Sex determination through X-Y heterogamety in *Salix nigra*

Brian J. Sanderson[1,8,9], Guanqiao Feng[1], Nan Hu[1], Craig H. Carlson[2], Lawrence B. Smart[2], Ken Keefover-Ring[3], Tongming Yin[4], Tao Ma[5], Jianquan Liu[5,6], Stephen P. DiFazio[7], Matthew S. Olson[1]

[1]Department of Biological Sciences, Texas Tech University, Lubbock, TX 79409-3131 USA

[2]Horticulture Section, School of Integrative Plant Science, Cornell University, Cornell AgriTech, Geneva, New York 14456 USA

[3]Departments of Botany and Geography, University of Wisconsin-Madison, Madison, WI 53706, USA

[4]Key Laboratory of Tree Genetics and Biotechnology of Jiangsu Province and Education Department of China, Nanjing Forestry University, Nanjing, China

[5]Key Laboratory of Bio-Resource and Eco-Environment of Ministry of Education & College of Life Sciences, Sichuan University, Chengdu 610065, China

[6]State Key Laboratory of Grassland Agro-Ecosystem, Institute of Innovation Ecology & College of Life Sciences, Lanzhou University, Lanzhou 730000, China

[7]Department of Biology, West Virginia University, Morgantown, WV, 26506 USA

[8]Author for correspondence: brian@biologicallyrelevant.com

[9]Current address: Department of Biology, West Virginia University, Morgantown, WV, 26506-6057 USA

Abstract

The development of non-recombining sex chromosomes has radical effects on the evolution of discrete sexes and sexual dimorphism. Although dioecy is rare in plants, sex chromosomes have evolved repeatedly throughout the diversification of angiosperms, and many of these sex chromosomes are relatively young compared to those found in vertebrates. The Salicaceae, the family of plants that includes poplars and willows, contains predominantly dioecious species and likely evolved separate sexes prior to the split of willows and poplars over 70 million years ago. Despite this shared origin of dioecy, both the chromosomal location of the sex determination region and sex chromosome heterogamety differ between and within these genera. In this study, we identified a novel sex-linked region (SLR) in *Salix nigra*, a basal species in the willow clade, and demonstrate that this species has XY heterogamety. We did not detect any genetic overlap with the previously characterized ZW SLRs in willows, which map to a different chromosome. The *S. nigra* SLR is characterized by strong recombination suppression across a 2 MB region. The non-recombining portion of the X chromosome also exhibits a pattern of low Tajima's D, consistent with a recent selective sweep of the X chromosome in this region.

Key words: Sex determination; sex chromosome; *Salix*; Salicaceae; targeted sequence capture

Introduction

Some groups of animals exhibit surprisingly frequent movement in the genomic locations of sex determination loci (Gammerdinger & Kocher, 2018; Miura, 2007), but it is not yet clear whether this is common in plant species, where the evolution of separate sexes (dioecy) is more frequent, compared with most animal groups (Beukeboom & Perrin, 2014). In mammals (Cortez et al., 2014; Lahn, 1999) and birds (Ellegren, 2010; Shetty et al., 1999; but see Sigeman et al., 2019), most species have maintained similar sex chromosomes for over 80-130 My. In other animal groups, however, sex chromosome transitions appear more common. For instance, among several genera of cichlid fishes the sex determination regions reside on 12 different sex chromosomes, with both male and female heterogametic systems (XY and ZW, respectively; Gammerdinger & Kocher, 2018). Similarly, Dipteran flies harbor more than six different sex chromosomes, and multiple sex chromosome transitions have occurred within the genus *Drosophila* (Vicoso & Bachtrog, 2015). However, in plants few taxonomic groups that share the same origin of dioecy have been studied in detail. One notable exception is octoploid *Fragaria* (strawberry), where a cassette of sex determining genes has moved among homeologs, acquiring linked genomic regions as it moved, likely all within the last 1 My (Tennessen et al., 2018). In this regard, the willows (genus *Salix*) are of particular interest because they are exclusively dioecious and diverged from a likely dioecious ancestor in the Cretaceous, over 70 million years ago (Zhang et al., 2018).

Chromosome 15 was originally identified as the location of the sex-linked region (SLR) in *Salix* using sex-linked SCAR markers (Temmel et al., 2007), which was later confirmed in three species of willow from subgenus Vetrix using genome resequencing and mapping (*Salix*

viminalis: Pucholt et al., 2015; *S. suchowensis*: Hou et al., 2015; and *S. purpurea*: Zhou et al., 2018). The most intensively studied of these species is *S. purpurea*, in which the SLR is a 6.7 MB nonrecombining region (Zhou et al., 2018, 2020). The location and pattern of heterogamety of the SLR in these species of *Salix* differ from those found in the sister genus *Populus*, where the SLR is commonly found on chromosome 19, with XY heterogamety (Geraldes et al., 2015; Pakull et al., 2011, 2009). These differences in the genomic location of the SLR and pattern of heterogamety have led to the hypothesis that dioecy evolved independently in these two genera (Hou et al., 2015). An alternative hypothesis of a single origin of dioecy and the movement of a common set of sex determination genes that are shared across both genera is supported by a set of three genes found in the SLRs of both *S. purpurea* and *P. trichocarpa* (Zhou et al., 2018, 2020).

The goal of the present study was to identify an SLR in a novel subgenus of *Salix* to address whether its location and heterogamety is common among widely diverged willow species. *Salix nigra* is a tree-form willow in the subgenus Protitea that is basal in the *Salix* lineage (Argus, 1997; Barkalov & Kozyrenko, 2014). Using targeted sequence capture to selectively genotype 24 males and 24 females with genome-wide markers, we show that a region on chromosome 7 exhibits highly divergent genotypes between males and females, which indicates that this interval harbors the SLR. A comparison of heterozygosity between males and females, as well as slow decay of linkage disequilibrium among males in this region is consistent with an XY sex determination system, which we confirmed with amplicon sequencing. These results suggest that sex chromosome transitions have occurred in the genus *Salix* and that differences in the chromosomal location of the SLR do not reflect independent origins of dioecy in the family, but rather a movement of the SLR between chromosomes.

Methods

Probe Design

To consistently genotype common sites in a reduced representation of the genome, we custom designed a targeted sequence capture array (RNA bait hybridization) to efficiently capture exonic regions throughout the genome across species in the genus *Salix*. As target regions diverge due to sequence polymorphism both the efficiency of RNA bait binding and capture efficiency are reduced (Lemmon & Lemmon, 2013). Thus, we quantified sequence polymorphism among whole-genome resequencing data from a diverse array of *Salix* species, including *S. arbutifolia*, *S. bebbiana*, *S. chaenomeloides*, *S. eriocephala*, *S. interior*, *S. scouleriana*, and *S. sitchensis*. Whole-genome short reads of the *Salix* species were aligned to the *S. purpurea* 94006 genome assembly version 1 (Salix purpurea v1.0, DOE-JGI, http://phytozome.jgi.doe.gov/pz/portal.html#!info?alias=Org_Spurpurea; Carlson et al., 2017; Zhou et al., 2018) using bwa mem (Li, 2013), single nucleotide polymorphisms (SNPs) and insertion-deletion mutations (indels) were identified using samtools mpileup (Li, 2011), and read depth for the variant calls was quantified using vcftools v. 0.1.15 (Danecek et al., 2011). A custom Python and script was used to identify variant and indel frequencies for the aligned species at all exons in the *S. purpurea* genome annotation.

