

Rapid report

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

Authors for correspondence:

Jos Käfer

Email: jos.kafer@univ-lyon1.fr

James Leebens-Mack

Email: jleebensmack@uga.edu

Received: 24 June 2021

Accepted: 28 July 2021

Jos Käfer¹ , Adam Bewick^{2,3}, Amélie Andres-Robin⁴, Garance Lapetoule¹, Alex Harkess^{5,6}, José Cañas⁷, Bruno Fogliani^{8,9}, Gildas Gâteblé⁸ , Paula Ralph¹⁰, Claude W. dePamphilis¹⁰, Franck Picard¹, Charlie Scutt^{4*} , Gabriel A. B. Marais^{1,11*}  and James Leebens-Mack^{2*} 

¹Laboratoire de Biométrie et Biologie Evolutive, CNRS UMR 5558, Université de Lyon, Université Lyon 1, Villeurbanne F-69622, France; ²Department of Plant Biology, University of Georgia, Athens, GA 30602-7271, USA; ³Bayer Crop Science, Chesterfield, MO 63017, USA; ⁴Laboratoire Reproduction et Développement des plantes, UMR 5667, Ecole Normale Supérieure de Lyon, CNRS, Lyon F-69364, France; ⁵Department of Crop, Soil, and Environmental Sciences, Auburn University, Auburn, AL 36849, USA; ⁶HudsonAlpha Institute for Biotechnology, Huntsville, AL 35806, USA; ⁷Institute of Plant Sciences, Plateforme transcriptomique de l'IPS2 (POPS), Université de Paris-Saclay, Gif-sur-Yvette F-91190, France; ⁸Institut Agronomique néo-Calédonien (IAC), BP 73 Port Laguerre, Païta 98890, New Caledonia; ⁹Institute of Exact and Applied Sciences (ISEA), Université de la Nouvelle-Calédonie, BP R4, Nouméa Cedex 98851, New Caledonia; ¹⁰Department of Biology and Huck Institutes of the Life Sciences, The Pennsylvania State University, University Park, PA 16802, USA; ¹¹LEAF- Linking Landscape, Environment, Agriculture and Food, Instituto Superior de Agronomia, Universidade de Lisboa, Lisbon 1349-017, Portugal

Summary

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

New Phytologist (2021)
doi: 10.1111/nph.17662

Key words: *Amborella trichopoda*, angiosperms, dioecy, sex chromosomes, sex determination.

Introduction

Dioecy (separate male and female plants) is one of the many sexual systems exhibited by flowering plant species. Dioecy has evolved repeatedly in angiosperms and *c.* 6% of all species (i.e. more than 15 000 species) are dioecious (Renner, 2014). Though detailed knowledge of the molecular mechanisms of sex determination in

dioecious plants is rather scarce, purely environmental sex determination seems to be rare (Renner, 2014; Charlesworth, 2016), and most dioecious species are therefore thought to have genetic sex determination. However, only *c.* 50 species with homomorphic or heteromorphic sex chromosomes have been reported, either using light microscopy or classical genetic inheritance studies, and more recently with genomic approaches (Ming *et al.*, 2011; Muyle *et al.*, 2017; Renner & Müller, 2021), and the exact sex-determining genes have been identified in even

*These authors contributed equally to this work.

fewer species (Akagi *et al.*, 2014, 2019; Harkess *et al.*, 2017, 2020; Müller *et al.*, 2020). Because dioecy evolved many times independently, different dioecious clades could have different mechanisms of sex determination, so these need to be identified in each clade separately. Very often, the sex chromosomes are only mildly diverged and homomorphic (with some notable exceptions; cf. Renner & Müller, 2021).

