bioRxiv preprint doi: https://doi.org/10.1101/2020.04.11.037556; this version posted April 13, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

1 # A general model to explain repeated turnovers of sex determination
2 # in the Salicaceae

3

4 Wenlu Yang[1#], Zhiyang Zhang[1#], Deyan Wang[1#], Yiling Li[1], Shaofei Tong[1],
5 Mengmeng Li[1], Xu Zhang[2], Lei Zhang[1], Liwen Ren[1], Xinzhi Ma[1], Ran Zhou[3], Brian J.
6 Sanderson[3,4], Ken Keefover-Ring[5], Tongming Yin[6], Lawrence B. Smart[7], Jianquan
7 Liu[1,2]*, Stephen P. DiFazio[3]*, Matthew Olson[4]*, Tao Ma[1]*

8

9 [1]Key Laboratory of Bio-Resource and Eco-Environment of Ministry of Education,
10 College of Life Sciences, Sichuan University, Chengdu 610065, China
11 [2]State Key Laboratory of Grassland Agro-Ecosystem, Institute of Innovation Ecology
12 & College of Life Sciences, Lanzhou University, Lanzhou 730000, China
13 [3]Department of Biology, West Virginia University, Morgantown, WV 25606, USA
14 [4]Department of Biological Sciences, Texas Tech University, Lubbock, TX
15 79409-3131 USA
16 [5]Departments of Botany and Geography, University of Wisconsin-Madison, 430
17 Lincoln Dr., Madison, WI 53706, USA
18 [6]The Key Laboratory of Tree Genetics and Biotechnology of Jiangsu Province and
19 Education Department of China, Nanjing Forestry University, Nanjing, China,
20 200137
21 [7]Horticulture Section, School of Integrative Plant Science, Cornell University, New
22 York State Agricultural Experiment Station, Geneva, New York 14456 USA

23

24 [#]These authors contributed equally to this work.
25 *Authors for correspondence: liujq@lzu.edu.cn; Stephen.DiFazio@mail.wvu.edu;
26 Matt.Olson@ttu.edu and matao.yz@gmail.com

27
28
29

bioRxiv preprint doi: https://doi.org/10.1101/2020.04.11.037556; this version posted April 13, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

30 **Abstract**

31 Dioecy, the presence of separate sexes on distinct individuals, has evolved repeatedly
32 in multiple plant lineages. However, the specific mechanisms through which sex
33 systems evolve and their commonalities among plant species remain poorly
34 understood. With both XY and ZW sex systems, the family Salicaceae provides a
35 system to uncover the evolutionary forces driving sex chromosome turnovers. In this
36 study, we performed a genome-wide association study to characterize sex
37 determination in two *Populus* species, *P. euphratica* and *P. alba*. Our results reveal an
38 XY system of sex determination on chromosome 14 of *P. euphratica*, and a ZW
39 system on chromosome 19 of *P. alba*. We further assembled the corresponding sex
40 determination regions, and found that their sex chromosome turnovers may be driven
41 by the repeated translocations of a *Helitron*-like transposon. During the translocation,
42 this factor may have captured partial or intact sequences that are orthologous to a
43 type-A cytokinin response regulator gene. Based on results from this and other
44 recently published studies, we hypothesize that this gene may act as a master regulator
45 of sex determination for the entire family. We propose a general model to explain how
46 the XY and ZW sex systems in this family can be determined by the same *RR* gene.
47 Our study provides new insights into the diversification of incipient sex chromosome
48 in flowering plants by showing how transposition and rearrangement of a single gene
49 can control sex in both XY and ZW systems.

50

51 **Keywords:** Dioecy, Sex determination, Sex chromosome turnover, Genome, *Populus*

52

bioRxiv preprint doi: https://doi.org/10.1101/2020.04.11.037556; this version posted April 13, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

53 **Introduction**

54 The origin and evolution of dioecy (separate sexes) has long been one of the most
55 fascinating topics for biologists (Henry et al., 2018; Feng *et al.*, 2020). The presence
56 of dioecy ensures outcrossing and optimal allocation of reproductive resources for
57 male and female sexual function, thereby providing them with certain advantages in
58 fertility, survival and evolution (Bawa, 1980). In flowering plants, dioecy occurs in
59 only ~6% of all species and has independently evolved thousands of times from
60 hermaphroditic ancestors (Renner and Ricklefs, 1995; Renner, 2014). Many of these
61 species have sex determined by a pair of heteromorphic sex chromosomes that differ
62 in morphology and/or sequence, in the form of male heterogamety (XY system) or
63 female heterogamety (ZW system) (Ming et al., 2011; Charlesworth, 2016). Theory
64 predicts that sex chromosomes evolve from ancestral autosomes via successive
65 mutations in two linked genes with complementary dominance (Charlesworth and
66 Charlesworth, 1978; Charlesworth, 1991). Subsequently, the suppression of
67 recombination between these two sex determination genes progressively spreads
68 along Y or W chromosomes, and permits the accumulation of repetitive elements and
69 duplication or translocation of genomic fragments, which in turn leads to the
70 formation of a sex-specific region and finally degeneration of the sex chromosome
71 (Bergero and Charlesworth, 2009; Charlesworth, 2012; Bachtrog, 2013).
72 Characterizing the genomic architecture of sex in dioecious species is critical for
73 understanding the origin of sex chromosomes, especially in their early stage of
74 evolution.

75 Over the past decade, impressive progress has been made in unraveling the
76 genetic basis of sex determination in several dioecious plants and the evolutionary
77 history of their sex chromosomes, including papaya (Wang *et al.*, 2012), persimmon
78 (Akagi *et al.*, 2014), asparagus (Harkess *et al.*, 2017), strawberry (Tennessen *et al.*,
79 2018), date palm (Torres *et al.*, 2018) and kiwifruit (Akagi *et al.*, 2018, 2019).
80 Consistent with the independent origins of sex chromosomes, the sex determination
81 genes identified in these species differ from each other, although most of them
82 function in similar hormone response pathways (Feng *et al.*, 2020). In addition, a

bioRxiv preprint doi: https://doi.org/10.1101/2020.04.11.037556; this version posted April 13, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

83 recent study found that the sex chromosome turnover in strawberries is driven by
84 repeated translocation of a female-specific sequence (Tennessen *et al.*, 2018). The
85 combined evidence from these studies demonstrates the high variation of plant sex
86 determination mechanisms, and so understanding the factors that drive the convergent
87 evolution of sex chromosomes in plants remains elusive (Zhang *et al.*, 2014).

88 The family Salicaceae provides an excellent system to study the drivers of sex
89 chromosome evolution. This family includes two sister genera, *Populus* and *Salix*,
90 which are composed exclusively of dioecious species (Peto, 1938; Zhang *et al.*, 2018;
91 Li *et al.*, 2019). Previous studies in multiple *Salix* species have consistently mapped
92 the sex determination regions (SDRs) to chromosome 15, and proposed a ZW system
93 in which females are the heterogametic sex (Pucholt *et al.*, 2015, 2017; Hou *et al.*,
94 2015; Chen *et al.*, 2016; Zhou e*t al.*, 2018, 2020). However, an XY system was
95 recently identified on chromosome 7 in *S. nigra* (Sanderson *et al.*, 2020). In
96 comparison, the SDR has been mapped to multiple locations in different *Populus*
97 species, indicating a dynamic evolutionary history of the sex chromosomes. The SDR
98 has been mapped to the proximal telomeric end of chromosome 19 in *P. trichocarpa*
99 and *P. nigra* (sections *Tacamahaca* and *Aigeiros*) (Gaudet *et al.*, 2007; Yin *et al.*, 2008;
100 Geraldes *et al.*, 2015), and to a pericentromeric region of chromosome 19 in *P.*
101 *tremula*, *P. tremuloides* and *P. alba* (section *Populus*) (Pakull *et al.*, 2009, 2014;
102 Paolucci *et al.*, 2010; Kersten *et al.*, 2014). Most *Populus* species display an XY sex
103 determination system, but there is some evidence that *P. alba* has a ZW system
104 (Paolucci *et al.*, 2010). Thus far, the only SDR that has been assembled in *Populus* is
105 that of *P. trichocarpa* and *P. deltoides*, and it appears to be much smaller than those
106 observed in *Salix* (Geraldes *et al.*, 2015; Xue *et al.*, 2020). Our recent study on the W
107 chromosome of *S. purpurea* showed intriguing palindromic structures, in which four
108 copies of the gene encoding a type A cytokinin response regulator (*RR*) were
109 identified (Zhou *et al.*, 2020). Interestingly, the ortholog of this gene has also been
110 reported to be associated with sex in *Populus* from section *Tacamahaca* (Geraldes *et*
111 *al.*, 2015; Bräutigam *et al.*, 2017; Melnikova *et al.*, 2019), which increases the
112 possibility that this gene is an excellent candidate for a common sex determination

bioRxiv preprint doi: https://doi.org/10.1101/2020.04.11.037556; this version posted April 13, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

113 mechanism in the Salicaceae. However, it is still unclear whether this candidate gene
114 is present in all of these SDRs. Most importantly, how the same gene functions in
115 both the XY and ZW systems remains elusive. Here, we identify the sex
116 determination systems of two additional *Populus* species, *P. euphratica* and *P. alba*,
117 which are from sects. *Turanga* and *Populus* respectively (Wang *et al*., 2020). We
118 report their complete SDR assemblies and propose a general model to illustrate the
119 potentially shared mechanism of sex determination in this family.
120
121 **Results**
122 **Genome assembly**
123 We have previously reported the assembly of the genomes of a male *P. euphratica*
124 (Zhang *et al*., 2020) and a male *P. alba* var. *pyramidalis* (a variety of *P. alba*) (Ma *et*
125 *al*., 2019). Here we further sequenced and *de novo* assembled female genomes for
126 both species using Oxford Nanopore reads. The assembly for the female *P. euphratica*
127 consists of 1,229 contigs with an N50 of 1.7 Mb and a total size of ~529.0 Mb, while
128 the female *P. alba* var. *pyramidalis* assembly has 357 contigs with an N50 of 3.08 Mb,
129 covering a total of ~358.5 Mb (**Table S1**). Both assemblies showed extensive synteny
130 with their respective male reference genomes, and therefore, based on their syntenic
131 relationships, the assembled contigs were anchored onto 19 pseudochromosomes
132 (**Figs. S1 and S2**). The chromosome identities were then assigned by comparison to *P.*
133 *trichocarpa* (Tuskan *et al*., 2006).
134
135 **XY sex determination on chromosome 14 in *P. euphratica***
136 In order to characterize the sex determination system of *P. euphratica*, we
137 resequenced the genomes of 30 male and 30 female individuals (**Table S2**) and
138 performed a genome-wide association study (GWAS). Using the male assembly as the
139 reference genome, a total of 24,651,023 high-quality single nucleotide polymorphisms
140 (SNPs) were identified. After Bonferroni correction, we recovered 310 SNPs
141 significantly associated with sex ($\alpha<0.05$; **Figs. 1A, S3A and Table S3**). In-depth
142 analysis found that almost all genotypes (99.99%) of these sex-associated loci are

bioRxiv preprint doi: https://doi.org/10.1101/2020.04.11.037556; this version posted April 13, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

143 homozygous in females, while 93.57% of the genotypes are heterozygous in males
144 (**Fig. 1B**). A similar pattern was observed when the sex association analysis was
145 performed by using the female assembly as the reference genome (**Figs. S3B and S4,**
146 **and Tables S4 and S5**). These results consistently indicate that an XY system is
147 involved in sex determination of *P. euphratica*.