We further screened candidate regions to exclude high-similarity duplicated regions by accepting only loci with single BLAST hits against the highly contiguous assembly of *S. purpurea* 94006 version 5 (Zhou et al., 2020), which is less fragmented than the *S. purpurea* 94006 version 1. For the remaining loci, we selected regions across which at least 360 base pairs (bp) of exon sequence contained 2-12% difference in single nucleotides and fewer than 2 indels

across all aligned species. For a majority of these genes, we identified multiple regions within each exon. These candidate regions were sent to Arbor Biosciences (Ann Arbor, MI, USA) for probe synthesis. Probes were designed with 50% overlap across the targeted regions, so that each targeted nucleotide position would potentially be captured by two probes. The final capture array consisted of 60,000 probes that target exonic sequences for 16,580 genes. This array is available from Arbor Biosciences (Ref#170623-30).

Library Preparation and Sequence Capture

The sex of 24 males and 24 females of *S. nigra* was identified from flowering catkins in a wild population near Dickens Springs, TX, USA. Leaf tissue was collected and dried on silica beads. DNA was extracted from leaf tissue using the Qiagen DNeasy Plant Minikit (Qiagen, Hilden, Germany) and fragmented using sonication with the Covaris E220 Focused Ultrasonicator (Covaris, Inc., Woburn, MA, USA). Libraries were prepared using the NEBNext Ultra II DNA Prep Kit (New England Biolabs, Ipswitch, MA, USA), and quantified using an Agilent Bioanalyzer 2100 DNA 1000 kit (Agilent Technologies, Santa Clara, CA, USA). Libraries were pooled at equimolar concentrations into sixteen pools of six prior to probe hybridization to targeted capture baits following the Arbor Biosciences myBaits protocol v 3.0.1. The hybridized samples were subsequently pooled at equimolar ratios and sequenced at the Oklahoma Medical Research Foundation (Oklahoma, OK, USA) using a HiSeq 3000 (Illumina, Inc., San Diego, CA, USA).

Variant calls and hard filtering

Reads were aligned to the version 5 assembly of *Salix purpurea* 94006 (female, ZW), using bwa mem (Li, 2013). We estimated the depth of read coverage across all targeted genes using bedtools

intersect v. 2.25.0 (Quinlan & Hall, 2010). The alignments were screened for optical duplicates following the Broad Institute's best practices recommendations (DePristo et al., 2011). Variants were identified using HaplotypeCaller in GATK v4.0.8.1 (McKenna et al., 2010) and initially called separately for each individual as GVCFs. The GVCFs were merged and variant sites and indels were called for all individuals using GenotypeGVCFs in GATK. Based on the distribution of quality scores, variant sites were screened with the hard filters: MQ < 40.00 || SOR > 3.000 || QD < 2.000 || FS > 60.000 || MQRankSum < -12.500 || ReadPosRankSum < -8.000 || ReadPosRankSum > 8.000. Individual genotypes were filtered out if supported by < 6 reads, or > 125 reads. Finally, genotypes that did not pass filtering criteria were assigned as no call genotypes (./.).

Location of the sex determination region

To determine the genomic regions exhibiting the greatest differences between males and females, we performed a genome-wide association study (GWAS) in which we regressed sex (male or female) on the filtered variants using the program emmax (Kang et al., 2010) with a Bonferroni-adjusted significance threshold of $P < 0.05$ to assess the significance of sex linkage. The results were visualized with a Manhattan plot using the R package ggplot2 (Wickham, 2016). We quantified heterozygosity of genotypes at the loci identified with significant sex linkage directly from the filtered VCF file using the program vcftools v. 0.1.15 (Danecek et al., 2011).

To confirm the sex association we analyzed an independent population of 16 male and 16 female *S. nigra* trees collected from the vicinity of Morgantown, WV. Primers were designed in conserved regions that flanked two loci that showed significant sex-linkage in the Texas population. These were sequenced using Sanger chemistry on an ABI3130XL sequencer. Base

calls were made using PHRED (Ewing & Green, 1998) and sequences were assembled using PHRAP and Consed (Gordon et al., 1998). Polymorphisms were identified using PolyPhred (Nickerson, 1997) and confirmed by visualization in Consed. The frequency of heterozygosity for males and females was compared for all loci with minor allele frequency > 0.35, which effectively screened out mutations that may have arisen since establishment of the SLR, as well as any instances of genotyping or sequencing error.

Population genomics statistics

Kinship was estimated using relatedness2 in vcftools, which applies the methods used in KING (Manichaikul et al., 2010). Linkage disequilibrium (LD) for all pairs of loci was calculated as the squared allele frequency correlation (r^2) using vcftools v. 0.1.15 (Danecek et al., 2011). LD decay was calculated both across the SLR in males and females, as well as in an upstream region of chromosome 7 that should be freely recombining, using formulas in Hill & Weir (1988) and Remington et al. (2001). To compare genetic diversity within the *S. nigra* SLR to diversity in other regions of the genome, we calculated per-site nucleotide diversity as π (Nei & Li, 1979) using vcftools v. 0.1.15 (Danecek et al., 2011) after filtering for only biallelic SNPs. We calculated π separately for males, which carry both X and Y chromosomes, and females, which carry two X chromosomes. The average π for 5 Kb sliding windows was calculated using R (R Core Team, 2018). To assess whether non-neutral processes may have influenced the SLR, we calculated values of Tajima's D (Tajima, 1989) using the sliding window method implemented in vcftools v. 0.1.15 (Carlson, 2005; Danecek et al., 2011), with a window size of 25 Kb.

The scripts described above as well as the full details of these analyses including the alignment and variant calling pipeline are available in Jupyter notebooks at https://github.com/BrianSanderson/salix-nigra-slr.

Results

The sequence capture reads aligned to 184 MB of the *S. purpurea* reference genome before filtering, of which 21,229,252 sites met our quality and depth filtering criteria. This corresponded to an average read depth of 6.17 $\pm$ 2.06X (mean $\pm$ sd) at sites genome-wide, and an average read depth of 44.68 $\pm$ 2.68X (mean $\pm$ sd) at on-target sites across the 16,580 targeted genes (Table S1). There were a total of 6,826,811 SNPs among the 48 individuals, of which 2,011,511 passed the filtering criteria, for an average SNP density in the sample of 95 SNPs/kb. The average kinship among individuals in the population was 0.176, indicating an average degree of kinship between half- sibs and first cousins in this population (Fig. S1).

