

Structural variation of a sex-linked region confers monoecy and implicates *GATA15* as a master regulator of sex in *Salix purpurea*

Brennan Hyden¹ , Junzhu Zou^{1,2} , Dustin G. Wilkerson¹ , Craig H. Carlson^{1,3} , Ayiana Rivera Robles¹ , Stephen P. DiFazio⁴  and Lawrence B. Smart¹ 

¹Horticulture Section, School of Integrative Plant Science, Cornell University, Cornell AgriTech, Geneva, NY 14456, USA; ²Research Institute of Forestry, Chinese Academy of Forestry, Dongxiao No. 1, Haidian District, Beijing 100091, China; ³Cereal Crops Research Unit, Edward T. Schafer Agricultural Research Center, USDA-ARS, Fargo, ND 58102, USA;

⁴Department of Biology, West Virginia University, Morgantown, WV 26506, USA

Author for correspondence
Lawrence B. Smart
Email: lbs33@cornell.edu

Received: 21 November 2022
Accepted: 21 February 2023

New Phytologist (2023) **238**: 2512–2523
doi: 10.1111/nph.18853

Key words: dioecy, genome assembly, monoecy, QTL, Salicaceae, sex determination, willow.

Summary

- The Salicaceae, including *Populus* and *Salix*, are dioecious perennials that utilize different sex determination systems. This family provides a useful system to better understand the evolution of dioecy and sex chromosomes.
- Here, a rare monoecious genotype of *Salix purpurea*, 94003, was self- and cross-pollinated and progeny sex ratios were used to test hypotheses on possible mechanisms of sex determination. To delimit genomic regions associated with monoecious expression, the 94003 genome sequence was assembled and DNA- and RNA-Seq of progeny inflorescences was performed.
- Based on alignments of progeny shotgun DNA sequences to the haplotype-resolved monoecious 94003 genome assembly and reference male and female genomes, a 1.15 Mb sex-linked region on Chr15W was confirmed to be absent in monoecious plants. Inheritance of this structural variation is responsible for the loss of a male-suppressing function in what would otherwise be genetic females (ZW), resulting in monoecy (ZW^H or WW^H), or lethality, if homozygous (W^HW^H).
- We present a refined, two-gene sex determination model for *Salix purpurea*, mediated by *ARR17* and *GATA15* that is different from the single-gene *ARR17*-mediated system in the related genus *Populus*.

Introduction

Dioecy, the occurrence of separate sexes, is almost universal in animals, but is relatively rare among plants, occurring in only 6% of angiosperm species (Renner, 2014). However, the genetic regulation of dioecy varies widely across taxa. In XY systems, the male is the heterogametic sex, such as in mammals (Painter, 1921), *Drosophila* (McKee & Karpen, 1990), and *Asparagus* (Harkess *et al.*, 2020), while in ZW systems, the female is heterogametic, as in birds (Irwin, 2018) and pistachio (*Pistacia vera* L.) (Kafkas *et al.*, 2022). Other, less common, examples of sex inheritance include the UV system in bryophytes where the gametophyte is the dominant generation, in *Apis* where males are hemizygous at the *csd* locus and females are homozygous (Cho *et al.*, 2006), and environmentally determined sexes in reptiles (Bachtrog *et al.*, 2014). For plants in particular, dioecy is present across nearly every major clade and thought to have evolved independently on as many as 5000 occasions (Renner, 2014). There are two prevailing hypotheses for the genetic regulation of dioecy:

the single-gene and two-gene models. In a two-gene model, one gene promotes the expression of one sex and another suppresses characteristics of the other sex (Charlesworth, 2002). Two-gene sex determination systems have been identified in *Asparagus*, where sex is regulated by the transcription factor *TDF1* and the domain of unknown function gene *SOFF* (Harkess *et al.*, 2020), and in *Actinidia*, which involves a cytokinin regulator and a *fasciclin*-like gene (Akagi *et al.*, 2018, 2019). In a single-gene model, one gene is entirely responsible for expression of both sexes, as is the case in *Diospyros* via a small RNA (Akagi *et al.*, 2014) and *Populus* via the cytokinin response regulator *ARR17* (Müller *et al.*, 2020). Sex chromosomes include a region of reduced recombination surrounding the sex determination genes, known as the sex determination region (SDR; Charlesworth *et al.*, 2005). Sex determination loci have also been shown to shift locations and heterogametic systems. The underlying causes have been hypothesized to be driven by linkage with a sexually antagonistic gene or emergence of an entirely new sex determination gene that confers a greater fitness advantage (van

Doorn & Kirkpatrick, 2010; but see also Lenormand & Roze, 2022).

Despite the maintenance of separate sexes in dioecious plants, monoecious and hermaphrodite plants have been observed in dioecious species. Examples range from populations where male, female, and hermaphroditic plants are all common, such as in domesticated *Carica papaya* L. (Wang *et al.*, 2012), to *Populus*, where monoecy is relatively rare (Stettler, 1971; Rowland *et al.*, 2002). In other species, such as *Cannabis sativa* L., monoecy can be induced by manipulating phytohormones, as can feminization of males and masculinization of females (Mohan Ram & Sett, 1982; Lubell & Brand, 2018). These monoecious and hermaphroditic genotypes within otherwise dioecious species constitute a powerful tool in understanding the regulation and evolution of sex determination. In *Vitis*, which is ancestrally dioecious, selection for hermaphrodites during domestication enabled more effective cultivation and, using sequencing of hermaphrodites, the precise mechanism and genes responsible for dioecy were identified (Massonnet *et al.*, 2020). In domesticated *C. papaya*, comparison of the Y^h chromosome responsible for hermaphroditism with the Y and X chromosomes has facilitated the study of sex determination in this genus, which is otherwise entirely dioecious (Wang *et al.*, 2012; Antunes Carvalho & Renner, 2015).

The Salicaceae family, which includes *Populus* and *Salix*, has emerged as a model system for understanding the evolution of sex determination in plants due to the presence of both XY and ZW systems as well as its economic importance for biomass production (Gouker *et al.*, 2021). Both *Populus* and *Salix* are thought to have diverged from a dioecious common ancestor and, throughout their evolution, speciation, and diversification, dioecy has been largely maintained (Wang *et al.*, 2022). In *Populus*, Chr19 XY, Chr19 ZW, and Chr14 XY systems have all been identified across species (Müller *et al.*, 2020; Yang *et al.*, 2020), while in *Salix*, sex is determined by Chr07 XY, Chr15 XY, and Chr15 ZW systems in different species (Sanderson *et al.*, 2021; Wang *et al.*, 2022; Wilkerson *et al.*, 2022). The Chr07 XY system, which is present in the tree willows including *S. chaenomeloides* Kimura (Wang *et al.*, 2022) and *S. nigra* Marshall (Sanderson *et al.*, 2021), is thought to represent the ancestral state of sex determination in *Salix*, with a shift to Chr15 followed by a switch to ZW heterogamety occurring before the divergence of the shrub willows, which include *S. purpurea* L. (Wang *et al.*, 2022).

The Z and W sex chromosomes have been sequenced and assembled from the *S. purpurea* female clone 94006 (Zhou *et al.*, 2020). Chr15W is 15.7 Mb in length and consists of a 6.73 Mb pericentromeric SDR flanked by pseudoautosomal regions (PARs) of 2.34 and 6.58 Mb, respectively. Chr15Z, on the contrary, is only 13.3 Mb, with an SDR region of *c.* 4 Mb (Zhou *et al.*, 2020). This was the first study to report both fully assembled and annotated sex chromosomes in *Salix*, and the release of this genome sequence assembly has enabled additional efforts to characterize sex determination in both *S. purpurea* and related species.

ARR17 has been implicated as the single-gene master regulator of sex in *Populus* (Müller *et al.*, 2020) and is expressed in ZW

and XX females, absent from ZZ males, and in XY males is silenced by small RNAs produced from exon 1 repeats on the Y chromosome (Yang *et al.*, 2020). However, in *Salix*, the precise genes involved in sex determination are less clear (Wang *et al.*, 2022). In some species, such as Chr15 XY *S. arbutifolia* Pall. and Chr07 XY *S. chaenomeloides*, *ARR17* is hypothesized to mediate sex determination via a similar system as XY poplars (Wang *et al.*, 2022). However, in *S. purpurea*, *ARR17* is expressed in both males and females and does not exhibit the same small RNA silencing mechanism (Hyden *et al.*, 2021), which suggests there is a unique sex determination mechanism in this species (Hyden *et al.*, 2021).