148 In addition, we found that the vast majority of the significantly sex-associated
149 SNPs were located at the proximal end of chromosome 14 (the un-anchored scaffold
150 '001598F' in male genome was located onto chromosome 14 based on its syntenic
151 relationship with *P. trichocarpa* genome), while a few other SNPs were present at
152 chromosomes 7, 9, 12 and 19 (**Figs. 1A, 1B and S4, and Table S5**). We then
153 attempted to use ultra-long nanopore reads generated from a male individual (**Table**
154 **S6**) to further reconstruct a new assembly with X and Y haplotypes as separate contigs.
155 This led to the identification of a contig that was highly similar to the sex-associated
156 regions and specifically contained Y-linked alleles (**Fig. S5**). The Y-linked region was
157 further determined by examining the relative depth of coverage when aligning male
158 versus female resequencing reads against the reference (**Fig. S6**). Based on the
159 syntenic relationship, the SDR of *P. euphratica* can be mapped to the proximal end of
160 chromosome 14 and the Y-linked region is about 658 kb in length, corresponding to
161 ~84 kb on the X chromosome (**Fig. 1C**). We found that two segments spanning 440
162 kb and 135 kb respectively, are specific to the Y-linked region (**Fig. 1C**), suggesting
163 the occurrence of significant chromosome divergence between the X and Y
164 haplotypes, which can be maintained by suppressed recombination.

165 We predicted a total of 37 protein-coding genes in the Y-linked region, many of
166 which have high similarity with genes on other autosomes and are considered as
167 translocated genes (**Table S7**). Among these, we found that 9 of the Y-specific genes
168 were annotated as members of the LONELY GUY (LOG) family, which encodes
169 cytokinin-activating enzymes that play a dominant role in the maintenance of the
170 shoot apical meristem and in the establishment of determinate floral meristems
171 (Kuroha *et al.*, 2009; Tokunaga *et al.*, 2012; Han and Jiao, 2015). Ten genes were
172 identified in both X and Y haplotypes. A phylogenetic analysis of these genes showed

bioRxiv preprint doi: https://doi.org/10.1101/2020.04.11.037556; this version posted April 13, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

173 that the X and Y alleles began to diverge after their split with *P. trichocarpa* and *P.*
174 *alba* (**Figs. 1D and S7**), suggesting that the SDR of *P. euphratica* appears to be
175 established relatively recently.
176

177 **ZW sex determination on chromosome 19 in *P. alba***

178 We used a similar GWAS strategy for 30 male and 30 female resequenced individuals
179 to characterize the sex determination system of *P. alba* (**Table S8**). When the male
180 and female assembly was used as a reference genome, respectively, 173 and 55 SNPs
181 that were significantly associated with sex were identified (**Figs. 2A, 2B, S8 and S9,**
182 **and Tables S9-S11**). Most of the sex-associated SNPs are heterozygous in females
183 and homozygous in males (**Fig. 2B and Table S10**), confirming the ZW sex
184 determination system in *P. alba*, which was also suggested based on genetic mapping
185 in a previous study (Paolucci *et al*., 2010).

186 We found that these sex-associated SNPs are mainly located on a non-terminal
187 region of chromosome 19 (**Figs. 2A, 2B and S8, and Table S10**). Next, we examined
188 the female-specific depth profile, combined with the support of ultra-long nanopore
189 reads (**Table S6**), to delineate the W haplotype of *P. alba* to a region of about 140 kb
190 on chromosome 19, with a corresponding Z haplotype that is only 33 kb in length
191 (**Figs. 2C, S10 and S11**). Compared to the Z haplotype and corresponding autosomal
192 regions of the other Salicaceae species, a specific insertion of 69 kb was observed in
193 the W haplotype, indicating a recent origin of the SDR in *P. alba*.

194 Sequence annotation predicted 18 protein-coding genes in the W haplotype, six of
195 which were also found in the Z haplotype (**Table S12**). The high identity of these
196 alleles between the W and Z haplotype suggests that recombination suppression
197 occurred very recently (**Fig. 2D**). We further found that the gene encoding
198 NAC-domain protein, *SOMBRERO* (*SMB*), which has a similar function to the
199 *VND*/*NST* transcription factors that regulate secondary cell wall thickening in woody
200 tissues and maturing anthers of *Arabidopsis* (Mitsuda *et al*., 2005; Bennett *et al*.,
201 2010), was expanded from one member in the Z haplotype to three copies in the W
202 haplotype ('HP2' in **Fig. 2D**). There are 12 genes specific to the W haplotype (**Table**

bioRxiv preprint doi: https://doi.org/10.1101/2020.04.11.037556; this version posted April 13, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

203 **S12**), including *DM2H* (*DANGEROUS MIX2H*), which encodes a nucleotide-binding
204 domain and leucine-rich repeat immune receptor protein (Chae *et al.*, 2014); *CCR2*
205 (Cinnamoyl CoA reductase), which is involved in lignin biosynthesis and plant
206 development (Thevenin *et al.*, 2011); and *STRS1* (*STRESS RESPONSE*
207 *SUPPRESSOR1*), a gene encoding a DEAD-box RNA helicase, which is involved in
208 epigenetic gene silencing related to stress responses (Khan *et al.*, 2014). More
209 interesting, we also identified three copies of the gene encoding a type A cytokinin
210 response regulator (*RR*) in the W-specific region (**Fig. 3A**), the ortholog of which has
211 also been identified to be associated with sex determination in poplar and willow
212 (Geraldes *et al.*, 2015; Bräutigam *et al.*, 2017; Melnikova *et al.*, 2019; Zhou *et al.*,
213 2020). Very little sequence differences were found among these three copies, and
214 combined with the fact that the ortholog of the *RR* gene is located at the distal end of
215 chromosome 19 in *P. trichocarpa* and *P. euphratica* (**Fig. 3**), we conclude that the *RR*
216 gene was translocated from the end of chromosome 19 to the W haplotype of *P. alba*
217 and then underwent at least two rounds of recent duplication.
218

219 **Evidence for SDR turnover in Salicaceae**

220 We have shown that *P. euphratica* and *P. alba* have different sex determination
221 systems, and that the SDRs are different from those reported in *P. trichocarpa* and *S.*
222 *purpurea*, indicating extraordinarily high diversity of sex determination in the
223 Salicaceae. In order to examine whether the sex determination regions originated
224 independently in each lineage, or evolved into the current SDRs separately after a
225 common ancient origin, we performed syntenic analysis on these SDRs in *P.*
226 *euphratica* and *P. alba*, and the corresponding autosomal regions in *P. trichocarpa*
227 and *S. purpurea*. We found that although the pseudo-autosomal regions of these sex
228 chromosomes are highly collinear with their corresponding autosomal regions in other
229 species, the sequences in the sex-specific regions are not alignable (**Figs. 1C and 2C**).
230 In contrast, although there was little collinearity among these SDRs, a homologous
231 sequence with multiple duplicates was identified between the Y haplotype of *P.*
232 *euphratica* and the W haplotype of *P. alba* **(Fig. 3A).** Interestingly, the locations of

bioRxiv preprint doi: https://doi.org/10.1101/2020.04.11.037556; this version posted April 13, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

233 the duplicates overlapped with the three predicted *RR* genes in *P. alba*. In the
234 corresponding regions of the Y haplotype of *P. euphratica*, we identified 10 partial
235 duplicates of the RR gene including four covering the first three exons (large
236 duplicate) and six covering only the first exon (small duplicate) of the *RR* gene (**Fig.**
237 **3**). Phylogenetic analysis of these duplicates showed that the three *RR* genes in *P.*
238 *alba* clustered together and are closely related to the intact orthologs of *P. euphratica*
239 and *P. trichocarpa*, while the partial duplicates from *P. euphratica* divided into two
240 main clades, one with only large duplicates and a second clade with only small
241 duplicates (**Fig. 3B**).

242 Since the *RR* duplicates were found in the SDRs of all of the current and
243 previously studied species, we believe that they may play important roles in sex
244 determination of the Salicaceae species. These results also lead to the hypothesis that
245 these species shared an ancient origin of sex chromosomes, followed by frequent
246 turnover events due to translocation of the *RR* duplicates. This is further supported by
247 the distant relationship between the partial and intact *RR* duplicates **(Fig. 3B)**, which
248 indicate that the partial duplicates originated before the divergence of these poplar
249 species and were repeatedly inserted into the SDRs of *P. euphratica*. We did not
250 detect any structurally intact long terminal repeat retrotransposons (LTR-RTs) around
251 these *RR* duplicates, which made it impossible to estimate their insertion time.
252 However, around the *RR* duplicates in *P. euphratica*, we identified a *Helitron*-like
253 transposable element upstream of each small duplicate except the second one
254 ('PeuY:S2'), and a *Copia*-like LTR fragment in the downstream region of each large
255 duplicate **(Fig. 3B)**. These two repetitive elements were also identified in all three *RR*
256 duplicates of *P. alba*, and are located upstream and in the third intron of the *RR* gene,
257 respectively, similar to that in *P. euphratica*. The phylogenetic trees of the two
258 elements and the *RR* duplicates exhibited a similar topological relationship,
259 suggesting that they may be transposed together as a unit (**Figs. 3C and 3D**). The
260 extremely high similarity of these sequences indicates that they were recently
261 transposed into the SDRs of *P. euphratica* and *P. alba*, respectively, consistent with
262 the observation that their sex chromosomes have not been severely degenerated. In

bioRxiv preprint doi: https://doi.org/10.1101/2020.04.11.037556; this version posted April 13, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

263 addition, we found that the *Helitron*-like element was not present in the upstream
264 region of the intact *RR* genes at chromosome 19 of *P. euphratica* and *P. trichocarpa*
265 (**Fig. 3B**), which led us to speculate that this element may be the main driving force
266 for gene replication during the evolution of SDRs in *P. euphratica* and *P. alba*.
267 However, we failed to detect the same pattern in *S. purpurea*, in which multiple *Copia*
268 LTR-RTs were predicted instead of the *Helitron* elements (Zhou *et al.*, 2020). This
269 implies that poplar and willow may have different SDR turnover mechanisms, which
270 requires further evidence from more species to confirm.
271

272 **Discussion**

273 It is notoriously difficult to assemble the complete sequence of SDRs or sex
274 chromosomes, which usually have a high repeat density and many translocated
275 segments from autosomes (Charlesworth, 2012; Bachtrog, 2013). In our study, the
276 sex-associated loci were initially mapped onto multiple different chromosomes **(Figs.
277 1 and 2)**, although they consistently revealed an XY sex determination system in *P.
278 euphratica* and a ZW system in *P. alba*. These results may be caused by the lack
279 and/or mis-assembly of SDRs in the reference genome, especially when the genome
280 from a homozygous (XX or ZZ) individual was used as reference, the reads from Y-
281 or W-specific regions of hemizygous (XY or ZW) individuals may be misaligned to
282 homologous sequences on autosomes and led to false associations. Similar
283 phenomena were also observed in the sex association analysis of *P. trichocarpa*, *P.
284 balsamifera* and *S. purpurea*, which may lead to an inaccurate localization of SDRs in
285 assemblies (Geraldes *et al*., 2015; Zhou *et al*., 2020). The high sequence similarity
286 between these sex-associated regions and the SDRs we finally established strongly
287 supports this possibility **(Figs. S5 and S10**). Therefore, our research emphasizes the
288 importance and necessity for precise assembly of SDRs using multiple
289 complementary methods, including the ultra-long read sequencing, haplotype phased
290 assembly and the sex-specific depth of read mapping.