A total of 38 SNPs on chromosome 7 between 4.88 Mb and 6.81 Mb exhibited significant associations with sex ($P_{FDR} < 0.05$; Fig. 1). We refer to this region as the sex-linked region (SLR). The average heterozygosity for males among these SNPs was 0.875 $\pm$ 0.218, and the average heterozygosity for females was 0.072 $\pm$ 0.148 (mean $\pm$ sd), suggesting that the SLR in this species exhibits XY heterogamety (Fig. 2; Table S2). Amplicons from two loci designed to target areas of this region confirmed these patterns in 32 independent individuals, with an average male heterozygosity of 0.959 $\pm$ 0.067 (mean $\pm$ sd), and an average female heterozygosity of 0.021 $\pm$ 0.088 for 18 SNPs with a MAF > 0.2 (Table S3).

Recombination is suppressed between the X and Y chromosomes in the SLR as shown by the slow decay in linkage disequilibrium exhibited in males (XY) compared to females (XX) and

the recombining end of chromosome 7 from 0.1-4 MB (Fig. 3). The distance at which LD decays by half across the SLR in males (1.7 MB) was two orders of magnitude greater than the SLR in females (0.08 MB), as well as the region of chromosome 7 from 0.1 MB - 4 MB upstream of the SLR in both males and females (0.06 MB and 0.05 MB, respectively). Based on the annotation of the *S. purpurea* genome to which we mapped our reads, the SLR includes 111 genes with *Arabidopsis* orthologs (Table S4), including homologs to cytokinin oxidase (AT5G56970.1) and a SAUR-like auxin responsive protein family gene (AT3G12830.1) which are both possible candidates in the metabolic pathway controlling sex determination (Feng et al., 2020).

In males the SLR exhibited elevated heterozygosity and pairwise nucleotide diversity compared to the average in autosomes (Fig. 2A), a pattern consistent with XY heterogamety. In females, the SLR exhibited high homozygosity and low pairwise nucleotide diversity compared to the average in autosomes (Fig. 2B), a pattern consistent with the female carrying two X chromosomes. Because the female carries two X chromosomes, we can use the population of females to assess deviations from neutral evolution for the SLR of the X chromosome. The SLR on the X chromosome exhibited a large negative deviation in Tajima's D compared to the rest of the genome (Figs. 4B & S2), indicating there is an excess of low-frequency polymorphisms in this region for females. When the population of males was analyzed, the SLR exhibited a large positive deviation in Tajima's D (Figs. 4B & S2), consistent with the lack of recombination between the X and Y chromosomes.

Discussion

We identified a new sex-linked region (SLR) in willows that exhibits XY heterogamety on chromosome 7 in *Salix nigra*. This result establishes that sex chromosomes in *Salix* have

undergone changes in the genomic location as well as patterns of heterogamety. The previously identified sex chromosomes in *Salix* have all been in the subgenus Vetrix, and the location and heterogamety of all of these SLRs was on chromosome 15 ZW (Carlson et al., 2017, 2019; Hou et al., 2015; Pucholt et al., 2015; Temmel et al., 2007; Zhou et al., 2018). The sex chromosomes of poplars (*Populus* spp.), the sister genus to *Salix*, have been mapped to two different SLRs on chromosome 19, with *P. trichocarpa* and *P. tremuloides* exhibiting XY sex determination (Geraldes et al., 2015; Pakull et al., 2011, 2009) and *P. alba* exhibiting ZW sex determination (Paolucci et al., 2010). We estimate that the size of the non-recombining SLR in *S. nigra* is approximately 2 MB and contains over 100 genes. However, these estimates are based on a comparison with the *S. purpurea* reference genome, and thus require confirmation by comparison with an assembled *S. nigra* genome. Nonetheless, this SLR appears to be substantially smaller than the 6.7 MB SLR in *S. purpurea* (Zhou et al., 2020) and 20 times larger than the 100kb SLR in *P. trichocarpa* (Geraldes et al., 2015).

Dioecy (separate male and female individuals) is found in only 5-6% of angiosperms (Renner & Ricklefs, 1995) and is unevenly distributed across the angiosperm orders (Renner, 2014). Because many dioecious species have congeners and sister species that are hermaphroditic, it has been posited the dioecy is an evolutionary dead end (reviewed in Käfer et al., 2017). The Salicaceae is among a small group of angiosperm families in which all members are dioecious, which provides an opportunity to test whether the locations of sex determination regions are evolutionarily stable, as in mammals (Cortez et al., 2014; Lahn, 1999) and birds (Ellegren, 2010; Shetty et al., 1999), or variable, as in cichlid fishes (Gammerdinger & Kocher, 2018), frogs (Miura, 2007), and strawberries (Tennessen et al., 2018). Previous studies have

hypothesized that dioecy and the sex determination regions in *Salix* and *Populus* evolved independently (Hou et al., 2015), which was previously supported by the observation that all *Salix* sex chromosomes in the subgenus Vetrix were 15 ZW and most *Populus* sex chromosomes were 19 XY. Thus, our discovery of the 7 XY SLR in *S. nigra* indicates a previously unknown dynamism in the movement of sex chromosomes in *Salix* and supports the hypothesis of a single origin of dioecy and sex chromosomes in the Salicaceae. The mechanism of sex determination in *Populus* is likely controlled by a Type A cytokinin response regulator (a homolog of the *Arabidopsis* gene ARR17) which shows sex association in both *Populus* and *Salix* (Bräutigam et al., 2017; Xue et al., 2020; Zhou et al., 2020). Although there is no homolog of this gene within the corresponding genome interval in *S. purpurea* to which the *S. nigra* reads mapped, it is intriguing that the region does contain genes involved with cytokinin and auxin signaling, raising the possibility that a different but related mechanism of sex determination may be present in this species.

The high level of heterozygosity in the 4.8-6.8 MB SLR region in *S. nigra* males is consistent with XY sex determination, and likely results from the evolutionary divergence between the non-recombining regions of the X and Y chromosomes. The suppression of recombination between the X and Y chromosomes in the SLR allows independent evolution of these two regions. Sequence differences that underlie the high levels of heterozygosity likely arose from both fixation of polymorphism that was present prior to recombination suppression and from the occurrence of new mutations since suppression. The pattern of elevated Tajima's D in the SLR of males directly reflects the high heterozygosity in this region and the independent evolution of the X and Y chromosomes. The females, however, carry two X chromosomes which

recombine freely at the same rate as the non-SLR region of chromosome 7. We found that Tajima's D among females was depressed relative to regions flanking the SLR on chromosome 7 and compared to the autosomes (Fig. 4; Fig. S2B), indicating that alleles at segregating sites of the SLR were rare compared to neutral expectations. This pattern is inconsistent with a recent reduction in the effective population size on the X, which would result in an increase in Tajima's D (Wakeley, 2009). The lower Tajima's D on the X also could not have been caused purely by demographic effects because these would have influenced the entire genome, including the autosomes. The elevated frequency of rare alleles underlying the reduced levels of Tajima's D is consistent with a recent selective sweep in the SLR on the X chromosome (Simonsen et al., 1995). Just after a selective sweep, Tajima's D is expected to decrease dramatically, and as new mutations are introduced into the region, Tajima's D rapidly recovers to neutral expectations (Simonsen et al., 1995).

