Title

Characterization of a large sex determination region in Salix purpurea L. (Salicaceae)

Permalink

https://escholarship.org/uc/item/3xd76169

Journal

Molecular Genetics and Genomics, 293(6)

ISSN

1617-4615

Authors

Zhou, Ran

Macaya-Sanz, David

Rodgers-Melnick, Eli

et al.

Publication Date

2018-12-01

DOI

10.1007/s00438-018-1473-y

Supplemental Material

https://escholarship.org/uc/item/3xd76169#supplemental

Peer reviewed

1

2

3

4 Characterization of a Large Sexually Dimorphic Genome Interval in *Salix purpurea* L.
5 (Salicaceae)

6

7

8

9

10 Ran Zhou[*], David Macaya-Sanz[*], Eli Rodgers-Melnick[*], Craig H. Carlson[†], Fred E. Gouker[†],

11 Luke M. Evans[*], Jeremy Schmutz[‡,**], Jerry W. Jenkins[‡], Juying Yan[**], Gerald A. Tuskan[**,††],

12 Lawrence B. Smart[†], and Stephen P. DiFazio[*]

13

14 [*] Department of Biology, West Virginia University, Morgantown, WV 26506-6057

15 [†] Horticulture Section, School of Integrative Plant Science, Cornell University, New York State

16 Agricultural Experiment Station, Geneva, NY 14456

17 [‡] HudsonAlpha Institute of Biotechnology, 601 Genome Way Northwest, Huntsville, AL 35806

18 USA

19 [**] Department of Energy Joint Genome Institute, 2800 Mitchell Drive, Walnut Creek, California

20 94598, USA

21 [††] Biosciences Division, Oak Ridge National Lab

22

23

24

25

26

27

28   Running Title: Genomic Sex Dimorphism in *S. purpurea*

29

30   Key Words: sex, *Salix*, genome, suppressed recombination, dioecy

31

32

33

34

35

36   Corresponding Author:

37

38   Stephen P. DiFazio

39   Department of Biology

40   West Virginia University

41   53 Campus Drive

42   Morgantown, WV 26506-6057, USA

43   (304) 293-5314

44   spdifazio@mail.wvu.edu

45

46

Abstract

47

48   Dioecy has evolved numerous times in plants, but heteromorphic sex chromosomes are

49   apparently rare. Sex determination has been studied in multiple *Salix* and *Populus* (Salicaceae)

50   species, and *Populus* typically has an XY sex determination system on chromosome 19, while *S.*

51   *suchowensis and S. viminalis* have a ZW system on chromosome 15. Here we use quantitative

52   trait locus mapping and a genome-wide association study to demonstrate that *Salix purpurea* also

53   has a ZW system on chromosome 15. This region is characterized by reduced recombination,

54   high structural polymorphism, and an abundance of transposable elements, as expected for a sex

55   chromosome. Genes from this region are known to be involved in sex expression in other plants.

56   We also show that chromosome 19 has some sex chromosome characteristics in *S. purpurea*,

57   including significant QTL and association peaks for sex. Due to a lack of such signatures on

58   chromosome 15 in *Populus*, we hypothesize that sex determination was originally on

59   chromosome 19 in this lineage.

Introduction

60

61   Nearly 90% of flowering plants are hermaphroditic (containing both male and female floral

62   parts in the same flower), and less than 6% are dioecious (separate male and female individuals)

63   (Renner 2014). Evolutionary factors favoring dioecy include inbreeding avoidance and the

64   ability to maximize reproductive output through unisexual resource partitioning (Charlesworth

65   and Charlesworth 1978; Charnov 1982; Ashman 2006). In angiosperms, dioecy has

66   independently evolved hundreds of times from hermaphroditic progenitors (Renner 2014).

67   Evolutionary pathways to dioecy include gynodioecious, heterostylous, and monoecious

68   intermediates (Lloyd 1979; Ainsworth 2000; Charlesworth 2006), but monoecious intermediates

69   tend to be the most common mechanism in woody angiosperms (Olson *et al.* 2017).

70   Trait divergence between females and males can be facilitated by the presence of sex

71   chromosomes, as these are the only genomic regions that consistently differ between the sexes

72   (Rice 1984; Mank 2009; Barrett and Hough 2013). Sex chromosomes usually have suppressed

73   recombination and increased haplotype divergence due to independently accumulating

74   mutations, leading to the development of sexually dimorphic regions (SDR, regions that

75   consistently differ between males and females). The SDR may comprise a majority of the

76   chromosome or only a small portion. Heterogametic SDRs may confer either maleness (XY

3

77 system), as in *Silene latifolia*, *Carica papaya*, *Phoenix dactylifera*, *Diospyros lotus*, and *Populus*

78 *trichocarpa*; or femaleness (ZW system), as in *Fragaria chiloensis*, *Silene ottites*, and *Pistacia*

79 *vera* (reviewed in Charlesworth 2016; Vyskot & Hobza 2015). Sex chromosomes also contain

80 pseudoautosomal regions (PAR) where sex chromosomes recombine freely and may often show

81 elevated recombination (Nicolas *et al.* 2005; Otto *et al.* 2011). Many plant sex chromosomes are

82 homomorphic, exhibiting no strong morphological differences, suggesting that these

83 chromosomes are at an early stage of development (Ming and Moore 2007).

84   The Salicaceae family is an excellent model system for exploring the ecological and

85 evolutionary dimensions of dioecy and sexual selection in plants. Widely distributed across

86 temperate, boreal, and arctic regions of the globe, these genera represent a diverse assemblage of

87 catkin-bearing trees and shrubs (Karp *et al.* 2011). There are approximately 30 *Populus* species,

88 most of which are trees that grow in the northern hemisphere (Slavov and Zhelev 2010). In

89 contrast, there are approximately 500 *Salix* species, most of which are shrubs (Dickmann and

90 Kuzovkina 2014). Nearly all species in *Salix* and *Populus* are dioecious, but none have obvious

91 heteromorphic sex chromosomes (Peto 1938). *Salix* is primarily insect pollinated (Karrenberg *et*

92 *al.* 2002), and produces complex volatiles and nectar rewards (Füssel *et al.* 2007). In contrast,

93 *Populus* is almost exclusively wind-pollinated. Furthermore, both lineages share a well-

94 preserved whole genome duplication (Tuskan *et al.* 2006; Hou *et al.* 2016) and both show an

95 ongoing propensity toward polyploid formation (Mock *et al.* 2012; Serapiglia *et al.* 2015), thus

96 facilitating exploration of the relationship between polyploidy and sex chromosome evolution

97 (Ashman *et al.* 2013; Glick *et al.* 2016).

98   There has been considerable work on characterizing sex determination in *Populus* over the

99 past decade. The SDR has been mapped to the proximal telomeric end of chromosome 19 in *P.*

100 *deltoides* and *P. nigra* (Gaudet *et al.* 2008; Yin *et al.* 2008) and to a pericentromeric region of

101 chromosome 19 in *P. tremuloides*, *P. tremula*, and *P. alba* (Pakull *et al.* 2009; Paolucci *et al.*

102 2010; Kersten *et al.* 2014). In both *P. deltoides* and *P. alba*, the SDR was mapped on a female

103 genetic map but not on a male genetic map, possibly supporting female heterogamety (Yin *et al.*

104 2008; Paolucci *et al.* 2010). In *P. tremuloides* and *P. nigra*, the SDR was mapped on the male

105 genetic map and not on the female genetic map, suggesting male heterogamety (Gaudet *et al.*

106 2008; Kersten *et al.* 2014). Recently, a genome-wide association study (GWAS) on 52 *P.*

107 *trichocarpa* and 34 *P. balsamifera* found 650 SNPs significantly associated with sex. These sex-

4

108 associated markers were nearly fixed heterozygous in males and homozygous in females, which

109 is consistent with an XY sex-determination system (Geraldes *et al.* 2015). However, the

110 significant marker associations were not confined to chromosome 19 but were scattered

111 throughout the genome, possibly due to problems with assembly of the structurally-complex

112 SDR (Geraldes *et al.* 2015).

113   In contrast to *Populus*, the SDR has been mapped to chromosome 15 in *S. viminalis* and *S.*

114 *suchowensis* (Temmel *et al.* 2007; Hou *et al.* 2015; Pucholt *et al.* 2015). Furthermore, there is a

115 preponderance of female heterozygosity in the SDR of these species, indicating a ZW sex

116 determination system, in contrast to *Populus* (Hou *et al.* 2015; Pucholt *et al.* 2015). However,

117 neither study identified candidate genes in the *Salix* SDR that were orthologous to genes in the

118 SDR of *Populus* (Hou *et al.* 2015; Pucholt *et al.* 2015). Thus, *Salix* and *Populus* appear to have

119 different sex determination mechanisms or sex-determining genes, and the nature of the SDR in

120 the *Salicaceae* family remains poorly-characterized. In this study, we sought to explore the SDR

121 in an additional Salicaceae species, *Salix purpurea*. Using robust linkage and association

122 analyses, we show that the principal SDR is on chromosome 15, and that the genotype

123 configuration in this region is consistent with a ZW system of sex determination. Furthermore,

124 we also present evidence that chromosome 19 may retain a residual SDR that has been

125 superseded by the chromosome 15 locus.

126 **Materials and Methods**
127
128 **Genome Assembly**
129
130   This work is based primarily on v1.0 of the *S. pupurea* genome, which is being described

131 more completely in a separate publication (Smart et al., in preparation). Briefly, a female diploid

132 genotype of *Salix purpurea* (clone 94006) was collected from the banks of the Fish Creek River

133 in Upstate New York in 1994 (43.2168 N, -75.6333 W). This clone has been an important parent

134 in *Salix* breeding programs, and is also the source of the reference genome that has been

135 developed by the Joint Genome Institute and a consortium of researchers (available at

136 http://phytozome.jgi.doe.gov). All DNA and RNA samples used for genomic and transcriptomic

137 sequencing were derived from clonally propagated individuals of this genotype. ALLPATHS-LG

138 was used to assemble sequences representing ~140X coverage of Illumina paired-end sequences,

139   as well as a set of mate-pair libraries (4.5 Kb, 5.3 Kb, 6.5 Kb), producing contigs with an

140   L50=46 kb and scaffolds with L50=191 kb. The ALLPATHS-LG assembly has a total length of

141   348 Mb and a total span of 392 Mb (including gaps) but is still relatively fragmented due to a

142   high level of heterozygosity (1 SNP per 120 bp, or 0.8%) and extensive structural variation.

143   Assessment of the assembly quality against willow BACs and transcripts suggested that ~ 78%

144   to 85% of the willow genome is captured in the current assembly. Gene annotations were

145   accomplished using the Phytozome pipeline (Goodstein *et al.* 2012). The RepeatModeler

146   (v1.0.8) package (http://www.repeatmasker.org) was used to identify and mask repetitive

147   elements.

148   **Genetic Mapping and Pseudomolecule Assembly**

149   An F_1 mapping population was produced by crossing two *S. purpurea* accessions, clone 94006

150   (female) and clone 94001 (male), and intercrossing two of the resulting progeny (female

151   'Wolcott' and male 'Fish Creek') to produce over 500 F_2 progeny (referred to as Family 317).

