

## **Title: A derived ZW chromosome system in *Amborella trichopoda*, the sister species to all other extant flowering plants**

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

### **Affiliations:**

- 10       1. Université Lyon 1, CNRS, Laboratoire de Biométrie et Biologie Évolutive UMR 5558, F-69622 Villeurbanne, France
2. University of Georgia, Department of Plant Biology, Athens, GA 30602-7271, USA
3. Ecole Normale Supérieure de Lyon, CNRS, Laboratoire Reproduction et Développement des plantes UMR 5667, F-69364 Lyon, France
- 15       4. Department of Crop, Soil, and Environmental Sciences, Auburn University, Auburn, AL, 36849, USA
5. HudsonAlpha Institute for Biotechnology, Huntsville, AL 35806, USA
6. Univ. Paris-Saclay, Institute of Plant Sciences, Plateforme transcriptomique de l'IP2S (POPS), F-91190 Gif-sur-Yvette, France
- 20       7. Institut Agronomique néo-Calédonien (IAC), BP 73 Port Laguerre, 98890 Paita, New Caledonia
8. Université de la Nouvelle-Calédonie (UNC), Laboratoire Insulaire du Vivant et de l'Environnement (LIVE), 98851 Nouméa, New Caledonia
9. Department of Biology and Huck Institutes of the Life Sciences, The Pennsylvania State University, University Park, PA 16802, USA
- 25       10. LEAF- Linking Landscape, Environment, Agriculture and Food, Instituto Superior de Agronomia, Universidade de Lisboa, 1349-017 Lisbon, Portugal

\*Correspondence to: [jos.kafer@univ-lyon1.fr](mailto:jos.kafer@univ-lyon1.fr), [jleebensmack@uga.edu](mailto:jleebensmack@uga.edu)

30       §Equal contribution

## Abstract

Sex determination is poorly understood in plants. *Amborella trichopoda* is a well-known plant model for evo-devo studies, which is also dioecious (has male and female individuals), with an unknown sex determination mechanism. *A. trichopoda* is a “sex switcher”, which points to possible environmental factors that act on sex, but populations grown from seed under greenhouse conditions exhibit a 50:50 sex ratio, which indicates the operation of genetic factors. Here, we use a new method (*SDpop*) to identify sex-linked genes from genotyping data of male and female individuals sampled in the field, and find that *A. trichopoda* has a ZW sex-chromosome system. The sex-linked genes map to a 4 Mb sex-determining region on chromosome 9. The low extent of ZW divergence suggests these sex chromosomes are of recent origin, which is consistent with dioecy being derived character in the *A. trichopoda* lineage. Our work has uncovered clearly formed sex chromosomes in a species in which both genetic and environmental factors can influence sex.

## One Sentence Summary:

*Amborella trichopoda*, a dioecious species in which both genetics and the environment influence sex, possesses a pair of quite recently evolved ZW chromosomes.

## Main Text:

Dioecy (separate male and female individuals) is one of the many sexual systems exhibited by flowering plants. This system has evolved repeatedly in angiosperms, and about 6% of angiosperm species are dioecious (1, 2). Genetic sex determination (GSD) has been reported in about 50 angiosperm species (3, 4) and master sex-determining genes have been identified in a few of these (5–9). Despite these great advances, we still lack a global understanding of the evolution of sex determination in dioecious plants. In particular, it is not clear how widespread GSD is, nor whether there are common genetic dynamics contributing to the evolution of dioecy in flowering plants. Environmental sex determination (ESD), defined as the possibility of a plant to switch sex during its lifetime (10, 11), has been poorly studied, probably because it is thought to be very rare (1, 12). However, even in species with well-established sex chromosomes, sex switching and incomplete sex determination are not uncommon (2, 13–15). It has been speculated that ESD is a consequence of gene expression changes elicited by environmental triggering of phytohormone signaling pathways (11), but we simply have no data on ESD, nor on the mechanisms responsible for environmentally induced floral phenotype changes in plants with GSD systems.