Conclusion

This report of a previously unidentified SLR in *S. nigra* on chromosome 7 with XY heterogamety illuminates the dynamic history of the shifting position and dominance relationships of sex chromosomes in the Salicaceae family, which includes both poplars (*Populus* spp.) and willows (*Salix* spp.). Patterns of Tajima's D suggest that the X chromosome was recently impacted by a selective sweep, underscoring the evolutionary dynamism in this region and suggesting that selective factors have influenced the movement of the SLRs in this plant family. This discovery indicates that the different locations of sex chromosomes in *Populus* and *Salix* more likely result from the movement of the SLR than multiple independent origins of dioecy in the Salicaceae. This study places the Salicaceae among a small number of taxonomic groups with high

dynamism in the turnover of sex chromosomes. Future research should focus on understanding whether additional shifts in the location and dominance relationships of the sex chromosomes have occurred within the family and what mechanisms are driving these changes.

Acknowledgements

We thank Rao Kottapalli, Pratibha Kottapalli, and Sandra Simon for assistance with library preparation, Julianne Grady for assistance with the amplicon analysis, and Haley Hale for assistance with DNA extraction. We acknowledge Quentin Cronk (UBC, Vancouver) for making available *Salix* genomic resources (supported through funding from Genome Canada, 168BIO). We thank the Dickens County Judge for permission to sample this *S. nigra* population. This research was supported by grants from the US National Science Foundation (1542509, 1542599, 1542479, 1542486) and the National Natural Science Foundation of China (31561123001).

Author Contributions

B.J.S., S.P.D, and M.S.O designed the experiment; S.P.D., L.B.S., K.K-R., T.Y., J.L., and M.S.O. secured funding for this project. L.B.S., T.M., and J.L. contributed whole-genome data for sequence capture design; B.J.S and S.P.D developed the targeted sequence capture array; M.S.O. collected plant materials; B.J.S., G.F., and N.H. carried out the experiment; B.J.S, S.P.D., G.F., C.H.C., and M.S.O. analyzed the data and prepared figures; B.J.S. and M.S.O. drafted the manuscript, and all authors contributed to revisions of the manuscript.

Data Accessibility Statement

Upon acceptance of this manuscript the short read data from the targeted sequence capture will be available at the Genbank Sequence Read Archive (Sanderson et al., 2020). All custom scripts used, as well as full Jupyter analysis notebooks are available at

https://github.com/BrianSanderson/salix-nigra-slr/, and upon acceptance of this manuscript a DOI will be generated from Zenodo to reference the state of the repository at the time of acceptance.

Literature Cited

Argus, G. W. (1997). Infrageneric Classification of Salix (Salicaceae) in the New World. *Systematic Botany Monographs*, *52*, 1. https://doi.org/10.2307/25096638

Barkalov, V. Y., & Kozyrenko, M. M. (2014). Phylogenetic relationships of Salix L. subg. Salix species (Salicaceae) according to sequencing data of intergenic spacers of the chloroplast genome and ITS rDNA. *Russian Journal of Genetics*, *50*(8), 828–837. https://doi.org/10.1134/S1022795414070035

Beukeboom, L., & Perrin, N. (2014). *The Evolution of Sex Determination*. Oxford University Press.

Bräutigam, K., Soolanayakanahally, R., Champigny, M., Mansfield, S., Douglas, C., Campbell, M. M., & Cronk, Q. (2017). Sexual epigenetics: gender-specific methylation of a gene in the sex determining region of Populus balsamifera. *Scientific Reports*, *7*(1), 45388. https://doi.org/10.1038/srep45388

Carlson, C. H., Choi, Y., Chan, A. P., Serapiglia, M. J., Town, C. D., & Smart, L. B. (2017). Dominance and Sexual Dimorphism Pervade the Salix purpurea L. Transcriptome. *Genome Biology and Evolution*, *9*(9), 2377–2394. https://doi.org/10.1093/gbe/evx174

Carlson, C. H., Gouker, F. E., Crowell, C. R., Evans, L., DiFazio, S. P., Smart, C. D., & Smart, L. B. (2019). Joint linkage and association mapping of complex traits in shrub willow (Salix purpurea L.). *Annals of Botany*, *124*(4), 701–715. https://doi.org/10.1093/aob/mcz047

Carlson, C. S. (2005). Genomic regions exhibiting positive selection identified from dense genotype data. *Genome Research*, *15*(11), 1553–1565. https://doi.org/10.1101/gr.4326505

Cortez, D., Marin, R., Toledo-Flores, D., Froidevaux, L., Liechti, A., Waters, P. D., Grützner, F., & Kaessmann, H. (2014). Origins and functional evolution of Y chromosomes across mammals. *Nature*, *508*(7497), 488–493. https://doi.org/10.1038/nature13151

Danecek, P., Auton, A., Abecasis, G., Albers, C. A., Banks, E., DePristo, M. A., Handsaker, R. E., Lunter, G., Marth, G. T., Sherry, S. T., McVean, G., & Durbin, R. (2011). The variant call format and VCFtools. *Bioinformatics*, *27*(15), 2156–2158. https://doi.org/10.1093/bioinformatics/btr330

DePristo, M. A., Banks, E., Poplin, R., Garimella, K. V., Maguire, J. R., Hartl, C., Philippakis, A. A., Angel, G. del, Rivas, M. A., Hanna, M., McKenna, A., Fennell, T. J., Kernytsky, A. M., Sivachenko, A. Y., Cibulskis, K., Gabriel, S. B., Altshuler, D., & Daly, M. (2011). A framework for variation discovery and genotyping using next-generation DNA sequencing data. *Nature Genetics*, *43*(5), 491–498. https://doi.org/10.1038/ng.806

Ellegren, H. (2010). Evolutionary stasis: the stable chromosomes of birds. *Trends in Ecology & Evolution*, *25*(5), 283–291. https://doi.org/10.1016/j.tree.2009.12.004

Ewing, B., & Green, P. (1998). Base-Calling of Automated Sequencer Traces Using Phred. II. Error Probabilities. *Genome Research*, *8*(3), 186–194. https://doi.org/10.1101/gr.8.3.186

Feng, G., Sanderson, B. J., Keefover-Ring, K., Liu, J., Ma, T., Yin, T., Smart, L. B., DiFazio, S. P., & Olson, M. S. (2020). Pathways to sex determination in plants: how many roads lead to Rome? *Current Opinion in Plant Biology*, *54*, 61–68. https://doi.org/10.1016/j.pbi.2020.01.004

Gammerdinger, W. J., & Kocher, T. D. (2018). Unusual Diversity of Sex Chromosomes in African Cichlid Fishes. *Genes*, *9*(10), 480. https://doi.org/10.3390/genes9100480

Geraldes, A., Hefer, C. A., Capron, A., Kolosova, N., Martinez-Nuñez, F., Soolanayakanahally, R. Y., Stanton, B., Guy, R. D., Mansfield, S. D., Douglas, C. J., & Cronk, Q. (2015). Recent Y chromosome divergence despite ancient origin of dioecy in poplars (Populus). *Molecular Ecology*, *24*(13), 3243–3256. https://doi.org/10.1111/mec.13126

Gordon, D., Abajian, C., & Green, P. (1998). Consed: A Graphical Tool for Sequence Finishing. *Genome Research*, *8*(3), 195–202. https://doi.org/10.1101/gr.8.3.195