291 Our results further indicate that the SDRs of poplar species are generally shorter
292 in length and contain relatively fewer genes than that recently reported in *S. purpurea*

bioRxiv preprint doi: https://doi.org/10.1101/2020.04.11.037556; this version posted April 13, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

293 (Zhou *et al*., 2020), though the size of this SDR may be inflated due to overlap with
294 the centromere (Zhou *et al.*, 2018). Although some specific insertions were observed
295 on the Y and W chromosomes, we found no obvious degeneration of sex
296 chromosomes at least in *P. euphratica* and *P. alba*. These results suggest that the
297 SDRs of these two species were established relatively recently, which is a common
298 feature of the sex chromosomes of the Salicaceae species studied so far (Geraldes *et*
299 *al*., 2015; Pucholt *et al.*, 2017; Zhou *et al*., 2018, 2020). Along with this, our results
300 also suggest that the Y and W chromosomes have expanded in content, a pattern that
301 is common in young sex chromosomes of plants (Hobza *et al*., 2015, 2017). Moreover,
302 our results simultaneously showed that the Salicaceae exhibit an extremely fast rate of
303 sex-chromosome turnover. In previous studies, SDRs have been reported only on
304 chromosome 15 with female heterogamety (ZW) in willow except *S. nigra* (Pucholt *et*
305 *al.*, 2015, 2017; Hou *et al.*, 2015; Chen *et al.*, 2016; Zhou e*t al.*, 2018, 2020;
306 Sanderson *et al.*, 2020), and on chromosome 19 of poplar with most species showing
307 male heterogamety (XY) (Gaudet *et al.*, 2007; Yin *et al.*, 2008; Pakull *et al.*, 2014;
308 Geraldes *et al.*, 2015). However, our study identified an XY system with the SDR on
309 chromosome 14 of *P. euphratica* for the first time, and confirmed a ZW system with
310 SDR on chromosome 19 of *P. alba*. These results highlight the complexity and
311 diversity of sex determination in this family. Comparative analysis showed that
312 translocation of genes from autosomes to the SDR and gene replication frequently
313 occurred both on the Y chromosomes of *P. euphratica* and on the W chromosomes of
314 *P. alba*, indicating that these two events are likely to be important contributors during
315 SDR turnover. The regulatory mechanisms and functions of these genes in sex
316 determination and sexual dimorphism in these two species need further investigation.

317 Among all genes on SDRs, the cytokinin response regulator is the most likely
318 candidate for controlling sex determination in the Salicaceae, not only because the
319 orthologs of this gene have been found to be sex-associated in most of the reported
320 species in the family, but also because it is the only homologous sequence found in
321 the sex chromosomes of *P. euphratica*, *P. alba*, *P. trichocarpa*, *P. deltoides* and *S.*
322 *purpurea* (**Fig. 3**), the only Salicaceae species with SDR precisely assembled (Zhou *et*

bioRxiv preprint doi: https://doi.org/10.1101/2020.04.11.037556; this version posted April 13, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

323 *al*., 2020; Xue *et al.*, 2020). Recent progress has revealed that the genes involved in
324 cytokinin signaling play important roles in the regulation of unisexual flower
325 development in plants (Wybouw *et al*., 2019; Kieber *et al*., 2018; Feng *et al*., 2020).
326 Specifically, a Y-specific type-C cytokinin response regulator (*Shy Girl*, *SyGI*) was
327 recently identified as a suppressor of carpel development and therefore is a strong
328 candidate of sex determination in kiwifruit (Akagi *et al*., 2018). Similar to the pattern
329 of the *RR* genes found in the Salicaceae species, in kiwifruit *SyGI* was duplicated
330 from an autosome and subsequently gained a new function on its Y chromosome.
331 However, the type-A *RR* genes we identified here are not orthologous to the *SyGI*
332 gene, so we speculate that they may have different functions in the cytokinin signaling
333 pathway. Based on our results, it is reasonable to suspect that the *RR* genes are more
334 likely to function as a dominate promoter of female function **(Fig. 4)**, as they exist on
335 the W chromosomes of both *P. alba* and *S. purpurea* in intact duplicates. In contrast,
336 the RR gene fragments on the Y chromosome of *P. euphratica* exist as two partial
337 duplicates with different sizes. This may serve as a female suppressor by encoding an
338 siRNA that targets the intact *RR* gene at the distal end of chromosome 19, possibly
339 through RNA-directed DNA methylation (Brautigam *et al.*, 2017; Xue *et al.*, 2020). It
340 should be noted that, although the intact *RR* gene has been reported to be associated
341 with sex in *P. trichocarpa*, there is still no evidence to support the gene's localization
342 on its Y chromosome. In the previous GWAS study (Geraldes *et al.*, 2015), most of
343 the sex-associated loci of *P. trichocarpa* were located on the proximal end of
344 chromosome 19. The associated signals scattered around the intact *RR* gene, which is
345 located at the distal end of chromosome 19, were most likely due to assembly errors
346 arising from the fact that this reference genome is derived from a female (XX)
347 individual (the major factor in misleading SDR localization as mentioned above).
348 Therefore, our findings consistently showed that Salicaceae species potentially share a
349 common mechanism of sex determination, in which the specific duplication of the *RR*
350 orthologs on SDRs may have played an important role in the acquisition of separate
351 sexes in these species.

352 More interestingly, we identified *Helitron*-like repetitive elements upstream of the

bioRxiv preprint doi: https://doi.org/10.1101/2020.04.11.037556; this version posted April 13, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

353 *RR* duplicate in both SDRs of *P. euphratica* and *P. alba*, regardless of whether the *RR*
354 duplicate is intact or partial (**Fig. 3**). As a major class of DNA transposons, *Helitrons*
355 were hypothesized to transpose by a rolling circle replication mechanism, and have
356 been found to frequently capture genes or gene fragments and move them around the
357 genome, which is believed to be important in the evolution of host genomes
358 (Morgante *et al.*, 2005; Kapitonov and Jurka, 2007). Our results suggest that the *RR*
359 fragments and intact gene sequences appear to have been captured by *Helitrons* in *P.*
360 *euphratica* and *P. alba*, and subsequently replicated in their SDRs **(Figs. 3 and 4)**.
361 Furthermore, our phylogenetic analysis indicated that the intact *RR* gene was captured
362 very recently in *P. alba*, at least after its split with *P. trichocarpa* (**Fig. 3B**). In contrast,
363 although we found high similarity among the *RR* partial duplicates of *P. euphratica*,
364 these sequences are quite different from the intact *RR* genes of other poplar species
365 **(Fig. 3B)**. These results indicate that the partial duplicates were present before the
366 diversification of poplar species, but only recently replicated on the Y chromosome of
367 *P. euphratica*. We found that the partial duplicate of the *RR* gene is lacking in *P. alba*,
368 which may be another key event in addition to the duplication of the intact *RR* gene,
369 in the transition of the sex determination system from XY to ZW **(Fig. 4)**. In addition,
370 the high nucleotide identity among intact *RR* genes of *S. purpurea* reflects another
371 possible SDR turnover event in willow, which might be driven by the replication of a
372 *Copia* LTR (Zhou *et al.*, 2020), rather than by a *Helitron* as we found in poplar.
373 Moreover, we also identified an inverted repeat of the first exon of the *RR* gene and an
374 intact copy on the chromosomes 15Z and 19 of *S. purpurea*, respectively **(Fig. 3)**.
375 This suggests a model whereby the inverted repeat is suppressing the *RR* gene of
376 chromosome 19 in males, but the SDR on the W chromosome may be dominant to
377 this effect in females, possibly due to increased dosage or another mechanism **(Fig. 4)**.
378 These observations further indicate that the sex determination system of *S. purpurea*
379 may have been changed from XY to ZW relatively recently, since the suppressing
380 mechanism from the *RR* partial duplication is still retained. This turnover was also
381 supported by the XY sex determination system of the basal *Salix* species, *S. nigra*
382 (Sanderson *et al.*, 2020). Therefore, our results suggest that the high activity of these

bioRxiv preprint doi: https://doi.org/10.1101/2020.04.11.037556; this version posted April 13, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

383 repetitive elements is the most likely cause of the recently established SDRs in these
384 species, and further indicate that at least three turnover events have occurred in the
385 evolution of sex chromosomes of the Salicaceae species **(Fig. 4).**

386 In conclusion, here we present an XY system of sex determination with the SDR
387 on the proximal end of chromosome 14 in *P. euphratica*, and a ZW system with the
388 SDR on a non-terminal region of chromosome 19 in *P. alba*. Both SDRs appear to
389 have evolved relatively recently and are characterized by frequent translocations from
390 autosomes and gene replication events. Our comparative analysis also demonstrated
391 an extremely fast rate of sex chromosome turnover among Salicaceae species, which
392 may be driven by *Helitron* transposons in poplar and by *Copia* LTRs in willow. Most
393 importantly, we propose a model showing that poplar and willow have a common
394 underlying mechanism of sex determination, which controls the XY and ZW systems
395 simultaneously through a type-A *RR* gene. In the future, it will be necessary to
396 conduct transgenic function experiments and comparative analysis from more species
397 in this family to further support our model.