152   The parents and progeny, were genotyped via "Genotyping by Sequencing" (GBS) using

153   *Eco*T221 and *Ape*KI restriction enzymes, and 96-fold multiplexed sequencing on an Illumina

154   HiSeq Genome Analyzer (Elshire *et al.* 2011). SNPs were identified using the reference based

155   pipeline of TASSEL (Glaubitz *et al.* 2014) using the *S. purpurea* v1.0 reference genome

156   (available at http://phytozome.jgi.doe.gov). SNPs were also called using the UNEAK pipeline

157   from TASSEL (Glaubitz *et al.* 2014). SNPs were filtered using the following parameters: -

158   hetFreq 0.75 -mnTCov 0.01 -mnSCov 0.2 -mnMAF 0.05 -hLD -mnR2 0.2 -mnBonP 0.005, and

159   <40% missing data. A total of 8,531 informative GBS markers obtained from 411 F_2 progeny

160   were used to derive separate maps for markers in the three informative configurations: male

161   backcross (n=2623), female backcross (n=2211), and intercross (n=3697). These genetic maps

162   were integrated with the reference genome assembly to produce a combined map on which 276

163   Mb (70%) of sequence scaffolds were anchored, with intervening gaps that were proportional to

164   distances between mapped markers in the scaffolds. The remaining unplaced scaffolds contained

165   another 116 Mb of sequence. Assuming that the unplaced scaffolds are not alternative haplotypes

166   of the mapped scaffolds, the total estimated genome size is approximately 392 Mb, an estimate

167   that was corroborated by kmer counting and flow cytometry (Smart et al., in preparation). The

168   assembly was compared to the *Populus trichocarpa* v3.0 reference genome with LASTZ

169   (v1.03.66), using parameters to exclude alignments between paralogous segments derived from

170   the most recent shared whole genome duplication (gapped, chain, transition, maxwordcount=4,

171   exact=100, step=20).

172   **Identification of Sexually Dimorphic Genome Regions**

173   Sex was scored for F_2 progeny by repeated observations during the spring of 2012, 2013,

174   and 2015 in common gardens at the New York State Agricultural Experiment Station (Cornell

175   University) in Geneva, NY. Quantitative Trait Locus (QTL) mapping was performed using the

176   r/QTL package in R with a binary phenotype model (Arends *et al.* 2010). Logarithm of odds

177   (LOD) support intervals or an approximate Bayesian credible interval were calculated using

178   r/QTL. QTL mapping was performed for all three genetic maps (female backcross, male

179   backcross and intercross).

180   We also performed a Genome-Wide Association Study (GWAS) on the sex trait using a

181   population of unrelated individuals collected from the wild. A population of 112 *Salix purpurea*

182   individuals was collected from upstate New York, Pennsylvania, Connecticut, and Vermont and

183   planted in common gardens at Cornell University in Geneva, NY and at West Virginia

184   University in Morgantown, WV. Sex was scored in the spring of 2013 and 2014 for six clonal

185   replicates at each site. Only individuals for which sex was consistently and unambiguously

186   scored as male or female were used for the analysis. The population was genotyped using GBS

187   with the *Ape*KI restriction enzyme and 48-fold multiplex sequencing on an Illumina HiSeq

188   Genome Analyzer. SNPs were called and filtered as described above, yielding 85,543 SNPs for

189   analysis. A kinship matrix was calculated using the scaled Identity-by-State (IBS) method

190   implemented in the EMMAX package (Kang *et al.* 2010). Clonal ramets were identified based

191   on pairwise IBS values in comparison to pairwise IBS of the F_2 population described above

192   (Figure S1). This resulted in removal of 34 ramets belonging to 9 clonal groups. Three

193   apparently hermaphroditic individuals and 12 individuals with inconsistent sex phenotypes were

194   also excluded from this analysis, leaving a total of 38 females and 22 males. To control for the

195   influence of population structure, a Principal Components Analysis (PCA) was performed using

196   smartPCA in the Eigenstrat package (Price *et al.* 2006). GWAS for sex was performed with the

197   first two principal components and the kinship matrix as covariates using a mixed linear model

198   implemented in the EMMAX package (Kang *et al.* 2010). We controlled for multiple testing

199   using a Bonferroni correction with an alpha value of 0.05.

200

201 **Characterization of the Genomic Composition of the SDR**

202   We defined the SDR intervals based on all GWAS loci that passed the Bonferroni

203 correction. SDR intervals were initially defined as +/- 5 kb around each significant GWAS locus.

204 Also, because the SDR region is structurally complex and repetitive, the genome assembly is

205 likely to be inaccurate in this region, thereby reducing the resolution of the GWAS and QTL

206 mapping. We therefore merged all SDR intervals that occurred within 1 Mb on the same

207 chromosome to include all intervening sequence.

208   To identify levels of polymorphism and divergence between the consensus reference and

209 female-specific haplotypes, we resequenced the F_1 female parent 94006 and the F_2 male parent

210 'Fish Creek' using 2×250 bp reads on an Illumina HiSeq sequencer. This yielded 106,305,281

211 paired reads (53 Gb) and 92,077,639 paired reads (46 Gb), respectively. These were aligned to

212 the 94006 reference genome using Bowtie2 with the parameters -D 15 -R 2 -N 0 -L 20 -i

213 S,1,0.75. SNPs were identified using the mpileup function of samtools, followed by bcftools with

214 the parameters -g 1 -O v –m. Since ALLPATHS-LG generates genome assemblies that consist of

215 chimeras of the two haplotypes from a heterozygous diploid genome (Gnerre *et al.* 2011), we

216 expected the *S. purpurea* assembly of chromosome 15 to include segments of Z and W

217 chromosomes. This should be apparent from the relative depth of coverage of female and male

218 sequences. For Z portions of the reference genome, male coverage should be roughly double that

219 of the female for divergent portions of the SDR, whereas for W portions of the reference,

220 coverage should be approximately 0.5X compared to the rest of the genome for the female, and

221 there should be zero coverage in males. We therefore used these alignments to evaluate depth of

222 coverage for the male and female sequences using raw output from the samtools mpileup

223 command to delineate putative Z and W portions of the reference. Similar expectations hold for

224 the GBS markers, which should be homozygous in females and null in males when mapped to

225 the divergent W portions of the reference genome.

226   We identified female-specific alleles for loci that were heterozygous in the female parent

227 clone 94006 and homozygous in the male offspring, Fish Creek. Sequences containing female-

228 specific polymorphisms (here called "W-type") were created using the

229 FastaAlternateReferenceMaker module of the GATK package (DePristo *et al.* 2011). For

230 comparison, we also used all polymorphisms from the resequencing of both genotypes to create

231 alternative haplotypes using the same approach. Genes with nonsense and frameshift mutations

232 were then removed as possible pseudogenes. Finally, synonymous (dS) substitution frequencies

233 were estimated for all pairs of predicted transcripts using the '*yn00*' module in the PAML

234 package (Yang 2007). The reference genome transcripts were compared to those containing

235 female-specific polymorphisms as well as to those containing all alternative alleles.

236   All predicted proteins in the *S. purpurea* reference genome annotation were compared to

237 the UniProt database (http://www.uniprot.org/) using blastp and against the Pfam database

238 (http://pfam.xfam.org/) using HMMER, with default parameters. Protein mapping results were

239 submitted to Argot[2] (Falda *et al.* 2012) to obtain Gene Ontology (GO) annotations, using a

240 stringent cut-off (Total Score=1500) to filter Type I errors. We used Fisher's Exact Test to

241 identify overrepresented GO terms for candidate genes in the SDR. All orthologs between *S.*

242 *purpurea* and *P. trichocarpa* were retrieved from Phytozome (https://phytozome.jgi.doe.gov/).

243 Synonymous (dS) and nonsynonymous (dN) substitution frequencies were estimated for each

244 pair of primary transcripts from each species using the '*yn00*' module in the PAML package

245 (Yang 2007). Pairs with dS>0.4 were dropped, assuming they were incorrectly defined as

246 orthologs. In total, 33,789 ortholog pairs were compared, including 27,118 genes from *S.*

247 *purpurea* and 24,000 genes from *P. trichocarpa*.

248 **Estimation of Recombination Rate**

249   As an indicator of recombination rate, we calculated the ratio of physical to genetic

250 distance between marker pairs using linkage groups with >30 markers. For each linkage group,

251 pairwise distances were calculated between every N loci, where N was 10% of the total number

252 of loci on the linkage group. For example, if the linkage group had 100 markers, the distance was

253 calculated between all pairs of loci that were separated by 10 loci. Negative and extreme values

254 (ratio>15) were removed for the purpose of visualization.

255 **Gene Expression**

256   RNA sequencing was performed for actively growing shoot tips for five male and five

257 female progeny from the family used for QTL analysis. Detailed methods are described in

258 Carlson et al. (2017). Briefly, total RNA was extracted using the Spectrum^{TM} Total Plant RNA

259 Kit. Libraries were constructed using the NEBNext Ultra Directional RNA Library Prep Kit.

260 Libraries were sequenced on the Illumina HiSeq platform (1x100 bp) yielding an average of 17.9

261 million mapped reads per sample. Reads were mapped to the *S. purpurea* reference genome v1.0

262 using the CLC Genomics Workbench, and differential expression analyses were performed using
263 EdgeR.

Results

Localization of the SDR to Chromosome 15 and Chromosome 19

267   Among the 396 phenotyped and genotyped individuals in the F_2 family, there were 234
268 females and 162 males. This ratio is significantly skewed toward females (F:M=1.44; χ^2=13.1;
269 df=1; P<0.001). QTL mapping identified sex-associated markers principally on chromosome 15
270 for all three maps (Figure 1; Table S1). On the female map, 125 markers were linked to sex, 105
271 of which were on chromosome 15, spanning from 225.42 cM to 240.17 cM (Table 1). On the
272 male map, only five markers were linked to sex, four of which were in the interval from 326.48
273 cM to 347.17 cM on chromosome 15 (Figure 1, Table 1). An additional 50 markers were linked
274 to sex on the intercross map, covering an interval of about 2.6 cM, all on chromosome 15 (Figure
275 1, Table 1). Based on anchoring mapped markers to physical positions in the *S. purpurea*
276 genome assembly, the potential SDR can be mapped to two regions on chromosome 15 ranging
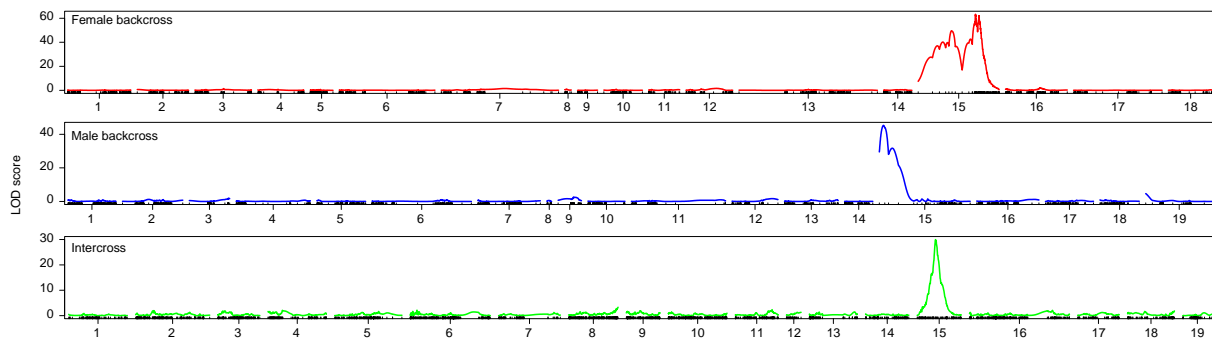277 from ~0.4 Mbp to 1.9 Mbp and from ~10.9 Mbp to ~15.1 Mbp.

Figure 1 QTL for sex in an F_2 *S. purpurea* cross. From top to bottom are LOD scans for
female backcross (red), male backcross (blue) and intercross (green) markers across the 19
major *S. purprea* linkage groups. Chromosome 15 has a very strong QTL sex in all three maps,
and the female backcross also shows a weak peak on chromosome 19 (LOD=4.68; table 1)**.**

278

Table 1 Bayesian credible intervals for sex QTL on chromosome 15.

	Physical Map		Genetic Map	
	Start (bp)	End (bp)	Start (cM)	End (cM)
Female Map	10,939,613	11,569,298	225.42	240.17
Male Map	372,445	1,881,243	326.48	347.17
Intercross	11,401,384	15,091,498	55.69	58.22

One additional sex-linked marker was located at the proximal end of chromosome 19 on the male map, with a LOD score of 4.68 (Figure 1; Table S1). However, mapping failed entirely for chromosome 19 for female backcross markers, the only chromosome for which this was the case. Chromosome 19 had the lowest density of GBS markers in the genome (Table S2). Furthermore, this chromosome had the lowest proportion of markers in a female-backcross configuration, and the highest proportion of markers with severe segregation distortion (Figure S2; Table S2).