Hill, W., & Weir, B. (1988). Variances and covariances of squared linkage disequilibria in finite populations. *Theoretical Population Biology*, *33*(1), 54–78. https://doi.org/10.1016/0040-5809(88)90004-4

Hou, J., Ye, N., Zhang, D., Chen, Y., Fang, L., Dai, X., & Yin, T. (2015). Different autosomes evolved into sex chromosomes in the sister genera of Salix and Populus. *Scientific Reports*, *5*(1), 9076. https://doi.org/10.1038/srep09076

Kang, H. M., Sul, J. H., Service, S. K., Zaitlen, N. A., Kong, S.-y., Freimer, N. B., Sabatti, C., & Eskin, E. (2010). Variance component model to account for sample structure in genome-wide association studies. *Nature Genetics*, *42*(4), 348–354. https://doi.org/10.1038/ng.548

Käfer, J., Marais, G. A. B., & Pannell, J. R. (2017). On the rarity of dioecy in flowering plants. *Molecular Ecology*, *26*(5), 1225–1241. https://doi.org/10.1111/mec.14020

Lahn, B. T. (1999). Four Evolutionary Strata on the Human X Chromosome. *Science*, *286*(5441), 964–967. https://doi.org/10.1126/science.286.5441.964

Lemmon, E. M., & Lemmon, A. R. (2013). High-Throughput Genomic Data in Systematics and Phylogenetics. *Annual Review of Ecology, Evolution, and Systematics*, *44*(1), 99–121. https://doi.org/10.1146/annurev-ecolsys-110512-135822

Li, H. (2013). *Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM*. *00*(00), 1–3. https://doi.org/arXiv:1303.3997

Li, H. (2011). A statistical framework for SNP calling, mutation discovery, association mapping and population genetical parameter estimation from sequencing data. *Bioinformatics*, *27*(21), 2987–2993. https://doi.org/10.1093/bioinformatics/btr509

Manichaikul, A., Mychaleckyj, J. C., Rich, S. S., Daly, K., Sale, M., & Chen, W.-M. (2010). Robust relationship inference in genome-wide association studies. *Bioinformatics*, *26*(22), 2867–2873. https://doi.org/10.1093/bioinformatics/btq559

McKenna, A., Hanna, M., Banks, E., Sivachenko, A. Y., Cibulskis, K., Kernytsky, A. M., Garimella, K. V., Altshuler, D., Gabriel, S. B., Daly, M., & DePristo, M. A. (2010). The Genome Analysis Toolkit: A MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Research*, *20*(9), 1297–1303. https://doi.org/10.1101/gr.107524.110

Miura, I. (2007). An Evolutionary Witness: the Frog Rana rugosa Underwent Change of Heterogametic Sex from XY Male to ZW Female. *Sexual Development*, *1*(6), 323–331. https://doi.org/10.1159/000111764

Nei, M., & Li, W. H. (1979). Mathematical model for studying genetic variation in terms of restriction endonucleases. *Proceedings of the National Academy of Sciences*, *76*(10), 5269–5273. https://doi.org/10.1073/pnas.76.10.5269

Nickerson, D. (1997). PolyPhred: automating the detection and genotyping of single nucleotide substitutions using fluorescence-based resequencing. *Nucleic Acids Research*, *25*(14), 2745–2751. https://doi.org/10.1093/nar/25.14.2745

Pakull, B., Groppe, K., Mecucci, F., Gaudet, M., Sabatti, M., & Fladung, M. (2011). Genetic mapping of linkage group XIX and identification of sex-linked SSR markers in a Populus tremula × Populus tremuloides cross. *Canadian Journal of Forest Research*, *41*(2), 245–253. https://doi.org/10.1139/X10-206

Pakull, B., Groppe, K., Meyer, M., Markussen, T., & Fladung, M. (2009). Genetic linkage mapping in aspen (Populus tremula L. and Populus tremuloides Michx.). *Tree Genetics & Genomes*, *5*(3), 505–515. https://doi.org/10.1007/s11295-009-0204-2

Paolucci, I., Gaudet, M., Jorge, V., Beritognolo, I., Terzoli, S., Kuzminsky, E., Muleo, R., Scarascia Mugnozza, G., & Sabatti, M. (2010). Genetic linkage maps of Populus alba L. and comparative mapping analysis of sex determination across Populus species. *Tree Genetics & Genomes*, *6*(6), 863–875. https://doi.org/10.1007/s11295-010-0297-7

Pucholt, P., Rönnberg-Wästljung, A.-C., & Berlin, S. (2015). Single locus sex determination and female heterogamety in the basket willow (Salix viminalis L.). *Heredity*, *114*(6), 575–583. https://doi.org/10.1038/hdy.2014.125

Quinlan, A. R., & Hall, I. M. (2010). BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics*, *26*(6), 841–842. https://doi.org/10.1093/bioinformatics/btq033

R Core Team. (2018). *R: A Language and Environment for Statistical Computing*. https://www.r-project.org/

Remington, D. L., Thornsberry, J. M., Matsuoka, Y., Wilson, L. M., Whitt, S. R., Doebley, J., Kresovich, S., Goodman, M. M., & Buckler, E. S. (2001). Structure of linkage disequilibrium and

phenotypic associations in the maize genome. *Proceedings of the National Academy of Sciences*, *98*(20), 11479–11484. https://doi.org/10.1073/pnas.201394398

Renner, S. S. (2014). The relative and absolute frequencies of angiosperm sexual systems: Dioecy, monoecy, gynodioecy, and an updated online database. *American Journal of Botany*, *101*(10), 1588–1596. https://doi.org/10.3732/ajb.1400196

Renner, S. S., & Ricklefs, R. E. (1995). Dioecy and Its Correlates in the Flowering Plants. *American Journal of Botany*, *82*(5), 596. https://doi.org/10.2307/2445418

[dataset] Sanderson, B. J., Feng, G., Hu, N., Carlson, C. H., Smart, L. B., Keefover-Ring, K., Yin, T., Ma, T., Liu, J., DiFazio, S. P., & Olson, M. S. (2020). *Mapping the sex-linked region of Salix nigra*. https://www.ncbi.nlm.nih.gov/sra/PRJNA614933

Shetty, S., Griffin, D. K., & Graves, J. A. (1999). Comparative painting reveals strong chromosome homology over 80 million years of bird evolution. *Chromosome Research*, *7*(4), 289–295. https://doi.org/10.1023/1009278914829

Sigeman, H., Ponnikas, S., Chauhan, P., Dierickx, E., Brooke, M. de L., & Hansson, B. (2019). Repeated sex chromosome evolution in vertebrates supported by expanded avian sex chromosomes. *Proceedings of the Royal Society B: Biological Sciences*, *286*(1916), 20192051. https://doi.org/10.1098/rspb.2019.2051

Simonsen, K. L., Churchill, G. A., & Aquadro, C. F. (1995). Properties of statistical tests of neutrality for DNA polymorphism data. *Genetics*, *141*(1), 413–429. http://www.ncbi.nlm.nih.gov/pubmed/8536987 http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=PMC1206737