398

399 **Methods**

400 **Genome sequencing**

401 We have previously reported the reference genome of a male *P. euphratica* (Zhang *et*
402 *al*., 2020) and a male *P. alba* (Ma *et al*., 2019). In this study, we further collected the
403 fresh leaves of a female *P. euphratica* and a female *P. alba* for genome sequencing
404 and assembly. Genomic DNA was extracted using the QIAGEN Genomic DNA
405 extraction kit (Qiagen, Hilden, Germany) following the manufacturer's protocol. To
406 generate Oxford Nanopore long reads, approximately 15 μg of genomic DNA was
407 size-selected using the BluePippin system (Sage Science, USA), and processed
408 according to the protocol of Ligation Sequencing Kit (SQK-LSK109). The final
409 library was sequenced on a PromethION sequencer (Oxford Nanopore Technologies,
410 UK) with a running time of 48 hours. The Oxford Nanopore proprietary base-caller,
411 Albacore v2.1.3, was used to perform base calling of the raw signal data and convert
412 the FAST5 files into FASTQ files.

bioRxiv preprint doi: https://doi.org/10.1101/2020.04.11.037556; this version posted April 13, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

413 In addition, paired-end libraries with insert size of ~300 bp were also constructed
414 using NEB Next® Ultra DNA Library Prep Kit (NEB, USA), with the standard
415 protocol provided by Illumina (San Diego, CA, USA). The library was sequenced on
416 an Illumina HiSeq X Ten platform (Illumina, San Diego, CA, USA). These
417 sequencing data were used for correction of errors inherent to long read data for
418 genome assembly.
419

420 **Genome Assembly**

421 For genome assembly, we first removed the Nanopore long reads shorter than 1 kb
422 and the low-quality reads with a mean quality ≤ 7. The long reads underwent
423 self-correction using the module 'NextCorrect' and then assembled into contigs using
424 'NextGraph' implemented in Nextdenovo v2.2.0
425 (https://github.com/Nextomics/NextDenovo) with default parameters. Subsequently,
426 the filtered Nanopore reads were mapped to the initial assembly using the program
427 Minimap2 v2.17-r941 (Li, 2018) and NextPolish v1.0
428 (https://github.com/Nextomics/NextPolish) was used with three iterations to polish
429 the genome. In addition, we further aligned the Illumina reads to the genome using
430 BWA-MEM v0.7.15 (Li and Durbin 2009) and corrected base-calling by an additional
431 three rounds of NextPolish runs with default parameters. Finally, the corrected
432 genome was aligned to their respective male reference genome using the LAST
433 program (Kielbasa *et al*., 2011) and the syntenic relationships were used to anchor the
434 assembled contigs onto 19 chromosomes.
435

436 **Population sample collection, resequencing and mapping**

437 Silica gel dried leaves of *P. euphratica* and *P. alba* were collected from wild
438 populations in western China. For each species, the sex of 30 male and 30 female
439 individuals was identified from flowering catkins. Genomic DNA of each sample was
440 extracted using the Qiagen DNeasy Plant Minikit (Qiagen, Hilden, Germany).
441 Paired-end libraries were prepared using the NEBNext Ultra DNA Library Prep Kit
442 (NEB, USA) and sequenced on an Illumina HiSeq X Ten platform, according to the

bioRxiv preprint doi: https://doi.org/10.1101/2020.04.11.037556; this version posted April 13, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

manufacturer's instructions.

The generated raw reads were first subjected to quality control and low-quality reads were removed if they met either of the following criteria (Ma *et al*., 2018): i) ≥10% unidentified nucleotides (N); ii) a phred quality ≤ 7 for > 65% of read length; iii) reads overlapping more than 10 bp with the adapter sequence, allowing < 2 bp mismatch. Reads shorter than 45 bp after trimming were also discarded. The obtained high-quality cleaned reads were subsequently mapped to the male and female reference genomes of each species, respectively, using BWA-MEM v0.7.15 with default parameters (Li and Durbin 2009). Then the alignment results and marked duplicate reads were sorted using SAMtools v0.1.19 (Li *et al*., 2009). Finally, Genome Analysis Toolkit (GATK) (DePristo *et al*., 2011) was performed to process base quality recalibrations to enhance alignments in regions around putative indels with two steps: i) 'RealignerTargetCreator' was applied to identify regions where realignment was needed; ii) 'IndelRealigner' was used to realign these regions.

**SNP calling, filtering and genome-wide association study (GWAS)**

To prevent biases in SNP calling accuracy due to the difference of samples size between groups, single-sample SNP and genotype calling were first implemented using GATK (DePristo *et al*., 2011) with 'HaplotypeCaller', and then multi-sample SNPs were identified after merging the results of each individual by 'GenotypeGVCFs'. A series of filtering steps were performed to reduce false positives (Yang *et al*., 2018), including removal of (1) indels with a quality scores < 30, (2) SNPs with more than two alleles, (3) SNPs at or within 5 bp from any indels, (4) SNPs with a genotyping quality scores (GQ) < 10, and (5) SNPs with extremely low (< one-third average depth) or extremely high (> threefold average depth) coverage. The identified SNPs were used for subsequent GWAS analysis. A standard case/control model between allele frequencies and sex phenotype was performed using Plink v1.9 (Purcell *et al*., 2007). For each species, associations at $\alpha < 0.05$ after Bonferroni correction for multiple testing were reported as the significantly

bioRxiv preprint doi: https://doi.org/10.1101/2020.04.11.037556; this version posted April 13, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

472 sex-associated SNPs. These sex-associated SNPs that occurred within 10 kb on the
473 same chromosome were merged into the same interval.
474
475 **Construction of *P. euphratica* Y contig and *P. alba* W contig**
476 To construct the Y contig of *P. euphratica* and the W contig of *P. alba*, we further
477 generated ultra-long sequences from a male (XY) *P. euphratica* and a female (ZW) *P.*
478 *alba*, using an optimized DNA extraction followed by modified library preparation
479 based on the Nanopore PromethION sequencer (Jain *et al*., 2018; Gong *et al*., 2019).
480 For *P. euphratica*, we did not find contigs that clearly contained Y-linked sequences
481 in its male genome, which may be due to assembly errors, so we used multiple
482 methods to determine its Y contig. At first, we attempted to find the male-specific
483 k-mers from the high-quality resequencing reads of both male and female samples.
484 Briefly, all 32 bp k-mers starting with the 'AG' dinucleotide were extracted from all
485 resequencing reads, and the number of occurrences of each specific subsequence in
486 female and male individuals was counted, respectively. The use of the 'AG'
487 dinucleotide is to reduce the number of k-mer sequences and effectively speed up the
488 analysis. The k-mer counts were then compared between male and female, and the
489 male-specific k-mers (female count was 0) were obtained. Next, we extracted the
490 ultra-long nanopore reads containing at least one of the identified male-specific
491 k-mers, and assembled these ultra-long reads using the software Canu v1.7 (Koren *et*
492 *al*., 2017), resulting in a 'male-specific contig' that was 450 kb in length.
493 Simultaneously, we also *de novo* assembled all of the ultra-long nanopore reads into a
494 draft male genome using Nextdenovo v2.2.0. By comparing the 'male-specific contig'
495 with the obtained male genome, we identified a candidate Y contig that contained a
496 large number of male-specific alleles and exhibited a widespread synteny and
497 continuity with the 'male-specific contig'. To further refine the sex determination
498 region along this candidate Y contig, we re-mapped the resequencing data to the draft
499 genome by BWA-MEM v0.7.15 (Li and Durbin, 2009), and extracted the average
500 depth of coverage using a non-overlapping sliding window (1 kb in length) by
501 SAMtools v0.1.19 (Li *et al*., 2009). Finally, we compared the relative depth of

bioRxiv preprint doi: https://doi.org/10.1101/2020.04.11.037556; this version posted April 13, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

502 coverage between male and female individuals, and found that the region between 0
503 and 658 kb of this contig showed male-specific depth and was therefore considered to
504 be the sex determination region on the Y chromosome of *P. euphratica*.

505 For *P. alba*, we first performed a whole genome alignment between its male and
506 female genome using the program LAST (Kielbasa *et al*., 2011). Fortunately, we
507 found that the sex-associated region in the female genome contained a large insert
508 compared to the corresponding region in the male genome. We used the same method
509 as above to count the relative depth of coverage between male and female individuals
510 of *P. alba*, and found that the region between 310 and 450 kb of this contig exhibited
511 female-specific depth. Therefore, this region was directly considered to be the sex
512 determination region on the W chromosome of *P. alba*, and the assembly accuracy of
513 this region was also confirmed by our ultra-long nanopore reads.

514 **Annotation and comparison of the Y and W contigs**

515 Transposable elements in our assembled Y and W contigs were identified and
516 classified using the software RepeatMasker (Tarailo-Graovac and Chen, 2009). Gene
517 annotation was conducted by combining the results of *de novo* prediction from the
518 program Augustus v.3.2.1 (Stanke *et al*., 2006), homology-based prediction using the
519 protein sequences of *A. thaliana*, *P. trichocarpa* and *S. purpurea* downloaded from
520 Phytozome 12 (https://phytozome.jgi.doe.gov/), as well as transcriptome data of *P.*
521 *euphratica* and *P. alba* generated from our previously studies (Ma *et al*., 2019; Hu *et*
522 *al*., 2020; Zhang *et al*., 2020). The predicted genes were searched against predicted
523 proteins from *P. trichocarpa*, *S. suchowensis* and *A. thaliana* to find the closest
524 homologous annotation.

525 To construct the phylogenetic relationships among the allelic genes on the X/Y or
526 Z/W contigs, we further identified their orthologous genes in *P. pruinosa* (Yang *et al*.,
527 2017), *P. ilicifolia* (Chen *et al*., 2020) and *S. suchowensis* (Dai *et al*., 2014) genomes
528 by combining reciprocal blast results and their syntenic relationships. The sequences
529 were aligned using ClustalW with default parameters provided in MEGA5 (Tamura *et*
530 *al*., 2011) and the resulting alignments were adjusted manually. A maximum
531 likelihood tree was built using MEGA5 with default parameters.

bioRxiv preprint doi: https://doi.org/10.1101/2020.04.11.037556; this version posted April 13, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

532

533 **Accession numbers**

534 The whole genome sequence data reported in this paper have been deposited in the
535 Genome Warehouse in BIG Data Center (BIG Data Center Members, 2019), Beijing
536 Institute of Genomics (BIG), Chinese Academy of Sciences, under accession number
537 PRJCA002485 that is publicly accessible at https://bigd.big.ac.cn/bioproject.

538

539 **Acknowledgements**

540 This research was supported by National Natural Science Foundation of China
541 (31561123001, 31922061, 41871044, 31500502), NSF Dimensions of Biodiversity
542 Program (1542509 to S.D. and 1542599 to M.O.), National Key Research and
543 Development Program of China (2016YFD0600101), Fundamental Research Funds
544 for the Central Universities (SCU2019D013).

bioRxiv preprint doi: https://doi.org/10.1101/2020.04.11.037556; this version posted April 13, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

