



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

Brian J. Sanderson^{1,8} · Guanqiao Feng¹ · Nan Hu¹ · Craig H. Carlson² · Lawrence B. Smart² · Ken Keefover-Ring³ · Tongming Yin⁴ · Tao Ma⁵ · Jianquan Liu^{5,6} · Stephen P. DiFazio⁷ · Matthew S. Olson¹

Received: 21 April 2020 / Revised: 1 December 2020 / Accepted: 14 December 2020
© The Author(s), under exclusive licence to The Genetics Society 2021

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. In this study, we designed and used a sequence capture array to identify a novel sex-linked region (SLR) in *Salix nigra*, a basal species in the willow clade, and demonstrated 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 and an excess of low-frequency alleles, resulting in a low Tajima's D compared to the remainder of the genome. We speculate that either a recent bottleneck in population size or factors related to positive or background selection generated this differential pattern of Tajima's D on the X and autosomes. This discovery provides insights into factors that may influence the evolution of sex chromosomes in plants and contributes to a large number of recent observations that underscore their dynamic nature.

Associate editor: Marc Stiff

Supplementary information The online version of this article (<https://doi.org/10.1038/s41437-020-00397-3>) contains supplementary material, which is available to authorized users.

✉ Brian J. Sanderson
brian@biologicallyrelevant.com

- 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, NY 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 26506-6057, USA
- ⁸ Present address: Department of Biology, West Virginia University, Morgantown, WV 26506-6057, USA

Introduction

Some groups of animals exhibit surprisingly frequent movement in the genomic locations of sex determination loci (Gammerdinger and 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 and Perrin 2014). In mammals (Cortez et al. 2014; Lahn and Page 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 million years. 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 and Kocher 2018). Similarly, Dipteran flies harbor more than six different sex chromosomes, and multiple sex chromosome transitions have occurred within the genus *Drosophila* (Vicoso and Bachtrog 2015). However, in plants few taxonomic groups that share the same origin of dioecy have been studied in detail (Charlesworth 2002). Two notable exceptions are 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 million year (Tennessen et al. 2018) and *Populus*, where sex determination genes have moved and switched from male to female heterogamety resulting from changes in the regulation of an *ARR17* ortholog, a type-A cytokinin response regulator (Muller et al. 2020; Yang et al. 2020). In this regard, the willows (genus *Salix*) are of particular interest because they are exclusively dioecious and diverged from their sister genus *Populus* in the mid-Eocene, 40–50 million years ago (Berlin et al. 2010; 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 (Alstrom-Rapaport et al. 1998; Temmel et al. 2007), which was later confirmed and characterized as ZW heterogamety 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 commonly exhibits XY heterogamety and is located on chromosome 19 (Gerald et al. 2015; Pakull et al. 2009, 2011).

Genetic diversity on sex chromosomes and autosomes is shaped by both demographic and selective factors (Charlesworth et al. 1987; Ellegren 2009; Vicoso and Charlesworth 2006). The expected effective population size (N_e) of the X chromosomes is $\frac{3}{4}$ that of the autosomes, resulting the stronger impacts of genetic drift on the X than on autosomes and a general expectation that diversity on the X should be $\frac{3}{4}$ that of diversity on autosomes (Ellegren 2009; Sayres 2018). At demographic equilibrium, however, the site frequency spectrum (SFS) for neutral alleles should be similar for both the X chromosomes and autosomes, resulting in similarity for statistics such as Tajima's D which is sensitive to the frequency of alleles in the population (Tajima 1989). However, because of its lower N_e , the X is more strongly impacted by bottlenecks in population size resulting in greater reductions in Tajima's D for the X than for autosomes (Gattepaille et al. 2013). Also, because X chromosomes spend less time in males than in females, differences in the variance of reproductive output and generation time between the sexes can differentially influence diversity on the X and autosomes (Amster and Sella 2020; Charlesworth 2001). Because the parts of the X chromosome aside from the pseudo-autosomal regions (PARs) do not recombine in males, linkage disequilibrium (LD) is often elevated compared to autosomes. Thus, sites on the non-recombining regions of the X chromosome are often more strongly affected by hitchhiking with sites that are directly impacted by either positive or purifying selection (Betancourt et al.

2004; Charlesworth et al. 1987). Also, although recessive or partially recessive alleles on autosomes can be sheltered from the effects of selection in females, because the X chromosome is hemizygous in males, alleles are exposed to selection (Charlesworth et al. 1987; Haldane 1924) resulting in a stronger response to selection for X-linked genes than is typical for autosomal genes (Vicoso and Charlesworth 2006).

The goals of this study were to investigate the location and heterogamety of the SLR in *Salix nigra*, a species in a subgenus of *Salix* within which sex chromosomes have not yet been mapped, and to investigate patterns of diversity in its SLR. This species is a tree-form willow in the subgenus *Protitea* that is basal in the *Salix* lineage (Argus 1997; Barkalov and 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 the identification of a pattern of slow decay of LD among males in this region, is consistent with an XY sex determination system, which we confirmed with amplicon sequencing. This is the first evidence that the location and heterogamety of the sex chromosomes are variable in the genus *Salix*. We also found a greater proportion of low frequency alleles in the SLR compared to the autosomes, which we argue likely results from either a recent selective sweep or rare recombination with the Y chromosome combined with background selection on the X.