Although it is rare for *Salix* to have both staminate and pistillate flowers borne on the same inflorescence, monoecious plants have been observed in nature (Mirski, 2014). One such monoecious individual, *S. purpurea* clone 94003, was collected from a natural stand in central New York (Carlson *et al.*, 2019). A spontaneous monoecious plant in this dioecious species presents an opportunity to refine existing models describing the mechanism and evolution of sex determination in *Salix* by investigating the perturbations that result in monoecy. Here, self- and cross-pollinated families were developed to: characterize monoecious expression across families and environments; determine genomic regions associated with sex- and biomass-related traits via linkage mapping; and build a refined model of sex determination using evidence obtained from pairwise alignments of whole-genome and transcriptome data from parents and segregating progeny.

Materials and Methods

Catkin sex incidence surveys

The monoecious clone 94003 was planted in three randomized 10-plant plots at Cornell University's Gates West farm in Geneva, NY, 18 randomized four-plant plots at the Cornell Lake Erie Research and Extension Laboratory in Portland, NY, and six randomized four-plant plots in Morgantown, WV, USA. A subset of shoots with floral buds were sampled from the Geneva trials in early March, *c.* 40 d before normal flowering, and forced in a glasshouse. A randomly sampled set of floral shoots were collected from field-grown plants in Portland and Morgantown in early April after catkins had emerged and immediately before anthesis (Supporting Information Fig. S1), the shoots were placed in water, and scoring of percent maleness by catkin was conducted in the glasshouse. Scoring for catkin percent maleness was done in the field on all floral shoots in the Geneva location in mid-April during catkin flowering. Percent maleness was measured by catkin on a continuous scale from 0 to 100% maleness based on the ratio of male-to-female flowers on an individual catkin (Fig. S1). To minimize rater variability, the same person made the counts for all floral shoots. For analysis and fitting of a generalized linear model for percent maleness, the data were divided by 100 to obtain values on the interval of zero to one and then adjusted to make the

data noninclusive of zero and one, following the beta distribution. The equation for the data adjustment is:

$$(y(n-1) + 0.5)/n$$

A generalized linear model was created using the function 'glmmTMB' in the R package GLMMTMB, with the link function set to beta (Brooks *et al.*, 2017). Glasshouse was included as a fixed covariate of one or zero depending on whether or not the shoots were brought into the glasshouse in March or left in the field, but it was not found to be significant and therefore was excluded from the final model. Field was included as a fixed effect to determine whether there were any macro-environment effects on percent maleness. Plot and shoot within plot were included as random effects to determine whether there were any microenvironment effects or plant physiological effects on percent maleness, respectively.

The following generalized linear model used:

$$Y_{ijkl} = \beta_0 + \Gamma_{ijk} + \Delta_{ik} + \zeta_l + \epsilon_{ijklm}$$

where Y is the percent maleness of catkin i on shoot j in plot k in field l .

Monoecious family crosses and sequencing

Field-grown shoots of 94003 were collected and forced in a glasshouse in February 2015. Catkins were self-pollinated by hand and seeds collected following maturation to produce the S_1 family, 15X-229. A total of 549 seedlings from this S_1 family were planted at Cornell AgriTech (Geneva, NY, USA) in June 2015 (Table 1). From 2017 to 2022, plants were scored visually each April to determine sex (male, monoecious, or female). In spring 2018, shoots from a single, monoecious S_1 individual, 15X-229-534 (referred to as 'Ze') were collected, forced in a glasshouse, and self-pollinated to generate the S_2 family 18X-551. A total of 437 seedlings from this S_2 family were planted at the Cornell AgriTech in June 2018 (Table 1). Cuttings from 345 18X-551 S_2 progeny were planted in three plant plots, organized as a randomized complete block design at Cornell AgriTech in June 2021. Height and leaf area were collected in June 2022. Leaf area was determined using the IMAGE analysis software LeafByte on a single leaf collected from each plot (Getman-Pickering *et al.*, 2020). Height data were measured in late June 2022 in the field on a single plant in each plot from the soil to the top of the tallest stem of the plant. Plants in the 18X-551 family were scored visually for sex annually each spring from 2019 to 2022. Families 19X-232 and 19X-562 were generated by crossing 94003 with the female clone 94006 and the male clone 'Fish Creek', respectively, and 222 and 328 individuals of 19X-232 and 19X-562, respectively, were planted at the Gates West Farm in June 2019 (Table 1). Plants in those families were scored for sex annually from 2020 to 2022. Plants were considered monoecious if over the course of multiple years of observations both staminate and pistillate flowers were observed at least once.

Sequencing of 94003 and progeny

High-molecular-weight DNA was extracted from 94003 using a protocol described by Mayjonade *et al.* (2016) followed by DNA clean up in 500 μ l, 24 : 1, chloroform : isoamyl alcohol, centrifugation at 8000 g , and ethanol precipitation. Resulting DNA had a peak at 15 509 bp as determined by analysis on a Femto Pulse (Agilent Technologies, Santa Clara, CA, USA). PacBio Sequencing was performed at the Mount Sinai Research Institute (New York, NY, USA) to generate *c.* 2.5 million HiFi reads and 140 $\times$ HiFi read coverage. HiFi reads were assembled using hifiasm (Cheng *et al.*, 2021) and aligned using MINIMAP2 (Li, 2018) to the *S. purpurea* 94006 v.5.1 reference genome to scaffold contigs (Goodstein *et al.*, 2011; Zhou *et al.*, 2020). Annotation was performed using Maker (Cantarel *et al.*, 2008) with snap *ab initio* gene prediction. Additional high-quality DNA for Illumina sequencing was extracted using a Qiagen DNeasy kit, and paired-end (2 $\times$ 150 bp) Illumina sequencing was performed at West Virginia University (Morgantown, WV, USA). Illumina adapters were removed using trimmomatic (Bolger *et al.*, 2014) and mapped to the *S. purpurea* 94006 v.5.1 reference genome using bwa-mem (Li & Durbin, 2009), with 11 $\times$ coverage.

A subset of monoecious and putative male and female individuals from the 18X-551 S_2 , 19X-232 female outcross, and 19X-562 male outcross families were selected for Illumina DNA and RNA sequencing. Monoecious individuals were selected from those that demonstrated the greatest growth and vigor in the field. Since monoecious plants producing a low number of catkins or only a single year of flowering data may be falsely identified as male or female, putative males and females for sequencing were chosen from those with the greatest number of catkins and years of flowering data, such that there was high confidence they were identified as true males and females. From 18X-551 (S_2 family), three monoecious individuals were selected for resequencing; from 19X-232 (female outcross family), five females, three males, and five monoecious plants were sequenced; and from 19X-562 (male outcross family), five monoecious and three male plants were sequenced. Paired-end Illumina DNA sequencing was performed at West Virginia University, generating *c.* 3 billion bp per sample and averaging 15 $\times$ coverage. Illumina reads were aligned to the *S. purpurea* 94006 v.5.1 reference, as well as the combined references of their respective parental genomes ('Fish Creek' v.3.1, https://phytozome-next.jgi.doe.gov/info/SpurpureaFishCreek_v3_1 (last accessed 4 November 2022); '94006' v.5.1 https://phytozome-next.jgi.doe.gov/info/Spurpurea_v5_1 (last accessed 4 November 2022)).

RNA-Seq of 94003 and progeny

RNA-Seq data were generated from monoecious parent clone 94003 and 23 progeny: Four monoecious clones from 18X-551 (S_2 family); three male, five monoecious, and five female clones from 19X-232 (female outcross family); and four monoecious and two male clones from 19X-562 (male outcross family). For RNA-Seq, emerging catkins before pistil

Table 1 Descriptions of two self-crossed and two outcrossed families generated using 94003 as a parent.

Family and cross type	Mother	Father	Progeny planted	Sex*	1 yr	2 yr	≥ 3 yr	Total	%
15X-229 (S ₁)	94003	94003	549	H	61	40	138	239	65.10
				M	22	22	48	92	25.10
				F*	34	2	0	36	9.80
18X-551 (S ₂)	'Ze'**	'Ze'	437	H	89	56	46	191	70.50
				M	47	17	12	76	28
				F*	3	1	0	4	1.50
19X-562 (F ₁ , Male outcross)	94003	'Fish Creek'	328	H	72	71	4	147	48.70
				M	61	74	8	143	47.40
				F*	11	1	0	12	4
19X-232 (F ₁ , Female outcross)	94006	94003	222	H	15	24	2	41	25
				M	9	13	4	26	15.90
				F	50	42	5	97	59.20

Distribution of observed sexes in each family and number of plants with one, two and three or more years of data supporting the observed sex.

*Putative monoecious plants misidentified as female.

