8):** 801–809.
- 234 **Amborella Genome Project. 2013.** The *Amborella* genome and the evolution of flowering plants. *Science*  
235 **342(6165):** 1241089.
- 236 **Anger N, Fogliani B, Scutt CP, and Gâteblé G. 2017.** Dioecy in *Amborella trichopoda*: evidence for  
237 genetically based sex determination and its consequences for inferences of the breeding system in early  
238 angiosperms. *Ann. Bot.* **119(4):** 591–597.
- 239 **Badouin H, Velt A, Gindraud F, Flutre T, Dumas V, Vautrin S, Marande W, Corbi J, Sallet**  
240 **E, Ganofsky J, et al. 2020.** The wild grape genome sequence provides insights into the transition from  
241 dioecy to hermaphroditism during grape domestication. *Genome Biol.* **21(1):** 223.
- 242 **Buzgo M, Soltis PS, and Soltis DE. 2004.** Floral developmental morphology of *Amborella trichopoda*  
243 (*Amborellaceae*). *Int. J. Plant Sci.* **165(6):** 925–947.
- 244 **Charlesworth B and Charlesworth D. 1978.** Model for evolution of dioecy and gynodioecy. *Am. Nat.*  
245 **112(988):** 975–997.
- 246 **Charlesworth D. 2016.** Plant sex chromosomes. *Annu. Rev. Plant Biol.* **67:** 397–420.
- 247 **Diggle PK, Di Stilio VS, Gschwend AR, Golenberg EM, Moore RC, Russell JR, and Sinclair JP.**  
248 **2011.** Multiple developmental processes underlie sex differentiation in angiosperms. *Trends Genet.* **27(9):**  
249 368–376.
- 250 **Dobin A, Davis CA, Schlesinger F, Drenkow J, Zaleski C, Jha S, Batut P, Chaisson M, and**  
251 **Gingeras TR. 2013.** STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* **29(1):** 15–21.

252 **Endress PK and Doyle JA. 2009.** Reconstructing the ancestral angiosperm flower and its initial specializa-  
253 tions. *Am. J. Bot.* **96**(1): 22–66.

254 **Endress PK and Igersheim A. 2000.** Reproductive structures of the basal angiosperm *Amborella trichopoda*  
255 (*Amborellaceae*). *Int. J. Plant Sci.* **161**(6): S237–S248.

256 **Garrison E and Marth G. 2012.** Haplotype-based variant detection from short-read sequencing. *arXiv*.

257 **Gayral P, Melo-Ferreira J, Glémin S, Bierne N, Carneiro M, Nabholz B, Lourenco JM, Alves**  
258 **PC, Ballenghien M, Faivre N, et al. 2013.** Reference-free population genomics from next-generation  
259 transcriptome data and the vertebrate-invertebrate gap. *PLoS Genet.* **9**(4): e1003457.

260 **Golenberg EM and West NW. 2013.** Hormonal interactions and gene regulation can link monoecy and  
261 environmental plasticity to the evolution of dioecy in plants. *Am. J. Bot.* **100**(6): 1022–1037.

262 **Grosse-Veldmann B, Korotkova N, Reinken B, Lobin W, and Barthlott W. 2011.** *Amborella tri-*  
263 *chopoda* – Cultivation of the most ancestral angiosperm in botanic gardens. *Sibbaldia* **9**: 143–155.

264 **Harkess A, Huang K, Van der Hulst R, Tissen B, Caplan JL, Koppula A, Batish M, Meyers BC,**  
265 **and Leebens-Mack J. 2020.** Sex Determination by Two Y-Linked Genes in Garden Asparagus. *Plant Cell*  
266 **32**(6): 1790–1796.

267 **Harkess A, Zhou J, Xu C, Bowers JE, Van der Hulst R, Ayyampalayam S, Mercati F, Riccardi P,**  
268 **McKain MR, Kakrana A, et al. 2017.** The asparagus genome sheds light on the origin and evolution  
269 of a young Y chromosome. *Nat. Commun.* **8**(1): 1279.

270 **Käfer J, Lartillot N, Marais GAB, and Picard F. 2021.** Detecting sex-linked genes using genotyped  
271 individuals sampled in natural populations. *Genetics* **218**(2).

272 **Li H and Durbin R. 2009.** Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioin-*  
273 *formatics* **25**(14): 1754–1760.

274 **Magallón S, Gómez-Acevedo S, Sánchez-Reyes LL, and Hernández-Hernández T. 2015.** A meta-  
275 calibrated time-tree documents the early rise of flowering plant phylogenetic diversity. *New. Phytol.* **207**(2):  
276 437–453.

277 **Ming R, Bendahmane A, and Renner SS. 2011.** Sex chromosomes in land plants. *Annu. Rev. Plant Biol.*  
278 **62**: 485–514.