Tajima, F. (1989). Statistical method for testing the neutral mutation hypothesis by DNA polymorphism. *Genetics*, *123*(3), 585–595. http://www.ncbi.nlm.nih.gov/pubmed/2513255 http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=PMC1203831

Temmel, N. A., Rai, H. S., & Cronk, Q. (2007). Sequence characterization of the putatively sex-linked Ssu72 -like locus in willow and its homologue in poplarThis note is one of a selection of papers published in the Special Issue on Poplar Research in Canada. *Canadian Journal of Botany*, *85*(11), 1092–1097. https://doi.org/10.1139/B07-058

Tennessen, J. A., Wei, N., Straub, S. C. K., Govindarajulu, R., Liston, A., & Ashman, T.-L. (2018). Repeated translocation of a gene cassette drives sex-chromosome turnover in strawberries. *PLOS Biology*, *16*(8), e2006062. https://doi.org/10.1371/journal.pbio.2006062

Vicoso, B., & Bachtrog, D. (2015). Numerous Transitions of Sex Chromosomes in Diptera. *PLOS Biology*, *13*(4), e1002078. https://doi.org/10.1371/journal.pbio.1002078

Wakeley, J. (2009). *Coalescent Theory: An introduction*. Roberts & Company.

Wickham, H. (2016). *ggplot2: Elegant Graphics for Data Analysis.* Springer-Verlag.

Xue, L., Wu, H., Chen, Y., Li, X., Hou, J., Lu, J., & Charlesworth, D. (2020). Two antagonistic effect genes mediate separation of sexes in a fully dioecious plant. *bioRxiv*. https://doi.org/https://doi.org/10.1101/2020.03.15.993022

Zhang, L., Xi, Z., Wang, M., Guo, X., & Ma, T. (2018). Plastome phylogeny and lineage diversification of Salicaceae with focus on poplars and willows. *Ecology and Evolution*, *8*(16), 7817–7823. https://doi.org/10.1002/ece3.4261

Zhou, R., Macaya-Sanz, D., Carlson, C. H., Schmutz, J., Jenkins, J. W., Kudrna, D., Sharma, A., Sandor, L., Shu, S., Barry, K., Tuskan, G. A., Ma, T., Liu, J., Olson, M. S., Smart, L. B., & DiFazio, S. P. (2020). A willow sex chromosome reveals convergent evolution of complex palindromic repeats. *Genome Biology*, *21*(1), 38. https://doi.org/10.1186/s13059-020-1952-4

Zhou, R., Macaya-Sanz, D., Rodgers-Melnick, E., Carlson, C. H., Gouker, F. E., Evans, L. M., Schmutz, J., Jenkins, J. W., Yan, J., Tuskan, G. A., Smart, L. B., & DiFazio, S. P. (2018). Characterization of a large sex determination region in Salix purpurea L. (Salicaceae). *Molecular Genetics and Genomics*, *293*(6), 1437–1452. https://doi.org/10.1007/s00438-018-1473-y

Figures

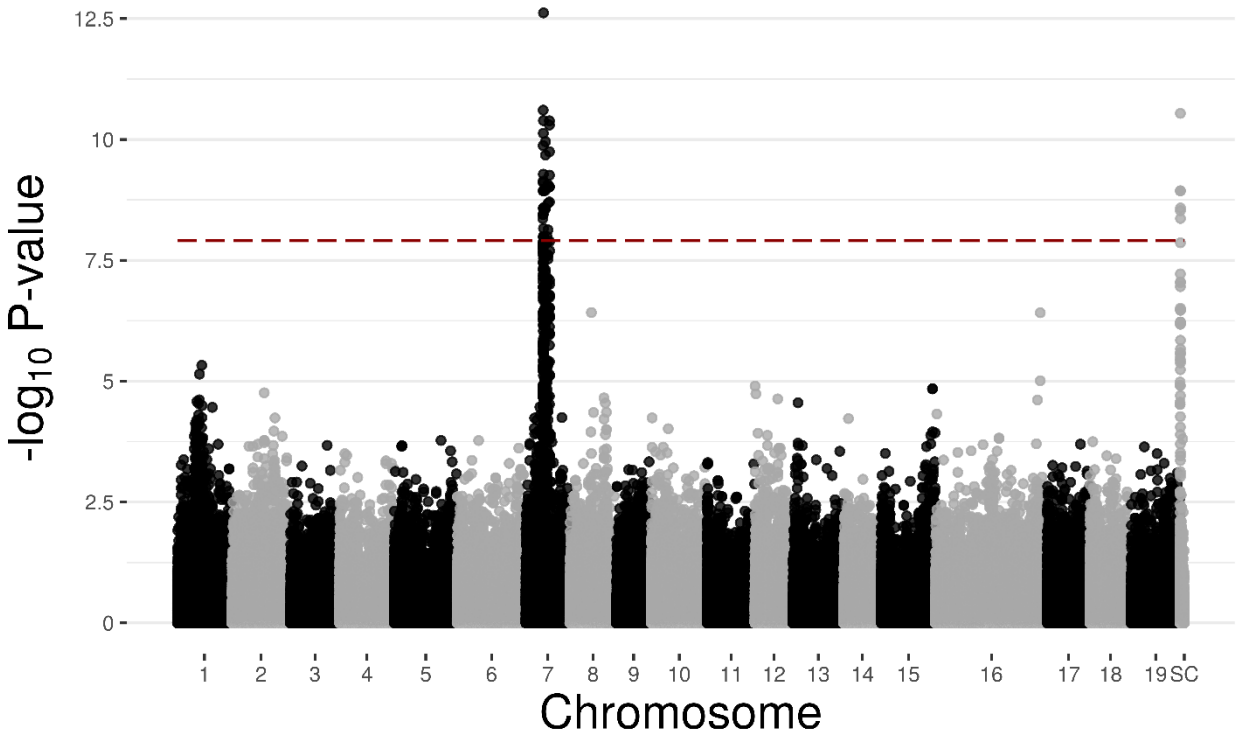

Figure 1. Genome wide associations with sex for 24 male and 24 female *Salix nigra* individuals, expressed in terms of $-\log_{10} P$-values. The relative positions of each SNP on each of the 19 chromosomes and scaffolds as mapped against the *S. purpurea* genome are shown on the X-axis. Alternating black and grey dots represent different chromosomes. Red dashed line indicates a Bonferroni-adjusted $P_{FDR} = 0.05$ ($-\log_{10} P_{FDR} = 7.91$).