**'Ze' is a monoecious S₁ individual derived from a self-cross of 94003.

receptivity and anther dehiscence were collected and flash-frozen in liquid nitrogen, between 10:00 h and 14:00 h, and then stored at -80°C . For RNA isolations, catkins were ground into a fine powder, and RNA isolated using the Spectrum™ Total Plant RNA Kit with DNase I digestion (Sigma). For 94003, we used paired-end, 2×150 bp RNA-Seq Illumina data that were generated as described in Hyden *et al.* (2021), using the same library prep and sequencing platform. Three-prime (3') RNA-Seq single-end Illumina reads (84 bp) were generated for the 23 progeny at Cornell University (Ithaca, NY, USA). RNA-Seq reads from 94003 and its progeny were aligned to the *S. purpurea* 94006 v.5.1 reference genome, using STAR (Dobin *et al.*, 2013). FEATURECOUNTS was used to calculate read count data by gene model (Liao *et al.*, 2014), and DESeq2 was used to test for differential expression between male ($n = 5$), female ($n = 5$), and monoecious progeny ($n = 13$; Love *et al.*, 2014). Log₂ fold-change of monoecious : female and monoecious : male expression between 94003 progeny and FDR-adjusted *P*-values were calculated in DESeq2. For 94003, log₂ fold-change was calculated using the same methods, but against 90 male and 90 female full-sib genets in F₂ *S. purpurea* family 317 family described by Hyden *et al.* (2021), since samples were collected at the same time and location as 94003, and used the same RNA isolation, library preparation, and sequencing methods.

Bisulfite sequencing analysis

For bisulfite sequencing of 94003, a single catkin was collected in April 2020 and flash-frozen. Genomic DNA was isolated using a Qiagen DNeasy Plant mini kit following the recommended protocol. Sequencing was performed using an Illumina NovaSeq 6000 system at the Joint Genome Institute (Berkeley, CA, USA). Read alignment to the *S. purpurea* v.5.1 reference was performed using Bismark (Krueger & Andrews, 2011), and methylated regions called using DMRfinder (Gaspar & Hart, 2017) and differential methylation determined relative to

six unrelated males and six unrelated females, as described in Hyden *et al.* (2021).

S₂ genotyping-by-sequencing data collection and QTL mapping

DNA was extracted from leaf tissue of 423 individuals in the S₂ family 18X-551, and genotyping-by-sequencing (GBS) was conducted using *ApeKI* at the University of Wisconsin Biotechnology Center. GBS reads were aligned to the *S. purpurea* 94006 v.5.1 reference genome using BWA-MEM (Li & Durbin, 2009) and the TASSEL GBS v.2 Discover Pipeline (Bradbury *et al.*, 2007) was used to call and filter SNPs ($\text{maf} \geq 0.05$; Call Rate > 50%). A linkage map was generated using intercross markers in R with a combination of R/QTL (Broman *et al.*, 2003), ASMAP (Taylor & Butler, 2017), and custom code (code and map data available at: <https://github.com/Willowpedia/Hyden-et-al-Monoecious-manuscript>). Briefly, this code first removes co-locating and extremely distorted markers using PULLCROSS and then forms linkage groups with MSTMAP. Next, simple error correction was performed to reduce the occurrence of double crossovers, deflating and stabilizing the linkage groups. Lastly, the quality of the linkage map was visually assessed using HEATMAP and by checking the marker order of the linkage group in relation to their physical position in the reference genome. The linkage map yielded 23 major linkage groups, numbered according to the physical 94006 v.5.1 reference genome. Considering a normal model for leaf area and plant height, and a binary model for sex (males = 0, monoecious/female = 1), scanone in R/QTL (Broman *et al.*, 2003) was used to perform linkage mapping ($n.\text{perm} = 1000$, $\text{conf} = 0.95$). In addition, Chr15 was also used as a covariate to identify additional loci associated with monoecy. LOD support intervals of QTL contained all markers with LOD scores greater than the genome-wide permutation score. Candidate genes were identified manually, by scanning annotations with functional relevance within a given LOD support interval or ≤ 1.5 Mb from the interval boundaries.

Results

Surveys of monoecious clone 94003 in different environments

After modeling catkin percent maleness using a generalized linear mixed model, field, plot, and shoot within plot were all found to have significant effects (Table S1). The average percent maleness was significantly greater in the Morgantown and Portland locations than in Geneva (Fig. S2).

Differential gene expression

A total of 36 661 genes were expressed in 94003 catkin tissue. Of these, 3248 genes showed significantly different expression in 94003 compared with females from an F_2 family and 1633 genes with differential expression compared with F_2 males (FDR < 0.05, Table S2; Hyden *et al.*, 2021). A total of 30 158 genes were expressed among the 94003 progeny, including 1947 differentially expressed genes (DEGs) in comparisons between male and female progeny, 411 DEGs between male and monoecious progeny, and 1000 DEGs between female and monoecious progeny (Table S3).

Bisulfite sequencing

There was differential methylation in 13 406 sites of the 94003 genome when compared to six unrelated females, including 8588 sites with greater methylation in 94003. There was differential methylation at 10 895 sites in the 94003 genome compared with six unrelated males, including 6439 with greater methylation in 94003. In total, 795 genes contained promoters with differentially methylated regions. (Table S4).

Sex ratios in offspring of monoecious clone *S. purpurea* 94003

Sex ratios in the S_1 , S_2 , and outcrossed families are summarized in Table 1. The small number of females identified and their

scarcity after three or more years of survey data suggests that plants visually scored as females may have been misidentified based on phenotypic observations due to limited sampling and are genetically monoecious. With this assumption, we observed an $\approx 3 : 1$ monoecious : male segregation ratio in both the S_1 and S_2 families. The 19X-562 male outcross family had a $1 : 1$ ratio of male to monoecious plants, while the 19X-232 female outcross family had a $2 : 1 : 1$, female : monoecious : male ratio.

Sequencing of monoecious clone 94003 and its progeny

The final diploid assembly of 94003 was 833 Mb with 447 Mb assigned to haplotype 1 and 406 Mb assigned to haplotype 2. The core eudicot complete BUSCO score of the diploid assembly was 98.3%. Alignment of contigs against the 94006 v.5.1 genome revealed four or fewer contigs per chromosome, including multiple fully assembled chromosomes (Fig. S3). The annotation consisted of 58,058 mRNA models across the diploid genome, with a BUSCO score of 92.9%. Assembly of *S. purpurea* 94003 HiFi PacBio contigs revealed a 1.153 Mb deletion in the Chr15W SDR in 94003, spanning 7.715–8.444 Mb and 8.664–9.088 Mb, relative to the *S. purpurea* 94006 v.5.1 reference genome (Fig. 1). The haplotype containing this deleted region, common to all monoecious plants, is referred to as Chr15W^H. The 94003 Chr15W^H was assembled into a single contig, with a length of 19.6 Mb, which is considerably longer than the 15.7 Mb 94006 Chr15W (Zhou *et al.*, 2020). When aligned to the 94006 Chr15W, the 94003 Chr15W^H exhibited high similarity across the PARs, and a series of inversions, repeats in the SDR, along with the 1.153 Mb deletion (Fig. 1). This unusual mapping pattern in the SDR region is likely a result of a misarrangement of the SDR contigs in the older 94006 reference, which did not use PacBio HiFi sequences and relied on a linkage map to scaffold the contigs (Zhou *et al.*, 2020). Since there is no recombination in the SDR, the authors had no way of determining the precise arrangement of contigs in the 94006 Chr15W SDR. Moreover, the accumulation of repetitive regions and mutations in the SDR resulting from a lack of recombination likely further

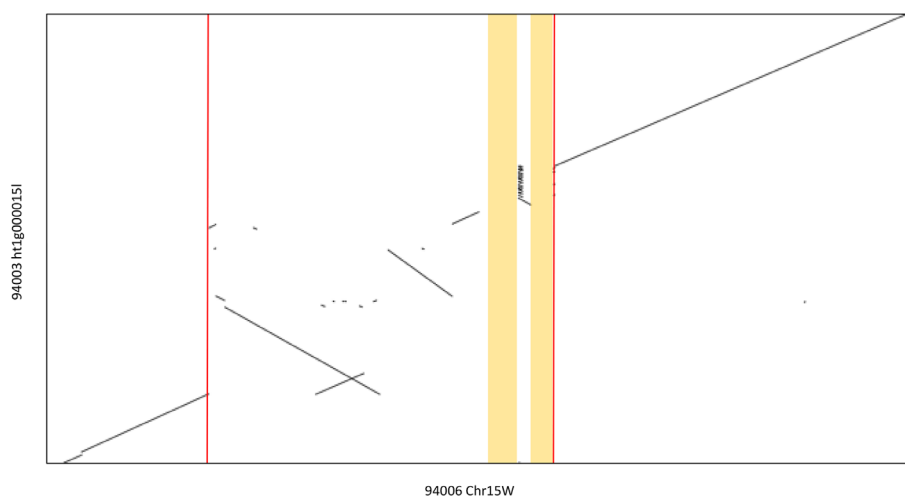


Fig. 1 Dot plot showing alignment of 94003 Chr15W^H sequence to the 94006 v.5.1 Chr15W sequence assembly. The 94003 ht1g0000151 sequence contains the Chr15W^H SDR. Highlighted regions indicate deletions in 94003, which include 68 genes in 94006. Red vertical lines indicate the boundaries of the SDR in 94006 Chr15W.