*Amborella trichopoda*, a dioecious shrub endemic to New Caledonia, is the sole species on the sister lineage to all other angiosperms, and as such has been used in evo-devo investigations of floral evolution (16–18). Rigorous ancestral state reconstruction suggests that the last common ancestor of all extant angiosperms had bisexual flowers (19) and independent transitions to dioecy have occurred repeatedly throughout angiosperm evolution (1), including within the *Amborella* lineage. Despite broad interest in *A. trichopoda* given its pivotal phylogenetic position, the nature and timing of the evolutionary shift to dioecy on the lineage leading to *Amborella* has remained enigmatic. Sex-switching is well documented in *A. trichopoda* (20–22) suggesting some degree of ESD. Whereas sex ratios in wild populations are often male-biased, plants grown from seeds exhibit a sex ratio close to 1:1 upon first

flowering, suggesting a combination of GSD and ESD (22). No evidence for sex chromosomes has been observed in the *A. trichopoda* karyotype, but many plant sex chromosomes are homomorphic, i.e. not distinguishable in light microscopy (3). Here we combine a novel approach to identify sex-linked markers in natural populations with genetic mapping to test hypothesized GSD and characterize the *A. trichopoda* sex chromosomes.

Identifying sex chromosomes from genomic data can be difficult, especially when these are homomorphic (4, 23). First, we investigated the presence of sex-linked markers in a sample of 10 female and 10 male individuals grown from seeds collected in the wild. RNA-seq data of these individuals were obtained from flower buds and used to genotype them. We used a novel method, *SDpop*, which uses modifications of the Hardy-Weinberg equilibrium expectation to identify sex-linkage (24). This allows probabilistic classification of genes as autosomal or sex-linked with *XY* or *ZW* heterogametic genotypes. The data obtained in this study strongly supported the *ZW* model (Table 1), inferring 49 *ZW* gametologs and 9 *Z*-hemizygous genes (lacking a detectable *W* copy).

We tested the *SDpop* inference by mapping putatively sex-linked genes to a chromosomal draft assembly of the *A. trichopoda* genome (V.6, genome\_ID = 50948, <https://genomeevolution.org/coge>) including 13 pseudomolecules corresponding to each chromosome pair (n=13). We mapped the 49 gametologous genes identified by *SDpop* onto this genome and found that 33 (67%) of them mapped uniquely to a single region of ~4 Mb on chromosome 9 of the assembly (Figure 1). This proportion increased to 89% when genomic contigs not included in the 13 chromosome scaffolds were excluded from the reference.

We investigated the extent of the sex-determining region further by mapping re-seq data from a male, a female and 33 F1 progeny for which flower sex had been scored (16 female and 17 male F1 offspring). First, we assessed whether 671 female-specific SNP variants identified in the wild population transcripts were only present in the seed parent and female progeny in the mapping population. Males in the mapping population were found to possess no more than five of the putatively female-specific SNP variants identified in the wild-collected population, whereas females were found to possess at least 465 of these variants (Table S1).

We also used the 33 genotyped and phenotyped progeny 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* (Figure 1). Finally, the male-versus-female coverage ratio(4, 23) was found to deviate significantly from one in the 4 Mb sex-determination region identified by *SDpop* (Figure 1).

*SDpop* and heterozygosity analyses performed on the wild-sampled plants and mapping population, together with an assessment of the mapping population read coverage data, provided consistently strong support for an approximately 4 Mb-long region on the right arm of chromosome 9 (45-49 Mb) as cosegregating perfectly with sex. However, we found another region, on the left arm of chromosome 9 (10-20 Mb), showing a weaker association with sex (Figure 1). The recombination landscape on chromosome 9 may explain this observation (Supplementary Text, Figure S1).

The individual that was sequenced in the *Amborella* genome project (*SantaCruz 1975.3*) is known to have switched from producing mostly male flowers (20) to all female flowers (18). The mapping of reads from the male and female progeny array described above identified the presence of both *Z* and *W* gametologs in the clearly *ZW*, genetically female reference genome. Further scrutiny of the sex-determining region revealed that the reference assembly of this region is chimeric with a mixture of *Z* and *W* segments. The male/female

coverage-ratio analysis mentioned above was used to identify Z and W segments and construct draft Z and W haplotypes by masking W-linked scaffolds (with only female reads, and thus a male/female coverage-ratio  $\sim 0$ ) and Z-linked scaffolds (with a male/female coverage-ratio  $\sim 2$ ), respectively (Supplementary Text, Figure S2). Remapping the reads on these masked assemblies reduced the regions with sex-biased coverage (DNA-reseq data) and yielded more sex-linked genes (Table 1).