Methods

Probe design

To consistently genotype common sites in a reduced representation of the genome, we designed a targeted sequence capture array (RNA probe 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 probe binding and capture efficiency are reduced (Lemmon and 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; Carlson et al. 2017; DOE-JGI 2016; 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 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 two 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 (Supplementary Table S1). 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 in April 2017. 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, Ipswich, 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 16 pools of 6 prior to probe hybridization to targeted capture probes following the Arbor Biosciences myProbes 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 and 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

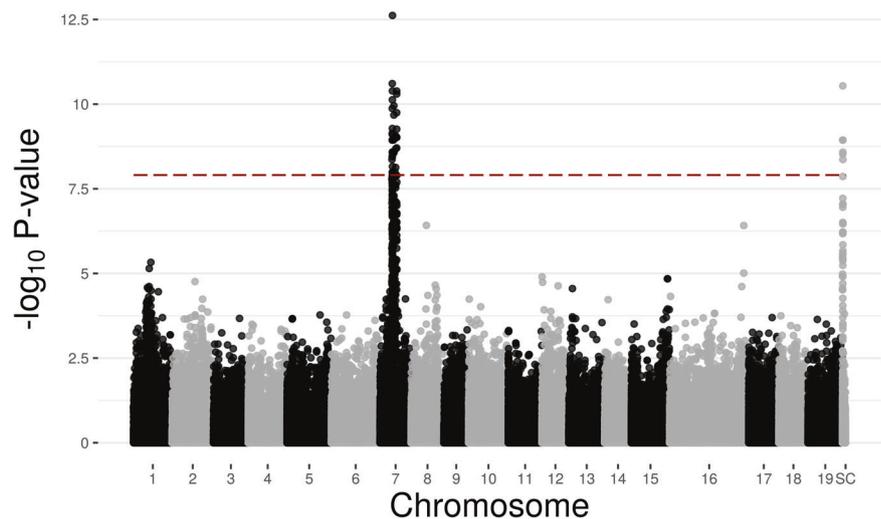
To determine the genomic regions exhibiting the greatest differences between males and females, we performed a genome-wide association study in which we regressed sex (male or female) on the filtered variants using the program emmax (Kang et al. 2010) with a Bonferroni-adjusted significance threshold 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 (Supplementary Table S2). These were sequenced using Sanger chemistry on an ABI3130XL sequencer. Base calls were made using PHRED (Ewing and Green 1998) and sequences were assembled using PHRAP and Consed (Gordon et al. 1998). Polymorphisms were identified using PolyPhred (Nickerson et al. 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). LD for all pairs of loci was calculated as the squared allele frequency correlation (r^2) using vcftools v. 0.1.15

Fig. 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 unplaced scaffolds (SC) as mapped against the *S. purpurea* genome are shown on the X-axis. Alternating black and gray dots represent different chromosomes. Red dashed line indicates a Bonferroni-adjusted $P_{\text{FDR}} = 0.05$ ($-\log_{10} P_{\text{FDR}} = 7.91$).



(Danecek et al. 2011). LD decay was calculated across the SLR in males and females as well as the upstream pseudo autosomal region PAR1 from the end of chromosome 7 to 1 MB from the beginning of the SLR (PAR1: 0.1–3.88 MB) and from 1 MB downstream of the SLR to the end of PAR2 (7.88–13.0 MB) using formulas in Hill and Weir (1988) and Remington et al. (2001). The 1 MB gaps at the beginning and end of the SLR were to avoid regions close to the edges of the SLR that may have intermediate recombination rates. 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 and 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 et al. 2005; Danecek et al. 2011), with a window size of 25 Kb. Windows with fewer than five segregating sites were removed from Tajima's D analyses.

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>.

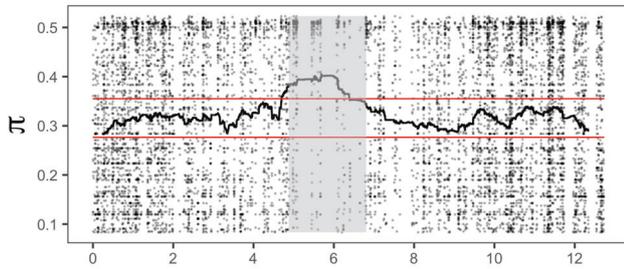
Results

The sequence capture reads aligned to an average of 47.82 ± 1.47 (mean \pm SD) Mb of the *S. purpurea* reference genome among the 48 libraries before filtering (approximately 14.5% of the 329.3 Mb genome). This corresponded

to an average read depth of $6.17 \pm 2.06\times$ (mean \pm SD) at sites genome-wide, and an average read depth of $44.68 \pm 2.68\times$ (mean \pm SD) at on-target sites across the 16 580 targeted genes (Supplementary Tables S1 and S3). The sequence capture probe set resulted in relatively even coverage across the genome (Supplementary Fig. S1), although unsequenced blocks of several thousand kb were present throughout the genome. This was especially true for the region close to the likely centromere on chromosome 7. Some regions were also sequenced that were not targeted (Supplementary Fig. S1). These off-target captures likely resulted from regions immediately up- and downstream of probe targets, as well as regions with weak homology to probes. A total of 6 826 811 SNPs were identified 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 (Supplementary Fig. S2) and suggesting that overall diversity estimated in this population may not be representative of the species as a whole.