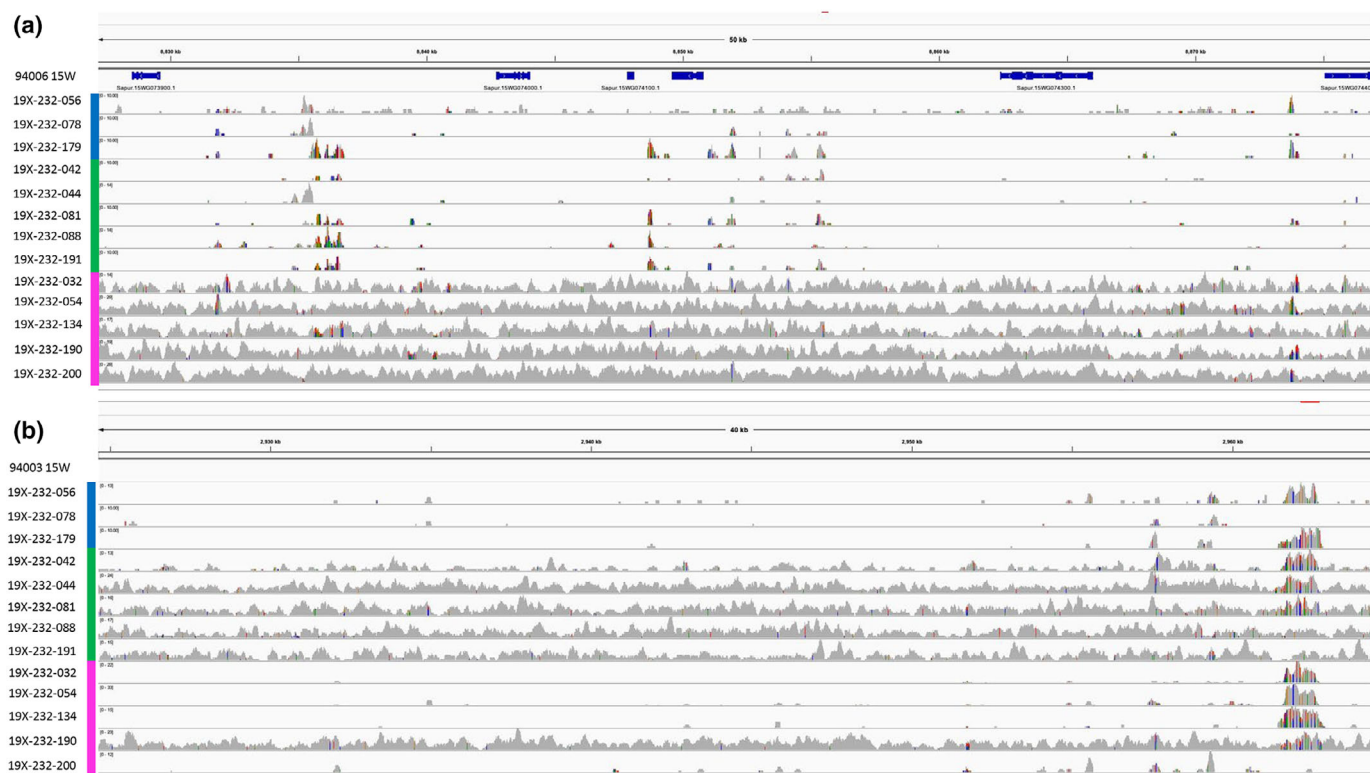


Fig. 2 IGV alignment of DNA sequences from 19X-232 (94006 × 94003) progeny to Chr15 sequence assemblies. (a) Alignment to the 94006 Chr15W region that is deleted in monoecious plants. (b) Alignment to the 94003 Chr15W^H region adjacent to the deleted region. Males (blue) do not significantly align to either 94006 or 94003 (i.e. ZZ genotype). Monoecious plants (green) align to 94003, but not to 94006 (i.e. ZW^H genotype). Four females (magenta) align to 94006, but not 94003 (i.e. ZW genotype). One female aligns to both 94003 15W^H and 94006 15W (i.e. WW^H genotype).

contributes to the observed differences in length and alignment between 94006 and 94003 (Zhang *et al.*, 2014).

Little-to-no read coverage in this portion of the SDR in the 94006 Chr15W assembly was observed for monoecious plants in the 18X-551, 19X-232, and 19X-562 families (Fig. 2a). Four female progeny had consistent read alignments to 94006 Chr15W, but not 94003 15W^H (Fig. 2b), which is consistent with a ZW genotype, and all males lacked substantial read coverage on either of the Chr15W assemblies, which suggests a ZZ genotype. However, one female, 19X-232-190, aligned to both 94003 Chr15W^H and 94006 Chr15W assemblies, indicating a WW^H genotype (Fig. 2).

The deleted region from 94003 contains 68 gene models in the 94006 Chr15W, including nearly all of the candidate master regulator genes of sex described by Hyden *et al.* (2021). The sole exception was a *GATA15* homolog that occurs 626 kb away from the deleted region in 94003.

QTL mapping of vegetative traits

A total of 423 individuals and 1019 markers across 23 linkage groups were included in the final linkage map of S₂ family 18X-551. A total of 13 QTL were detected for vegetative traits (height and leaf area) on linkage groups 1, 4, 5, 6, 7, 8, 16, 17, and 19 (Fig. 3). Thirteen candidate genes for height and one candidate gene for leaf area were identified within QTL confidence

intervals, including homologs of several auxin response genes, as well as *CLAVATA* and *PIN* (Table 2).

QTL mapping of monoecious and male sexes

In addition to mapping vegetative traits, the S₂ family linkage map was also used to confirm the genetic inheritance of the monoecious and male sexes. LOD support for sex was restricted to Chr15, however, when the Chr15 genotype was used as a covariate, a single QTL on Chr07 was observed (Fig. 3; Table S5). Homologs of *GATA18* and *AP3* were identified as candidates for epistatic interactions with the sex determination region for the regulation of sex (Table 2).

Discussion

The stable monoecious phenotype of 94003 is the result of a natural disruption of dioecy in *S. purpurea* and is a powerful resource for understanding how sex is controlled in this species. This study is the first effort in the Salicaceae family that sought to understand the genetic basis of monoecy by developing novel self- and cross-pollinated families segregating for monoecy, as well as genome sequencing of a monoecious plant. The heritability of monoecious phenotypic expression across environments confirms its genetic stability and its unique potential in future shrub willow breeding and genetics efforts.

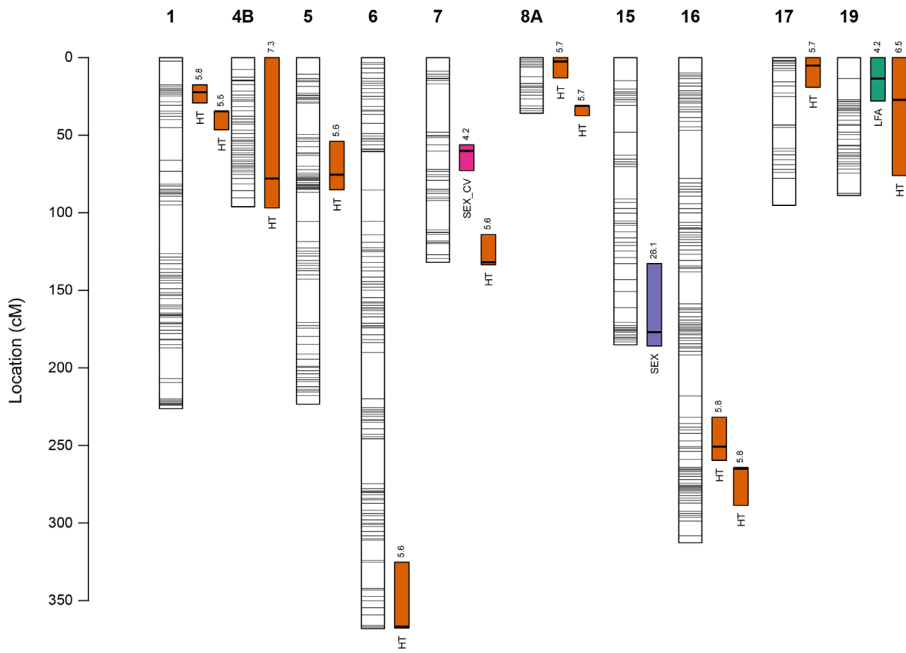


Fig. 3 QTL confidence intervals placed on a genetic linkage map of *Salix purpurea* derived from the 18X-229 S₂ family. Bars represent linkage groups in centiMorgans (cM) with genotyping-by-sequencing markers indicated by lines. Shaded bars represent the confidence interval with a line indicating the most significant marker. Numbers above each bar are the maximum LOD scores for each analyzed trait. Bars are colored according to trait group: orange, height (HT); green, leaf area (LFA); purple, sex; magenta, sex with Chr15 used as a covariate (SEX_CV).