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

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

Materials and Methods

For the initial inference of sex-linkage, we obtained RNA-sequencing (RNA-Seq) data of 20 individuals (10 females, 10 males; see Supporting Information Table S1; Methods S1) grown from seeds collected in the wild (hereafter referred to as the 'wild population'). Sequence reads were aligned with STAR (Dobin *et al.*, 2013) to the draft chromosomal assembly of the *A. trichopoda* genome (V6.1, available on CoGe, genome ID 50948), consisting of 13 pseudomolecules containing 95% of the genomic scaffolds. Single nucleotide polymorphisms (SNPs) were inferred with READS2SNP (Gayral *et al.*, 2013) and analyzed with SDPOP, a probabilistic framework based on the modeling of allele and genotype frequencies in populations (Käfer *et al.*, 2021). These RNA-Seq data from flower buds (between stages 5 and 6 according to Buzgo *et al.*, 2004) were also used to infer sex-biased expression with EDGER (Robinson *et al.*, 2010).

We furthermore used shotgun DNA sequencing data from an F1 mapping population: parents along with 16 female and 17 male F1 offspring; these data will be referred to as the 'F1 mapping population', see Table S2. These data were also aligned to the V6.1 draft chromosomal assembly (using BWA; Li & Durbin, 2009). Genotypes were inferred with freebayes (Garrison & Marth, 2012). Putative W- and Z-specific regions in the assembly were identified

using coverage differences in genomic windows of 20 Kbp between females and males. These were used to construct masks of the W- and Z-specific sequences for remapping the sequencing data on the W- and Z-masked genomes (see Methods S1).

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

Results

We applied SDPOP (Käfer *et al.*, 2021) to the wild population RNA-Seq data, and found Z/W segregation of sex-linked SNPs. SDPOP inferred 49 ZW gametologs and nine putatively Z-hemizygous genes (Notes S1; Table S3), which represents 0.28% of all transcribed genes. Of the ZW gametologs, 33 (67%) were located in a single region of c. 4 Mb on chromosome 9 of the reference genome assembly (Fig. 1). Twelve of the genes inferred as ZW occurred on scaffolds that were not assigned to 13 chromosomal pseudomolecules, suggesting that they may reside on alternative haplotypes for the nonrecombining region of chromosome 9 in the primary chromosomal assembly.

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

We then used the mapping population to estimate the frequency of the sites that were heterozygous in females and homozygous in males, and found a clear excess of female-heterozygous SNPs in the sex-determining region identified by SDPOP (Fig. 1). Furthermore, the male-vs-female coverage ratio showed a remarkable pattern in this region, with some genes having equal coverage for both sexes,