A total of 38 SNPs on chromosome 7 between 4.88 and 6.81 Mb exhibited significant associations with sex ($P_{\text{FDR}} < 0.05$; Fig. 1 and Supplementary Fig. S3; Supplementary Table S4). We refer to this region as the SLR. An additional two SNPs on scaffold 197 and five SNPs on scaffold 257 also exhibited significant associations with sex ($P_{\text{FDR}} < 0.05$; Fig. 1 and Supplementary Fig. S3; Supplementary Table S4). We suspect that with additional data these scaffolds will be assembled to the SLR on chromosome 7, as was the case with early assemblies of *P. trichocarpa* (Geraldes et al. 2015). An additional 78 SNPs on chr7, sc197, and sc257 exhibited high association with sex but were not below the statistical significance threshold; these did not change the boundaries of the SLR as defined by the

A. Males only (XY)



B. Females only (XX)

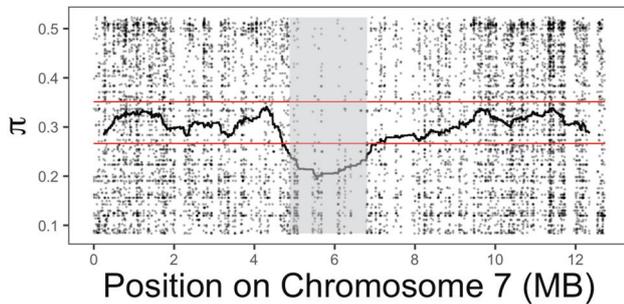


Fig. 2 Patterns of nucleotide diversity and heterozygosity on *Salix nigra* chromosome 7. **A** Dots represent Nei's π for all SNPs in males across chromosome 7. Darkness of the dots represents the relative density of overlapping SNPs plotted in the same area on the figure. The black line represents 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. Gray area represents the proposed sex-linked region.

38 SNPs with significant associations (Supplementary Table S4). The average heterozygosity for males among the 38 SNPs was 0.875 ± 0.218 , and the average heterozygosity for females was 0.072 ± 0.148 (mean \pm SD), suggesting that the SLR in this species exhibits XY heterogamety (Fig. 2; Supplementary Tables S5 and S6). 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 ± 0.067 (mean \pm SD), and an average female heterozygosity of 0.021 ± 0.088 for 18 SNPs with a MAF > 0.2 (Supplementary Table S6). Finally, read counts for males in the SLR were slightly lower for males than for females (Supplementary Fig. S1; males: mean = 24.8, median = 16; females: mean = 25.9, median = 17; Kruskal–Wallis $\chi^2 = 49.2$, $P < 0.001$), but they were not half, as would be expected the absence of loci through degeneration of the Y chromosome.

Recombination is suppressed between the X and Y chromosomes in the SLR as shown by the slow decay in LD exhibited in males (XY; Fig. 3). Females (XX) exhibited faster LD decay than males, but slower than PAR1 and PAR2 (Fig. 3). The distance at which LD decays by half across the SLR in males (1.7 MB) was two orders of

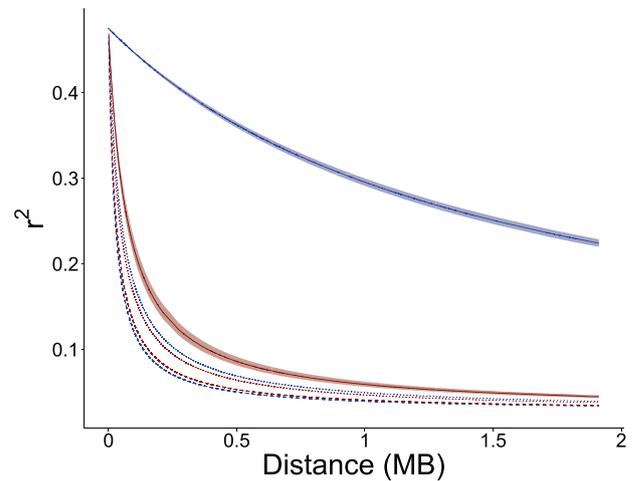


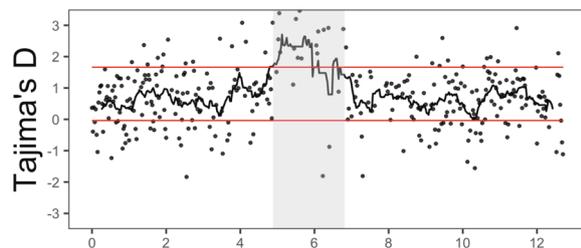
Fig. 3 Patterns of linkage disequilibrium decay on *Salix nigra* chromosome 7. LD is expressed as the squared allele frequency correlation (r^2) between two points, with the distance between points indicated on the X-axis. Blue lines represent LD-decay for chromosome 7 in males and red lines represent LD-decay for chromosome 7 in females. The solid lines represent LD-decay for the SLR, and the recombining regions of chromosome 7 upstream (PAR1) and downstream (PAR2) are represented by dotted and dashed lines, respectively. Ribbons around lines represent the 95% confidence intervals of estimates of r^2 . Note that the 95% CIs for PAR1 and PAR2 are plotted, but the intervals around the center line are very small.