Table 2 Candidate *Salix purpurea* genes for height, leaf area, and sex located in QTL confidence intervals.

Salix gene model	Chr	Description
<i>Plant height</i>		
Sapur.001G016500	01	<i>SLO1</i> homolog, slow growth 1
Sapur.004G101300	04	<i>CDC2</i> homolog, cell division control
Sapur.004G117700	04	<i>SAUR</i> -like auxin-responsive protein
Sapur.005G098100	05	<i>GATA18</i> homolog, regulates SAM
Sapur.006G042900	06	<i>ROTUNDIFOLIA</i> like 7, meristem maintenance and shoot development
Sapur.007G103300	07	<i>SAUR</i> -like auxin-responsive protein
Sapur.016G008100	16	Auxin-responsive family protein
Sapur.016G031200	16	<i>PIN1</i> homolog, specifies shoot/root auxin gradient and development
Sapur.016G159400	16	<i>CLE22 CLAVATA3/ESR-RELATED 22</i> homolog
Sapur.016G166000	16	Lateral Organ Boundaries (LOB) protein
Sapur.019G051200	19	<i>PIN5</i> homolog, auxin efflux carrier protein
Sapur.019G086600	19	<i>CLE4</i> homolog, <i>CLAVATA3/ESR-RELATED 4</i>
Sapur.019G090900	19	Auxin-responsive family protein
<i>Leaf area</i>		
Sapur.019G051200	19	<i>PIN5</i> homolog, auxin efflux carrier protein
<i>Sex covariate</i>		
Sapur.007G014400	07	<i>AP3</i> homolog, specifies stamen identity
Sapur.007G021900	07	<i>GATA18</i> homolog, regulates SAM/floral development

Gene descriptions were obtained from TAIR (Lamesch *et al.*, 2011) and UniProt (UniProt Consortium, 2018). chr, chromosome.

Environment had a significant effect on the expression of individual floral sex ratio in 94003 across different sites. Catkins exhibited a broad range of male : female sex ratios, including whole catkins and floral shoots that had entirely pistillate or staminate flowers. This is noteworthy as it suggests monoecious plants can be misidentified as either male or female, particularly

if there is only a small sample size of catkins. Notably, hermaphrodite flowers were not observed in any of the environments. The results from the general linearized model show a significant association between field, plot, and shoot conditions on the floral sex ratio and suggest that both macro and microenvironmental factors may be influencing percent maleness (Table S1). In Central New York, *S. purpurea* forms floral meristems in July, while differentiation of anther primordia occurs in August and microspore mother cells differentiate in September (Zhang & Fernando, 2005). Data obtained from NOAA (Vose *et al.*, 2014) and the USDA Web Soil Survey (<https://websoilsurvey.sc.egov.usda.gov/>, accessed 21 May 2019) for the three trial locations during these months in 2018 show both greater soil drainage and higher temperatures at both the Morgantown and Portland sites compared with Geneva during summer 2018 (Table S6). Catkins at these locations also exhibited a greater percent maleness in this study. Warm temperatures and low water availability have been found to promote maleness in other monoecious species (Grant *et al.*, 1994). However, further evaluation across multiple sites and growing seasons would be required to elucidate the precise role that specific environmental factors have on floral sex ratios.

As expected, a single, high-effect QTL for sex was observed in the SDR of Chr15. However, when Chr15W was included as a covariate, a new QTL was identified on Chr07 (Fig. 3; Table S5). This is particularly interesting as it may suggest a role of this locus in the monoecious phenotype that acts downstream of the SDR. This QTL contains *GATA18* and *APETALA3* homologs (Table 2). *APETALA3* is a MADS-box gene that associates with *PISTILLATA* to specify stamen identity (Krizek & Meyerowitz, 1996), while *GATA18* is thought to be involved in floral development (Manfield *et al.*, 2007; Kanei *et al.*, 2012). Moreover, Chr07 is the location of the SDR in tree willows and may be the ancestral SDR locus for all *Salix* (He *et al.*, 2021; Sanderson *et al.*, 2021; Wang *et al.*, 2022).

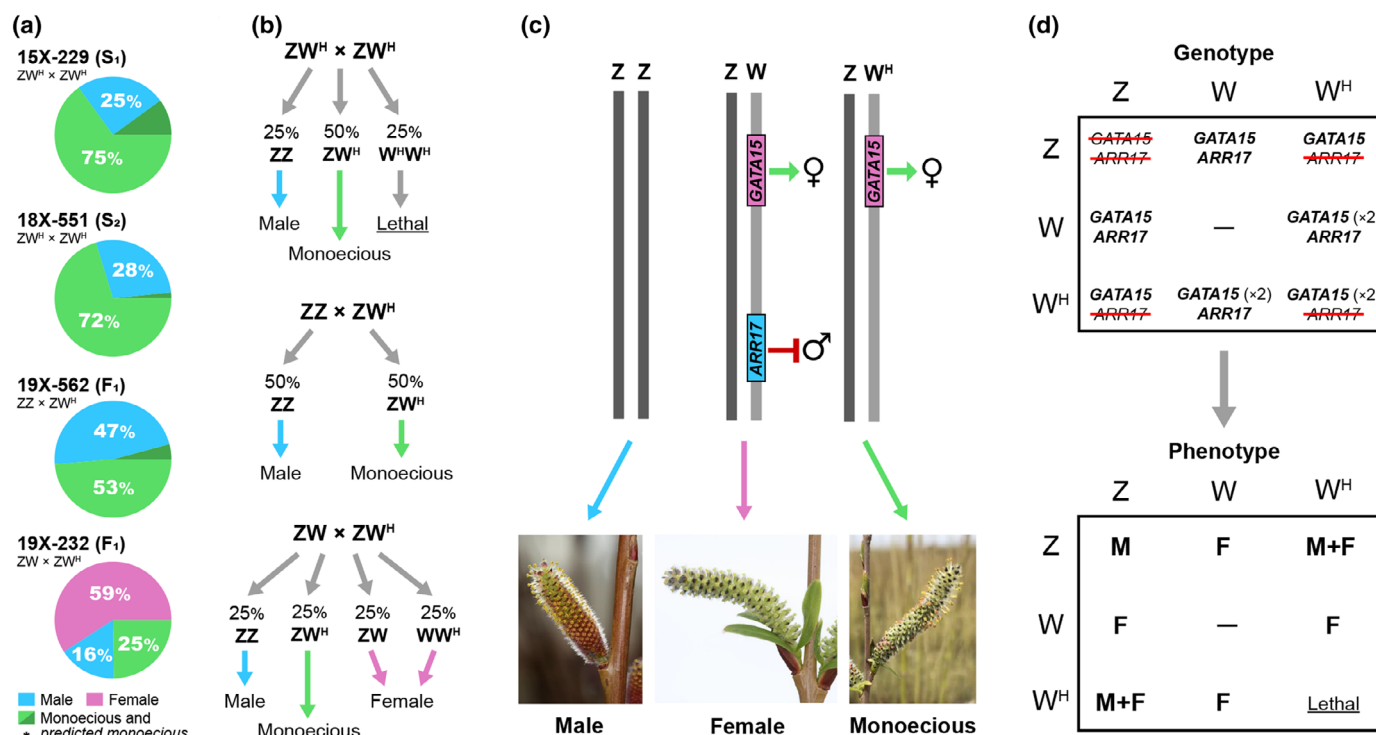


Fig. 4 Segregation ratios for sex in monoecious families and models of dioecious and monoecious sex determination mechanisms. (a) Pie charts showing observed sex segregation ratios for monoecious families. (b) Proposed inheritance models for the Chr15W SDR region in the self-crossed, male outcross, and female outcross families. (c) Hypothesized model for sex determination in *Salix purpurea*. ZZ plants lack a 15 W chromosome and produce male flowers. ZW plants contain both sex determination genes, where ARR17 suppresses male floral development and GATA15 simultaneously induces female flowers, resulting in female plants. Monoecious plants contain GATA15, but lack other candidate sex determination genes, causing induction of female flowers without suppression of male flowers, resulting in mixed catkins; (d) Punnett square model showing genotypic and phenotypic expression of sex in each sex chromosome combination. Pie charts and arrows in (a), (b) and lower panel of (c) are colored to indicate sex: blue, male; magenta, female; green, monoecious. Arrows in the upper panel of (c) are colored green to indicate positive regulation and red to indicate negative regulation.