The sex-determining region includes  $\sim 150$  sex-linked genes and functional analyses will be required to determine how many of these may influence floral phenotypes. The ability of *A. trichopoda* genotypes (including *SantaCruz 1975.3*) to produce both male and female flowers or change sex completely (20–22) suggests that both the Z and W regions include genes that singly or collectively can act as a switch to induce male or female flower developmental programs. We focused on i) ZW gene pairs with differential expression in males and females, and ii) genes present in one haplotype and not the other (W-specific or Z-hemizygous genes). We used the RNA-seq data from flower buds mentioned above to conduct a male vs female differential expression analysis (Figure S3). Although a few genes in the sex-determining region had sex-biased expression, none were expressed exclusively in males or females (Data S1). We identified 37 genes exhibiting sex-biased expression. We then looked at the annotation of these genes (Data S2) and found that four of them were involved in the auxin pathway, a known sex-related phytohormone (25) (Table S2). The fact that all four genes are male-biased suggest a possible masculinizing effect. We also looked at the putative W-specific and Z-hemizygous genes that we identified (Data S1) and found that a W-specific gene had a DMY domain that is characteristic of PPR genes (Data S2), a class of genes involved in cytonuclear interactions, and particularly in cytoplasmic male sterility (26) (Table S2).

We used masked assemblies to calculate the synonymous divergence ( $dS$ ) between the two copies of all predicted genes in the assembly and then used the maximum Z-W  $dS$  values obtained to estimate the age of the sex-determining region (4, 27). The highest  $dS$  values are estimated from the mapping population re-seq data, and are around 0.14, but these values have large standard errors. The values estimated from the natural population RNA-seq have lower maximum values and standard errors. Among-gene variation in  $dS$  is high, as expected due to sampling variation (28). We therefore estimate that the maximum  $dS$  value is between 0.04 and 0.06 (Figure 2). This is 20 to 30 times higher than the level of genome-wide nucleotide polymorphism (0.002), suggesting that recombination suppression occurred 9.5 to 14.5 times  $4*Ne$  generations ago. Pairwise sequentially Markovian coalescent (PSMC) (29) analysis of SNP variation among extant *A. trichopoda* populations yield a  $Ne$  approximation of 37,500 and past population size fluctuations (18). When using the extant  $Ne$  estimate and a generation time of four years (18), we estimate suppression of recombination in the *A. trichopoda* sex-determination region of the *Amborella* chromosome 5.7 to 8.7 million years ago. Four years is probably a lower bound estimate of the *A. trichopoda*'s generation time in the wild as individuals of 40-50 years old have been observed (30). A much more realistic estimate is probably around 20-25 years and the sex-determining region could be as old as 54.3 millions years. The divergence of the *Amborella* lineage from that of all remaining extant angiosperm is estimated to have occurred at least 125 million years ago (31, 32), and the mean  $dS$  value between *A. trichopoda* and other angiosperms was estimated to be between 1 and 2.2 (33). We found a much smaller  $dS$  for the sex-determining region, which points to a much more recent origin than the split of the *Amborella* lineage from the other angiosperms.

Our results clearly point to the existence of a *ZW* system with a small (~4 Mb) sex-determining region on chromosome 9 of *A. trichopoda*. Thus, *Amborella* has genetic sex determination, and all of the 55 individuals used in this study had a phenotypic sex that corresponded with their genetic sex. However, as in many plant species with sex chromosomes, sex changes can occur, as illustrated by the individual used for the reference genome and observations described in previous studies (22). The reasons why the plant used for the reference genome did not initially correspond to its genetic sex, and why it subsequently switched to its genetic sex, are both unknown. It is, however, known that local heating of branches of an otherwise female *Amborella* plant can induce male flowering in these branches; thus, genetic sex determination doesn't rule out any influence from the environment on sex. A previous study suggested that *A. trichopoda* has inconstant males, which could indicate that dioecy had evolved from a gynodioecious intermediate in this species (22,34). However, the observation of sex-switching and the influence of the environment on sex could indicate that genetic sex determination evolved from environmentally controlled monoecy, in which plant hormones might have played an important role (25). The presence of distinct Z and W haplotypes in *Amborella* could be used to test how often phenotypic and genetic sex differ in nature, and whether inconstant individuals always have the same genetic sex.