magnitude greater than the SLR in females (0.08 MB), PAR1 (males: 0.06 MB, females: 0.05 MB) and PAR2 (males: 0.03, females: 0.03).

Based on homology of our probes to the annotation of the *S. purpurea* genome to which we mapped our reads, our probes targeted 25 genes in the region of chromosome 7 we identified as the SLR plus the two sex-linked scaffolds (Supplementary Table S7), of which 12 genes carried SNPs significantly associated with sex and 4 genes carried SNPs with associations greater than any genomic location outside the SLR (Supplementary Table S7). This set of 16 genes included a homolog to cytokinin oxidase (AT5G56970.1), a possible candidate for the metabolic pathway controlling sex determination (Feng et al. 2020). The entire region between 4.88 and 6.81 Mb of chromosome 7 contains 111 genes with *Arabidopsis* orthologs (Supplementary Table S7) in *S. purpurea*, but the common observation of dramatic restructuring of sex-specific regions when they experience movement in other species (Charlesworth 2016; Yang et al. 2020; Zhou et al. 2020) suggests caution when interpreting too much from homology across this entire interval.

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 lower effective population size of the two X chromosomes

A. Males only (XY)



B. Females only (XX)

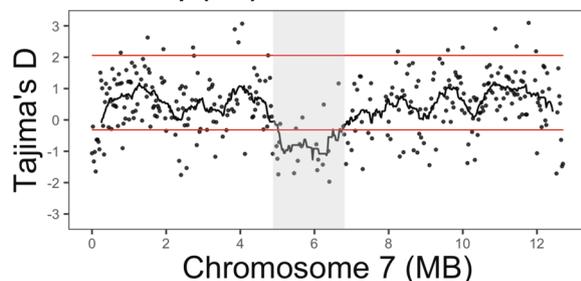


Fig. 4 Patterns of Tajima's *D* on *Salix nigra* chromosome 7. **A** Tajima's *D* for 25 kb windows for males (both X and Y chromosomes) across chromosome 7. Dots represent the mean Tajima's *D* for non-overlapping 25 kb windows across the chromosome. The black line is the rolling mean of 22–25 kb windows, the approximate size of the non-recombining sex-linked region (SLR). Red horizontal lines indicate the 99 and 1% quantiles of Tajima's *D* calculated from non-overlapping 25 kb windows across the genome including all of chr7. Gray area represents the proposed sex-linked region. **B** Tajima's *D* for 25 kb windows for females (only X chromosomes) across chromosome 7.

carried by females, relative to autosomes. The SLR on the X chromosome also exhibited a large negative deviation in Tajima's *D* (mean of 25 kb windows = -0.786 , 95% CI: -1.270 to -0.232) compared to the autosomes (mean = 0.635 , 95% CI: 0.288 to 0.988) and PARs (PAR1 mean = 0.522 , 95% CI: 0.156 to 0.852 ; PAR2 mean = 0.573 , 95% CI: 0.196 to 0.938 ; Fig. 4B and Supplementary Fig. S4), indicating there is an excess of low-frequency polymorphisms in this region. When the population of males was analyzed, the SLR exhibited a large positive deviation in Tajima's *D* (Fig. 4B and Supplementary Fig. S4), consistent with the lack of recombination between the X and Y chromosomes.

Discussion

We identified a new SLR in willows that exhibits XY heterogamety on chromosome 7 in *S. 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; 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 (Geraldès et al. 2015; Pakull et al. 2009, 2011) and *P. alba* exhibiting ZW sex determination (Paolucci et al. 2010) and a third SLR on chromosome 14 with XY heterogamety in *P. euphratica* (Yang et al. 2020). Based on the assumption of strong homology between *S. nigra* and *S. purpurea* on chromosome 7 between 4.88 and 6.81 Mb, the size of the non-recombining SLR in *S. nigra* may be around 2 MB and may contain 111 or more genes. This assumption is not unreasonable given the common chromosome numbers and strong synteny among genomes across *Populus* and *Salix* (Chen et al. 2019; Dai et al. 2014); however, the sex determination regions in *Populus* and *Salix* are known to be highly dynamic (Yang et al. 2020; Zhou et al. 2018, 2020). Thus, we expect that these estimates may change dramatically once the SLR is assembled in *S. nigra*. If this estimate is correct, the SLR in *S. nigra* would be intermediate in size between the 6.7 MB SLR in *S. purpurea* (Zhou et al. 2020) and the 100 kb SLR in *P. trichocarpa* (Geraldès et al. 2015).