To confirm the location of the SDR in a diverse population, a GWAS for sex was performed using naturalized *S. purpurea* accessions collected from northeastern North America. Of the 60 genets that were unambiguously phenotyped for sex, 38 were female and 22 were male, which is a significantly female-biased sex ratio (F:M=1.73; χ^2=4.3; df=1; P=0.02). Of the 85,543 SNP markers that passed filtering, 72 were significantly associated with sex ($P<5.85 \times 10^{-7}$, Figure 2; Figure S3). Among these markers, 41 were located on chromosome 15, from 10.7 Mb to 15.3 Mb, and four were located at the distal portion of chromosome 15 (1.9 Mbp). Thus, the primary SDR identified by GWAS overlaps with those mapped by QTL in the F_2 family (Figure 3). In addition, six markers from chromosome 19 at ~69 kb were also significantly associated with sex (Figure 2), which also corresponds with the QTL results. Additionally, there were minor peaks on chromosomes 1,2,3, and 5, and there were six scaffolds containing a total of 13 significant sex-associated markers that were not anchored to the genetic maps (Table S3).

To evaluate whether these secondary chromosomal peaks could have been due to assembly errors, we aligned these SDR sequences to the *S. purpurea* reference genome using blastn. None of these chromosomal loci shared homology with the chromosome 15 SDR (Table S4). We also compared these SDR sequences to the *Populus trichocarpa* v3.0 reference genome using blastn. The SDRs on chromosomes 1,2, and 5 had best hits to the same chromosomes in *P. trichocarpa*. However, the SDRs on chromosomes 3 and 19 had best hits to scaffold_25 in *P.*
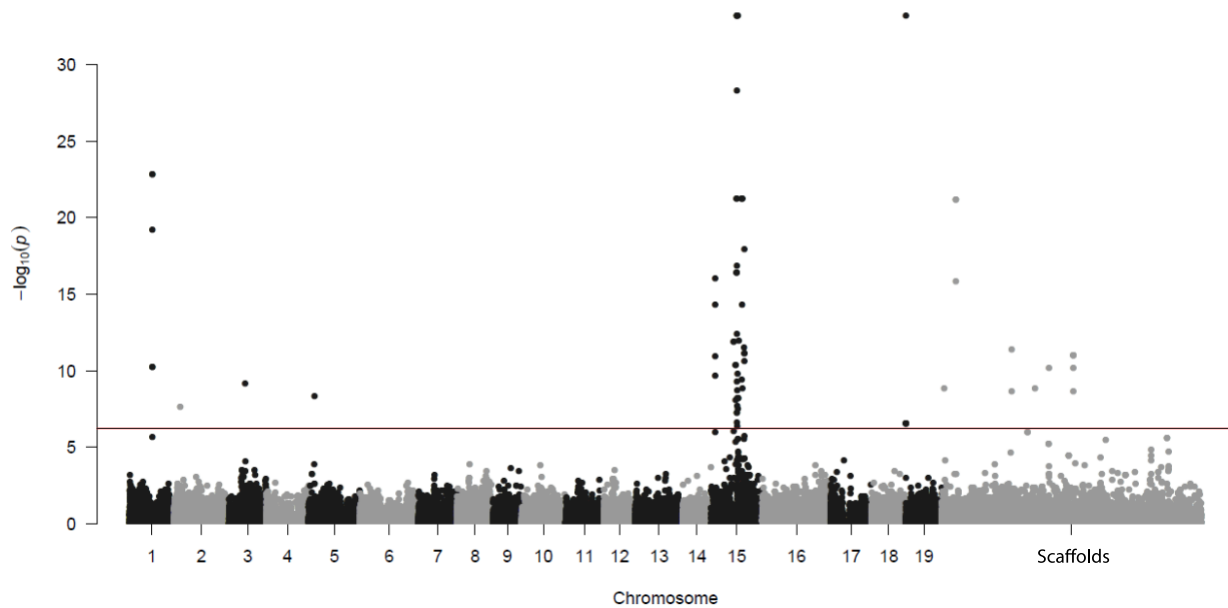
Figure 2 Manhattan plot derived from genome-wide association analysis for sex determination. The Y-axis shows the strength of association ($-\log_{10}$(P value)) for each SNP ordered by chromosome and SNP position (x axis). The horizontal line indicates significance after a Bonferroni correction for multiple testing.

307   *trichocarpa* (Table S4). Because the SDR is known to be poorly assembled in the *P. trichocarpa*

308   v3.0 assembly (Geraldes *et al.* 2015), we aligned scaffold_25 to the *P. trichocarpa* v1.0

309   assembly and found that it matched primarily to chromosome 19, positions 751 to 1040 kb,

310   which coincides with the main *P. trichocarpa* SDR (Geraldes *et al.* 2015). Therefore, the QTL

311   and GWAS results both indicate that sequences homologous to the *P. trichocarpa* SDR retain

312   evidence of sex dimorphism in *S. purpurea*.

313   ***S. purpurea* Has a ZW System of Sex Determination**

314   Under Mendelian segregation, the frequency of heterozygotes should be 0.5 for both male

315   and female F$_2$ progeny. However, the frequency of heterozygosity was 0.64 for female progeny

316   and only 0.12 for males (Table S1). The skewed heterozygosity occurred in blocks in the vicinity

317   of the sex QTL peaks (Figure 3). Furthermore, females in the association population had an

318   average observed heterozygosity of 0.79 for the sex-associated SNP loci, while males had an

319   observed heterozygosity of only 0.05 for these same loci (Figure 4a, Table S3, Figure S4). This

320   difference was significant based on a t-test ($P < 2.2 \times 10^{-16}$). Both observations are consistent with

321   a female heterogametic (ZW) system of sex determination, where females should be nearly fixed

322   heterozygous for female-specific portions of the SDR, while males should be homozygous for
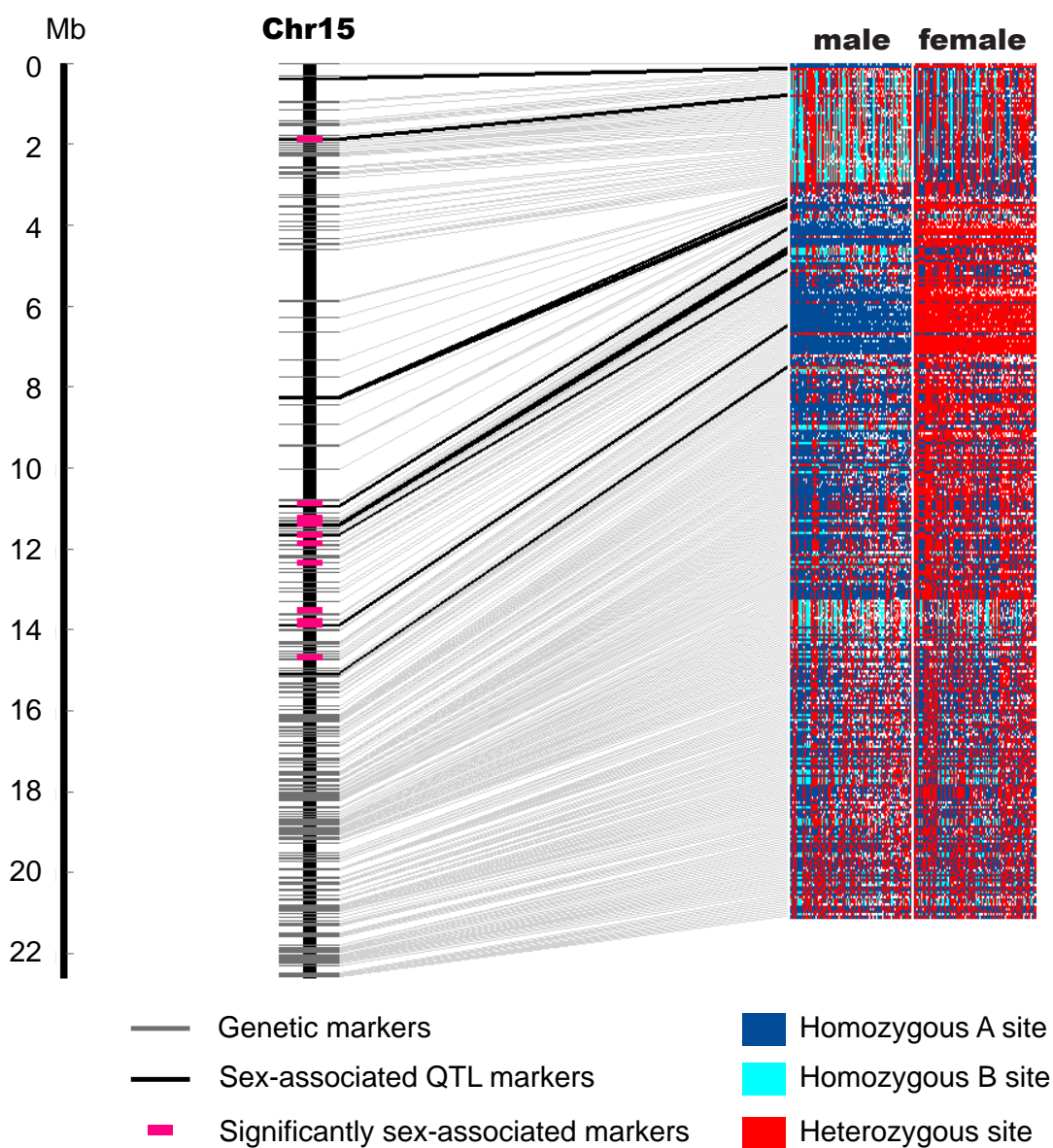
12

Figure 3 Genotype configurations in males and females from the F$_2$ family on chromosome 15. Markers from all three genetic maps are shown as horizontal lines corresponding to their physical positions on the chromosome 15 physical assembly. Markers with top LOD scores in each map are colored as black. Significantly associated markers from the GWAS analysis with P < 1x10^{-7} are indicated by fuschia marks on the physical map. Each marker is connected between physical map and its genotype configurations with 100 selected progeny of each sex. Genotypes of QTL markers are colored according to their homozygosity or heterozygosity.

323

324   those same loci. This is due to the typically biallelic nature of SNP polymorphisms, where

325   polymorphic alleles from the W chromosome are identical by descent and therefore only occur in

326   females. The discrepancy between the observed values and the expected fixed heterozygosity in

327   females is likely due to null alleles caused by allele dropout and/or inadequate sequencing depth

328   for the GBS markers (Andrews *et al.* 2016).

329   Since our reference sequence was derived from a female, we expected that the assembly

330   could contain hemizygous or highly divergent portions of the W chromosome. We used two

331   complementary approaches to determine the size and extent of these regions: allelic

332   configurations of the GBS markers, and relative depth of sequence coverage in male and female

333   clones. Candidate W segments contained a large proportion of GBS markers that were

334   homozygous in females and mostly lacking genotype calls (i.e., double null markers) in males in

335   the association population (Figure S5). We identified 231 of these W-type markers (0.27%)

336   (Figure 4a; Table S5). Of these, 51 occurred on chromosome 15, another 158 occurred on 20

337   unanchored scaffolds, and the remaining 22 occurred on small segments of chromosomes 3, 5,

338   and 7. The genotype configuration for these markers was consistent with a ZW system, such that

339   the average observed homozygosity for females was 0.80 (presumably due to hemizygosity or

340   divergence of W segments) whereas 85% of males had null alleles at these loci, on average

341   (Figure 4a, Table S5). The putative W haplotypes were interspersed along chromosome 15,

342   suggesting that the genome assembly is a chimeric representation of the Z and W haplotypes

343   (Figure 4a; Table S5).