*A. trichopoda* is the sole extant species in the sister lineage to all other extant flowering plants. This species and other members of the ANA grade (including Nymphaeales and Austrobaileyales) are of pivotal importance for reconstructing evolution of flower development and the characteristics of the ancestor of all angiosperms (16,18). Based on extant angiosperm species, it was inferred that the most likely ancestral breeding system was bisexuality (19). However, several ANA-grade species such as some *Trithuria* spp. (Nymphaeales) and *Schisandra chinensis* (Austrobaileyales) are dioecious (35,36). Dating the origin of dioecy in these plants is very important to accurately reconstruct the ancestral sexual system (22). Our estimate of the age of the *ZW* chromosomes of *A. trichopoda* points to relatively recent sex chromosomes compared to the chronology of angiosperm evolution, which does not support the view that dioecy is ancient in *A. trichopoda*. However, we cannot completely rule out this possibility, as frequent sex chromosome turnover is possible in some lineages (37,38). The structure of the *A. trichopoda* flower, however, does not fit well with ancient dioecy, as non-functional male organs are present in female flowers, while a pronounced central dome in male flowers may be a remnant of the gynoeceium (39). Overall, our results tend to support the idea that the most recent ancestor of all extant angiosperms was indeed bisexual.

## References and Notes:

1. S. S. Renner, The relative and absolute frequencies of angiosperm sexual systems: dioecy, monoecy, gynodioecy, and an updated online database. *Am. J. Bot.* 101, 1588–1596 (2014).
2. J. Käfer, G. A. B. Marais, J. R. Pannell, On the rarity of dioecy in flowering plants. *Mol. Ecol.* 26, 1225–1241 (2017).
3. R. Ming, A. Bendahmane, S. S. Renner, Sex chromosomes in land plants. *Annu. Rev. Plant Biol.* 62, 485–514 (2011).
4. A. Muyle, R. Shearn, G. A. Marais, The Evolution of Sex Chromosomes and Dosage Compensation in Plants. *Genome Biol. Evol.* 9, 627–645 (2017).
5. T. Akagi, I. M. Henry, R. Tao, L. Comai, A Y-chromosome-encoded small RNA acts as a sex determinant in persimmons. *Science* (2014)
6. T. Akagi, S. M. Pilkington, E. Varkonyi-Gasic, I. M. Henry, S. S. Sugano, M. Sonoda, A. Firl, M. A. McNeilage, M. J. Douglas, T. Wang, R. Rebstock, C. Voogd, P. Datson, A. C. Allan, K. Beppu, I. Kataoka, R. Tao, Two Y-chromosome-encoded genes determine sex in kiwifruit. *Nat Plants.* 5, 801–809 (2019).