In addition to its use for confirming the inheritance of sex in the S₂ family, the 18X-551 linkage map was also utilized to map vegetative traits and identify candidate genes, which can be informative for biomass yield (Carlson *et al.*, 2019). Many QTL identified in the S₂ family for plant height and leaf area contained genes whose homologs in Arabidopsis have known roles in meristem and organ development, including multiple auxin response genes, *CLAVATA*, and *PIN* homologs (Feraru & Friml, 2008; Somssich *et al.*, 2016; Table 2). These mapping results from self- and cross-pollinated families are proof-of-concept as a tool for genetics studies in *Salix*.

In families segregating for monoecy, the low incidence of females identified in both the S₁ and S₂ families suggests that plants scored as females were misidentified based on initial phenotypic observations but later identified as genetically monoecious. Furthermore, in the S₂ family there were only two genotypes present across the SDR of Chr15 based on SNP analysis: major allele homozygotes (ZZ) and heterozygotes (ZW). Monoecious plants were all heterozygous, while the homozygotes were phenotypically male. The absence of homozygous WW genotypes in the GBS data from the S₂ family strongly suggests that WW plants in this family were not viable (Fig. 4). The monoecious : male ratio in both the S₁ and S₂ families was *c.* 3 : 1 (Fig. 4), rather than a 2 : 1 ratio expected under a WW lethal model, despite the GBS resequencing

data, which only identified ZW and ZZ progeny in the S₂ family. One possible explanation for this skew is that female sex ratio bias, which has been observed in *S. purpurea* (Gouker *et al.*, 2021), may be responsible for creating a 'monoecious bias' in these populations, resulting in a greater than expected number of monoecious plants. Together, these observations support a model in which Chr15W^H in 94003 carries a mutation resulting in ZW plants expressing monoecy instead of being dioecious females. We describe this mutated Chr15W as Chr15W^H according to the nomenclature used to describe the hermaphrodite sex chromosome in *C. papaya* (Wang *et al.*, 2012). The nearly 1 : 1 segregation ratio for males and monoecious plants in the 19X-562 male outcross family confirms the inheritance observed in the S₂ family and supports the hypothesis that inheritance of the 94003 Chr15W^H results in monoecious phenotypic expression (Fig. 4). The *c.* 2 : 1 : 1 female : male : monoecious ratio observed in the 19X-232 female outcross family (Fig. 4), and the DNA mapping results from 19X-232-190 (Fig. 2) suggest that WW^H individuals in this family are viable and result in phenotypically female plants. This, along with the lack of W^HW^H progeny genotypes in the S₂ family, suggests that the mutation in Chr15W^H is lethal in the homozygous state. Under this model, plants that are heterozygous for the 94003 Chr15W^H do survive, and the female phenotype can be 'restored' by a wild-type Chr15W (Fig. 4).

Among the features missing from the 94003 Chr15W^H is the entire palindromic region, characterized by Zhou *et al.* (2020), which contains several candidate genes for the regulation of sex, including homologs of *ARR17*, *AGO4*, and *DRB1* (Zhou *et al.*, 2020; Hyden *et al.*, 2021). Moreover, of the candidate sex determination genes proposed by Hyden *et al.* (2021), only the *GATA15* homolog (Sapur.15WG062800) is present and supported by RNA-Seq gene expression data in 94003 and its monoecious progeny (Fig. 2). Our data suggest that a male suppressor gene has been deleted in the Chr15W^H found in the monoecious clone 94003, leading to male floral development and the monoecious phenotype (Fig. 4). This indicates a role of *GATA15* as a 'female inducer' gene that is normally co-inherited with a gene present in the Chr15W SDR that acts as a male suppressor in a two-gene model. Notably, *GATA15* has been shown to have a role in floral organ initiation and meristem formation in Arabidopsis (Ranftl *et al.*, 2016) and floral organ development in *Lagerstroemia speciosa* L. (Hu *et al.*, 2019), further supporting its role in sex determination. The defect in 94003 offers support for a two-gene model of sex determination in *S. purpurea*, which is different from the single-gene, *ARR17*-mediated sex determination system identified in *Populus* (Müller *et al.*, 2020; Renner & Müller, 2022). In the *S. purpurea* two-gene system, we suggest that *ARR17* serves as an active suppressor of male floral development. RNA-Seq results from 94003 and 13 monoecious progeny confirmed a high level of expression of *GATA15*, as well as a lack of expression for any other sex determination genes, including *ARR17*, *AGO4*, and *DRB1*. These results are consistent with the Illumina and PacBio DNA sequencing results, which showed a gap in the 94003 assembly in the region containing these genes (Fig. 1).

The existence of separate staminate and pistillate flowers rather than hermaphrodite flowers indicates that, early in development, the precursor tissue for individual flowers must commit to differentiate into either androecium or gynoecium, not both. While the precise mechanisms of how *GATA15* mediates the male-to-female switch during early floral development are unknown, particularly in the absence of the male suppressor *ARR17*, it is possible that a dosage threshold of *GATA15* must be reached to produce a female flower rather than a male flower in monoecious plants. This *GATA15* dosage may be stochastic, possibly as a result of random variations in expression, protein degradation, and DNA binding activity, where even subtle differences could affect the differentiation of floral tissue and might explain the wide range of male : female flower ratios observed within the 94003 clone across environments.

With regard to methylation, widespread differences were observed between catkins of 94003 and six diverse male and female *S. purpurea* genotypes. The role of methylation in sex determination in willow has been previously implicated, with nearly threefold differences in methylated regions between male and female catkins being reported (Hyden *et al.*, 2021). Among the differentially methylated promoter regions in 94003 are homologs of six known floral development genes: *UNUSUAL FLORAL ORGANS* (*UFO*), *PISTILLATA* (*PI*), *AGL104*, *AGL44*, *AGL62*, and *AGL32*. These genes are of interest since their transcriptional regulation and role in floral development

may indicate involvement in monoecious phenotypic expression. Sapur.005G143300 is a homolog of *PISTILLATA*, a gene necessary for male floral organ development with high expression in male catkins (Hyden *et al.*, 2021). Sapur.005G143300 shows exceptionally high expression in 94003 (monoecious : male $\log_2FC = 5.21$; monoecious : female $\log_2FC = 11.43$) while also being undermethylated relative to males and females in the distal promoter region 538–1036 bp upstream of the transcription start site (Table S1). In Arabidopsis, *PISTILLATA* forms a heterodimer with *APETALA3* (*AP3*) and an *AP3* homolog is present on the Chr07 QTL peak for sex in the *S. purpurea* S₂ family (Table 2). In 94003, *UFO* shows overmethylation in its promoter region as well as overall lower expression levels relative to F₂ family males and females. In Arabidopsis, *PISTILLATA* expression is initiated by *UFO* binding in the distal promoter region (Honma & Goto, 2000). In *Populus*, *arr17* CRISPR mutants show increased expression of both *UFO* and *PI* homologs and *ARR17* is thought to directly repress the expression of both genes (Montalvão *et al.*, 2022). Taken together, these results suggest a failure to suppress *PISTILLATA* through the deletion of *ARR17* may be responsible for the expression of male flowers on ZW^H plants, resulting in the monoecious phenotype (Fig. 4).

Monoecious willows are a novel tool for investigating fundamental questions regarding sex determination and a valuable resource for breeding and genetics. Here, we showed that crosses of monoecious willows can generate male, female, and monoecious progeny using breeding methods, which was previously considered impossible. The utility of such a cross was demonstrated by QTL analysis for sex- and biomass-related traits. Importantly, using next-generation sequencing, we discovered that several candidate master regulator genes for sex were absent in monoecious plants and present a refined model of sex determination in *S. purpurea*. In this model, we detail evidence which supports a two-gene sex determination system, including *ARR17* and *GATA15*, and implicate *ARR17* as a suppressor of male function. From sex chromosome turnover to shifts between sex heterogametic systems, the dynamic nature of sex in the willows is reinforced by our results and contributes a unique story to the study of sex chromosome evolution in plants.