Dioecy (separate male and female individuals) is found in only 5–6% of angiosperms (Renner and Ricklefs 1995) and is unevenly distributed across the angiosperm orders (Renner 2014). Many dioecious species have congeners and sister species that are hermaphroditic, and recent analyses support the hypothesis that dioecy is a key innovation that results in increased diversification within plant clades (Kafer et al. 2014). Previous studies have hypothesized that dioecy and the sex determination regions in *Salix* and *Populus* evolved independently (Hou et al. 2015), which was formerly supported by the observation that all *Salix* sex chromosomes in the subgenus *Vetrix* were 15 ZW and most *Populus* sex chromosomes were 19 XY. Recent reports indicate that *Populus* also has sex chromosomes in multiple locations, where a common mechanism of sex determination that includes a cytokinin response regulator (*RR* gene) suggests movement of the sex chromosome within a dioecious lineage (Yang et al. 2020). The discovery of the 7 XY SLR in *S. nigra* indicates a previously unknown dynamism in the evolution of sex chromosomes in *Salix*. Our sequence capture library did not include probes for the *RR* gene that is associated with sex determination in both *Populus* (Muller et al. 2020) and *Salix* (Zhou et al. 2020), so confirmation of the presence or absence of an *RR* homolog or miRNA that targets an *RR* gene in *S. nigra* will require assembly of both the X and Y chromosomes. Such a confirmation of a common sex determination mechanism across both *Salix* and *Populus* would support an alternative hypothesis that

dioecy evolved once prior to the split of the genera and the subsequent movement of the sex determination region among dioecious species.

We found that Tajima's *D* in the SLR of the X chromosome (chr 7) was negative, whereas Tajima's *D* was positive on the autosomes and the PARs (Fig. 4B and Supplementary Fig. S4B), indicating that minor alleles in the SLR were rare. Several mechanisms may account for different aspects of these patterns including our sampling method, the demographic history of the populations, and selection targeted to genes in the SLR. First, we sampled among individuals from a single small and isolated population in west Texas that exhibited generally high relatedness (Supplementary Fig. S2). Sampling strategy is well known to influence the SFS and Tajima's *D*, and these influences vary depending on the population structure and migration rates among populations (Städler et al. 2009). The key implication here is that we should not over-interpret the high value of Tajima's *D* in the autosomal sample as necessarily reflecting a species-wide pattern. Second, theoretical studies indicate that the lower effective population size for the X chromosome than the autosomes (the N_e of the X is 75% of the autosome) results in the X chromosome exhibiting lower values of Tajima's *D* than the autosomes during recovery from a population bottleneck (Gattepaille et al. 2013). For a bottleneck to produce the strongly divergent patterns between the X chromosome and autosomes (and PARs) observed in *S. nigra*, however, this bottleneck must have been strong, and occurred in the relatively recent past within the very small window of time that is expected to generate both a positive Tajima's *D* in autosomes and a negative Tajima's *D* on the X chromosome (Gattepaille et al. 2013). Interestingly, a similar pattern is apparent when comparing the effects of the recent bottleneck in humans on the SFS for the mitochondrial and autosomal genomes (Fay and Wu 1999), but the N_e of the mitochondrial genome is $\frac{1}{4}$ that of the autosomes, so differences between Tajima's *D* are expected to be much more pronounced and long lived than for differences between the X chromosomes and autosomes. Finally, some form of selection, such as a recent selective sweep or rare recombination associated with purifying selection (Simonsen et al. 1995), may have contributed to this pattern in Tajima's *D*. These hypotheses can be partially discriminated based on pattern of alleles private to the X and Y chromosomes. A selective sweep on the X without recombination with the Y would maintain and possibly increase the number of alleles unique to either the X or Y, whereas recombination between the X and Y would decrease the numbers of alleles unique to the X or Y. The genotypes of the 38 SNPs significantly associated with sex in the SLR indicated that no alleles were unique to the X or the Y (Supplementary Table S8), supporting the hypothesis that recombination between the X

and Y within the SLR may be contributing to variation on the X. The question remains, however, why are these alleles in low frequency on the X chromosome? If the recombination events have been occurring repeatedly over time, the SFS would be expected to reach an equilibrium, with Tajima's *D* that is similar to the remainder of the genome. One possibility is that the low frequency of recombinant alleles is influenced by ongoing background selection on linked sites (Vicoso and Charlesworth 2006). Greater insight into the processes generating the deviation from a neutral SFS on the X chromosome will be possible after assembly of the X and Y chromosomes in *S. nigra*.