7. A. Harkess, J. Zhou, C. Xu, J. E. Bowers, R. Van der Hulst, S. Ayyampalayam, F. Mercati, P. Riccardi, M. R. McKain, A. Kakrana, H. Tang, J. Ray, J. Groenendijk, S. Arikrit, S. M. Mathioni, M. Nakano, H. Shan, A. Telgmann-Rauber, A. Kanno, Z. Yue, H. Chen, W. Li, Y. Chen, X. Xu, Y. Zhang, S. Luo, H. Chen, J. Gao, Z. Mao, J. C. Pires, M. Luo, D. Kudrna, R. A. Wing, B. C. Meyers, K. Yi, H. Kong, P. Lavrijsen, F. Sunseri, A. Falavigna, Y. Ye, J. H. Leebens-Mack, G. Chen, The asparagus genome sheds light on the origin and evolution of a young Y chromosome. *Nat. Commun.* 8, 1279 (2017).
8. A. Harkess, K. Huang, R. van der Hulst, B. Tissen, J. L. Caplan, A. Koppula, M. Batish, B. C. Meyers, J. Leebens-Mack, Sex Determination by Two Y-linked Genes in Garden Asparagus. *The Plant Cell* (2020), p. tpc.00859.2019.
9. N. A. Müller, B. Kersten, A. P. Leite Montalvão, N. Mähler, C. Bernhardsson, K. Bräutigam, Z. Carracedo Lorenzo, H. Hoenicka, V. Kumar, M. Mader, B. Pakull, K. M. Robinson, M. Sabatti, C. Vettori, P. K. Ingvarsson, Q. Cronk, N. R. Street, M. Fladung, A single gene underlies the dynamic evolution of poplar sex determination. *Nat Plants.* 6, 630–637 (2020).
10. E. L. Charnov, J. Bull, When is sex environmentally determined? *Nature.* 266 (1977), pp. 828–830.
11. J. R. Pannell, Plant Sex Determination. *Curr. Biol.* 27, R191–R197 (2017).
12. M. A. Schlessman, Gender diphasy (“sex choice”). *Plant reproductive ecology: patterns and strategies*, 139–153 (1988).
13. J. Sneep, The significance of andromonoecism for the breeding of *Asparagus officinalis* L. II. *Euphytica.* 2 (1953), pp. 224–228.
14. D. C. Freeman, K. T. Harper, E. L. Charnov, Sex change in plants: Old and new observations and new hypotheses. *Oecologia.* 47, 222–232 (1980).
15. C. Ainsworth, Boys and Girls Come Out to Play: The Molecular Biology of Dioecious Plants. *Ann. Bot.* 86, 211–221 (2000).
16. C. P. Scutt, G. Theissen, C. Ferrandiz, The Evolution of Plant Development: Past, Present and Future: Preface. *Ann. Bot.* 100, 599–601 (2007).
17. G. Arnault, A. C. M. Vialette, A. Andres-Robin, B. Fogliani, G. Gâteblé, C. P. Scutt, Evidence for the Extensive Conservation of Mechanisms of Ovule Integument Development Since the Most Recent Common Ancestor of Living Angiosperms. *Front. Plant Sci.* 9, 1352 (2018).
18. Amborella Genome Project, The Amborella genome and the evolution of flowering plants. *Science.* 342, 1241089 (2013).
19. H. Sauquet, M. von Balthazar, S. Magallón, J. A. Doyle, P. K. Endress, E. J. Bailes, E. Barroso de Morais, K. Bull-Hereñu, L. Carrive, M. Chartier, G. Chomicki, M. Coiro, R. Cornette, J. H. L. El Ottra, C. Epicoco, C. S. P. Foster, F. Jabbour, A. Haevermans, T. Haevermans, R. Hernández, S. A. Little, S. Löfstrand, J. A. Luna, J. Massoni, S. Nadot, S. Pamperl, C. Prieu, E. Reyes, P. Dos Santos, K. M. Schoonderwoerd, S. Sontag, A. Soulebeau, Y. Staedler, G. F. Tschan, A. Wing-Sze Leung, J. Schönenberger, The ancestral flower of angiosperms and its early diversification. *Nat. Commun.* 8, 16047 (2017).
20. M. Buzgo, P. S. Soltis, D. E. Soltis, Floral Developmental Morphology of *Amborella trichopoda* (Amborellaceae). *Int. J. Plant Sci.* 165, 925–947 (2004).
21. B. Große-Veldmann, N. Korotkova, B. Reinken, W. Lobin, W. Barthlott, *Amborella trichopoda*: Cultivation of the most ancestral angiosperm in botanic gardens. *Sibbaldia: The Journal of Botanic Garden Horticulture*, 143–154 (2011).
22. N. Anger, B. Fogliani, C. P. Scutt, G. Gâteblé, 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* (2017), p. mcw278.
23. D. H. Palmer, T. F. Rogers, R. Dean, A. E. Wright, How to identify sex chromosomes and their turnover. *Mol. Ecol.* 28, 4709–4724 (2019).
24. J. Käfer, N. Lartillot, G. A. B. Marais, F. Picard, Detecting sex-linked genes using genotyped individuals sampled in natural populations (2020), p. 2020.01.31.928291.
25. Golenberg, E. M., & West, N. W. (2013). Hormonal interactions and gene regulation can link monoecy and environmental plasticity to the evolution of dioecy in plants. *American journal of botany*, 100(6), 1022–1037. <https://doi.org/10.3732/ajb.1200544>
26. Rovira, A. G., & Smith, A. G. (2019). PPR proteins - orchestrators of organelle RNA metabolism. *Physiologia plantarum*, 166(1), 451–459. <https://doi.org/10.1111/ppl.12950>
27. R. Bergero, D. Charlesworth, The evolution of restricted recombination in sex chromosomes. *Trends Ecol. Evol.* 24, 94–102 (2009).
28. N. Takahata, M. Nei, Gene genealogy and variance of interpopulational nucleotide differences. *Genetics.* 110, 325–344 (1985).
29. H. Li, R. Durbin, Inference of human population history from individual whole-genome sequences. *Nature.* 475, 493–496 (2011).
30. B. Fogliani, C. P. Scutt, G. Gâteblé. Personal communication.