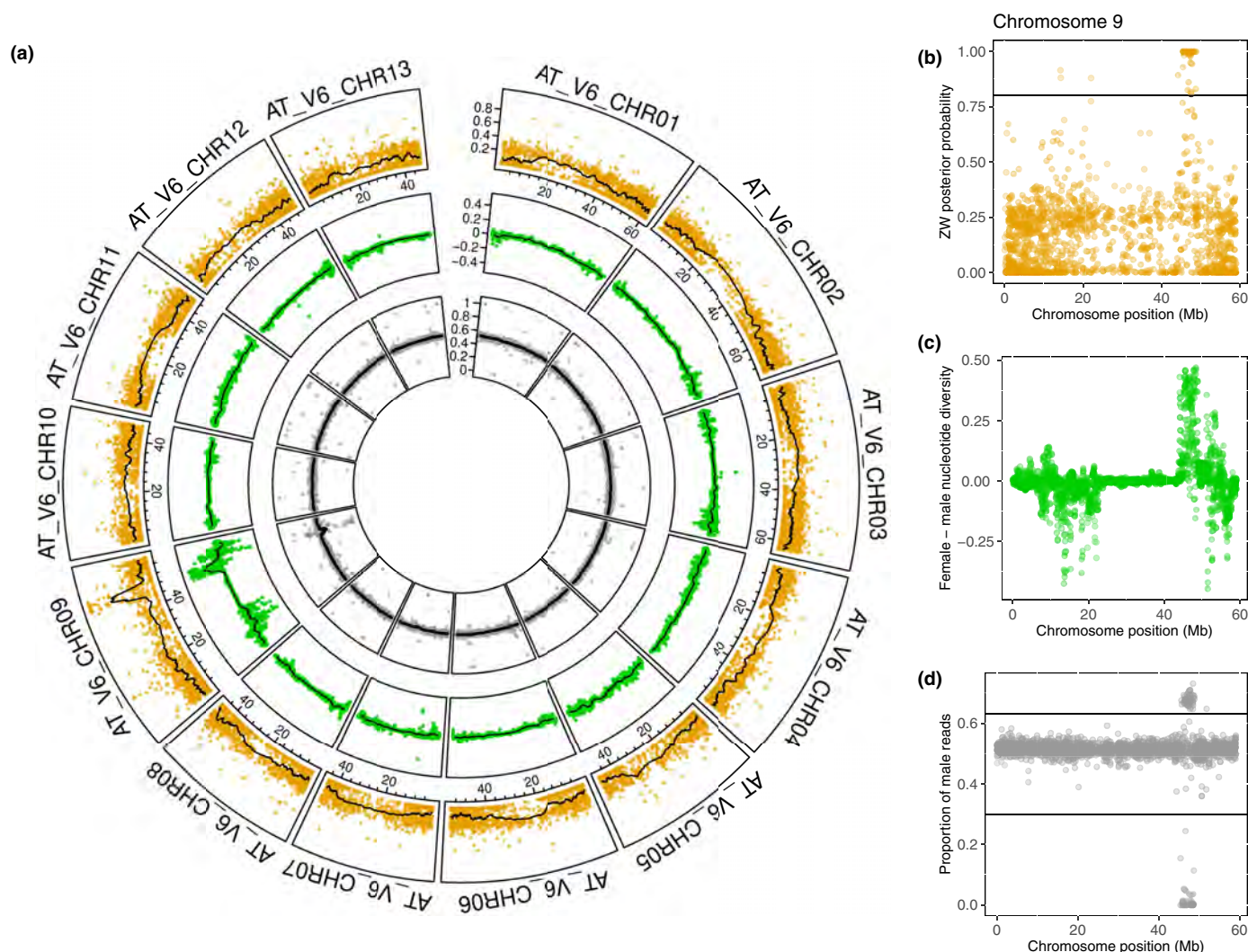


Fig. 1 Sex-linkage inferred from sequencing data in the whole *Amborella trichopoda* genome (a) and on chromosome 9 (b, c, d). Orange (a, outer graphs; panel b): SDPOP posterior scores for the ZW segregation type per gene. The black line in (b) indicates the posterior score threshold (0.8) for classification as sex-linked. Green: (a, middle graphs; c): difference in average nucleotide diversity per gene between females and males. Gray (a, inner graphs; 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.

others no male coverage, and yet others for which two-thirds of the reads were from males. As the existence of sex chromosomes (or any structural heterozygosity) was not considered in the haploid assembly process, the assembled sex determination region placed on chromosome 9 is a chimeric mix of fragments derived from the W and Z chromosomes likely together with regions of consensus sequence for the W and Z haplotypes. Accordingly, divergent regions of the W and Z haplotypes not included in the chromosomal pseudomolecules were assembled in unassigned scaffolds. For the inference of divergence and sex-biased gene expression, we thus obtained two masks for the assembly based on the coverage, one discarding W-specific regions, and the other Z-specific regions of the chromosome 9 assembly and scaffolds that were not included in the chromosomal pseudomolecules (see Notes S1; Figs S1, S2; Datasets S1, S2).