545 **References**


546 Akagi, T., Henry, I.M., Ohtani, H., Morimoto, T., Beppu, K., Kataoka, I., and Tao, R.
547 (2018). A Y-Encoded Suppressor of Feminization Arose via Lineage-Specific
548 Duplication of a Cytokinin Response Regulator in Kiwifruit. The Plant cell
549 30:780-795.
550 Akagi, T., Henry, I.M., Tao, R., and Comai, L. (2014). Plant genetics. A
551 Y-chromosome-encoded small RNA acts as a sex determinant in persimmons.
552 Science 346:646-650.
553 Akagi, T., Pilkington, S.M., Varkonyi-Gasic, E., Henry, I.M., Sugano, S.S., Sonoda,
554 M., Firl, A., McNeilage, M.A., Douglas, M.J., Wang, T., et al. (2019). Two
555 Y-chromosome-encoded genes determine sex in kiwifruit. Nature plants
556 5:801-809.
557 Bachtrog, D. (2013). Y-chromosome evolution: emerging insights into processes of
558 Y-chromosome degeneration. Nature reviews Genetics 14:113-124.
559 Bawa, K.S. (1980). Evolution of Dioecy in Flowering Plants. Annual Review of
560 Ecology and Systematics 11:15-39.
561 Bennett, T., van den Toorn, A., Sanchez-Perez, G.F., Campilho, A., Willemsen, V.,
562 Snel, B., and Scheres, B. (2010). SOMBRERO, BEARSKIN1, and BEARSKIN2
563 regulate root cap maturation in *Arabidopsis*. The Plant cell 22:640-654.
564 Bergero, R., and Charlesworth, D. (2009). The evolution of restricted recombination
565 in sex chromosomes. Trends in ecology & evolution 24:94-102.
566 Brautigam, K., Soolanayakanahally, R., Champigny, M., Mansfield, S., Douglas, C.,
567 Campbell, M.M., and Cronk, Q. (2017). Sexual epigenetics: gender-specific
568 methylation of a gene in the sex determining region of *Populus balsamifera*.
569 Scientific reports 7:45388.
570 Chae, E., Bomblies, K., Kim, S.T., Karelina, D., Zaidem, M., Ossowski, S.,
571 Martin-Pizarro, C., Laitinen, R.A., Rowan, B.A., Tenenboim, H., et al. (2014).
572 Species-wide genetic incompatibility analysis identifies immune genes as hot
573 spots of deleterious epistasis. Cell 159:1341-1351.
574 Charlesworth, B. (1991). The evolution of sex chromosomes. Science 251:1030-1033.
575 Charlesworth, B., and Charlesworth, D. (1978). A Model for the Evolution of Dioecy
576 and Gynodioecy. The American Naturalist 112:975-997.
577 Charlesworth, D. (2012). Plant sex chromosome evolution. Journal of Experimental
578 Botany 64:405-420.
579 Charlesworth, D. (2016). Plant Sex Chromosomes. Annual Review of Plant Biology
580 67:397-420.
581 Chen, Y., Wang, T., Fang, L., Li, X., and Yin, T. (2016). Confirmation of Single-Locus
582 Sex Determination and Female Heterogamety in Willow Based on Linkage
583 Analysis. PloS one 11:e0147671.
584 Chen, Z., Ai, F., Zhang, J., Ma, X., Yang, W., Wang, W., Su, Y., Wang, M., Yang, Y.,
585 Mao, K., et al. (2020) Survival in the Tropics despite isolation, inbreeding and
586 asexual reproduction: insights from the genome of the world's southernmost
587 poplar (*Populus ilicifolia*). The Plant Journal doi:10.1111/tpj.14744.

bioRxiv preprint doi: https://doi.org/10.1101/2020.04.11.037556; this version posted April 13, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

588 Dai, X., Hu, Q., Cai, Q., Feng, K., Ye, N., Tuskan, G.A., Milne, R., Chen, Y., Wan, Z.,
589 Wang, Z., et al. (2014). The willow genome and divergent evolution from poplar
590 after the common genome duplication. Cell Research 24:1274-1277.
591 DePristo, M.A., Banks, E., Poplin, R., Garimella, K.V., Maguire, J.R., Hartl, C.,
592 Philippakis, A.A., del Angel, G., Rivas, M.A., Hanna, M., et al. (2011). A
593 framework for variation discovery and genotyping using next-generation DNA
594 sequencing data. Nature genetics 43:491-498.
595 Feng, G., Sanderson, B.J., Keefover-Ring, K., Liu, J., Ma, T., Yin, T., Smart, L.B.,
596 DiFazio, S.P., and Olson, M.S. (2020). Pathways to sex determination in plants:
597 how many roads lead to Rome? Current Opinion in Plant Biology 54:61-68.
598 Gaudet, M., Jorge, V., Paolucci, I., Beritognolo, I., Mugnozza, G.S., and Sabatti, M.
599 (2008). Genetic linkage maps of *Populus nigra* L. including AFLPs, SSRs, SNPs,
600 and sex trait. Tree Genetics & Genomes 4:25-36.
601 Geraldes, A., Hefer, C.A., Capron, A., Kolosova, N., Martinez-Nuñez, F.,
602 Soolanayakanahally, R.Y., Stanton, B., Guy, R.D., Mansfield, S.D., Douglas, C.J.,
603 et al. (2015). Recent Y chromosome divergence despite ancient origin of dioecy
604 in poplars (*Populus*). Molecular Ecology 24:3243-3256.
605 Gong, L., Wong, C.H., Idol, J., Ngan, C.Y., and Wei, C.L. (2019). Ultra-long Read
606 Sequencing for Whole Genomic DNA Analysis. Journal of visualized
607 experiments: JoVE. 10.3791/58954.
608 Han, Y., and Jiao, Y. (2015). APETALA1 establishes determinate floral meristem
609 through regulating cytokinins homeostasis in *Arabidopsis*. Plant signaling &
610 behavior 10:e989039.
611 Harkess, A., Zhou, J., Xu, C., Bowers, J.E., Van der Hulst, R., Ayyampalayam, S.,
612 Mercati, F., Riccardi, P., McKain, M.R., Kakrana, A., et al. (2017). The
613 asparagus genome sheds light on the origin and evolution of a young Y
614 chromosome. Nature communications 8:1279.
615 Henry, I.M., Akagi, T., Tao, R., and Comai, L. (2018). One Hundred Ways to Invent
616 the Sexes: Theoretical and Observed Paths to Dioecy in Plants. Annual Review
617 of Plant Biology 69:553-575.
618 Hobza, R., Cegan, R., Jesionek, W., Kejnovsky, E., Vyskot, B., and Kubat, Z. (2017).
619 Impact of repetitive elements on the Y chromosome formation in plants. Genes
620 8:302.
621 Hobza, R., Kubat, Z., Cegan, R., Jesionek, W., Vyskot, B., and Kejnovsky, E. (2015).
622 Impact of repetitive DNA on sex chromosome evolution in plants. Chromosome
623 Research 23:561-570.
624 Hou, J., Ye, N., Zhang, D., Chen, Y., Fang, L., Dai, X., and Yin, T. (2015). Different
625 autosomes evolved into sex chromosomes in the sister genera of *Salix* and
626 *Populus*. Scientific reports 5:9076.
627 Hu, H., Yang, W., Zheng, Z., Niu, Z., Yang, Y., Wan, D., Liu, J., and Ma, T. (2020).
628 Analysis of Alternative Splicing and Alternative Polyadenylation in *Populus alba*
629 var. *pyramidalis* by Single-Molecular Long-Read Sequencing. Frontiers in
630 genetics 11:48.
631 Jain, M., Koren, S., Miga, K.H., Quick, J., Rand, A.C., Sasani, T.A., Tyson, J.R.,

bioRxiv preprint doi: https://doi.org/10.1101/2020.04.11.037556; this version posted April 13, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

632 Beggs, A.D., Dilthey, A.T., Fiddes, I.T., et al. (2018). Nanopore sequencing and
633 assembly of a human genome with ultra-long reads. Nature Biotechnology
634 36:338-345.
635 Kapitonov, V.V., and Jurka, J. (2007). Helitrons on a roll: eukaryotic rolling-circle
636 transposons. Trends in genetics: TIG 23:521-529.
637 Kersten, B., Pakull, B., Groppe, K., Lueneburg, J., and Fladung, M. (2014). The
638 sex-linked region in *Populus tremuloides* Turesson 141 corresponds to a
639 pericentromeric region of about two million base pairs on *P. trichocarpa*
640 chromosome 19. Plant biology 16:411-418.
641 Khan, A., Garbelli, A., Grossi, S., Florentin, A., Batelli, G., Acuna, T., Zolla, G., Kaye,
642 Y., Paul, L.K., Zhu, J.K., et al. (2014). The *Arabidopsis* STRESS RESPONSE
643 SUPPRESSOR DEAD-box RNA helicases are nucleolar- and
644 chromocenter-localized proteins that undergo stress-mediated relocalization and
645 are involved in epigenetic gene silencing. The Plant journal 79:28-43.
646 Kieber, J.J., and Schaller, G.E. (2018). Cytokinin signaling in plant development.
647 Development 145.
648 Kiełbasa, S.M., Wan, R., Sato, K., Horton, P., and Frith, M.C. (2011). Adaptive seeds
649 tame genomic sequence comparison. Genome research 21:487-493.
650 Koren, S., Walenz, B.P., Berlin, K., Miller, J.R., Bergman, N.H., and Phillippy, A.M.
651 (2017). Canu: scalable and accurate long-read assembly via adaptive k-mer
652 weighting and repeat separation. Genome research 27:722-736.
653 Kuroha, T., Tokunaga, H., Kojima, M., Ueda, N., Ishida, T., Nagawa, S., Fukuda, H.,
654 Sugimoto, K., and Sakakibara, H. (2009). Functional analyses of LONELY GUY
655 cytokinin-activating enzymes reveal the importance of the direct activation
656 pathway in *Arabidopsis*. The Plant cell 21:3152-3169.
657 Li, H. (2018). Minimap2: pairwise alignment for nucleotide sequences.
658 Bioinformatics 34:3094-3100.
659 Li, H., and Durbin, R. (2009). Fast and accurate short read alignment with
660 Burrows-Wheeler transform. Bioinformatics 25:1754-1760.
661 Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G.,
662 Abecasis, G., and Durbin, R. (2009). The Sequence Alignment/Map format and
663 SAMtools. Bioinformatics 25:2078-2079.
664 Li, M., Wang, D., Zhang, L., Kang, M., Lu, Z., Zhu, R., Mao, X., Xi, Z., and Ma, T.
665 (2019). Intergeneric Relationships within the Family Salicaceae *s.l.* Based on
666 Plastid Phylogenomics. International journal of molecular sciences 20:3788.
667 Ma, J., Wan, D., Duan, B., Bai, X., Bai, Q., Chen, N., and Ma, T. (2019). Genome
668 sequence and genetic transformation of a widely distributed and cultivated poplar.
669 Plant biotechnology journal 17:451-460.
670 Ma, T., Wang, K., Hu, Q., Xi, Z., Wan, D., Wang, Q., Feng, J., Jiang, D., Ahani, H.,
671 Abbott, R.J., et al. (2018). Ancient polymorphisms and divergence hitchhiking
672 contribute to genomic islands of divergence within a poplar species complex.
673 Proceedings of the National Academy of Sciences of the United States of
674 America 115:E236-e243.
675 Melnikova, N.V., Kudryavtseva, A.V., Borkhert, E.V., Pushkova, E.N., Fedorova,

bioRxiv preprint doi: https://doi.org/10.1101/2020.04.11.037556; this version posted April 13, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