We used a targeted sequence capture method to sequence a broad portion of the gene space within the *S. nigra* genome because funds were limited for this project, and this meant that it was only possible to discover variants at loci homologous to our probes. This method allowed us to include 48 individuals within our sample, which increased the power and accuracy to detect alleles associated with sex as well as differences in diversity and Tajima's *D* compared to what we would have been able to afford if we had taken a whole genome sequencing approach. Because the density of probes around the SLR was low, we did not capture large regions of the SLR that we may have sampled using whole-genome sequencing (WGS), and this limited our ability to address patterns of diversity across the entire SLR. However, SLRs are notoriously challenging to sequence and assemble, due to the common occurrence of repetitive regions, so a low-coverage WGS approach also may not have recovered a larger portion of the SLR in this novel species. A second limitation of our work was the absence of a reference assembly from *S. nigra*, and instead we relied on a *S. purpurea* reference to map our loci. This is especially limiting when mapping an SLR, which are known to have complex histories of rapid evolutionary change, including translocations, deletions, and inversions (Furman et al. 2020). Nonetheless, all of the SNPs with high association to sex mapped to one of only three locations on the *S. purpurea* reference; either to chromosome 7 or to one of two scaffolds that remain unassembled on the *S. purpurea* v5 assembly. We predict that these scaffolds will map to chromosome 7 as assemblies of *S. purpurea* and *S. nigra* improve. Despite these limitations of the sequence capture approach, we believe these results highlight the utility of this method for inexpensively generating new insights via mapping experiments in novel species.

Conclusion

This report of a previously unidentified SLR in *S. nigra* on chromosome 7 with XY heterogamety illuminates the dynamic history of the shifting positions 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 either *S. nigra* is recovering from a recent population size bottle neck or that the X chromosome was impacted by either a selective sweep or ongoing background selection. It is intriguing to consider that the excess of low-frequency alleles may have resulted from selection induced by the recent the movement of the SLRs (Saunders et al. 2018) in this plant family, but given that the effects of selection on the SFS are fleeting, this SLR may be sufficiently old that the footprint of the sex chromosome transition on the SFS may no longer be present. This study further supports that the Salicaceae are among a small number of taxonomic groups with high dynamism in the turnover of sex chromosomes (Furman et al. 2020). 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.

Data availability

The raw short read data from the targeted sequence capture array are available at the National Center for Biotechnology Information Sequence Read Archive (NCBI SRA) under the BioProject accession number PRJNA614933. All custom scripts used, as well as full Jupyter analysis notebooks, are available at <https://github.com/BrianSanderson/salix-nigra-slr> (Sanderson 2020).

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. We are also grateful to Marc Stift and three anonymous reviewers for their valuable editorial feedback, which greatly improved this manuscript. 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 BJS, SPD, and MSO designed the experiment; SPD, LBS, KK-R, TY, JL, and MSO secured funding for this project. LBS, TM, and JL contributed whole-genome data for sequence capture design; BJS and SPD developed the targeted sequence capture array; MSO collected plant materials; BJS, GF, and NH carried out the experiment; BJS, SPD, GF, CHC, and MSO analyzed the data and prepared figures; BJS and MSO drafted the manuscript, and all authors contributed to revisions of the manuscript.

Compliance with ethical standards

Conflict of interest The authors declare that they have no conflict of interest.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

References

- Alstrom-Rapaport C, Lascoux M, Wang YC, Roberts G, Tuskan GA (1998) Identification of a RAPD marker linked to sex determination in the basket willow (*Salix viminalis* L.). *J Hered* 89 (1):44–49
- Amster G, Sella G (2020) Life history effects on neutral diversity levels of autosomes and sex chromosomes. *Genetics* 215 (4):1133–1142
- Argus GW (1997) Infrageneric classification of *Salix* (Salicaceae) in the new world. *Syst Bot Monogr* 52:1–121
- Barkalov VY, Kozyrenko MM (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 J Genet* 50(8):828–837
- Berlin S, Lagercrantz U, von Arnold S, Öst T, Rönnerberg-Wästljung AC (2010) High-density linkage mapping and evolution of paralogs and orthologs in *Salix* and *Populus*. *BMC Genomics* 11 (1):129
- Betancourt AJ, Kim Y, Orr HA (2004) A pseudohitchhiking model of x vs. autosomal diversity. *Genetics* 168(4):2261–2269
- Beukeboom LE, Perrin N (2014) The evolution of sex determination. Oxford University Press, New York, NY
- Carlson CH, Choi Y, Chan AP, Serapiglia MJ, Town CD, Smart LB (2017) Dominance and sexual dimorphism pervade the *Salix purpurea* L. transcriptome. *Genome Biol Evol* 9(9):2377–2394
- Carlson CS, Thomas DJ, Eberle MA, Swanson JE, Livingston RJ, Rieder MJ et al. (2005) Genomic regions exhibiting positive selection identified from dense genotype data. *Genome Res* 15 (11):1553–1565
- Charlesworth B (2001) The effect of life-history and mode of inheritance on neutral genetic variability. *Genet Res* 77(2):153–166
- Charlesworth B, Coyne JA, Barton NH (1987) The relative rates of evolution of sex-chromosomes and autosomes. *Am Nat* 130 (1):113–146
- Charlesworth D (2002) Plant sex determination and sex chromosomes. *Heredity* 88:94–101
- Charlesworth D (2016) Plant sex chromosomes. *Annu Rev Plant Biol* 67(1):397–420
- Chen J-H, Huang Y, Brachi B, Yun Q-Z, Zhang W, Lu W et al. (2019) Genome-wide analysis of Cushion willow provides insights into alpine plant divergence in a biodiversity hotspot. *Nat Commun* 10 (1):5230
- Cortez D, Marin R, Toledo-Flores D, Froidevaux L, Liechti A, Waters PD et al. (2014) Origins and functional evolution of Y chromosomes across mammals. *Nature* 508(7497):488
- Dai X, Hu Q, Cai Q, Feng K, Ye N, Tuskan GA et al. (2014) The willow genome and divergent evolution from poplar after the common genome duplication. *Cell Res* 24(10):1274–1277
- Danecek P, Auton A, Abecasis G, Albers CA, Banks E, DePristo MA et al. (2011) The variant call format and VCFtools. *Bioinformatics* 27(15):2156–2158
- DePristo MA, Banks E, Poplin R, Garimella KV, Maguire JR, Hartl C et al. (2011) A framework for variation discovery and genotyping using next-generation DNA sequencing data. *Nat Genet* 43 (5):491
- DOE-JGI (2016) http://phytozome.jgi.doe.gov/pz/portal.html#info?alias=Org_Spurpurea
- Ellegren H (2009) The different levels of genetic diversity in sex chromosomes and autosomes. *Trends Genet* 25(6):278–284
- Ellegren H (2010) Evolutionary stasis: the stable chromosomes of birds. *Trends Ecol Evol* 25(5):283–291