Realigining the wild population RNA-Seq data on these masked assemblies allowed detection of slightly more sex-linked genes using

SDPOP (Table S3). As these RNA-Seq data only contain expressed genes (*c.* 50% of all genes), we used the DNA-resequencing (DNA-Reseq) data from the mapping population to expand our inference of putatively sex-linked genes. First, we analysed the data with SEX-DETECTOR (Muyle *et al.*, 2016; Badouin *et al.*, 2020) that assigns genes as sex-linked or not based on the transmission of alleles from parents to progeny. When the nonrecombining region is small, as is the case here, many genes are in linkage disequilibrium with this region due to the limited number of crossovers in one generation, and SEX-DETECTOR overestimates the size of the nonrecombining sex-determination region (Fig. S3). We thus used the divergence between the predicted Z and W haplotypes as a second criterion for inference of sex-linkage: genes with dS values higher than 0.005 (*i.e.* 2.5 times higher than the genome-wide heterozygosity) were considered as potential ZW gametologs. Second, we used the coverage data from the mapping-population sequencing: we considered as potentially sex-linked the genes that had the lowest

and highest proportions of male reads (Fig. 1(d) and Fig. S2). Using these criteria, we identified 211 potentially sex-linked genes, of which 140 are on chromosome 9 and 69 on the unassigned scaffolds (Dataset S3). Among these, stringent analyses of the wild-population and F1 mapping populations data implicated four gene models as W-specific (AmTr_v6.0_c9.20670.1, AmTr_v6.0_c9.20870.1, AmTr_v6.0_s593.A.10.1, 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 AmTr_v6.0_s2752.A.10.1; Dataset S3). These assignments are conservative and it is likely that more W- and Z-specific genes may be identified in the future as high-fidelity long read data are generated and used to construct a fully phased diploid assembly.

We calculated the dS value using the mapping on the two masked assemblies, and using the RNA-Seq as well as the DNA-Reseq data (Fig. 2). As the sampling variance of these values is predictably high, we used the average of the two sequence datasets as our best estimate for the dS value of each gene (cf. Takahata & Nei, 1985). The maximum dS value is between 0.05 and 0.06, depending on the masking of the reference. We calculated the time since recombination suppression based on these values in two ways, either using a population genetics-based (i.e. comparing gametolog divergence to the average level of nucleotide polymorphism) or phylogenetically based (i.e. comparing gametolog divergence to divergence estimates with other species; see Notes S1). The

phylogeny-based estimate yields a wider range of divergence-time estimates (3.2 to 16.5 Myr ago) than the population genetics-based estimate (7.2 to 8.7 Myr ago), but both estimates rely heavily on their respective underlying assumptions (e.g. molecular clock, generation time, past population size) which cannot be directly inferred nor tested.

Using our RNA-Seq data (from the wild population), we quantified sex-related expression biases. Whereas the number of female-biased genes is approximately equal to the number of male-biased genes genome-wide, male-biased genes seem to be overrepresented in the nonrecombining region (Table S4; Fig. S4). Of the sex-linked female-biased genes, only two almost completely lacked male expression (AmTr_v6.0_c9.19430.1 and AmTr_v6.0_s1359.A.20.1; Notes S1). These genes could have a role in female sex determination, but their function is currently not known.

The mapping population data showed another intriguing pattern on chromosome 9: male offspring were more often heterozygous than female offspring in a region from *c.* 10 to *c.* 20 Mb (Fig. 1). Upon closer inspection, it appears that this occurs particularly at SNP positions for which the mother, but not the father, was heterozygous (Fig. S5). This pattern is not typical of pseudoautosomal transmission in regions closely linked to the nonrecombining region, nor can it be explained by allele loss (e.g. through lethality) or by increased heterozygosity in the mother (Notes S1; Fig. S6). An obligate recombination event between the