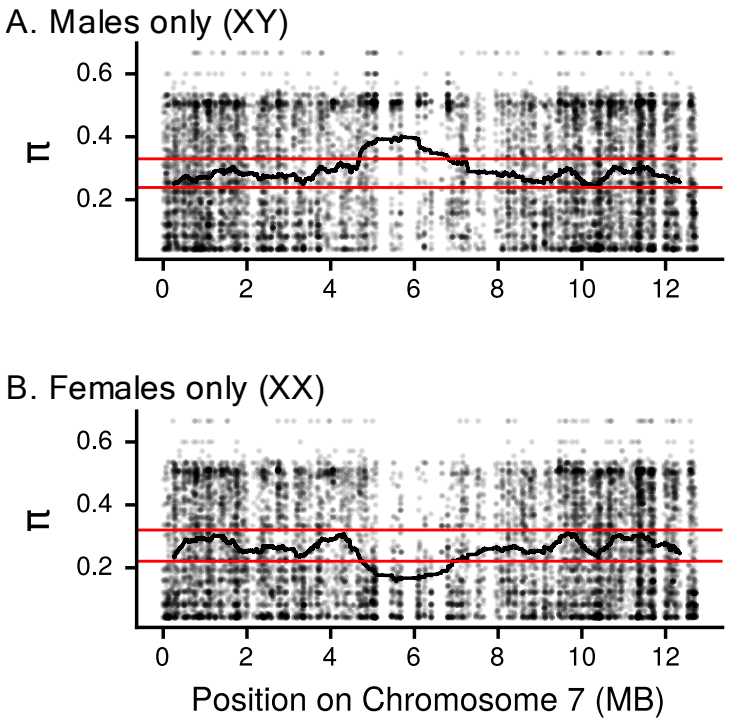

Figure 2. Patterns of nucleotide diversity on *Salix nigra* chromosome 7. A) Dots represent Nei's π for all SNPs in males across chromosome 7. Darkness of the dots represent the relative density of overlapping SNPs plotted in the same are on the figure. The black line represents the rolling mean of 5000 base pair windows. Red horizontal lines indicate the 99% and 1% quantiles of non-overlapping 5000 base pair windows across the genome. B) π for all SNPs in females across chromosome 7.

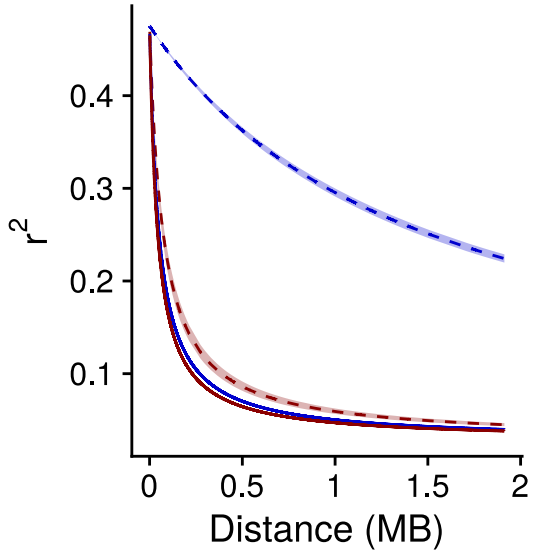

Figure 3. Patterns of linkage disequilibrium decay on *Salix nigra* chromosome 7. LD is expressed as the squared allele frequency correlation (r^2) between two points, with the distance between points indicated on the X-axis. Blue lines are the estimated LD-decay calculated in the SLR (dashed; X & Y) and recombining regions (solid) of chromosome 7 in males. Red lines are the estimated LD-decay calculated in the SLR (dashed; only X) and recombining regions (solid) of chromosome 7 in females. Ribbons around lines represent the 95% confidence intervals of estimates of r^2.

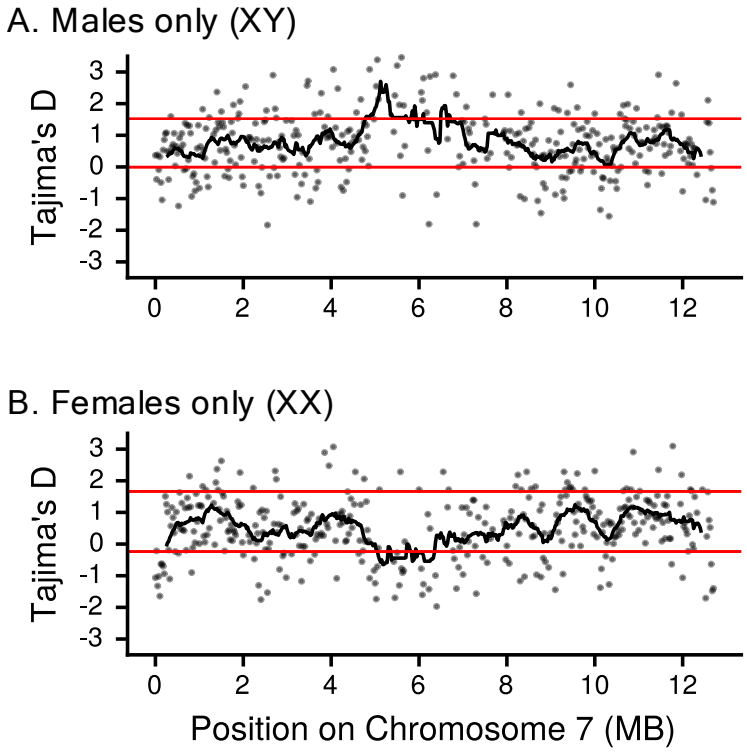

Figure 4. Patterns of Tajima's D on *Salix nigra* chromosome 7. A) Tajima's D for 25 kb windows for males (both X and Y chromosomes) across chromosome 7. Dots represent the mean Tajima's D for non-overlapping 25kb windows across the chromosome. The black line is the rolling mean of 22 25kb windows, the approximate size of the non-recombining sex-linked region (SLR). Red horizontal lines indicate the 99% and 1% quantiles of the Tajima's D calculated from non-overlapping 25kb windows across the genome. B) Tajima's D for 25 kb windows for females (only X chromosomes) across chromosome 7.