31. H.-T. Li, T.-S. Yi, L.-M. Gao, P.-F. Ma, T. Zhang, J.-B. Yang, M. A. Gitzendanner, P. W. Fritsch, J. Cai, Y. Luo, H. Wang, M. van der Bank, S.-D. Zhang, Q.-F. Wang, J. Wang, Z.-R. Zhang, C.-N. Fu, J. Yang, P. M. Hollingsworth, M. W. Chase, D. E. Soltis, P. S. Soltis, D.-Z. Li, Origin of angiosperms and the puzzle of the Jurassic gap. *Nat Plants*. 5, 461–470 (2019).
- 5 32. C. J. van der Kooij, J. Ollerton, The origins of flowering plants and pollinators. *Science*. 368, 1306–1308 (2020).
33. One Thousand Plant Transcriptomes Initiative., Leebens-Mack, J.H., Barker, M.S. et al. One thousand plant transcriptomes and the phylogenomics of green plants. *Nature* 574, 679–685 (2019). <https://doi.org/10.1038/s41586-019-1693-2>
- 10 34. Ehlers, B. K., & Bataillon, T. (2007). 'Inconstant males' and the maintenance of labile sex expression in subdioecious plants. *The New phytologist*, 174(1), 194–211. <https://doi.org/10.1111/j.1469-8137.2007.01975.x>
- 15 35. Iles WJ, Rudall PJ, Sokoloff DD, Remizowa MV, Macfarlane TD, Logacheva MD, Graham SW. Molecular phylogenetics of Hydatellaceae (Nymphaeales): sexual-system homoplasy and a new sectional classification. *Am J Bot*. 2012 Apr;99(4):663-76.
36. Endress PK (2001) The flowers in extant basal angiosperms and inferences on ancestral flowers. *Int J Plant Sci* 162:1111–1140.
37. Palmer, D. H., Rogers, T. F., Dean, R., & Wright, A. E. (2019). How to identify sex chromosomes and their turnover. *Molecular ecology*, 28(21), 4709–4724. <https://doi.org/10.1111/mec.15245>
- 20 38. Martin, H., Carpentier, F., Gallina, S., Godé, C., Schmitt, E., Muyle, A., Marais, G., & Touzet, P. (2019). Evolution of Young Sex Chromosomes in Two Dioecious Sister Plant Species with Distinct Sex Determination Systems. *Genome biology and evolution*, 11(2), 350–361. <https://doi.org/10.1093/gbe/evz001>
39. Endress, P. K., & Doyle, J. A. (2015). Ancestral traits and specializations in the flowers of the basal grade of living angiosperms. *Taxon*, 64(6), 1093–1116.
- 25

**Acknowledgments:** we 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 for performing some of the earlier work in genome assembly on the *Amborella trichopoda* genome used in this study, and Elise Lucotte for discussions. **Funding:** this work was supported by an ANR grant to GABM and CS (ANR-14-CE19-0021-01) and NSF-PGRP grant 0922742 ; **Author contributions:** Conceptualization: GABM, CS, JK, JLM, AB Methodology: JK, FP, GABM, JLM, AB Software: JK, FP, Formal analysis: JK, AB, GL, AB, JB, AH Investigation: JK, AB, AAR, GL, JC, CS, GABM, JLM Resources: BF, GG, PR, CWdP Writing - Original Draft: JK, GABM, JLM, CS Writing - Review & Editing: all authors, Visualization: JK, AB, Supervision: GABM, JLM, CS, JK Project administration: GABM, JLM, CS, Funding acquisition: GABM, CS, JLM, CWdP ; **Competing interests:** Authors declare no competing interests; and **Data and materials availability:** The *A. trichopoda* genome is in CoGe (V.6, genome\_ID = 50948, <https://genomevolution.org/coge>), all of the genome re-seq data are in SRA under the bioproject PRJNA212863, the RNA-seq data are in SRA under the bioproject PRJNA687004.