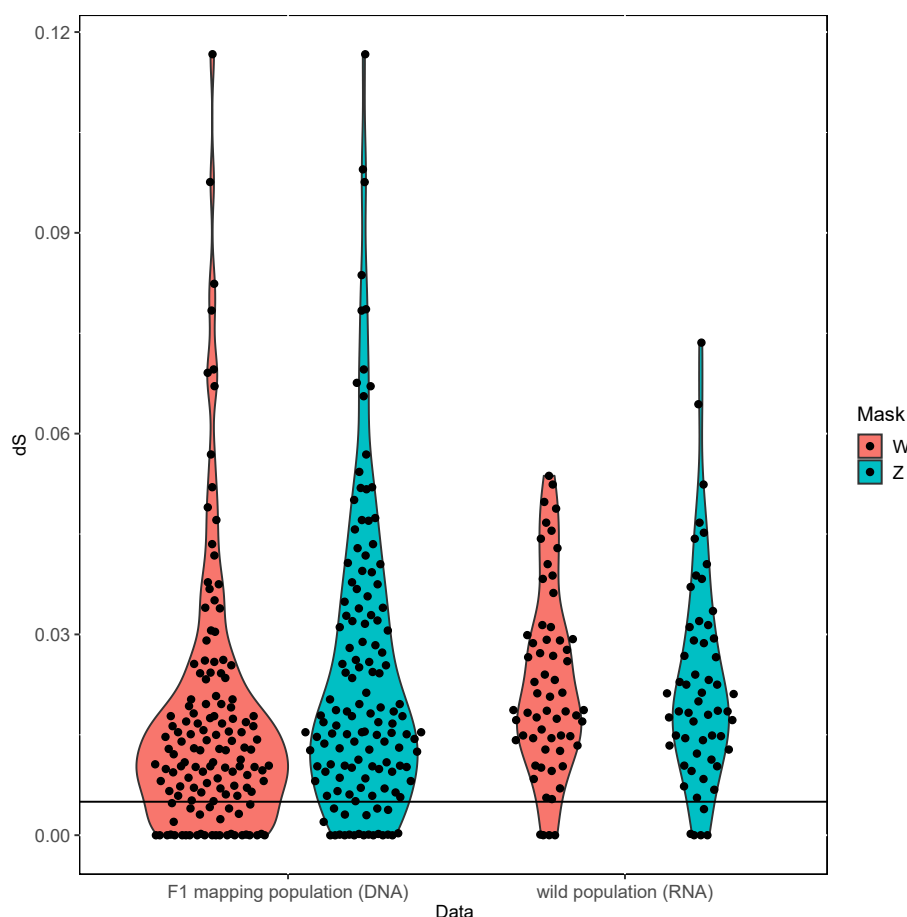


Fig. 2 Distribution of synonymous divergence (dS) between the Z and W copies of gametologous genes of *Amborella trichopoda*. For each dataset (mapping population DNA-Reseq or wild population RNA-Seq), Z and W haplotypes were obtained after mapping on the 'W masked' (red) and 'Z masked' (blue) genomes. 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.

left and the right arms of chromosome 9 during meiosis in the mother would be able to cause such a pattern, although this also would imply that the less frequent (minor) alleles all occur on the same copy of chromosome 9. Although a weak deviation from purely autosomal allele-genotype equilibria was also detected in the wild population (Fig. 1), more data (e.g. using different parents for an F1) are needed before a precise mechanistic hypothesis for this pattern can be formulated and tested.

Discussion

We here show that *A. trichopoda* has sex chromosomes of the ZW type, with female heterogamy. Based on the current haploid genome assembly, the nonrecombining region is estimated to comprise only *c.* 4 Mb of the 59 Mb sex chromosome. The maximum synonymous divergence of *c.* 5 to 6% indicates that recombination suppression occurred rather recently, at least much more recently than the divergence of *Amborella* and other living angiosperms. Although it might be possible that the first genes for which recombination was suppressed are not expressed or have been lost, the fact that Z and W haplotype sequences were collapsed in the current assembly and that we were not able to unambiguously identify more than a handful of Z- or W-specific genes also suggests that the *A. trichopoda* sex chromosomes are rather young.