344   We also examined depth of coverage in male and female reference-based assemblies to

345   identify putative hemizygous W chromosome segments in the reference genome. If females are

346   heterogametic, then there should be regions in the female reference that are not covered by reads

347   from a male individual. Aligning paired 250 bp Illumina sequences from a male offspring ('Fish

348   Creek') of clone 94006 back to the female reference assembly, yielded a very high alignment

349   rate of 95.19% compared to 96.67% when clone 94006 was aligned to itself. Nevertheless, after

350   excluding known repeats and gaps, there were 22,733 regions totaling 7.69 Mb on chromosomes

351   and another 6.87 Mb of unanchored scaffolds that had coverage in the female but lacked

352   coverage in the male (Table 2; Figure S6). These analyses identified 222 scaffolds comprised of

353   >30% female-specific sequences (Table S5). Some of these are likely caused by

354   insertion/deletion polymorphisms that are not sex-specific. However, we identified 11 scaffolds
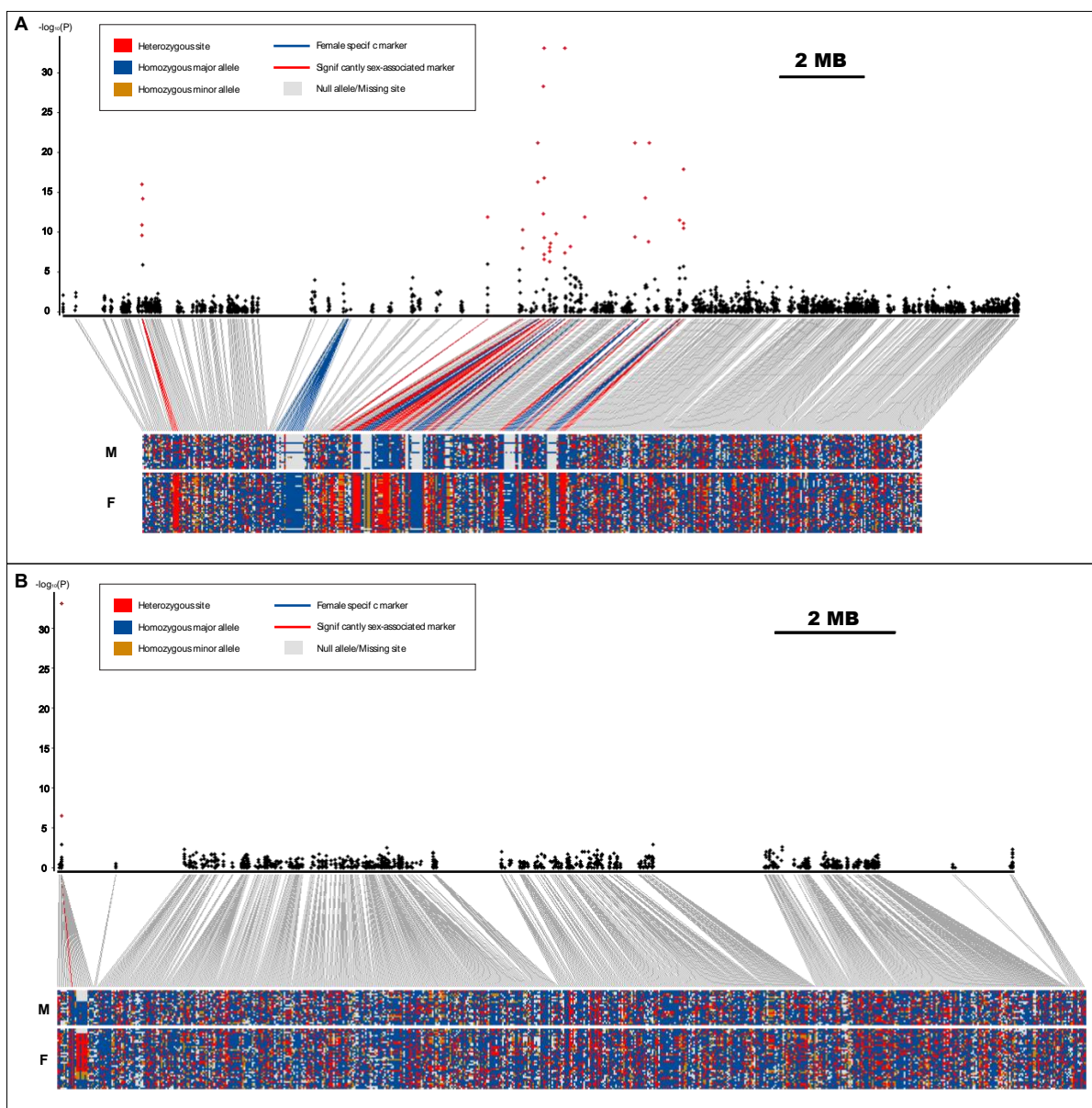
Figure 4 Genotype configurations of markers on chromosome 15 (A) and chromosome 19 (B) from the *S. purpurea* association population. The top is a blowup of chromosome 15 from the Manhattan plot in Figure 2, with significantly sex-associated markers colored red. The bottom shows the genotype configurations of 22 males and 38 females from the association population, where each row represents an individual. "Major alleles" are here defined as those with higher frequency in males, and are shaded blue where homozygous, while homozygotes for male minor alleles are shaded gold. Heterozygous sites are shaded red, and missing data is light gray. Lines connect each plotted marker to its physical position. Red lines indicate that markers are significantly associated with sex while blue lines indicate the markers were identified as female-specific (putatively derived from the W haplotype).

355

15

356 that were also identified as putative W segments based on allelic configurations (see above).

357 Portions of five of these scaffolds had high sequence similarity to chromosome 15, supporting

358 the contention that these are alternate haplotypes from the SDR. For example, Scaffold0265 is

359 298 kb in length and contains 38.9% female-specific sequence and 20 W-type GBS markers

360 (Table S6). This scaffold also contains three sex-associated markers identified in the GWAS.

361 Cumulatively, these 11 scaffolds covered 1.04 Mb, which is a reasonable lower limit for the size

362 of the divergent portions of the SDR.

363

364 **Table 2** Length of intervals that lacked coverage in alignments of 2x250 bp reads against the

365 reference genome assembly (also derived from female clone 94006). Number in the parentheses

366 is the percentage of the total genome composition in that category that lacked coverage.

367

	Whole Genome	Fish Creek (♂)	94006 (♀)
Total Length	348,745,509	14,564,089 (4.18)	562,813 (0.16)
Chromosomes	251,661,964	7,693,428 (3.06)	303,356 (0.12)
Scaffolds	97,083,545	6,870,661 (7.08)	259,457 (0.27)
Repeats	98,506,863	5,328,429 (5.41)	260,598 (0.26)
Genes	120,852,638	2,654,305 (2.20)	78,325 (0.06)
SDR	3,073,122	480,360 (15.63)	4,814 (0.16)

368
369

The SDR is Highly Repetitive, Has Repressed Recombination, and is Divergent from the

371 *Populus* **SDR**

372   The SDR on chromosome 15 of *S. purpurea* overlaps with a large region (9.8 Mb to 16.2

373 Mb) with elevated physical-to-genetic distance ratio of 0.867 Mb/cM, compared to the genome-

374 wide average of 0.172 Mb/cM (Figure 5), which indicates reduced recombination. This interval

375 contained high repeat abundance relative to the rest of the genome (Figure S7). A portion of the

376 SDR in *S. purpurea* is homologous to the SDR in *S. suchowensis*. The *S. suchowensis* SDR

377 primarily occurs on scaffold64, an ~900 kb scaffold that maps to chromosome 15 (Hou *et al.*

378 2015). Aligning this sequence to the *S. purpurea* genome with lastz, we observed homology

379 from 6.2 to 7.3 Mb and from 14.1 and 15.1 Mb on *S. purpurea* chromosome 15 (Figure S8). The

380 latter sequence overlaps with a portion of the *S. purpurea* SDR. In contrast, the *S. viminalis* SDR

381   matches from 5.9 to 8.4 Mb on *S. purpurea* chromosome 15, which is outside the *S. purpurea*

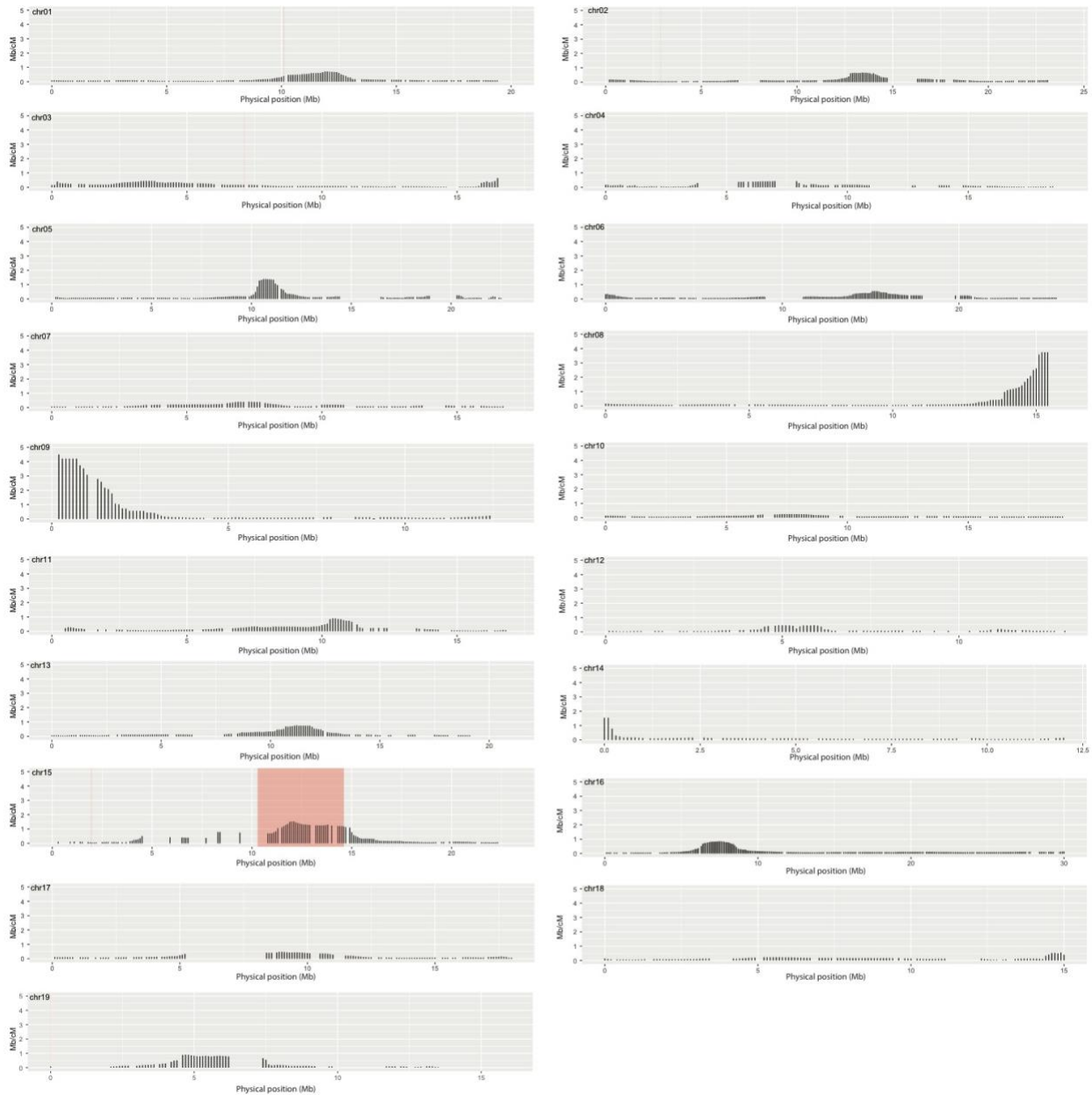382   SDR (Pucholt, Wright, *et al.* 2017).

383

Figure 5 Recombination across the *S. purpurea* genome, as inferred from physical:genetic distance ratio. Bar plots represent the physical:genetic distance ratio (Mb/cM) in 100 kb windows for the 19 chromosomes. The position of the SDRs are indicated by vertical red shading.