30

35

40

## Supplementary Materials:

Materials and Methods

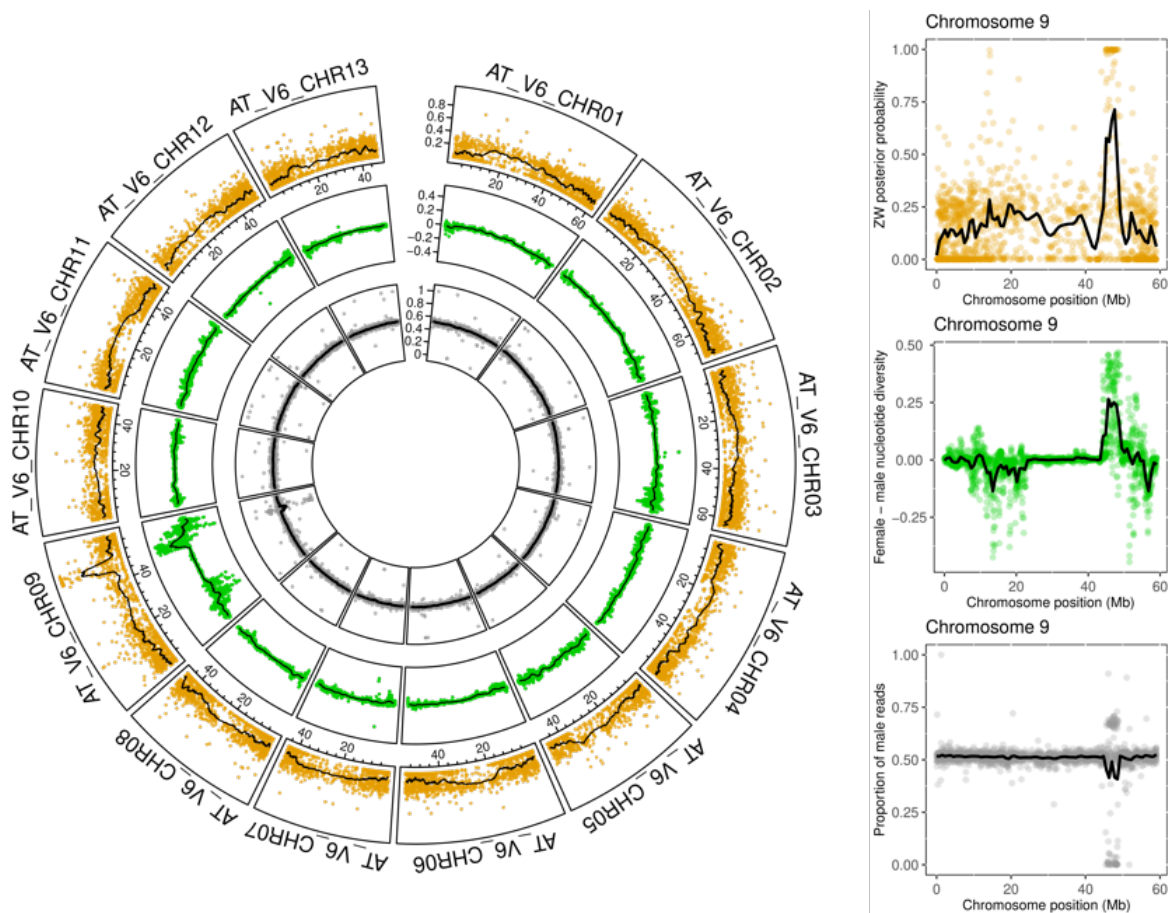
45 Supplementary Text

Figures S1-S3

Tables S1-S2

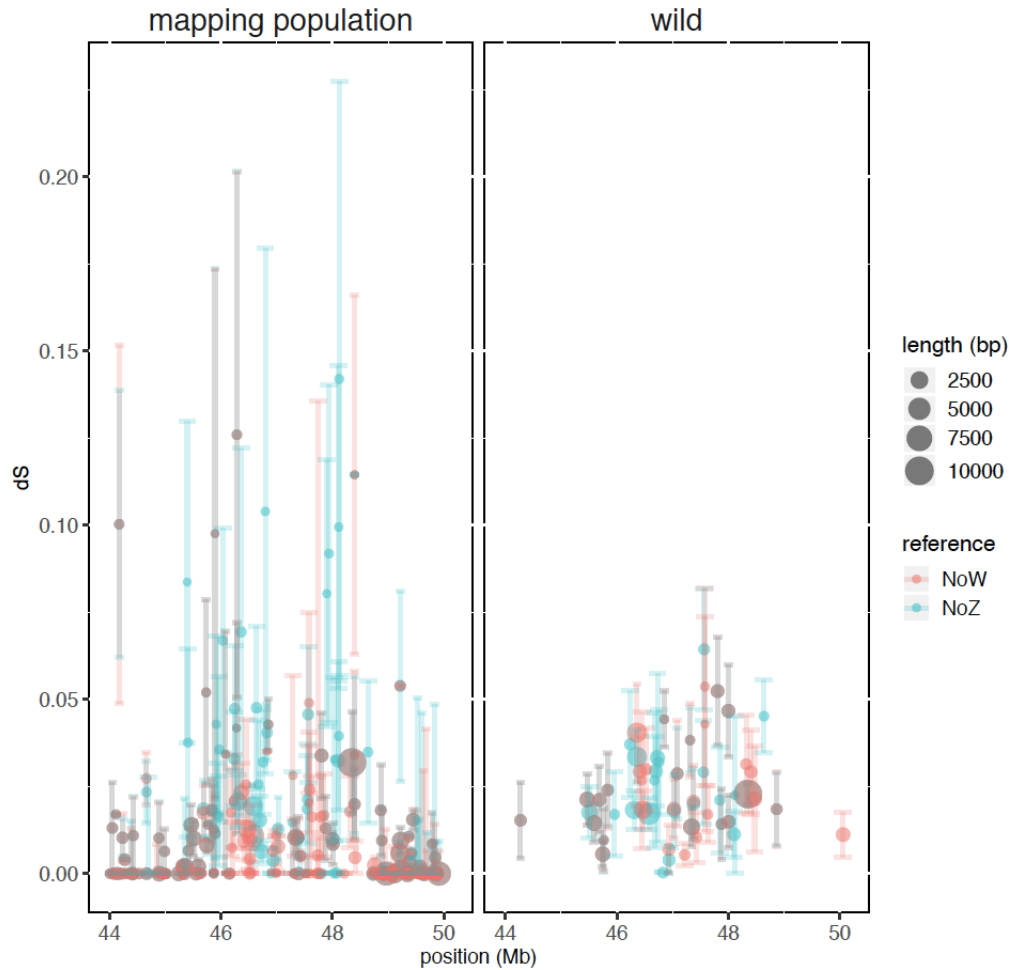
50

**Fig. 1.** Evidence for a ZW sex-determining region in *A. trichopoda*. (A) Circular representation of the *A. trichopoda* genome. Chromosome number (1 to 13) is indicated. Outer circle indicates ZW posterior probability, inner circle indicates female – male nucleotide diversity, in-between circle indicates proportion of male reads. (B) Zoom in for chromosome 9. The sex-determining region is located at positions 45–49 Mb (see text for more details). Region 10–20 Mb shows an excess of male heterozygous – female homozygous sites reminiscent of a XY locus, which explains a drop in female – male nucleotide diversity. This region is in fact not associated to sex and is discussed in Supplementary Text.





5 **Fig. 2.** Synonymous divergence between Z and W gametologs along the sex-determining region. A) in the mapping population and B) in the collection of plants from a wild population. The synonymous divergence is the  $dS$  values obtained using codeml on coding sequences (see Material and Methods). Coding sequence length is indicated. Data are shown using both assemblies (Z masked and W masked, see text for more details).



**Table 1.** Results of the *SDpop* analysis of the RNA-seq data from a natural population

	# genes with expression <sup>*</sup>	# genes with informative SNPs <sup>□</sup>	#SNPs	BIC without sex chromosomes	BIC XY	BIC ZW	# autosomal <sup>§</sup>	# z-hemizygous <sup>§</sup>	# ZW <sup>§</sup>
<b>unmasked</b>	24460	17223	420101	5599123	5594811	5590632	7518	9	49
<b>NoW</b>	24415	17195	420766	5601968	5597448	5590920	7741	9	62
<b>NoZ</b>	24418	17206	420047	5443303	5438903	5432196	7696	9	57

\* at least 1 read

5 □ at least 1 individual of each sex had more than three reads, and the gene had at least 1 SNP

§ based on the ZW model, posterior probability threshold for classification is 0.8, BIC = Bayes Information Criterion