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

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

Our inference that the ZW sex determination system is a recently derived trait in *A. trichopoda* underscores the fact that the species is not ancestral nor 'basal'. The *Amborella* lineage and the ancestral lineage of all other flowering plants diverged from their most recent common ancestor (MRCA) in the early Cretaceous or even late Jurassic (Magallón *et al.*, 2015; Silvestro *et al.*, 2021). Thus, the biology of extant *Amborella* populations has been molded by more than 130 Myr of evolution, since the early diversification of the extant flowering plants. However, *A. trichopoda* has retained some characteristics that are hypothesized to be shared with the MCRA of living angiosperms (e.g. free floral organs, including an

undifferentiated perianth of tepals, and carpels that contain an apical, secretion-filled canal for pollen-tube growth; Endress & Igersheim, 2000; Sauquet *et al.*, 2017), and no genome duplication has occurred in the *Amborella* lineage since it diverged from the other angiosperms (Amborella Genome Project, 2013). Thus, further study of the genes involved in sex determination in *A. trichopoda* might shed light on the regulatory pathways that underlie development of unisexual flowers in other angiosperms.

Acknowledgements

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

Author contributions

Conceptualization: GABM, CS, JK, JL-M, AB; Methodology: JK, FP, GABM, JL-M, AB; Software: JK, FP; Formal analysis: JK, AB, GL, AH; Investigation: JK, AB, AAR, GL, JC, CS, GABM, JL-M; Resources: BF, GG, PR, CWdP; Writing – original draft: JK, GABM, JL-M, CS; Writing – review & editing: all authors; Visualization: JK, AB; Supervision: GABM, JL-M, CS, JK; Project administration: GABM, JL-M, CS; Funding acquisition: GABM, CS, JL-M, CWdP. CS, GABM, and JL-M contributed equally to this work.

ORCID

Gildas Gâteblé  <https://orcid.org/0000-0003-2831-6384>

Jos Käfer  <https://orcid.org/0000-0002-0561-8008>

James Leebens-Mack  <https://orcid.org/0000-0003-4811-2231>

Gabriel A. B. Marais  <https://orcid.org/0000-0003-2134-5967>

Charlie Scutt  <https://orcid.org/0000-0003-2970-3810>

Data availability

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

References

Ainsworth C. 2000. Boys and girls come out to play: the molecular biology of dioecious plants. *Annals of Botany* **86**: 211–221.