384   *P. trichocarpa* is another member of the Salicaceae and has a fairly-well characterized

385   XY system of sex determination (Geraldes *et al.* 2015). In general, *S. purpurea* and *P.*

386   *trichocarpa* have high synteny at the chromosome scale (Figure 6), but chromosome 15 in *S.*

387   *purpurea* stands out in several ways. First, the SDR on chromosome 15 of *S. purpurea* is not

388   syntenic with chromosome 15 or any other chromosome of *P. trichocarpa* (Figure 6). Second,

389   the proportion of repeats is significantly elevated in the *S. purpurea* SDR, with an average of

390   37% repeat composition, compared to the genome-wide average of 24.8% (Welch's Two-Sample

391   T = -4.6 *P*5948, <0.0001; Table S7; Figure S7). Chromosome 19, which contains the SDR in *P.*

392   *trichocarpa*, also had the highest average repeat content in *S. purpurea* (33.5%, compared to

393   25.1% genome-wide average) (Table S7).

394   **Gene Content of the SDR**

395   We identified 251 protein-coding genes within the *S. purpurea* SDR (Table S8). A GO

396   enrichment analysis based on 203 genes annotated with GO terms identified 4 significantly

397   enriched terms (Bonferroni adjusted $P < 2.45 \times 10^{-4}$), all of which were related to microtubule

398   functions. These include microtubule-based movement (GO:0007018), microtubule motor

399   activity (GO:0003777) and microtubule binding (GO:0008017), as well as kinesin complex

400   (GO:0005871) (Table 3). This enrichment is partly due to two pairs of tandemly-duplicated

401   kinesin-like genes in the SDR (Table S8). Since there is only one homolog of these kinesin-like

402   genes in *P. trichocarpa*, it appears that this expansion occurred after the divergence of the two

403   genera, a scenario supported by high sequence conservation between the tandem duplicates

404   (Figure S9).

405

406   **Table 3** Significantly overrepresented GO terms of candidate genes from SDR.

407

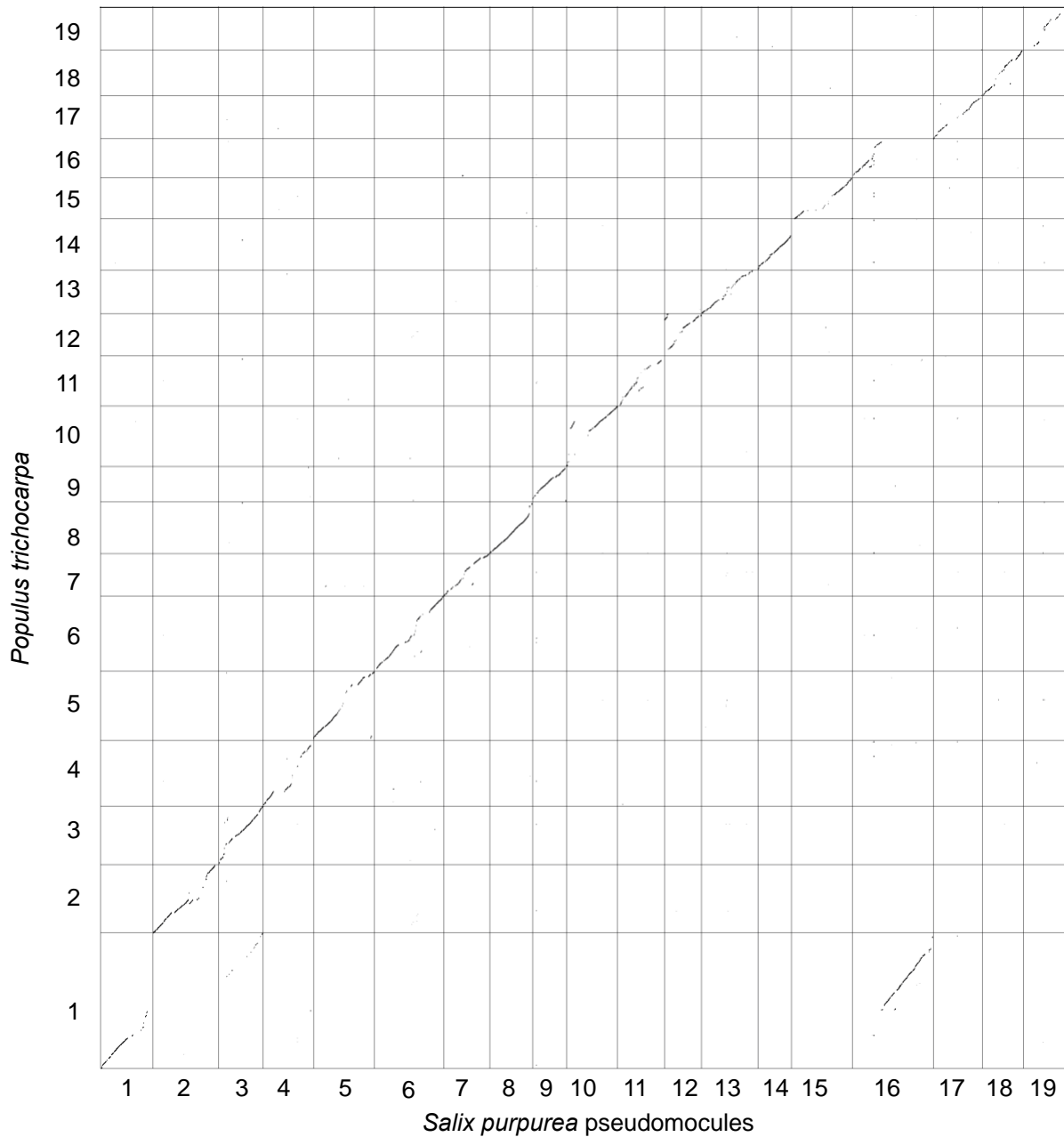| Description	GO term	Number of genes in SDR	Number of genes outside SDR	P value
Microtubule motor activity	GO:0003777	7	91	4.73×10^{-6}
Kinesin complex	GO:0005871	7	92	5.07×10^{-6}
Microtubule-based movement	GO:0007018	7	92	5.07×10^{-6}
Microtubule binding	GO:0008017	7	133	4.84×10^{-5}

408

Figure 6 Comparison between the *S. purpurea* (x-axis) and *P. trichocarpa* (y-axis) genomes. The two genomes are largely syntenic based on genome-scale alignments using LASTZ, with parameters set to exclude paralogous segments derived from the most recent whole genome duplication.

409

410   The SDR contains 20 genes that have >70% female-specific sequence, and many of these

411   genes also show sex-biased expression in stem tissue in *S. purpurea* (Table S8; Carlson *et al.*

412   2017). These include an extracellular calcium-sensing receptor (SapurV1A.0301s0080), an auxin

413   response factor (SapurV1A.0718s0100), a peptidase M50B-like protein (SapurV1A.0475s0170),

414   a zinc finger C3hC4 type transcription factor (SapurV1A.0301s0170), and a reticulon-like

415   protein (SapurV1A.0530s0130). Among these, only the reticulon-like protein showed an

416   elevated dN/dS ratio when compared to *P. trichocarpa* (0.687, versus a genome-wide average of

417   0.406). Of the 14 genes that showed significant female-biased expression in the SDR, only one

418   lacked female-specific sequence (SapurV1A.1386s0030, a small heat shock protein). No genes

419   showed significant male-biased expression after Bonferroni correction.

420   The chromosome 19 SDR is of particular interest, since it overlaps with the SDR of *P.*

421   *trichocarpa*. This region spans approximately 10 kb in the current assembly, and harbors three

422   small genes. SapurV1A.1005s0060 contains a Small MutS-Related (SMR) domain. A second

423   gene, SapurV1A.1005s0050, is a calcium-dependent kinase with two EF-Hand domains. The

424   third gene, SapurV1A.1005s0070, encodes a hypothetical protein (Table S8). None of these

425   genes have sex-biased expression or unusual dN/dS ratios compared to *Populus* (Table S8).

426   We attempted to estimate the relative age of the region of suppressed recombination by

427   estimating the rate of synonymous substitution of W alleles compared to Z alleles in the SDR.

428   Calculated this way, the Z-W synonymous substitution rate within the SDR was 0.00343 while

429   the rate calculated the same way outside of the SDR was 0.00151. These differences were

430   statistically significant (t = -4.099; *df* = 249; P = 5.63e-05). For comparison, we also calculated

431   divergence between alleles using all observed polymorphisms. Genes within the SDRs showed

432   similar overall divergence (dS=0.00616) compared to genes outside the SDRs (dS=0.00607), and

433   the difference was not significant (t = -0.077; *df* = 235; P = 0.938). There was no evidence of

434   evolutionary strata in the SDR based on lack of clustering of genes with similar dS values.

435   **Discussion**

436

437   **The *S. purpurea* SDR is Similar to Other *Salix* Species and Divergent from *Populus***

438

439   In all three of the *Salix* species studied thus far, *S. viminalis* (Pucholt *et al.* 2015), *S.*

440   *suchowensis* (Hou *et al.* 2015; Chen *et al.* 2016), and now *S. purpurea*, the largest SDR is on

441 chromosome 15, and shows clear female heterogamety. Furthermore, the *S. suchowensis* SDR

442 overlaps with a portion of the *S. purpurea* SDR, but the *S. viminalis* SDR does not. This may

443 reflect the evolutionary distinctness of *S. viminalis* from the other two taxa. Based on

444 morphological characters, *S. viminalis* belongs to section *Viminella*, which is strongly

445 differentiated from section *Helix*, which contains *S. purpurea* (Argus 1997). This is similar to the

446 situation in *Populus*, where the location of the sex determination region varies across different

447 sections of the genus, though all are located on chromosome 19 (Gaudet *et al.* 2008; Pakull *et al.*

448 2009, 2014; Paolucci *et al.* 2010; Tuskan *et al.* 2012; Kersten *et al.* 2014; Geraldes *et al.* 2015).

449 Comparison of the sequence composition of the *Salix* SDRs and the *P. trichocarpa* SDR

450 revealed no extensive stretches of homology, suggesting a largely independent evolution of these

451 genome regions (Hou *et al.* 2015; Pucholt, Hallingbäck, *et al.* 2017). Clearly, the SDR is highly

452 dynamic within this family.

453 The alternative peaks from the GWAS analysis on chromosomes 1, 2, 3, and 5 were not

454 upheld by the QTL analysis, and mainly consisted of isolated markers. This is unlikely to

455 represent a case of multi-locus sex determination (Moore and Roberts 2013), as the evidence is

456 weak since there is little other corroborating information. The peaks on chromosomes 2, 3, and 5

457 consisted of solitary markers, while that on chromosome 1 included 5 markers that occurred

458 within a 1 kb interval. Our results are similar to those in *P. trichocarpa*, which also contained

459 multiple secondary GWAS peaks in a sex determination GWAS (Geraldes *et al.* 2015).

460 While some of the secondary *Populus* peaks appear to be assembly and/or alignment artifacts

461 (Geraldes *et al.* 2015), we found no evidence of assembly errors in these regions for *S. purpurea*

462 based on examining the sequence assembly itself as well as the underlying genetic map.

463 Problems with assembly of SDRs are common, presumably due to strong haplotype divergence

464 and high repeat composition, which impede assembly of short-read sequence data (Miller *et al.*

465 2010). Furthermore, the suppressed recombination in these regions inhibits map-based assembly

466 methods. An alternative explanation for the secondary peaks is recent translocation from those

467 chromosomes to the W chromosome in *S. purpurea*. If the W haplotype is not represented in the

468 reference genome assembly, then the reads derived from the recently-translocated regions could

469 align to their original locations. Short-read sequence aligners like Bowtie2 do not handle

470 repetitive sequences well, and commonly misalign reads derived from such regions (Lian *et al.*

471 2016).