Acknowledgements

We are grateful for excellent technical and field trial support from Matt Christiansen, Matt Hall, Pete Griner, Curt Carter, Rebecca Wilk, Dawn Fishback, Lauren Carlson, Jane Petzoldt, Deanna Gentner, Allison DeSario, Savanna Shelnut, Alexander Wares, McKenzie Schessl, Michael Quade, and staff at the Cornell Lake Erie Research and Extension Laboratory and for research advice and assistance from Drs. Eric Fabio and Fred Gouker. Support for this research was provided by grants (1542486 and 1542509) from the National Science Foundation Division of Environmental Biology. BH was partially funded by a predoctoral research fellowship from the United States Department of Agriculture (USDA) National Institute for Food and Agriculture (award no. 2021-67034-35116). JZ was partially supported by China Scholarship Council. We acknowledge the WVU Genomics Core Facility, Morgantown, WV, for support provided to help make

this publication possible and CTSI Grant no. U54 GM104942, which in turn provides financial support to the Core Facility. The work by CHC was supported in part by the USDA Agricultural Research Service through project 3060-21000-038-000-D.

Competing interests

None declared.

Author contributions

BH designed and conducted all experiments unless otherwise noted, collected and analyzed data, and wrote the manuscript. JZ created the S_2 linkage map. DGW and CHC provided code for genetic mapping. ARR collected phenotype data. SPD collected data and provided project management and funding. LBS conceived the project, and provided project management and funding support. All authors contributed to manuscript preparation.

ORCID

Craig H. Carlson  <https://orcid.org/0000-0003-2050-5455>
 Stephen P. DiFazio  <https://orcid.org/0000-0003-4077-1590>
 Brennan Hyden  <https://orcid.org/0000-0002-4896-0030>
 Ayiana Rivera Robles  <https://orcid.org/0000-0001-7660-0203>
 Lawrence B. Smart  <https://orcid.org/0000-0002-7812-7736>
 Dustin G. Wilkerson  <https://orcid.org/0000-0003-3996-7430>
 Junzhu Zou  <https://orcid.org/0000-0002-3715-1158>

Data availability

Sequence data used in this study can be found on the NCBI sequence read archive (SRA) under the following BioProjects: PRJNA373058 (94003 RNA-Seq), PRJNA887422 (94003 and progeny Illumina reads), PRJNA887710 (94003 progeny RNA-Seq), and PRJNA887746 (94003 PacBio reads). 94003 bisulfite sequencing data can be found on the JGI genome portal under Project ID: 1287375. The 94003 genome sequence assembly haplotype 1 and 2 assignments can be found on the NCBI Genome database under BioProjects PRJNA8893215 and PRJNA889819, respectively. The 94003 genome sequence unassigned haplotype assembly and annotation is available from the corresponding author upon request. The 94003 genome sequence assembly and annotation have been deposited on Dryad at doi: [10.5061/dryad.0k6djhb4d](https://doi.org/10.5061/dryad.0k6djhb4d). Code used for analysis and the S_2 genetic map are available at: <https://github.com/Willowpedial/Hyden-et-al-Monoecious-manuscript>.

References

- Akagi T, Henry IM, Ohtani H, Morimoto T, Beppu K, Kataoka I, Tao R. 2018. A Y-encoded suppressor of feminization arose via lineage-specific duplication of a cytokinin response regulator in kiwifruit. *Plant Cell* 30: 780–795.
- Akagi T, Pilkington SM, Varkonyi-Gasic E, Henry IM, Sugano SS, Sonoda M, Firl A, McNeilage MA, Douglas MJ, Wang T *et al.* 2019. Two Y-chromosome-encoded genes determine sex in kiwifruit. *Nature Plants* 5: 801–809.
- Antunes Carvalho F, Renner SS. 2015. IV. A dated phylogeny of the papaya family (Caricaceae) reveals the crop's closest relatives and the family's biogeographic history. In: Antunes Carvalho F, ed. *Molecular phylogeny, biogeography and an e-monograph of the papaya family (Caricaceae) as an example of taxonomy in the electronic age*. Wiesbaden, Germany: Springer Spektrum, 49–81.
- Bachtrog D, Mank JE, Peichel CL, Kirkpatrick M, Otto SP, Ashman TL, Hahn MW, Kitano J, Mayrose I, Ming R *et al.* 2014. Sex determination: why so many ways of doing it? *PLoS Biology* 12: e1001899.
- Bolger AM, Lohse M, Usadel B. 2014. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* 30: 2114–2120.
- Bradbury PJ, Zhang Z, Kroon DE, Casstevens TM, Ramdoss Y, Buckler ES. 2007. TASSEL: software for association mapping of complex traits in diverse samples. *Bioinformatics* 23: 2633–2635.
- Broman KW, Wu H, Sen S, Churchill GA. 2003. R/qtl: QTL mapping in experimental crosses. *Bioinformatics* 19: 889–890.
- Brooks ME, Kristensen K, Van Benthem KJ, Magnusson A, Berg CW, Nielsen A, Skaug HJ, Machler M, Bolker BM. 2017. GLMMTMB balances speed and flexibility among packages for zero-inflated generalized linear mixed modeling. *The R Journal* 9: 378–400.
- Cantarel BL, Korff I, Robb SM, Parra G, Ross E, Moore B, Holt C, Alvarado AS, Yandell M. 2008. MAKER: an easy-to-use annotation pipeline designed for emerging model organism genomes. *Genome Research* 18: 188–196.
- Carlson CH, Gouker FE, Crowell CR, Evans L, DiFazio SP, Smart CD, Smart LB. 2019. Joint linkage and association mapping of complex traits in shrub willow (*Salix purpurea* L.). *Annals of Botany* 124: 701–715.
- Charlesworth D. 2002. Plant sex determination and sex chromosomes. *Heredity* 88: 94–101.
- Charlesworth D, Charlesworth B, Marais G. 2005. Steps in the evolution of heteromorphic sex chromosomes. *Heredity* 95: 118–128.
- Cheng H, Concepcion GT, Feng X, Zhang H, Li H. 2021. Haplotype-resolved *de novo* assembly using phased assembly graphs with hifiasm. *Nature Methods* 18: 170–175.
- Cho S, Huang ZY, Green DR, Smith DR, Zhang J. 2006. Evolution of the complementary sex-determination gene of honey bees: balancing selection and trans-species polymorphisms. *Genome Research* 16: 1366–1375.
- 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.
- van Doorn GS, Kirkpatrick M. 2010. Transitions between male and female heterogamety caused by sex-antagonistic selection. *Genetics* 186: 629–645.
- Feraru E, Friml J. 2008. PIN polar targeting. *Plant Physiology* 147: 1553–1559.
- Gaspar JM, Hart RP. 2017. DMRfinder: efficiently identifying differentially methylated regions from MethylC-seq data. *BMC Bioinformatics* 18: 1–8.
- Getman-Pickering ZL, Campbell A, Affitto N, Grele A, Davis JK, Ugine TA. 2020. LeafByte: A mobile application that measures leaf area and herbivory quickly and accurately. *Methods in Ecology and Evolution* 11: 215–221.
- Goodstein DM, Shu S, Howson R, Neupane R, Hayes RD, Fazo J, Mitros T, Dirks W, Hellsten U, Putnam N *et al.* 2011. Phytozome: a comparative platform for green plant genomics. *Nucleic Acids Research* 40: D1178–D1186.
- Gouker FE, Carlson CH, Zou J, Evans L, Crowell CR, Smart CD, DiFazio SP, Smart LB. 2021. Sexual dimorphism in the dioecious willow *Salix purpurea*. *American Journal of Botany* 108: 1374–1387.
- Grant S, Houben A, Vyskot B, Siroky J, Pan WH, Macas J, Saedler H. 1994. Genetics of sex determination in flowering plants. *Developmental Genetics* 15: 214–230.
- 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.
- He L, Jia K-H, Zhang R-G, Wang Y, Shi T-L, Li Z-C, Zeng S-W, Cai X-J, Wagner ND, Hörandl E *et al.* 2021. Chromosome-scale assembly of the