- Akagi T, Henry IM, Tao R, Comai L. 2014. Plant genetics. A Y-chromosome-encoded small RNA acts as a sex determinant in persimmons. *Science* **346**: 646–650.
- Akagi T, Pilkington SM, Varkonyi-Gasic E, Henry IM, Sugano SS, Sonoda M, Firl A, Mc-Neillage MA, Douglas MJ, Wang T *et al.* 2019. Two Y-chromosome-encoded genes determine sex in kiwifruit. *Nature Plants* **5**: 801–809.
- Amborella Genome Project. 2013. The Amborella genome and the evolution of flowering plants. *Science* **342**: 1241089.
- Anger N, Fogliani B, Scutt CP, Gâtéblé G. 2017. Dioecy in *Amborella trichopoda*: evidence for genetically based sex determination and its consequences for inferences of the breeding system in early angiosperms. *Annals of Botany* **119**: 591–597.
- Badouin H, Velt A, Gindraud F, Flutre T, Dumas V, Vautrin S, Marande W, Corbi J, Sallet E, Ganofsky J *et al.* 2020. The wild grape genome sequence provides insights into the transition from dioecy to hermaphroditism during grape domestication. *Genome Biology* **21**: 223.
- Buzgo M, Soltis PS, Soltis DE. 2004. Floral developmental morphology of *Amborella trichopoda* (Amborellaceae). *International Journal of Plant Sciences* **165**: 925–947.
- Charlesworth B, Charlesworth D. 1978. Model for evolution of dioecy and gynodioecy. *American Naturalist* **112**: 975–997.
- Charlesworth D. 2016. Plant sex chromosomes. *Annual Review of Plant Biology* **67**: 397–420.
- Diggle PK, Di Stilio VS, Gschwend AR, Golenberg EM, Moore RC, Russell JR, Sinclair JP. 2011. Multiple developmental processes underlie sex differentiation in angiosperms. *Trends in Genetics* **27**: 368–376.
- Dobin A, Davis CA, Schlesinger F, Drenkow J, Zaleski C, Jha S, Batut P, Chaisson M, Gingeras TR. 2013. STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* **29**: 15–21.
- Endress PK, Doyle JA. 2009. Reconstructing the ancestral angiosperm flower and its initial specializations. *American Journal of Botany* **96**: 22–66.
- Endress PK, Igersheim A. 2000. Reproductive structures of the basal angiosperm *Amborella trichopoda* (Amborellaceae). *International Journal of Plant Sciences* **161**: S237–S248.
- Garrison E, Marth G. 2012. Haplotype-based variant detection from short-read sequencing. *arXiv*: 1207.3907.
- Gayral P, Melo-Ferreira J, Glémin S, Bierre N, Carneiro M, Nabholz B, Lourenco JM, Alves PC, Ballenghien M, Faivre N *et al.* 2013. Reference-free population genomics from nextgeneration transcriptome data and the vertebrate-invertebrate gap. *PLoS Genetics* **9**: e1003457.
- Golenberg EM, West NW. 2013. Hormonal interactions and gene regulation can link monoecy and environmental plasticity to the evolution of dioecy in plants. *American Journal of Botany* **100**: 1022–1037.
- Grosse-Veldmann B, Korotkova N, Reinken B, Lobin W, Barthlott W. 2011. *Amborella trichopoda* – cultivation of the most ancestral angiosperm in botanic gardens. *Sibbaldia* **9**: 143–155.
- Harkess A, Huang K, Van der Hulst R, Tissen B, Caplan JL, Koppula A, Batish M, Meyers BC, Leebens-Mack J. 2020. Sex determination by two Y-linked genes in garden asparagus. *Plant Cell* **32**: 1790–1796.
- Harkess A, Zhou J, Xu C, Bowers JE, Van der Hulst R, Ayyampalayam S, Mercati F, Riccardi P, McKain MR, Kakrana A *et al.* 2017. The asparagus genome sheds light on the origin and evolution of a young Y chromosome. *Nature Communications* **8**: 1279.
- Käfer J, Lartillot N, Marais GAB, Picard F. 2021. Detecting sex-linked genes using genotyped individuals sampled in natural populations. *Genetics* **218**: iyab053.
- Li H, Durbin R. 2009. Fast and accurate short read alignment with Burrows–Wheeler transform. *Bioinformatics* **25**: 1754–1760.
- Magallón S, Gómez-Acevedo S, Sánchez-Reyes LL, Hernández-Hernández T. 2015. A metacalibrated time-tree documents the early rise of flowering plant phylogenetic diversity. *New Phytologist* **207**: 437–453.
- Ming R, Bendahmane A, Renner SS. 2011. Sex chromosomes in land plants. *Annual Review of Plant Biology* **62**: 485–514.
- Müller NA, Kersten B, Leite Montalvão AP, Mähler N, Bernhardsson C, Bräutigam K, Carracedo Lorenzo Z, Hoenicka H, Kumar V, Mader M *et al.* 2020. A single gene underlies the dynamic evolution of poplar sex determination. *Nature Plants* **6**: 630–637.
- Muyle A, Käfer J, Zemp N, Mousset S, Picard F, Marais GA. 2016. SEX-DETECTOR: a probabilistic approach to study sex chromosomes in non-model organisms. *Genome Biology and Evolution* **8**: 2530–2543.
- Muyle A, Shearn R, Marais GA. 2017. The evolution of sex chromosomes and dosage compensation in plants. *Genome Biology and Evolution* **9**: 627–645.
- Nei M, Li WH. 1979. Mathematical model for studying genetic variation in terms of restriction endonucleases. *Proceedings of the National Academy of Sciences, USA* **76**: 5269–5273.
- One Thousand Plant Transcriptomes Initiative. 2019. One thousand plant transcriptomes and the phylogenomics of green plants. *Nature* **574**: 679–685.
- Renner SS. 2014. The relative and absolute frequencies of angiosperm sexual systems: dioecy, monoecy, gynodioecy, and an updated online database. *American Journal of Botany* **101**: 1588–1596.
- Renner SS, Müller NA. 2021. Plant sex chromosomes defy evolutionary models of expanding recombination suppression and genetic degeneration. *Nature Plants* **7**: 392–402.
- Robinson MD, McCarthy DJ, Smyth GK. 2010. edgeR: a bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics* **26**: 139–140.
- Salomo K, Smith JF, Feild TS, Samain MS, Bond L, Davidson C, Zimmers J, Neinhuis C, Wanke S. 2017. The emergence of earliest angiosperms may be earlier than fossil evidence indicates. *Systematic Botany* **42**: 607–619.
- Sauquet H, Magallón S. 2018. Key questions and challenges in angiosperm macroevolution. *New Phytologist* **219**: 1170–1187.
- Sauquet H, von Balthazar M, Magallón S, Doyle JA, Endress PK, Bailes EJ, Barroso de Morais E, Bull-Hereñu K, Carrive L, Chartier M *et al.* 2017. The ancestral flower of angiosperms and its early diversification. *Nature Communications* **8**: 16047.
- Silvestro D, Bacon CD, Ding W, Zhang Q, Donoghue PCJ, Antonelli A, Xing Y. 2021. Fossil data support a pre-Cretaceous origin of flowering plants. *Nature Ecology & Evolution* **5**: 449–457.
- Takahata N, Nei M. 1985. Gene genealogy and variance of interpopulational nucleotide differences. *Genetics* **110**: 325–344.
- Yang Z. 2007. PAML 4: phylogenetic analysis by maximum likelihood. *Molecular Biology and Evolution* **24**: 1586–1591.