472   The GWAS peak on chromosome 19 are especially interesting because it coincides with

473   the position of one of the SDRs in *Populus*. This peak also has more corroborating evidence than

474   the other secondary peaks because it had one of the lowest observed P-values, and it is

475   recapitulated in the QTL analysis. Furthermore, the peak on chromosome 3 best matches a

476   scaffold from the SDR region of *Populus* on chromosome 19, so at least two independent

477   association results point to sex-specific genotypes in genomic segments with homology to the

478   *Populus* SDR. If these represent recent translocations, then this could be a clue to the origin of

479   the chromosome 15 SDR in the *Salix* lineage.

480   **Recombination Suppression and Relative Age of the SDR**

481   Reduced recombination is a crucial component of sex chromosome evolution which

482   ensures that male and female sterility factors do not co-occur in the zygote (Bergero and

483   Charlesworth 2009; Ming *et al.* 2011). As expected, we observed reduced recombination across

484   most of the SDR in *S. purpurea* (Figure 5). This could be caused by large-scale structural

485   polymorphisms and reinforced by the accumulation of nonhomologous sequences in the female-

486   specific haplotype (Ming *et al.* 2011; Charlesworth 2015). The SDR also shows a higher

487   proportion of repetitive elements, as expected in regions with reduced recombination. Similar

488   features are also apparent within the SDR of *S. suchowensis* and *S. viminalis* (Hou *et al.* 2015;

489   Pucholt *et al.* 2015; Chen *et al.* 2016), but are not as apparent for the *P. trichocarpa* SDR, which

490   is estimated to be quite small (Geraldes *et al.* 2015). If this is accurate, it could indicate that the

491   *P. trichocarpa* region has not yet developed these features, or that it is highly dynamic. In the

492   case of *S. purpurea*, the SDR is quite large, with a lower limit of 1.04 Mb (based on the

493   cumulative length of female-specific scaffolds), and an upper limit of approximately 5 Mb, based

494   on suppressed recombination and the occurrence of SNPs that are significantly associated with

495   sex. It is possible that the SDR overlaps with the centromere on chromosome 15, and this could

496   contribute to the large apparent size of the region of suppressed recombination. However, the

497   SDR does not contain any of the tandem minisatellite repeats that are apparently characteristic of

498   the *S. purpurea* centromeres, as identified in a previous study (Melters *et al.* 2013). It remains to

499   be seen if the lack of these repeats is due to poor assembly, or if the centromere is located

500   elsewhere on this chromosome.

501   Divergence between Z and W transcripts in the *S. purpurea* SDR is relatively low,

502   suggesting that suppression of recombination is incomplete or recently established. This is

503 similar to the SDRs of *P. trichocarpa* (Geraldes *et al.* 2015) and *S. viminalis* (Pucholt, Wright, *et*
504 *al.* 2017), which also show low divergence of sex-specific sequences. Furthermore, we saw no
505 evidence of the presence of evolutionary strata within or around the *S. purpurea* SDR. Such
506 features occur due to the establishment of regions of suppressed recombination at different times
507 during sex chromosome evolution (Charlesworth 2016). Evolutionary strata are apparent in well-
508 established SDRs of other plants, including *Silene latifolia* (Bergero *et al.* 2007) and *Carica*
509 *papaya* (Wang *et al.* 2012). However, no such regions were detected in *S. suchowensis* (Pandey
510 and Azad 2016). Given the low divergence, lack of strata, and the frequent movement of the
511 SDR within the family, it is reasonable to conclude that the SDR is highly dynamic in this
512 family, and that sex determination loci frequently translocate to new positions and/or are
513 superseded by other loci on autosomes, as predicted by theoretical models of SDR movement
514 (van Doorn and Kirkpatrick 2007, 2010).

515
516 **Candidate Genes and Their Function**
517
518   The SDRs are genomic regions that are statistically associated with gender. This
519 association must be due to the presence of loci that control sex determination, but the regions
520 also likely harbor loci that are under sexually antagonistic selection (van Doorn and Kirkpatrick
521 2007; Bachtrog *et al.* 2014). The gene content of these regions could therefore provide insights
522 about mechanisms of sex determination as well as sex dimorphism. We identified 251 protein-
523 coding genes in the SDRs of *S. purpurea* (Table S8). Most have not been functionally annotated,
524 but clues can be inferred based on conserved domains and their predicted function in model
525 organisms. It is also important to note that the assembly problems mentioned previously have
526 probably prevented full enumeration of the gene content of the SDRs. This problem may be
527 particularly challenging for female-specific portions of the W chromosome (Pucholt *et al.* 2015).
528 Nevertheless, there are several genes in this region that could plausibly be involved in floral
529 development and sex-specific regulation that are worthy of consideration.
530   Since floral morphology is the most striking difference between the sexes, it is reasonable
531 to expect that genes involved in floral development would be located in the SDRs. Indeed, the
532 SDR contains SapurV1A.0718s0010, an ortholog of WUSCHEL-related homeotic genes (e.g.,
533 *WOX1*). Orthologs in other species, including STF in *Medicago truncatula*, LAM1 in *Nicotiana*
534 *sylvestris*, and MAW in *Petunia*, are key regulators of the lateral outgrowth of leaf blades and

535   floral organs (Lin *et al.* 2013). This gene showed slightly elevated expression in male shoot tips

536   compared to female shoot tips (Table S7).

537   Several genes in the SDR may be involved specifically with male development and

538   function. For example, our analysis of GO term over-representation highlighted the presence of

539   seven genes containing the kinesin motor domain (PF00225), which is involved in microtubule-

540   based movement or organelles, including during pollen tube growth (Cai and Cresti 2009). For

541   example, loss-of-function mutants of the closest homolog of SapurV1A.0530s0110 in

542   *Arabidopsis thaliana* (*NACK*1) showed reduced growth and prematurely-terminated petals,

543   pistils, and stamens (Nishihama *et al.* 2002).

544   Two other genes in the SDR may be related to pollen function. First,

545   SapurV1A.1741s0030, is a homolog of *PLANT INTRACELLULAR RAS GROUP-RELATED*

546   *LRR 3* (*PIRL3*), which has been implicated in pollen development in *Arabidopsis* (Forsthoefel *et*

547   *al.* 2013). The second, SapurV1A.0301s0130 is a homolog of the *Arabidopsis* gene

548   AT3G01570.1, a member of the oleosin family (Kim *et al.* 2002). This gene occurs in a female-

549   specific portion of the SDR and has female-biased gene expression, perhaps reflecting a potential

550   role in sex dimorphism in *Salix*. Other members of the oleosin family have been associated with

551   pollen wall development, and eight of these have been shown to have tapetal-specific expression

552   in *Arabidopsis* (Kim *et al.* 2002; Hsieh and Huang 2004; Yang *et al.* 2007). The expression of

553   one of these oleosin genes is regulated by *MALE STERILITY1* (*MS1*) in *Arabidopsis*, which

554   controls pollen and tapetal development (Yang *et al.* 2007), raising the possibility that altered

555   regulation of oleosin genes could provide a pathway to male sterility.

556   The SDR on chromosome 19 deserves special attention due to its shared homology with

557   the *Populus* SDR. One particularly interesting gene in this region is SapurV1A.1005s0060,

558   which contains a Small MutS-Related (SMR) domain and a domain of unknown function

559   (DUF1771). These domains frequently occur together in eukaryotes, but the function of

560   DUF1771 has yet to be characterized (Fukui and Kuramitsu 2011). Proteins with the SMR

561   domain, such as MutS2, can suppress (Fukui *et al.* 2007; Fukui and Kuramitsu 2011) or promote

562   (Burby and Simmons 2017) homologous recombination by endonucleolytic digestion, and are

563   involved in mismatch repair in diverse prokaryotes (Kunkel and Erie 2005). The roles of the

564   SMR domain in plants are not fully characterized, but when coupled with the pentatricopeptide

565   repeat motif, the SMR domain shows sequence-specific RNA endonuclease activity and affects

566 chloroplast function (Zhou *et al.* 2017). Due to its potential roles in recombination, mismatch

567 repair, and regulation of organellar function, this gene is an intriguing candidate in the context of

568 sex determination as well as mediation of the female-biased sex ratios that are commonly

569 observed in *Salix* (Alliende and Harper 1989; Alstrom-Rapaport *et al.* 1998; Ueno *et al.* 2007;

570 Pucholt, Hallingbäck, *et al.* 2017), including in *S. purpurea*, as reported here.

571 **Sex Chromosome Evolution in the Salicaceae**

572   *Populus* and *Salix* are closely-related genera that share many key characteristics, the most

573 notable of which is that they are both nearly fixed for dioecy. *Populus* first appears in the fossil

574 record between 40 and 60 MYA, apparently slightly earlier than *Salix* (Boucher *et al.* 2003).

575 However, *Populus* and *Salix* exhibit much less divergence in nucleotide sequence and

576 chromosome structure than expected, presumably due to long average generation times (Sterck *et*

577 *al.* 2005; Hou *et al.* 2016). It may therefore seem surprising that the chromosomal location and

578 gene content of the SDRs are so different, and that they have different heterogametic

579 configurations (Hou *et al.* 2015; Pucholt *et al.* 2015). In fact, movement of sex determination

580 loci and transitions between XY and ZW systems are well-known in organisms that lack

581 strongly-differentiated, heteromorphic sex chromosomes (Bachtrog *et al.* 2014).

582   A striking finding of this study is the existence of multiple sexually dimorphic regions in

583 the *S. purpurea* genome, one of which is on chromosome 15 and shared with other *Salix* species

584 (Pucholt *et al.* 2015; Chen *et al.* 2016), and one on chromosome 19, which harbors the SDR of

585 multiple *Populus* species (Tuskan *et al.* 2012; Kersten *et al.* 2014; Geraldes *et al.* 2015). There

586 are several lines of evidence to support a model whereby the original sex determination locus

587 was located on chromosome 19 in the common ancestor of *Salix* and *Populus*. First, *Salix*

588 chromosome 19 shows overall high synteny with *P. trichocarpa* chromosome 19, and both

589 species apparently have an SDR in the same region of this chromosome. In contrast,

590 chromosome 15 shows no sex chromosome characteristics in *Populus*, and the composition of

591 chromosome 15 is quite different in the *Salix* SDR. Second, we were unable to create a female

592 backcross map for *S. purpurea* chromosome 19, the only chromosome for which mapping failed.

593 This was due to a paucity of genetic markers in this region, particularly in the female backcross

594 configuration. Notably, such a configuration would be absent in the SDR of an XY sex

595 determination system. On the other hand, the overall lack of genetic markers is probably caused

596 in part by the high repeat content of this chromosome, which can inhibit genotyping based on

597   short read sequences (Treangen and Salzberg 2013). High repeat content is also expected in

598   regions of reduced recombination, commonly found in sex chromosomes (Ming and Moore

599   2007; Bergero and Charlesworth 2009). In summary, because chromosome 19 shows

600   characteristics of an SDR in *S. purpurea* and in multiple *Populus* species, but chromosome 15

601   only shows such characteristics in *Salix*, it is logical to hypothesize that chromosome 19 is the

602   ancestral sex chromosome in the Salicoid lineage.

603   In the present study, it is important to reemphasize that the locus mapped to chromosome

604   19 may be an assembly or alignment artifact. This could be caused by a recent translocation from

605   chromosome 19 to the W haplotype of chromosome 15, which would result in incorrect

606   alignment of GBS reads to the original chromosome 19 locus if the W haplotype is not in the

607   main genome assembly. However, because the locus matches a portion of the SDR of

608   chromosome 19 in *Populus*, and the gene content of these regions is similar between the taxa,

609   this finding would still provide valuable clues about sex determination and/or sex dimorphism in

610   this family even if it is caused by a recent translocation. It is also noteworthy that the *S. purpurea*

611   *de novo* genome assembly did not use the *P. trichocarpa* genome assembly as a reference to

612   guide placement of scaffolds in pseudomolecules (Smart et al., in preparation), so the results

613   reported here are not caused by carryover of biases or errors from the original *P. trichocarpa*

614   assembly.