- Ewing B, Green P (1998) Base-calling of automated sequencer traces using phred. II. Error probabilities. *Genome Res* 8(3):186–194
- Fay JC, Wu CI (1999) A human population bottleneck can account for the discordance between patterns of mitochondrial versus nuclear DNA variation. *Mol Biol Evol* 16(7):1003–1005
- Feng G, Sanderson BJ, Keefover-Ring K, Liu J, Ma T, Yin T et al. (2020) Pathways to sex determination in plants: how many roads lead to Rome? *Curr Opin Plant Biol* 54:61–68
- Furman BLS, Metzger DCH, Darolti I, Wright AE, Sandkam BA, Almeida P et al. (2020) Sex chromosome evolution: so many exceptions to the rules. *Genome Biol Evol* 12(6):750–763
- Gammerding WJ, Kocher TD (2018) Unusual diversity of sex chromosomes in African Cichlid fishes. *Genes (Basel)* 9:10
- Gattepaille LM, Jakobsson M, Blum MGB (2013) Inferring population size changes with sequence and SNP data: lessons from human bottlenecks. *Heredity* 110(5):409–419
- Geraldes A, Hefer CA, Capron A, Kolosova N, Martinez-Nunez F, Soolanayakanahally RY et al. (2015) Recent Y chromosome divergence despite ancient origin of dioecy in poplars (*Populus*). *Mol Ecol* 24(13):3243–3256
- Gordon D, Abajian C, Green P (1998) Consed: a graphical tool for sequence finishing. *Genome Res* 8(3):195–202
- Haldane JBS (1924) A mathematical theory of natural and artificial selection. *Transactions of the Cambridge philosophical Society* 1:19–41
- Hill WG, Weir BS (1988) Variances and covariances of squared linkage disequilibria in finite populations. *Theor Popul Biol* 33(1):54–78
- Hou J, Ye N, Zhang D, Chen Y, Fang L, Dai X et al. (2015) Different autosomes evolved into sex chromosomes in the sister genera of *Salix* and *Populus*. *Sci Rep* 5:9076–9076
- Kafer J, de Boer HJ, Mousset S, Kool A, Dufay M, Marais GAB (2014) Dioecy is associated with higher diversification rates in flowering plants. *J Evolut Biol* 27(7):1478–1490
- Kang HM, Sul JH, Service SK, Zaitlen NA, Kong SY, Freimer NB et al. (2010) Variance component model to account for sample structure in genome-wide association studies. *Nat Genet* 42(4):348–U110
- Lahn BT, Page DC (1999) Four evolutionary strata on the human X chromosome. *Science* 286(5441):964–967
- Lemmon EM, Lemmon AR (2013) High-throughput genomic data in systematics and phylogenetics. *Annu Rev Ecol Evol Syst* 44(1):99–121
- 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
- Li H (2013) Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. *arXiv: Genomics*
- Manichaikul A, Mychaleckyj JC, Rich SS, Daly K, Sale M, Chen WM (2010) Robust relationship inference in genome-wide association studies. *Bioinformatics* 26(22):2867–2873
- McKenna A, Hanna M, Banks E, Sivachenko A, Cibulskis K, Kernysky A et al. (2010) The Genome Analysis Toolkit: A MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Res* 20(9):1297–1303
- Miura I (2007) An evolutionary witness: the frog *Rana rugosa* underwent change of heterogametic sex from XY male to ZW female. *Sex Dev* 1(6):323–331
- Muller NA, Kersten B, Montalvao APL, Mahler N, Bernhardtsson C, Brautigam K et al. (2020) A single gene underlies the dynamic evolution of poplar sex determination. *Nat Plants* 6(6):630
- Nei M, Li WH (1979) Mathematical-model for studying genetic-variation in terms of restriction endonucleases. *Proc Natl Acad Sci USA* 76(10):5269–5273
- Nickerson DA, Tobe VO, Taylor SL (1997) PolyPhred: automating the detection and genotyping of single nucleotide substitutions using fluorescence-based resequencing. *Nucleic Acids Res* 25(14):2745–2751
- 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* x *Populus tremuloides* cross. *Can J For Res* 41(2):245–253
- Pakull B, Groppe K, Meyer M, Markussen T, Fladung M (2009) Genetic linkage mapping in aspen (*Populus tremula* L. and *Populus tremuloides* Michx.). *Tree Genet Genomes* 5(3):505–515