279 **Müller NA, Kersten B, Leite Montalvão AP, Mähler N, Bernhardsson C, Bräutigam K, Carracedo**  
280 **Lorenzo Z, Hoenicka H, Kumar V, Mader M, et al. 2020.** A single gene underlies the dynamic  
281 evolution of poplar sex determination. *Nat. Plants* **6**(6): 630–637.

282 **Muyle A, Käfer J, Zemp N, Mousset S, Picard F, and Marais GA. 2016.** SEX-DETECTOR: a probabilistic  
283 approach to study sex chromosomes in non-model organisms. *Genome Biol. Evol.* **8**(8): 2530–2543.

284 **Muyle A, Shearn R, and Marais GA. 2017.** The evolution of sex chromosomes and dosage compensation  
285 in plants. *Genome Biol. Evol.* **9**(3): 627–645.

286 **Nei M and Li WH. 1979.** Mathematical model for studying genetic variation in terms of restriction endonu-  
287 cleases. *Proc. Natl. Acad. Sci. U.S.A.* **76**(10): 5269–5273.

288 **One Thousand Plant Transcriptomes Initiative. 2019.** One thousand plant transcriptomes and the phy-  
289 logenomics of green plants. *Nature* **574**(7780): 679–685.

290 **Renner SS. 2014.** The relative and absolute frequencies of angiosperm sexual systems: dioecy, monoecy,  
291 gynodioecy, and an updated online database. *Am. J. Bot.* **101**(10): 1588–1596.

- 292 **Renner SS and Müller NA. 2021.** Plant sex chromosomes defy evolutionary models of expanding recombina-  
293 tion suppression and genetic degeneration. *Nat. Plants* **7**(4): 392–402.
- 294 **Robinson MD, McCarthy DJ, and Smyth GK. 2010.** edgeR: a Bioconductor package for differential  
295 expression analysis of digital gene expression data. *Bioinformatics* **26**(1): 139–140.
- 296 **Salomo K, Smith JF, Feild TS, Samain MS, Bond L, Davidson C, Zimmers J, Neinhuis C, and**  
297 **Wanke S. 2017.** The Emergence of Earliest Angiosperms may be Earlier than Fossil Evidence Indicates.  
298 *Syst. Bot.* **42**(4): 607–619.
- 299 **Sauquet H and Magallon S. 2018.** Key questions and challenges in angiosperm macroevolution. *New. Phytol.*  
300 **219**(4): 1170–1187.
- 301 **Sauquet H, Von Balthazar M, Magallon S, Doyle JA, Endress PK, Bailes EJ, Barroso de Morais**  
302 **E, Bull-Herenu K, Carrive L, Chartier M, et al. 2017.** The ancestral flower of angiosperms and its  
303 early diversification. *Nat. Commun.* **8**: 16047.
- 304 **Silvestro D, Bacon CD, Ding W, Zhang Q, Donoghue PCJ, Antonelli A, and Xing Y. 2021.** Fossil  
305 data support a pre-Cretaceous origin of flowering plants. *Nat. Ecol. Evol.* **5**(4): 449–457.
- 306 **Takahata N and Nei M. 1985.** Gene genealogy and variance of interpopulational nucleotide differences.  
307 *Genetics* **110**(2): 325–344.
- 308 **Yang Z. 2007.** PAML 4: phylogenetic analysis by maximum likelihood. *Mol. Biol. Evol.* **24**(8): 1586–1591.

## 309 **Supporting Information**

- 310 Additional supporting information may be found in the online version of this article.
- 311 **Materials and Methods S1** Additional information about the data and analyses of sex linkage.
- 312 **Table S1** RNAseq statistics for the twenty individuals grown from seeds collected in the wild.
- 313 **Table S2** Read depths for the individuals from the F1 family sequenced using DNA-reseq data.
- 314 **Results S1** Detailed description of the results of the analyses.
- 315 **Table S3** Summary of sex-linkage inference using the RNAseq data from a wild population.
- 316 **Fig. S1** Masking of putative W- and Z-specific regions in the non-recombining region of chromosome 9.
- 317 **Fig. S2** Re-seq coverage per gene using mapping on the unmasked and masked genome assemblies.
- 318 **Fig. S3** Results of the sex-linkage analysis in the mapping population data.
- 319 **Table S4** Sex-biased gene expression analysis.
- 320 **Fig. S4** Sex-biased gene expression inferred from RNAseq data from flower buds.
- 321 **Fig. S5** Distribution of heterozygosity between parents and offspring in the mapping population.
- 322 **Fig. S6** Allele transmission on chromosome 9.
- 323 **Data File S1** bed file indicating the putative W-specific regions for masking.
- 324 **Data File S2** bed file indicating the putative Z-specific regions for masking.
- 325 **Data File S3** Table (xlsx format) with all putatively sex-linked gene models and gene-level statistics.