615   Unfortunately, a definitive comparison of the Salicaceae sex chromosomes is not possible

616   with the currently-available genome sequences. The SDRs of *Salix* and *Populus* are typical in

617   that they have complex structural polymorphisms, high repeat content, and low recombination

618   rates, all of which contribute to fragmentary and erroneous genome assemblies (Geraldes *et al.*

619   2015). Efforts are underway to assemble these regions using long read sequencing and dense

620   genetic mapping in multiple pedigrees. This will facilitate analyses that can date the origin of

621   these regions based on differentiation of sex-specific haplotypes in the non-recombining portions

622   of the SDR (Otto *et al.* 2011). Furthermore, elucidation of the sex determination system in

623   additional Salicaceae taxa should help to determine the ancestral state. This family should

624   therefore be instrumental in advancing our knowledge of the evolution and ecological

625   significance of sex chromosomes as genetic and genomic resources continue to accumulate.

Supplementary Materials

Figure S1 Pairwise Scaled Identity by State (IBS) for the (a) complete association population (N=112), (b) the complete F_2 full sib Family (N=497), and (c) the association population with clones removed (N=75). The IBS cutoff used for identifying clonal pairs was 0.9.

Figure S2 Frequency of mapped markers with and without segregation distortion in family 317 for males and females. A. Markers in female-backcross configuration. B. Markers in male-backcross configuration. Notice the lack of undistorted (normal) markers on chromosome 19 in female backcross configuration.

Figure S3 Quantile–Quantile (Q–Q) plots of observed and expected P-values for the GWAS for sex. Red line indicates X = Y.

Figure S4 Stacked histogram of average observed heterozygosity for males, females, and hermaphrodites for sex-associated loci in the *S. purpurea* association population.

Figure S5 Distribution of differences in null allele frequency between females and males in the association population. Extreme values are shaded in red.

Figure S6 Proportion of reference sequence gaps ("assembly Ns") in regions that showed no coverage in the female (a) or male (b) reference-based alignments. The male had 0 coverage primarily in regions with minimal reference gaps, suggesting that these are regions that are present in the female sequence and absent in the male.

Figure S7 Box plot showing that the proportion of repeat elements is elevated in the SDR.

Figure S8 Dot plot derived from aligning the *S. suchowensis* SDR (primarily located on scaffold64) to *S. purpurea* chromosome 15 using lastz.

657   **Figure S9** Alignment of Kinesin genes from the SDR of *S. purpurea* and their closest ortholog in

658   *P. trichocarpa.* SapurV1A.1267s0010 is artificially truncated due to an assembly gap

659   overlapping with the gene. Conserved domains are highlighted and labeled. Tandem duplicate

660   pairs are 1.) SapurV1A.0719s0080 and SapurV1A.0719s0090; and 2.) SapurV1A.1267s0010 and

661   SapurV1A.1267s0020.

662

663   **Table S1** Significant markers (LOD>3.5) from QTL mapping of sex. The table includes linkage

664   group (LG), map positions (in centimorgans), map type (female backcross, F, male backcross,

665   M, and intercross, IC), the physical scaffold from the genome assembly, the physical position of

666   the marker in the genome assembly, and the frequency of different genotype configurations in

667   the progeny.

668   **Table S2** Number of unfiltered GBS markers produced by the Tassel pipeline for the F_2 family

669   317. Markers/100kb is the average number of markers per 100 kb interval. F:M Backcross is the

670   ratio of markers in a Female Backcross configuration (heterozygous in the female parent,

671   homozygous in the male parent) to markers in the Male Backcross configuration (homozygous in

672   female parent, heterozygous in male parent).

673   **Table S3** Results of GWAS for sex. The table includes all significant markers ($p<1 \times 10^{-7}$).

674   **Table S4** Best matches for secondary *S. purpurea* SDRs to the *S. purpurea* and *P. trichocarpa*

675   genomes. "Secondary Blast Hit" is the best blastn hit to the *S. purpurea* genome, after excluding

676   self hits.

677   **Table S5** Markers showing a female-specific genotype configuration (one allele observed in

678   females, none in males). These are presumably derived from W segments included in the genome

679   assembly.

680   **Table S6** Scaffolds with >30% female-specific sequence. "Proportion W" is a calculation based

681   on the proportion of the scaffold, after excluding gaps, that is present in the female sequence but

682   absent in the male sequence (Female-Specific).

683   **Table S7** Repeat composition of the *S. purpurea* chromosomes.

684   **Table S8** Predicted genes found within the SDR of *S. purpurea*. "W Overlap" and "W

685   proportion" represent the intersection of the location of the gene with female-specific genome

686   segments. Omega values are the ratio of nonsynonymous (dN) to synonymous (dS) substitutions

687   between the *S. purpurea* and *P. trichocarpa* orthologs. Multiple values are provided in cases

688   with multiple *Populus* orthologs, presumably due to lineage-specific expansion.

689

690   **Literature Cited**

691

692   Ainsworth, C., 2000 Boys and girls come out to play: The molecular biology of dioecious plants.

693   Ann. Bot. 86: 211–221.

694   Alliende, M. C., and J. L. Harper, 1989 Demographic studies of a dioecious tree. I. Colonization,

695   sex and age structure of a population of *Salix cinerea*. J. Ecol. 77: 1029–1047.

696   Alstrom-Rapaport, C., M. Lascoux, Y. C. Wang, G. Roberts, and G. A. Tuskan, 1998

697   Identification of a RAPD marker linked to sex determination in the basket willow (*Salix*

698   *viminalis* L.). J. Hered. 89: 44–49.

699   Andrews, K. R., J. M. Good, M. R. Miller, G. Luikart, and P. A. Hohenlohe, 2016 Harnessing

700   the power of RADseq for ecological and evolutionary genomics. Nat Rev Genet 17: 81–92.

701   Arends, D., P. Prins, R. C. Jansen, and K. W. Broman, 2010 R/qtl: High-throughput multiple

702   QTL mapping. Bioinformatics 26: 2990–2992.

703   Argus, G. W., 1997 Infrageneric classification of *Salix* (Salicaceae) in the new world. Syst. Bot.

704   Monogr. 52: 1–121.

705   Ashman, T.-L., 2006 The evolution of separate sexes: a focus on the ecological context, pp. 370

706   in *Ecology and evolution of flowers*, edited by L. D. Harder and S. C. H. Barrett. Oxford

707   University Press, Oxford.

708   Ashman, T.-L., A. Kwok, and B. C. Husband, 2013 Revisiting the Dioecy-Polyploidy

709   Association: Alternate Pathways and Research Opportunities. Cytogenet. Genome Res. 140:

710   241–255.

711   Bachtrog, D., J. Mank, C. L. Peichel, M. Kirkpatrick, S. P. Otto *et al.*, 2014 Sex Determination:

712   Why So Many Ways of Doing It? PLoS Biol. 12: e1001899.

713   Barrett, S. C. H., and J. Hough, 2013 Sexual dimorphism in flowering plants. J. Exp. Bot. 64:

714   67–82.

715   Bergero, R., and D. Charlesworth, 2009 The evolution of restricted recombination in sex
716   chromosomes. Trends Ecol. Evol. 24: 94–102.

717   Bergero, R., A. Forrest, E. Kamau, and D. Charlesworth, 2007 Evolutionary strata on the X
718   chromosomes of the dioecious plant *Silene latifolia*: Evidence from new sex-linked genes.
719   Genetics 175: 1945–1954.

720   Boucher, L. D., S. R. Manchester, and W. S. Judd, 2003 An extinct genus of Salicaceae based on
721   twigs with attached flowers, fruits, and foliage from the Eocene Green River Formation of
722   Utah and Colorado, USA. Am. J. Bot. 90: 1389–1399.

723   Burby, P. E., and L. A. Simmons, 2017 MutS2 promotes homologous recombination in *Bacillus*
724   *subtilis*. J. Bacteriol. 199: e00682-16.

725   Cai, G., and M. Cresti, 2009 Organelle motility in the pollen tube: a tale of 20 years. J. Exp. Bot.
726   60: 495–508.

727   Carlson, C. H., Y. Choi, A. P. Chan, M. J. Serapiglia, C. D. Town *et al.*, 2017 Dominance and
728   Sexual Dimorphism Pervade the *Salix purpurea* L. Transcriptome. Genome Biol. Evol. 9:
729   2377–2394.

730   Charlesworth, D., 2006 Evolution of Plant Breeding Systems. Curr. Biol. 16: 726–735.

731   Charlesworth, D., 2015 Plant contributions to our understanding of sex chromosome evolution.
732   New Phytol. 208: 52–65.

733   Charlesworth, D., 2016 Plant Sex Chromosomes. Annu. Rev. Plant Biol. 67: 397–420.

734   Charlesworth, D., and B. Charlesworth, 1978 Population genetics of partial male-sterility and the
735   evolution of monoecy and dioecy. Heredity 41: 137–153.

736   Charnov, E. L., 1982 The theory of sex allocation. Monogr. Popul. Biol. 18: 1–355.

737   Chen, Y., T. Wang, L. Fang, X. Li, and T. Yin, 2016 Confirmation of single-locus sex
738   determination and female heterogamety in willow based on linkage analysis. PLoS One 11:
739   e0147671.

740   DePristo, M. a, E. Banks, R. Poplin, K. V Garimella, J. R. Maguire *et al.*, 2011 A framework for
741   variation discovery and genotyping using next-generation DNA sequencing data. Nat.
742   Genet. 43: 491–8.

743 Dickmann, D. I., and J. Kuzovkina, 2014 Poplars and willows of the world, with emphasis on
744   silviculturally important species., pp. 8–91 in *Poplars and willows: trees for society and the*
745   *environment*, CABI, Wallingford.

746 van Doorn, G. S., and M. Kirkpatrick, 2010 Transitions between male and female heterogamety
747   caused by sex-antagonistic selection. Genetics 186: 629–645.

748 van Doorn, G. S., and M. Kirkpatrick, 2007 Turnover of sex chromosomes induced by sexual
749   conflict. Nature 449: 909–912.

750 Elshire, R. J., J. C. Glaubitz, Q. Sun, J. a. Poland, K. Kawamoto *et al.*, 2011 A robust, simple
751   genotyping-by-sequencing (GBS) approach for high diversity species. PLoS One 6: e19379.

752 Falda, M., S. Toppo, A. Pescarolo, E. Lavezzo, B. Di Camillo *et al.*, 2012 Argot2: a large scale
753   function prediction tool relying on semantic similarity of weighted Gene Ontology terms.
754   BMC Bioinformatics 13: S14.

755 Forsthoefel, N., K. Klag, B. Simeles, R. Reiter, L. Brougham *et al.*, 2013 The *Arabidopsis* Plant
756   Intracellular Ras-group LRR (PIRL) Family and the Value of Reverse Genetic Analysis for
757   Identifying Genes that Function in Gametophyte Development. Plants 2: 507–520.

758 Fukui, K., H. Kosaka, S. Kuramitsu, and R. Masui, 2007 Nuclease activity of the MutS
759   homologue MutS2 from *Thermus thermophilus* is confined to the Smr domain. Nucleic
760   Acids Res. 35: 850–860.

761 Fukui, K., and S. Kuramitsu, 2011 Structure and Function of the Small MutS-Related Domain.
762   Mol. Biol. Int. 2011: 1–9.

763 Füssel, U., S. Dötterl, A. Jürgens, and G. Aas, 2007 Inter- and Intraspecific Variation in Floral
764   Scent in the Genus *Salix* and its Implication for Pollination. J. Chem. Ecol. 33: 749–765.

765 Gaudet, M., V. Jorge, I. Paolucci, I. Beritognolo, G. S. Mugnozza *et al.*, 2008 Genetic linkage
766   maps of *Populus nigra* L. including AFLPs, SSRs, SNPs, and sex trait. Tree Genet.
767   Genomes 4: 25–36.

768 Geraldes, A., C. A. Hefer, A. Capron, N. Kolosova, F. Martinez-Nuñez *et al.*, 2015 Recent Y
769   chromosome divergence despite ancient origin of dioecy in poplars (*Populus*). Mol. Ecol.
770   24: 3243–3256.