676 M.S., Snezhkina, A.V., Krasnov, G.S., and Dmitriev, A.A. (2019). Sex-specific
677 polymorphism of *MET1* and *ARR17* genes in *Populus* × *sibirica*. Biochimie
678 162:26-32.
679 Ming, R., Bendahmane, A., and Renner, S.S. (2011). Sex chromosomes in land plants.
680 Annual Review of Plant Biology 62:485-514.
681 Mitsuda, N., Seki, M., Shinozaki, K., and Ohme-Takagi, M. (2005). The NAC
682 transcription factors NST1 and NST2 of *Arabidopsis* regulate secondary wall
683 thickenings and are required for anther dehiscence. The Plant cell 17:2993-3006.
684 Morgante, M., Brunner, S., Pea, G., Fengler, K., Zuccolo, A., and Rafalski, A. (2005).
685 Gene duplication and exon shuffling by helitron-like transposons generate
686 intraspecies diversity in maize. Nature genetics 37:997-1002.
687 Pakull, B., Groppe, K., Meyer, M., Markussen, T., and Fladung, M. (2009). Genetic
688 linkage mapping in aspen (*Populus tremula* L. and *Populus tremuloides* Michx.).
689 Tree Genetics & Genomes 5:505-515.
690 Pakull, B., Kersten, B., Lüneburg, J., and Fladung, M. (2015). A simple PCR-based
691 marker to determine sex in aspen. Plant biology 17:256-261.
692 Palmer, D.H., Rogers, T.F., Dean, R., and Wright, A.E. (2019). How to identify sex
693 chromosomes and their turnover. Molecular Ecology 28:4709-4724.
694 Paolucci, I., Gaudet, M., Jorge, V., Beritognolo, I., Terzoli, S., Kuzminsky, E., Muleo,
695 R., Scarascia Mugnozza, G., and Sabatti, M. (2010). Genetic linkage maps of
696 *Populus alba* L. and comparative mapping analysis of sex determination across
697 *Populus* species. Tree Genetics & Genomes 6:863-875.
698 Peto, F. (2011). Cytology of poplar species and natural hybrids. Canadian Journal of
699 Research 16:445-455.
700 Pucholt, P., Rönnberg-Wästljung, A.C., and Berlin, S. (2015). Single locus sex
701 determination and female heterogamety in the basket willow (*Salix viminalis* L.).
702 Heredity 114:575-583.
703 Pucholt, P., Wright, A.E., Conze, L.L., Mank, J.E., and Berlin, S. (2017). Recent Sex
704 Chromosome Divergence despite Ancient Dioecy in the Willow *Salix viminalis*.
705 Molecular Biology and Evolution 34:1991-2001.
706 Purcell, S., Neale, B., Todd-Brown, K., Thomas, L., Ferreira, M.A., Bender, D.,
707 Maller, J., Sklar, P., de Bakker, P.I., Daly, M.J., et al. (2007). PLINK: a tool set
708 for whole-genome association and population-based linkage analyses. American
709 journal of human genetics 81:559-575.
710 Renner, S.S. (2014). The relative and absolute frequencies of angiosperm sexual
711 systems: dioecy, monoecy, gynodioecy, and an updated online database.
712 American journal of botany 101:1588-1596.
713 Renner, S.S., and Ricklefs, R.E. (1995). Dioecy and its correlates in the flowering
714 plants. American journal of botany:596-606.
715 Sanderson, B., Feng, G., Hu, N., Grady, J., Carlson, C., Smart, B., Keefover-Ring, K.,
716 Yin, T., Ma, T., Liu, J., DiFazio, S., et al. (2020). Sex determination through X-Y
717 heterogamety in *Salix nigra*. bioRxiv.
718 Stanke, M., Keller, O., Gunduz, I., Hayes, A., Waack, S., and Morgenstern, B. (2006).
719 AUGUSTUS: ab initio prediction of alternative transcripts. Nucleic acids

bioRxiv preprint doi: https://doi.org/10.1101/2020.04.11.037556; this version posted April 13, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

720 research 34:W435-439.
721 Tamura, K., Peterson, D., Peterson, N., Stecher, G., Nei, M., and Kumar, S. (2011).
722 MEGA5: molecular evolutionary genetics analysis using maximum likelihood,
723 evolutionary distance, and maximum parsimony methods. Molecular Biology
724 and Evolution 28:2731-2739.
725 Tarailo-Graovac, M., and Chen, N. (2009). Using RepeatMasker to identify repetitive
726 elements in genomic sequences. Current protocols in bioinformatics Chapter
727 4:Unit 4.10.
728 Tennessen, J.A., Wei, N., Straub, S.C.K., Govindarajulu, R., Liston, A., and Ashman,
729 T.L. (2018). Repeated translocation of a gene cassette drives sex-chromosome
730 turnover in strawberries. PLoS biology 16:e2006062.
731 Thévenin, J., Pollet, B., Letarnec, B., Saulnier, L., Gissot, L., Maia-Grondard, A.,
732 Lapierre, C., and Jouanin, L. (2011). The simultaneous repression of CCR and
733 CAD, two enzymes of the lignin biosynthetic pathway, results in sterility and
734 dwarfism in *Arabidopsis thaliana*. Molecular plant 4:70-82.
735 Tokunaga, H., Kojima, M., Kuroha, T., Ishida, T., Sugimoto, K., Kiba, T., and
736 Sakakibara, H. (2012). *Arabidopsis* lonely guy (LOG) multiple mutants reveal a
737 central role of the LOG-dependent pathway in cytokinin activation. The Plant
738 journal: 69:355-365.
739 Torres, M.F., Mathew, L.S., Ahmed, I., Al-Azwani, I.K., Krueger, R., Rivera-Nuñez,
740 D., Mohamoud, Y.A., Clark, A.G., Suhre, K., and Malek, J.A. (2018).
741 Genus-wide sequencing supports a two-locus model for sex-determination in
742 *Phoenix*. Nature communications 9:3969.
743 Tuskan, G.A., and Difazio, S., and Jansson, S., and Bohlmann, J., and Grigoriev, I.,
744 and Hellsten, U., and Putnam, N., and Ralph, S., and Rombauts, S., and Salamov,
745 A., et al. (2006). The genome of black cottonwood, *Populus trichocarpa* (Torr. &
746 Gray). Science 313:1596-1604.
747 Wang, J., Na, J.K., Yu, Q., Gschwend, A.R., Han, J., Zeng, F., Aryal, R., VanBuren, R.,
748 Murray, J.E., Zhang, W., et al. (2012). Sequencing papaya X and $Y^h$
749 chromosomes reveals molecular basis of incipient sex chromosome evolution.
750 Proceedings of the National Academy of Sciences of the United States of
751 America 109:13710-13715.
752 Wang, M., Zhang, L., Zhang, Z., Li, M., Wang, D., Zhang, X., Xi, Z., Keefover-Ring,
753 K., Smart, L.B., DiFazio, S.P., et al. (2020). Phylogenomics of the genus *Populus*
754 reveals extensive interspecific gene flow and balancing selection. The New
755 phytologist 225:1370-1382.
756 Wybouw, B., and De Rybel, B. (2019). Cytokinin - A Developing Story. Trends in
757 plant science 24:177-185.
758 Xue, L., Wu, H., Chen, Y., Li, X., Hou, J., Lu, J., Wei, S., Dai, X., Olson, M., Liu, J.
759 (2020). Two antagonistic effect genes mediate separation of sexes in a fully
760 dioecious plant. bioRxiv.
761 Yang, W., Wang, K., Zhang, J., Ma, J., Liu, J., and Ma, T. (2017). The draft genome
762 sequence of a desert tree *Populus pruinosa*. GigaScience 6:1-7.
763 Yang, Y., Ma, T., Wang, Z., Lu, Z., Li, Y., Fu, C., Chen, X., Zhao, M., Olson, M.S.,

bioRxiv preprint doi: https://doi.org/10.1101/2020.04.11.037556; this version posted April 13, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

764 and Liu, J. (2018). Genomic effects of population collapse in a critically
765 endangered ironwood tree *Ostrya rehderiana*. Nature communications 9:5449.
766 Yin, T., Difazio, S.P., Gunter, L.E., Zhang, X., Sewell, M.M., Woolbright, S.A., Allan,
767 G.J., Kelleher, C.T., Douglas, C.J., Wang, M., et al. (2008). Genome structure
768 and emerging evidence of an incipient sex chromosome in *Populus*. Genome
769 research 18:422-430.
770 Zhang, J., Boualem, A., Bendahmane, A., and Ming, R. (2014). Genomics of sex
771 determination. Current Opinion in Plant Biology 18:110-116.
772 Zhang, L., Xi, Z., Wang, M., Guo, X., and Ma, T. (2018). Plastome phylogeny and
773 lineage diversification of Salicaceae with focus on poplars and willows. Ecology
774 and evolution 8:7817-7823.
775 Zhang, Z., Chen, Y., Zhang, J., Ma, X., Li, Y., Li, M., Wang, D., Kang, M., Wu, H.,
776 Yang, Y., et al. (2020). Improved genome assembly provides new insights into
777 genome evolution in a desert poplar (*Populus euphratica*). Molecular ecology
778 resources.
779 Zhou, R., Macaya-Sanz, D., Carlson, C.H., Schmutz, J., Jenkins, J.W., Kudrna, D.,
780 Sharma, A., Sandor, L., Shu, S., Barry, K., et al. (2020). A willow sex
781 chromosome reveals convergent evolution of complex palindromic repeats.
782 Genome biology 21:38.
783 Zhou, R., Macaya-Sanz, D., Rodgers-Melnick, E., Carlson, C.H., Gouker, F.E., Evans,
784 L.M., Schmutz, J., Jenkins, J.W., Yan, J., Tuskan, G.A., et al. (2018).
785 Characterization of a large sex determination region in *Salix purpurea* L.
786 (Salicaceae). Molecular genetics and genomics: MGG 293:1437-1452.
787
788
789

bioRxiv preprint doi: https://doi.org/10.1101/2020.04.11.037556; this version posted April 13, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

790 **Figure Legends**

791

792 **Fig. 1 XY sex determination on chromosome 14 in *P. euphratica*. (A)** Manhattan
793 plot of *P. euphratica* based on the results of genome-wide association study (GWAS)
794 with the male genome as reference. The y-axis represents the strength of association
795 ($-\log_{10}$($P$ value)) for each SNP sorted by chromosomes and scaffolds (SC; x-axis).
796 The red line indicates the significance after Bonferroni multiple corrections ($\alpha < 0.05$).
797 Note that the scaffold '001598F' is located on chromosome 14 based on its syntenic
798 relationship with the proximal end of chromosome 14 of *P. trichocarpa*. (**B**) Summary
799 of male *P. euphratica* genome regions containing SNPs significantly associated with
800 sex. SNP*, significantly associated SNPs; Homo, Homozygous; Hete, Heterozygosis.
801 **(C)** Synteny relationships between our assembled Y-contig and X chromosome of *P.*
802 *euphratica*, as well as the corresponding region of chromosome 14 for *P. alba*, *P.*
803 *trichocarpa* and *S. purpurea*. The highlighted part represents the sex determination
804 region (SDR), yellow for Y-SDR and green for X-SDR. Schematic diagram showing
805 the corresponding position of the SDR on chromosome 14 of *P. euphratica*. **(D)**
806 Phylogenetic relationships of the homolog pairs (HP) shared between Y- and X-SDR
807 of *P. euphratica* and their orthologous genes in other Salicaceae species. Detailed
808 information about these genes is listed in Table S7 and additional phylogenetic trees
809 are shown in Fig. S7. Note that only the orthologous genes located on the
810 corresponding region of chromosome 14 were used for phylogenetic analysis.