- genome of *Salix dunnii* reveals a male-heterogametic sex determination system on chromosome 7. *Molecular Ecology Resources* 21: 1966–1982.
- Honma T, Goto K. 2000. The Arabidopsis floral homeotic gene *PISTILLATA* is regulated by discrete cis-elements responsive to induction and maintenance signals. *Development* 127: 2021–2030.
- Hu L, Zheng T, Cai M, Pan H, Wang J, Zhang Q. 2019. Transcriptome analysis during floral organ development provides insights into stamen petaloidy in *Lagerstroemia speciosa*. *Plant Physiology and Biochemistry* 142: 510–518.
- Hyden B, Carlson CH, Gouker FE, Schmutz J, Barry K, Lipzen A, Sharma A, Sandor L, Tuskan GA, Feng G *et al.* 2021. Integrative genomics reveals paths to sex dimorphism in *Salix purpurea* L. *Horticulture Research* 8: 170.
- Irwin DE. 2018. Sex chromosomes and speciation in birds and other ZW systems. *Molecular Ecology* 27: 3831–3851.
- Kafkas S, Ma X, Zhang X, Topçu H, Navajas-Perez R, Wai CM, Tang H, Xu X, Khodaeiaminjan M *et al.* 2022. Pistachio genomes provide insights into nut tree domestication and ZW sex chromosome evolution. *Plant Communications* 4: 100497.
- Kanei M, Horiguchi G, Tsukaya H. 2012. Stable establishment of cotyledon identity during embryogenesis in Arabidopsis by *ANGUSTIFOLIA3* and *HANABA TARANU*. *Development* 139: 2436–2446.
- Krizek BA, Meyerowitz EM. 1996. The Arabidopsis homeotic genes *APETALA3* and *PISTILLATA* are sufficient to provide the B class organ identity function. *Development* 122: 11–22.
- Krueger F, Andrews SR. 2011. Bismark: a flexible aligner and methylation caller for Bisulfite-Seq applications. *Bioinformatics* 27: 1571–1572.
- Lamesch P, Berardini TZ, Li D, Swarbreck D, Wilks C, Sasidharan R, Müller R, Dreher K, Alexander DL *et al.* 2011. The Arabidopsis Information Resource (TAIR): improved gene annotation and new tools. *Nucleic Acids Research* 40: D1202–D1210.
- Lenormand T, Roze D. 2022. Y recombination arrest and degeneration in the absence of sexual dimorphism. *Science* 375: 663–666.
- Li H. 2018. MINIMAP2: pairwise alignment for nucleotide sequences. *Bioinformatics* 34: 3094–3100.
- Li H, Durbin R. 2009. Fast and accurate short read alignment with Burrows–Wheeler transform. *Bioinformatics* 25: 1754–1760.
- Liao Y, Smyth GK, Shi W. 2014. featureCounts: an efficient general purpose program for assigning sequence reads to genomic features. *Bioinformatics* 30: 923–930.
- Love MI, Huber W, Anders S. 2014. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biology* 15: 1–21.
- Lubell JD, Brand MH. 2018. Foliar sprays of silver thiosulfate produce male flowers on female hemp plants. *HortTechnology* 28: 743–747.
- Manfield IW, Devlin PF, Jen CH, Westhead DR, Gilmartin PM. 2007. Conservation, convergence, and divergence of light-responsive, circadian-regulated, and tissue-specific expression patterns during evolution of the Arabidopsis GATA gene family. *Plant Physiology* 143: 941–958.
- Massonnet M, Cochetel N, Minio A, Vondras AM, Lin J, Muyle A, Garcia JF, Zhou Y, Delledonne M, Riaz S *et al.* 2020. The genetic basis of sex determination in grapes. *Nature Communications* 11: 2902.
- Mayjonade B, Gouzy J, Donnadiu C, Pouilly N, Marande W, Callot C, Langlade N, Muñoz S. 2016. Extraction of high-molecular-weight genomic DNA for long-read sequencing of single molecules. *Biotechniques* 61: 203–205.
- McKee BD, Karpen GH. 1990. Drosophila ribosomal RNA genes function as an XY pairing site during male meiosis. *Cell* 61: 61–72.
- Mirski P. 2014. Exceptions from dioecy and sex lability in genus *Salix*. *Dendrobiology* 71: 167–171.
- Mohan Ram HY, Sett R. 1982. Reversal of ethephon-induced feminization in male plants of *Cannabis sativa* by ethylene antagonists. *Zeitschrift für Pflanzenphysiologie* 107: 85–89.
- Montalvão APL, Kersten B, Kim G, Fladung M, Müller NA. 2022. ARR17 controls dioecy in *Populus* by repressing B-class MADS-box gene expression. *Philosophical Transactions of the Royal Society of London. Series B: Biological Sciences* 377: 20210217.
- Müller NA, Kersten B, Leite Montalvão AP, Mahler N, Bernhardtsson C, Brautigam 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.
- Painter TS. 1921. The Y-chromosome in mammals. *Science* 53: 503–504.
- Ranfil QL, Bastakis E, Klermund C, Schwechheimer C. 2016. LLM-domain containing B-GATA factors control different aspects of cytokinin-regulated development in *Arabidopsis thaliana*. *Plant Physiology* 170: 2295–2311.
- 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. 2022. Sex determination and sex chromosome evolution in land plants. *Philosophical Transactions of the Royal Society B* 377: 20210210.
- Rowland DL, Garner ER, Jespersen M. 2002. A rare occurrence of seed formation on male branches of the dioecious tree, *Populus deltoides*. *The American Midland Naturalist* 147: 185–187.
- Sanderson BJ, Feng G, Hu N, Carlson CH, Smart LB, Keefover-Ring K, Yin T, Ma T, Liu J, DiFazio SP. 2021. Sex determination through X–Y heterogamety in *Salix nigra*. *Heredity* 126: 630–639.
- Somssich M, Je BI, Simon R, Jackson D. 2016. CLAVATA-WUSCHEL signaling in the shoot meristem. *Development* 143: 3238–3248.
- Stettler RF. 1971. Variation in sex expression of black cottonwood and related hybrids. *Silvae Genetica* 20: 42–46.
- Taylor J, Butler D. 2017. R package ASMAP: efficient genetic linkage map construction and diagnosis. *Journal of Statistical Software* 79: 1–29.
- The UniProt Consortium. 2018. UniProt: a worldwide hub of protein knowledge. *Nucleic Acids Research* 47: D506–D515.
- Vose RS, Applequist S, Squires M, Durre I, Menne MJ, Williams CN, Fenimore C, Gleason K, Arndt D. 2014. Improved historical temperature and precipitation time series for U.S. climate divisions. *Journal of Applied Meteorology and Climatology* 53: 1232–1251.
- Wang D, Li Y, Li M, Yang W, Ma X, Zhang L, Wang Y, Feng Y, Zhang Y, Zhou R *et al.* 2022. Repeated turnovers keep sex chromosomes young in willows. *Genome Biology* 23: 200.
- Wang J, Na JK, Yu Q, Gschwend AR, Han J, Zeng F, Aryal R, VanBuren R, Murray JE, Zhang W *et al.* 2012. Sequencing papaya X and Y chromosomes reveals molecular basis of incipient sex chromosome evolution. *Proceedings of the National Academy of Sciences, USA* 109: 13710–13715.
- Wilkerson DG, Taskiran B, Carlson CH, Smart LB. 2022. Mapping the sex determination region in the *Salix* F₁ hybrid common parent population confirms a ZW system in six diverse species. *G3: Genes Genomes Genetics* 12: jkac071.
- Yang W, Wang D, Li Y, Zhang Z, Tong S, Li M, Zhang X, Zhang L, Ren L, Ma X *et al.* 2020. A general model to explain repeated turnovers of sex determination in the Salicaceae. *Molecular Biology and Evolution* 38: 968–980.
- Zhang J, Boualem A, Bendahmane A, Ming R. 2014. Genomics of sex determination. *Current Opinion in Plant Biology* 18: 110–116.
- Zhang S, Fernando DD. 2005. Structural, histochemical, and protein analysis of male reproductive development in willow. *Sexual Plant Reproduction* 18: 37–46.
- Zhou R, Macaya-Sanz D, Carlson CH, Schmutz J, Jenkins JW, Kudrna D, Sharma A, Sandor L, Shu S, Barry K *et al.* 2020. A willow sex chromosome reveals convergent evolution of complex palindromic repeats. *Genome Biology* 21: 38.

Supporting Information

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

Fig. S1 Monoecious *Salix purpurea* catkins during anthesis (top) and after pollination (bottom).

Fig. S2 Histogram showing the distribution of floral sex ratio data.

Fig. S3 Dotplot alignment of 94003 assembly (y -axis) to the 94006 v.5.1 assembly (x -axis).

Table S1 Parameter estimates, standard errors ($\pm$ SE), and P -values for the generalized linear model used to estimate percent maleness by catkin.

Table S2 RNA-Seq differential expression results for 94003 compared with 317 F2 family males and females.

Table S3 RNA-Seq differential expression results for the 94003 male, female, and monoecious progeny.

Table S4 Differential methylation results of 795 gene promoter regions showing significantly different methylation between 94003 and diverse male and female *Salix purpurea*.

Table S5 Physical QTL intervals and LOD scores for sex, height, and leaf area.

Table S6 Monthly precipitation and temperature conditions along with soil characteristics at the three field locations during the summer before flowering.

Please note: Wiley is 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.