771   Glaubitz, J. C., T. M. Casstevens, F. Lu, J. Harriman, R. J. Elshire *et al.*, 2014 TASSEL-GBS: A

772   high capacity genotyping by sequencing analysis pipeline. PLoS One 9: e0090346.

773   Glick, L., N. Sabath, T. L. Ashman, E. Goldberg, and I. Mayrose, 2016 Polyploidy and sexual

774   system in angiosperms: Is there an association? Am. J. Bot. 103: 1223–1235.

775   Gnerre, S., I. MacCallum, D. Przybylski, F. J. Ribeiro, J. N. Burton *et al.*, 2011 High-quality

776   draft assemblies of mammalian genomes from massively parallel sequence data. Proc. Natl.

777   Acad. Sci. U. S. A. 108: 1513–1518.

778   Goodstein, D. M., S. Shu, R. Howson, R. Neupane, R. D. Hayes *et al.*, 2012 Phytozome: a

779   comparative platform for green plant genomics. Nucleic Acids Res. 40: D1178–D1186.

780   Hou, J., N. Ye, Z. Dong, M. Lu, L. Li *et al.*, 2016 Major chromosomal rearrangements

781   distinguish willow and poplar after the ancestral "Salicoid" genome duplication. Genome

782   Biol. Evol. 8: 1868–1875.

783   Hou, J., N. Ye, D. Zhang, Y. Chen, L. Fang *et al.*, 2015 Different autosomes evolved into sex

784   chromosomes in the sister genera of *Salix* and *Populus*. Sci. Rep. 5: e9076.

785   Hsieh, K., and A. H. C. Huang, 2004 Endoplasmic reticulum, oleosins, and oils in seeds and

786   tapetum cells. Plant Physiol. 136: 3427–3434.

787   Kang, H. M., J. H. Sul, S. K. Service, N. a Zaitlen, S.-Y. Kong *et al.*, 2010 Variance component

788   model to account for sample structure in genome-wide association studies. Nat. Genet. 42:

789   348–354.

790   Karp, A., S. J. Hanley, S. O. Trybush, W. Macalpine, M. Pei *et al.*, 2011 Genetic Improvement

791   of Willow for Bioenergy and Biofuels. J. Integr. Plant Biol. 53: 151–165.

792   Karrenberg, S., J. Kollmann, and P. J. Edwards, 2002 Pollen vectors and inflorescence

793   morphology in four species of *Salix*. Plant Syst. Evol. 235: 181–188.

794   Kersten, B., B. Pakull, K. Groppe, J. Lueneburg, and M. Fladung, 2014 The sex-linked region in

795   *Populus tremuloides* Turesson 141 corresponds to a pericentromeric region of about two

796   million base pairs on *P. trichocarpa* chromosome 19. Plant Biol. 16: 411–418.

797   Kim, H. U., K. Hsieh, C. Ratnayake, and A. H. C. Huang, 2002 A novel group of oleosins is

798   present inside the pollen of *Arabidopsis*. J. Biol. Chem. 277: 22677–22684.

799　Kunkel, T. a, and D. a Erie, 2005 DNA mismatch repair. Annu. Rev. Biochem. 74: 681–710.

800　Lian, S., T. Liu, K. Gong, X. Chen, and G. Zheng, 2016 A Complete and Accurate Short
801　　　Sequence Alignment Algorithm for Repeats. J. Biosci. Med. 4: 144–151.

802　Lin, H., L. Niu, N. a McHale, M. Ohme-Takagi, K. S. Mysore *et al.*, 2013 Evolutionarily
803　　　conserved repressive activity of WOX proteins mediates leaf blade outgrowth and floral
804　　　organ development in plants. Proc. Natl. Acad. Sci. U. S. A. 110: 366–371.

805　Lloyd, D. G., 1979 Evolution towards dioecy in heterostylous populations. Plant Syst. Evol. 131:
806　　　71–80.

807　Mank, J. E., 2009 Sex chromosomes and the evolution of sexual dimorphism: lessons from the
808　　　genome. Am. Nat. 173: 141–150.

809　Melters, D. P., K. R. Bradnam, H. a Young, N. Telis, M. R. May *et al.*, 2013 Comparative
810　　　analysis of tandem repeats from hundreds of species reveals unique insights into centromere
811　　　evolution. Genome Biol. 14: R10.

812　Miller, J. R., S. Koren, and G. Sutton, 2010 Assembly algorithms for next-generation sequencing
813　　　data. Genomics 95: 315–27.

814　Ming, R., A. Bendahmane, and S. S. Renner, 2011 Sex chromosomes in land plants. Annu. Rev.
815　　　Plant Biol. 62: 485–514.

816　Ming, R., and P. H. Moore, 2007 Genomics of sex chromosomes. Curr. Opin. Plant Biol. 10:
817　　　123–130.

818　Mock, K. E., C. M. Callahan, M. N. Islam-Faridi, J. D. Shaw, H. S. Rai *et al.*, 2012 Widespread
819　　　triploidy in western North American aspen (*Populus tremuloides*). PLoS One 7: e48406.

820　Moore, E. C., and R. B. Roberts, 2013 Polygenic sex determination. Curr. Biol. 23: R510-2.

821　Nicolas, M., G. Marais, V. Hykelova, B. Janousek, V. Laporte *et al.*, 2005 A gradual process of
822　　　recombination restriction in the evolutionary history of the sex chromosomes in dioecious
823　　　plants. PLoS Biol. 3: e4.

824　Nishihama, R., T. Soyano, M. Ishikawa, S. Araki, H. Tanaka *et al.*, 2002 Expansion of the cell
825　　　plate in plant cytokinesis requires a kinesin-like protein/MAPKKK complex. Cell 109: 87–
826　　　99.

827   Olson, M. S., J. L. Hamrick, and R. C. Moore, 2017 Breeding systems, mating systems, and

828   gender determination in angiosperm trees, in *Comparative and Evolutionary Genomics of*

829   *Angiosperm Trees*, edited by A. Groover and Q. C. B. Cronk. Springer International

830   Publishing, Switzerland.

831   Otto, S. P., J. R. Pannell, C. L. Peichel, T.-L. Ashman, D. Charlesworth *et al.*, 2011 About PAR:

832   the distinct evolutionary dynamics of the pseudoautosomal region. Trends Genet. 27: 358–

833   367.

834   Pakull, B., K. Groppe, M. Meyer, T. Markussen, and M. Fladung, 2009 Genetic linkage mapping

835   in aspen (*Populus tremula* L. and *Populus tremuloides* Michx.). Tree Genet. Genomes 5:

836   505–515.

837   Pakull, B., B. Kersten, J. Lüneburg, and M. Fladung, 2014 A simple PCR-based marker to

838   determine sex in aspen. Plant Biol. 17: 256–261.

839   Pandey, R. S., and R. K. Azad, 2016 Deciphering evolutionary strata on plant sex chromosomes

840   and fungal mating-type chromosomes through compositional segmentation. Plant Mol. Biol.

841   90: 359–373.

842   Paolucci, I., M. Gaudet, V. Jorge, I. Beritognolo, S. Terzoli *et al.*, 2010 Genetic linkage maps of

843   *Populus alba* L. and comparative mapping analysis of sex determination across Populus

844   species. Tree Genet. Genomes 6: 863–875.

845   Peto, F. H., 1938 Cytology of poplar species and natural hybrids. Can. J. Res. 16: 446–455.

846   Price, A. L., N. J. Patterson, R. M. Plenge, M. E. Weinblatt, N. a Shadick *et al.*, 2006 Principal

847   components analysis corrects for stratification in genome-wide association studies. Nat.

848   Genet. 38: 904–909.

849   Pucholt, P., H. R. Hallingbäck, and S. Berlin, 2017 Allelic incompatibility can explain female

850   biased sex ratios in dioecious plants. BMC Genomics 18: 251.

851   Pucholt, P., A.-C. Rönnberg-Wästljung, and S. Berlin, 2015 Single locus sex determination and

852   female heterogamety in the basket willow (*Salix viminalis* L.). Heredity 114: 575–583.

853   Pucholt, P., A. E. Wright, L. L. Conze, J. E. Mank, and S. Berlin, 2017 Recent Sex Chromosome

854   Divergence despite Ancient Dioecy in the Willow, *Salix viminalis*. Mol. Biol. Evol. 22:

855   522–525.

856   Renner, S. S., 2014 The relative and absolute frequencies of angiosperm sexual systems: dioecy,

857   monoecy, gynodioecy, and an updated online database. Am. J. Bot. 101: 1588–1596.

858   Rice, W. W. R., 1984 Sex chromosomes and the evolution of sexual dimorphism. Evolution 38:

859   1416–1424.

860   Serapiglia, M. J., F. E. Gouker, J. F. Hart, F. Unda, S. D. Mansfield *et al.*, 2015 Ploidy Level

861   Affects Important Biomass Traits of Novel Shrub Willow (*Salix*) Hybrids. BioEnergy Res.

862   8: 259–269.

863   Slavov, G. T., and P. Zhelev, 2010 Salient Biological Features, Systematics, and Genetic

864   Variation of *Populus*, pp. 15–38 in *Genetics and Genomics of* Populus, edited by S.

865   Jansson, R. P. Bhalerao, and A. Groover. Springer New York, New York, NY.

866   Sterck, L., S. Rombauts, S. Jansson, F. Sterky, P. Rouzé *et al.*, 2005 EST data suggest that poplar

867   is an ancient polyploid. New Phytol. 167: 165–170.

868   Temmel, N. A., H. S. Rai, and Q. C. B. Cronk, 2007 Sequence characterization of the putatively

869   sex-linked Ssu72 -like locus in willow and its homologue in poplar. Can. J. Bot. 85: 1092–

870   1097.

871   Treangen, T. J., and S. L. Salzberg, 2013 Repetitive DNA and next-generation sequencing:

872   computational challenges and solutions. Nat Rev Genet. 13: 36–46.

873   Tuskan, G. A., S. DiFazio, P. Faivre-Rampant, M. Gaudet, A. Harfouche *et al.*, 2012 The

874   obscure events contributing to the evolution of an incipient sex chromosome in *Populus*: a

875   retrospective working hypothesis. Tree Genet. Genomes 8: 559–571.

876   Tuskan, G. A., S. DiFazio, S. Jansson, J. Bohlmann, I. Grigoriev *et al.*, 2006 The genome of

877   black cottonwood, *Populus trichocarpa* (Torr. & Gray). Science 313: 1596–1604.

878   Ueno, N., Y. Suyama, and K. Seiwa, 2007 What makes the sex ratio female-biased in the

879   dioecious tree *Salix sachalinensis*? J. Ecol. 95: 951–959.

880   Vyskot, B., and R. Hobza, 2015 The genomics of plant sex chromosomes. Plant Sci. 236: 126–

881   135.

882   Wang, J., J. Na, Q. Yu, A. R. Gschwend, J. Han *et al.*, 2012 Sequencing papaya X and Y h

883   chromosomes reveals molecular basis of incipient sex chromosome evolution. Proc. Natl.

884   Acad. Sci. 109: 13710–13715.

885   Yang, Z., 2007 PAML 4: phylogenetic analysis by maximum likelihood. Mol. Biol. Evol. 24:

886   1586–1591.

887   Yang, C., G. Vizcay-Barrena, K. Conner, and Z. a Wilson, 2007 MALE STERILITY1 is

888   required for tapetal development and pollen wall biosynthesis. Plant Cell 19: 3530–3548.

889   Yin, T., S. P. DiFazio, L. E. Gunter, X. Zhang, M. M. Sewell *et al.*, 2008 Genome structure and

890   emerging evidence of an incipient sex chromosome in *Populus*. Genome Res. 18: 422–430.

891   Zhou, W., Q. Lu, Q. Li, L. Wang, S. Ding *et al.*, 2017 PPR-SMR protein SOT1 has RNA

892   endonuclease activity. Proc. Natl. Acad. Sci. U. S. A. 114: E1554–E1563.

893