811

812

813 **Fig. 2 ZW sex determination on chromosome 19 in *P. alba*. (A)** Manhattan plot of *P.*
814 *alba* based on the results of GWAS with female genome as reference respectively. The
815 y-axis represents the strength of association ($-\log_{10}$($P$ value)) for each SNP sorted by
816 chromosomes and scaffolds (SC; x-axis). The red line indicates the significance after
817 Bonferroni multiple corrections ($\alpha < 0.05$). (**B**) Summary of female *P. alba* genome
818 regions containing SNPs significantly associated with sex. SNP*, significantly
819 associated SNPs; Homo, Homozygous; Hete, Heterozygosis. **(C)** Synteny

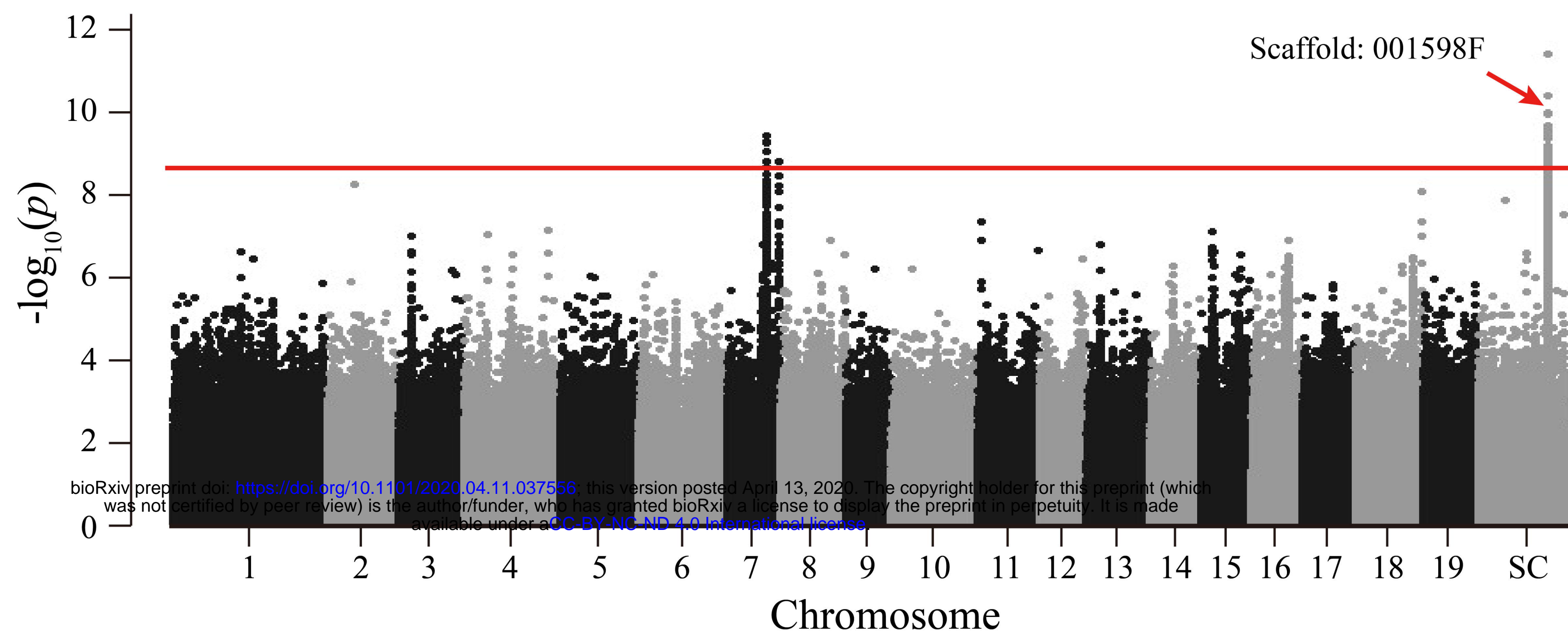
bioRxiv preprint doi: https://doi.org/10.1101/2020.04.11.037556; this version posted April 13, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

820 relationships between our assembled W-contig and Z chromosome of *P. alba*, as well
821 as the corresponding region of chromosome 19 for *P. euphratica*, *P. trichocarpa* and *S.*
822 *purpurea*. The highlighted part represents SDR, red for W-SDR and blue for Z-SDR.
823 Schematic diagram showing the corresponding position of the SDR on chromosome
824 19 of *P. alba*. **(D)** Phylogenetic relationships of the homolog pairs (HP) shared
825 between W- and Z-SDR of *P. alba* and their orthologous genes in other Salicaceae
826 species. The detail information of these genes is listed in Table S12. Note that there
827 are 3 copies for 'HP2' on the W-SDR of *P. alba*, and only the orthologous genes
828 located on the corresponding region of chromosome 19 were used for phylogenetic
829 analysis.
830

831 **Fig. 3 Evidence for SDR turnover in Salicaceae.** **(A)** Synteny relationships among
832 the Y-SDR of *P. euphratica* (yellow) and the W-SDRs of *P. alba* (red) and *S. purpurea*
833 (blue), showing the copies of *RR* intact gene ('C') and partial duplicates ('S': small
834 duplicate; 'L': large duplicate) on their SDRs. For each species, corresponding
835 positions for other *RR* gene copies or partial duplicates on the autosome are also
836 shown. **(B)** Phylogenetic relationship of the *RR* sequences (including intact genes and
837 partial duplicates) identified in the four species. The tree was rooted by a paralogous
838 gene '*RR16*'. The gene structures and relative positions of *Helitron* and *Copia*-like
839 LTR are also shown. Phylogenetic relationships of the *Helitron* **(C)** and *Copia*-like
840 LTR **(D)** around the *RR* sequences. All the sequences were named according to Fig.
841 3A. Peu: *P. euphratica*; Pal: *P. alba*; Ptr: *P. trichocarpa*; Spur: *S. purpurea*.
842

843 **Fig. 4 Hypothetical model for sex system turnovers in Salicaceae.** The W
844 chromosomes of *P. alba* and *S. purpurea* both carry several intact *RR* genes and are
845 likely to serve as a dominate promoter of female function. On the Y chromosome of *P.*
846 *euphratica*, partial duplicates of the *RR* gene are like to serve as a female suppressor
847 by encoding an siRNA that targets the intact RR gene through RNA-directed DNA
848 methylation. Note that Y-SDR of *P. trichocarpa* has not yet been assembled, so
849 whether a similar pattern should be found in this species remains to be confirmed.

A

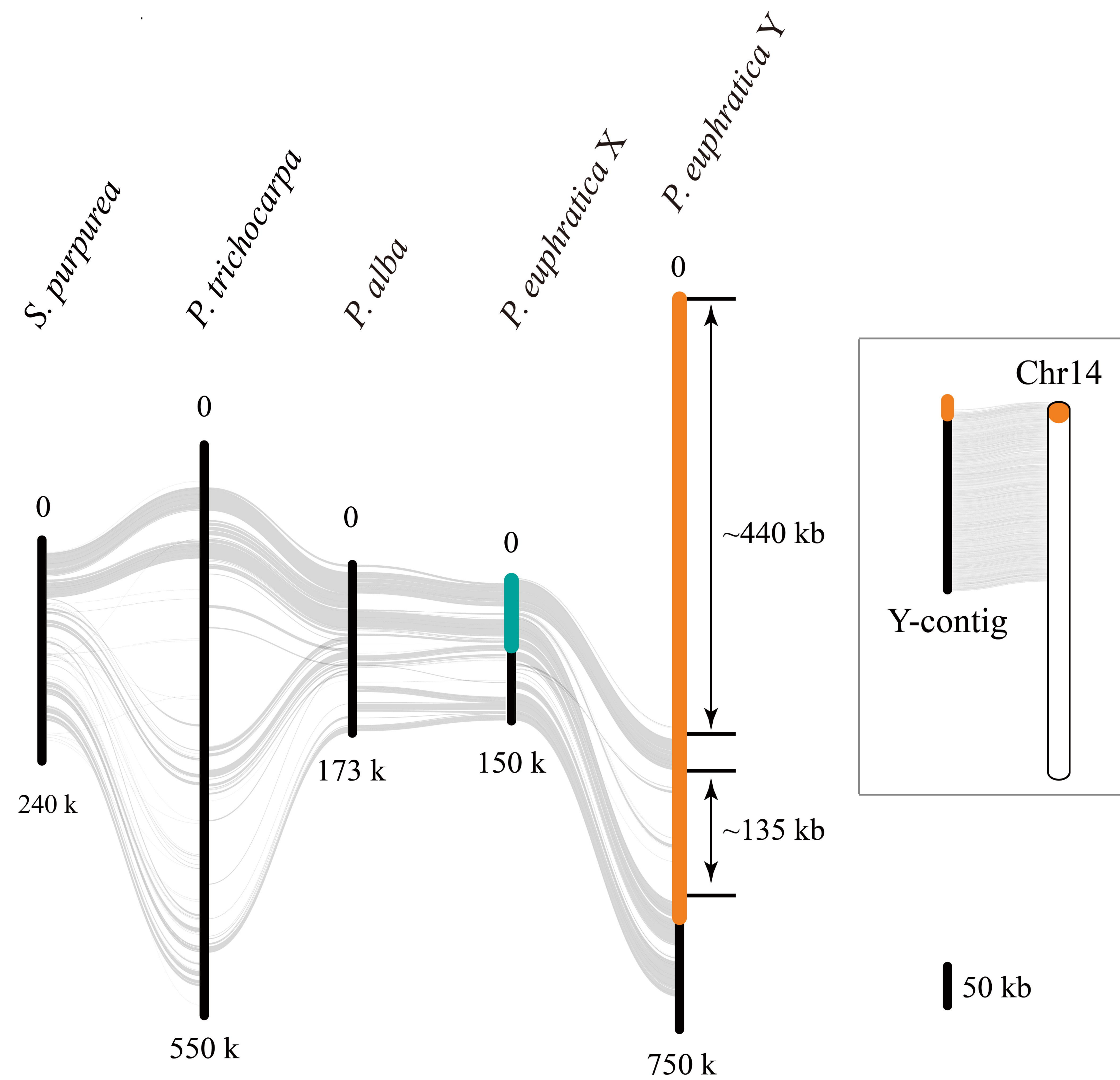



bioRxiv preprint doi: https://doi.org/10.1101/2020.04.11.037564; this version posted April 13, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