Supporting Information

Additional Supporting Information may be found online in the Supporting Information section at the end of the article.

Dataset S1 Bed file indicating the putative W-specific regions for masking.

Dataset S2 Bed file indicating the putative Z-specific regions for masking.

Dataset S3 Table with all putatively sex-linked gene models and gene-level statistics.

Fig. S1 Masking of putative W- and Z-specific regions in the nonrecombining region of chromosome 9.

Fig. S2 Resequencing coverage per gene using mapping on the unmasked and masked genome assemblies.

Fig. S3 Results of the sex-linkage analysis in the mapping population data.

Fig. S4 Sex-biased gene expression inferred from RNA-Seq data from flower buds.

Fig. S5 Distribution of heterozygosity between parents and offspring in the mapping population.

Fig. S6 Allele transmission on chromosome 9.

Methods S1 Additional information about the data and analyses of sex linkage.

Notes S1 Detailed description of the results of the analyses.

Table S1 RNA-Seq statistics for the 20 individuals grown from seeds collected in the wild.

Table S2 Read depths for the individuals from the F1 family sequenced using DNA-Reseq data.

Table S3 Summary of sex-linkage inference using the RNA-Seq data from a wild population.

Table S4 Sex-biased gene expression analysis.

Please note: Wiley Blackwell are not responsible for the content or functionality of any Supporting Information supplied by the authors. Any queries (other than missing material) should be directed to the *New Phytologist* Central Office.



About *New Phytologist*

- *New Phytologist* is an electronic (online-only) journal owned by the New Phytologist Foundation, a **not-for-profit organization** dedicated to the promotion of plant science, facilitating projects from symposia to free access for our Tansley reviews and Tansley insights.
- Regular papers, Letters, Viewpoints, Research reviews, Rapid reports and both Modelling/Theory and Methods papers are encouraged. We are committed to rapid processing, from online submission through to publication 'as ready' via *Early View* – our average time to decision is <26 days. There are **no page or colour charges** and a PDF version will be provided for each article.
- The journal is available online at Wiley Online Library. Visit **www.newphytologist.com** to search the articles and register for table of contents email alerts.
- If you have any questions, do get in touch with Central Office (np-centraloffice@lancaster.ac.uk) or, if it is more convenient, our USA Office (np-usaoffice@lancaster.ac.uk)
- For submission instructions, subscription and all the latest information visit **www.newphytologist.com**