Supplemental Information

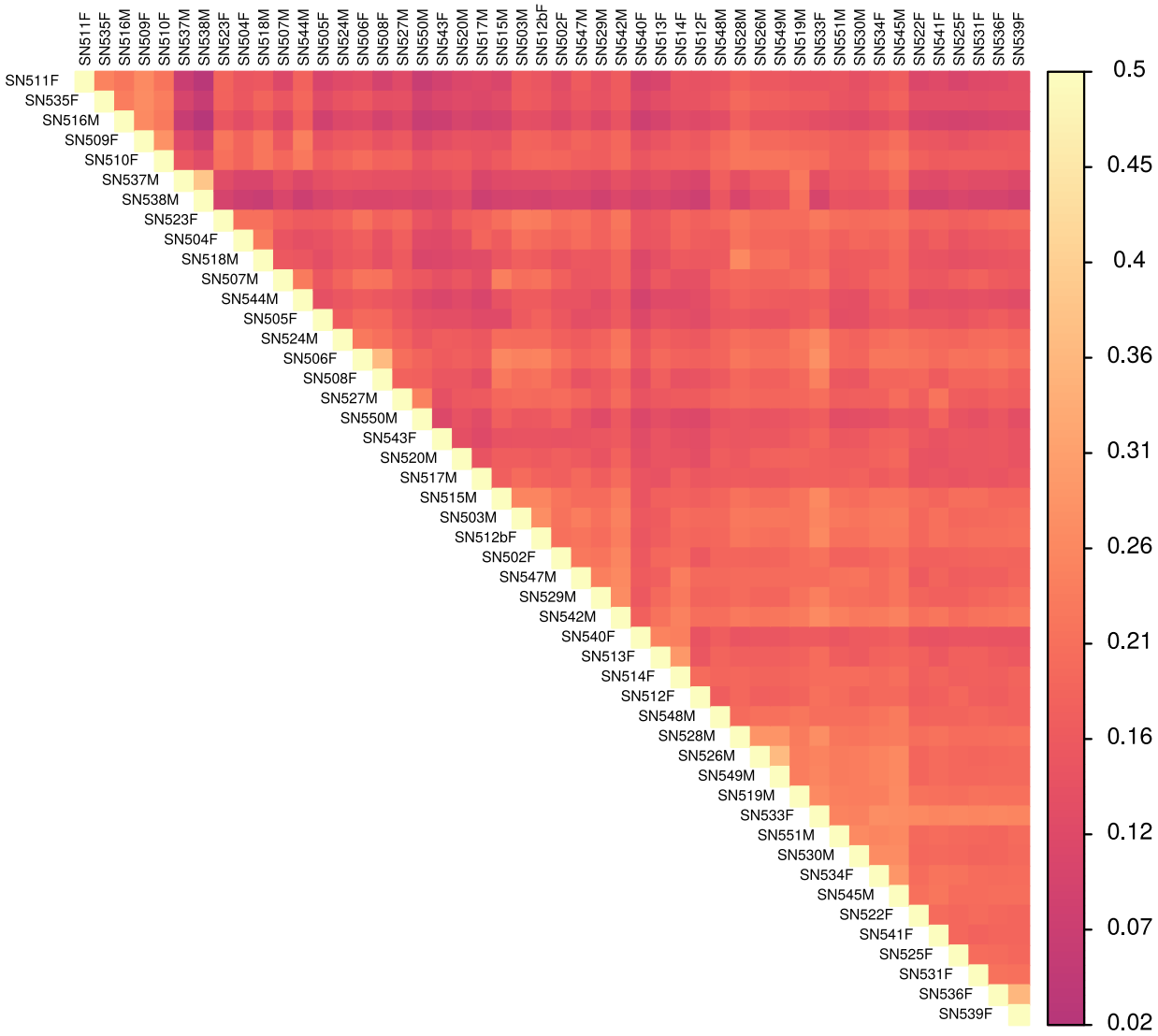

Figure S1. Pairwise kinship estimates between males and females of *Salix nigra* in this study. Kinship was estimated using the relatedness2 algorithm in vcftools, which applies the methods used in KING (Manuchaikul et al. 2010).

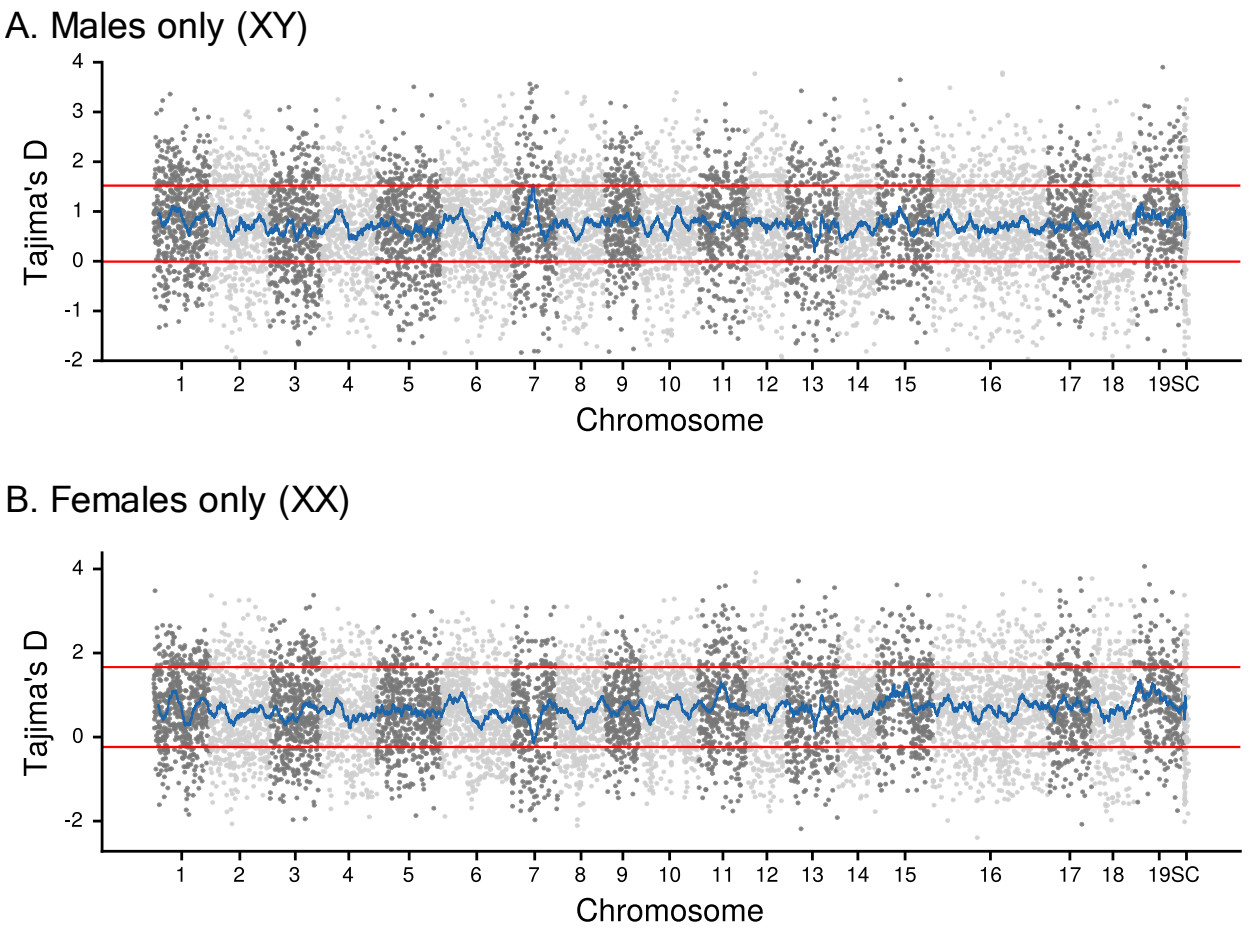

Figure S2. Patterns of Tajima's D across the *Salix nigra* genome. A) Tajima's D for males. Dots represent values of Tajima's D within 25kb windows across the genome. The blue line is the rolling mean of 22 25kb windows, the approximate size of the non-recombining sex linked region (SLR). Red horizontal lines indicate the 99% and 1% quantiles of the Tajima's D calculated from non-overlapping 25kb windows across the genome. B) Tajima's D for females only.

Supplemental Table Legends (uploaded as separate spreadsheet document)

Table S1: Read depth for filtered BAM files (on-target) and filtered VCF (whole genome)

Table S2. Patterns of heterozygosity for males and females for SNPs with significant sex association on chromosome 7

Table S3 Results of amplicon sequencing of 16 male and 16 female *S. nigra* trees from New York and West Virginia. Results are shown only for loci that had a minor allele frequency > 0.2.

Table S4. Genes between 4.88MB and 6.88 MB in the *S. purpurea* genome. This region is the location of the non-recombining sex linked region in *Salix nigra*