- Paolucci I, Gaudet M, Jorge V, Beritognolo I, Terzoli S, Kuzminsky E et al. (2010) Genetic linkage maps of *Populus alba* L. and comparative mapping analysis of sex determination across *Populus* species. *Tree Genet Genomes* 6(6):863–875
- Pucholt P, Ronnberg-Wastljung AC, Berlin S (2015) Single locus sex determination and female heterogamety in the basket willow (*Salix viminalis* L.). *Heredity* 114(6):575–583
- Quinlan AR, Hall IM (2010) BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics* 26(6):841–842
- R Core Team (2018) R: a language and environment for statistical computing
- Remington DL, Thornsberry JM, Matsuoka Y, Wilson LM, Whitt SR, Doeblay J et al. (2001) Structure of linkage disequilibrium and phenotypic associations in the maize genome. *Proc Natl Acad Sci U S A* 98(20):11479–11484
- Renner SS (2014) The relative and absolute frequencies of angiosperm sexual systems: dioecy, monoecy, gynodioecy, and an updated online database. *Am J Bot* 101(10):1588–1596
- Renner SS, Ricklefs RE (1995) Dioecy and its correlates in the flowering plants. *Am J Bot* 82(5):596–606
- Sanderson, BJ (2020) BrianSanderson/salix-nigra-sl: Publication. Release 1.0. Available at Zenodo repository <https://doi.org/10.5281/zenodo.4300582>
- Saunders PA, Neuenschwander S, Perrin N (2018) Sex chromosome turnovers and genetic drift: a simulation study. *J Evolut Biol* 31(9):1413–1419
- Sayres MAW (2018) Genetic diversity on the sex chromosomes. *Genome Biol Evol* 10(4):1064–1078
- Shetty S, Griffin DK, Graves JAM (1999) Comparative painting reveals strong chromosome homology over 80 million years of bird evolution. *Chromosome Res* 7(4):289–295
- Sigeman H, Ponnikas S, Chauhan P, Dierickx E, Brooke MD, Hansson B (2019) Repeated sex chromosome evolution in vertebrates supported by expanded avian sex chromosomes. *P R Soc B-Biol Sci* 286:1916
- Simonsen KL, Churchill GA, Aquadro CF (1995) Properties of statistical tests of neutrality for DNA polymorphism data. *Genetics* 141(1):413–429
- Städler T, Haubold B, Merino C, Stephan W, Pfaffelhuber P (2009) The impact of sampling schemes on the site frequency spectrum in nonequilibrium subdivided populations. *Genetics* 182(1):205–216
- Tajima F (1989) Statistical-method for testing the neutral mutation hypothesis by DNA polymorphism. *Genetics* 123(3):585–595
- Tammel NA, Rai HS, Cronk QCB (2007) Sequence characterization of the putatively sex-linked Ssu72-like locus in willow and its homologue in poplar. *Can J Bot* 85(11):1092–1097
- Tenessen JA, Wei N, Straub S, Govindarajulu R, Liston A, Ashman TL (2018) Repeated translocation of a gene cassette drives sex-chromosome turnover in strawberries. *PLoS Biol* 16:8
- Vicoso B, Bachtrog D (2015) Numerous transitions of sex chromosomes in diptera. *PLoS Biol* 13:4
- Vicoso B, Charlesworth B (2006) Evolution on the X chromosome: unusual patterns and processes. *Nat Rev Genet* 7(8):645–653

- Wickham H (2016) *ggplot2: Elegant graphics for data analysis*. Springer-Verlag New York
- Yang W, Wang D, Li Y, Zhang Z, Tong S, Li M et al. (2020) A general model to explain repeated turnovers of sex determination in the Salicaceae. *Mol Biol Evol* <https://doi.org/10.1093/molbev/msaa261>
- Zhang L, Xi ZX, Wang MC, Guo XY, Ma T (2018) Plastome phylogeny and lineage diversification of Salicaceae with focus on poplars and willows. *Ecol Evolution* 8(16):7817–7823
- Zhou R, Macaya-Sanz D, Carlson CH, Schmutz J, Jenkins JW, Kudrna D et al. (2020) A willow sex chromosome reveals convergent evolution of complex palindromic repeats. *Genome Biol* 21(1):38
- Zhou R, Macaya-Sanz D, Rodgers-Melnick E, Carlson CH, Gouker FE, Evans LM et al. (2018) Characterization of a large sex determination region in *Salix purpurea* L. (Salicaceae). *Mol Genet Genomics* 293(6):1437–1452