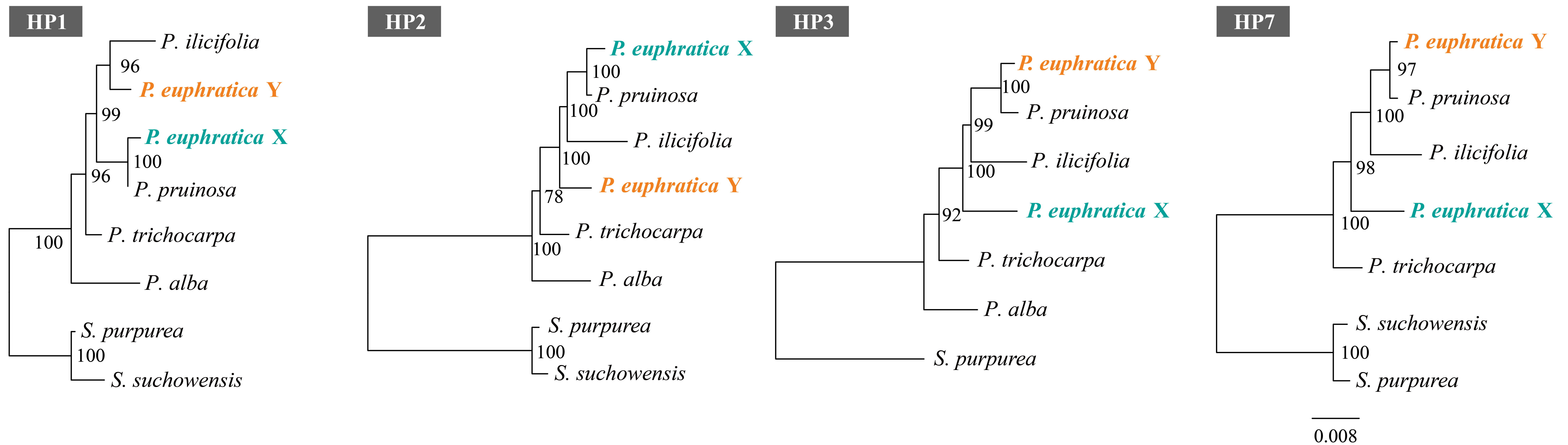
B

| Reference genome | Scaffold ID | Chr ID | Position (bp) | | SNP* | Female | | Male | |
|---|---|---|---|---|---|---|---|---|---|
| | | | Start | End | | Homo(%) | Hete(%) | Homo(%) | Heter(%) |
| Male | 001598F | 14 | 595 | 45,911 | 296 | 99.99 | 0.01 | 6.06 | 93.94 |
| | Lachesis_group10 | 7 | 17,334,724 | 17,391,719 | 8 | 100.00 | 0.00 | 21.10 | 78.90 |
| | Lachesis_group10 | 7 | 22,941,832 | 22,956,851 | 6 | 100.00 | 0.00 | 6.67 | 93.33 |
| | **Total** | - | - | - | **310** | **99.99** | **0.01** | **6.43** | **93.57** |

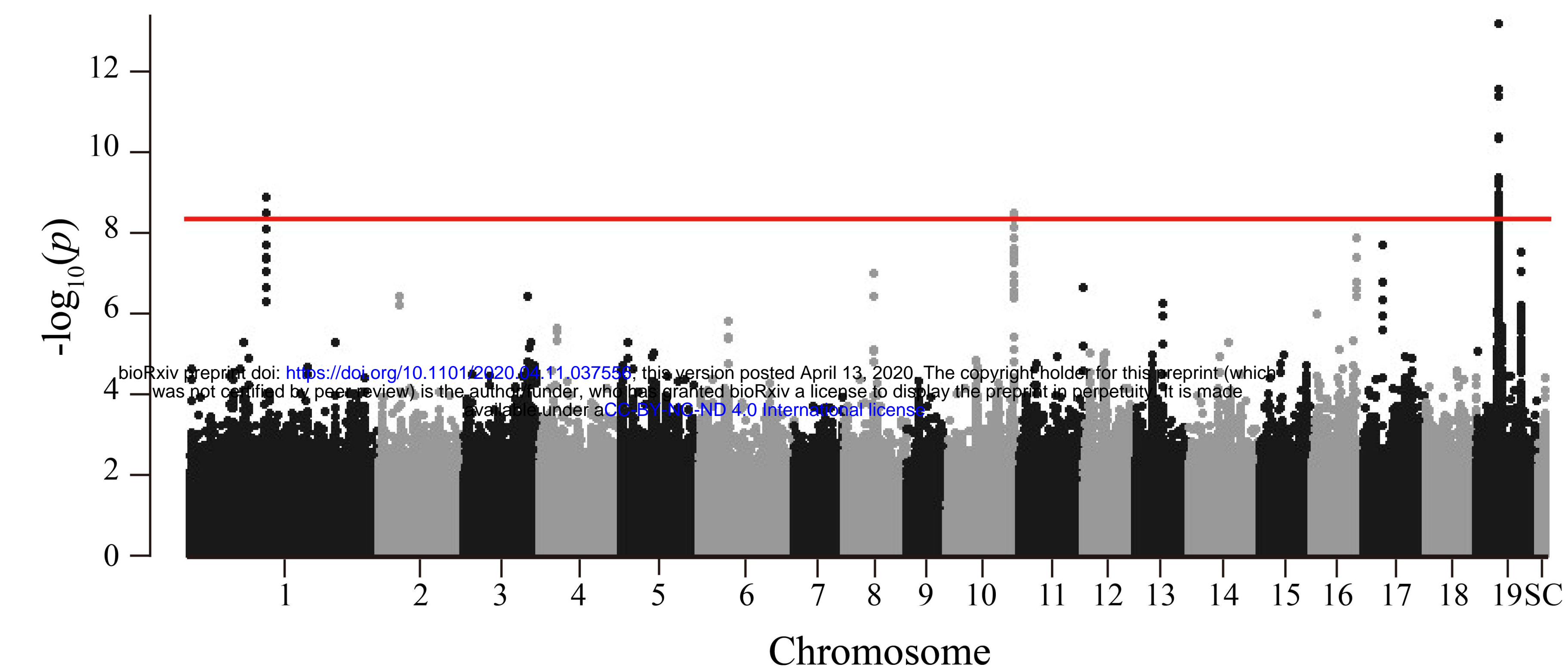
C

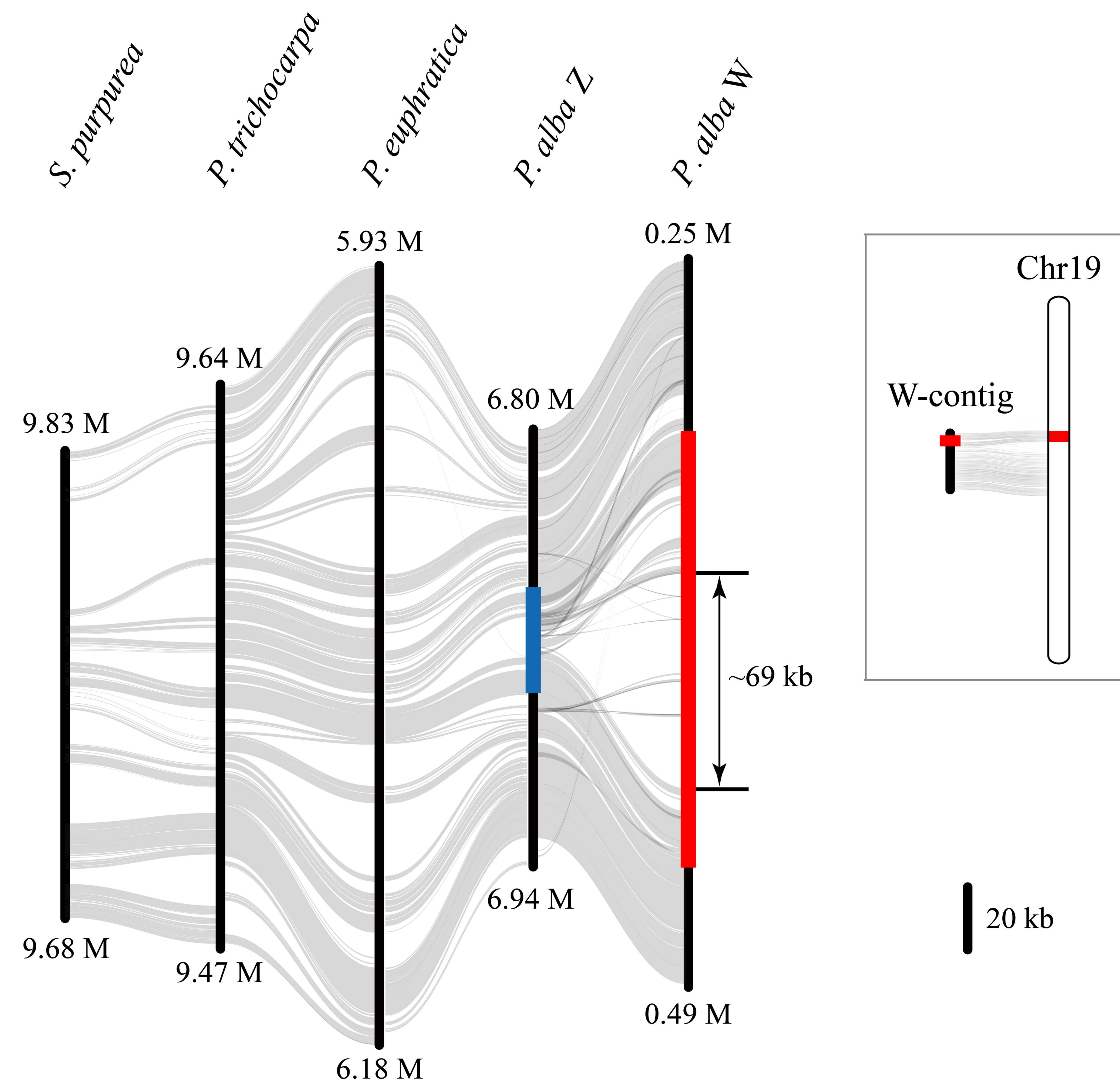



D



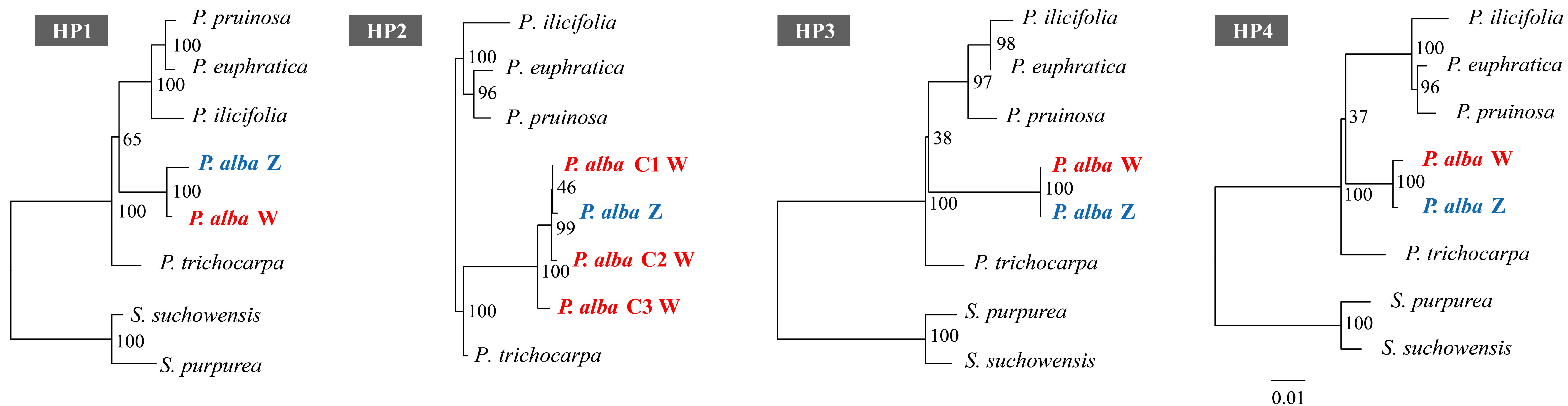
bioRxiv preprint doi: https://doi.org/10.1101/2020.04.11.037556; this version posted April 13, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

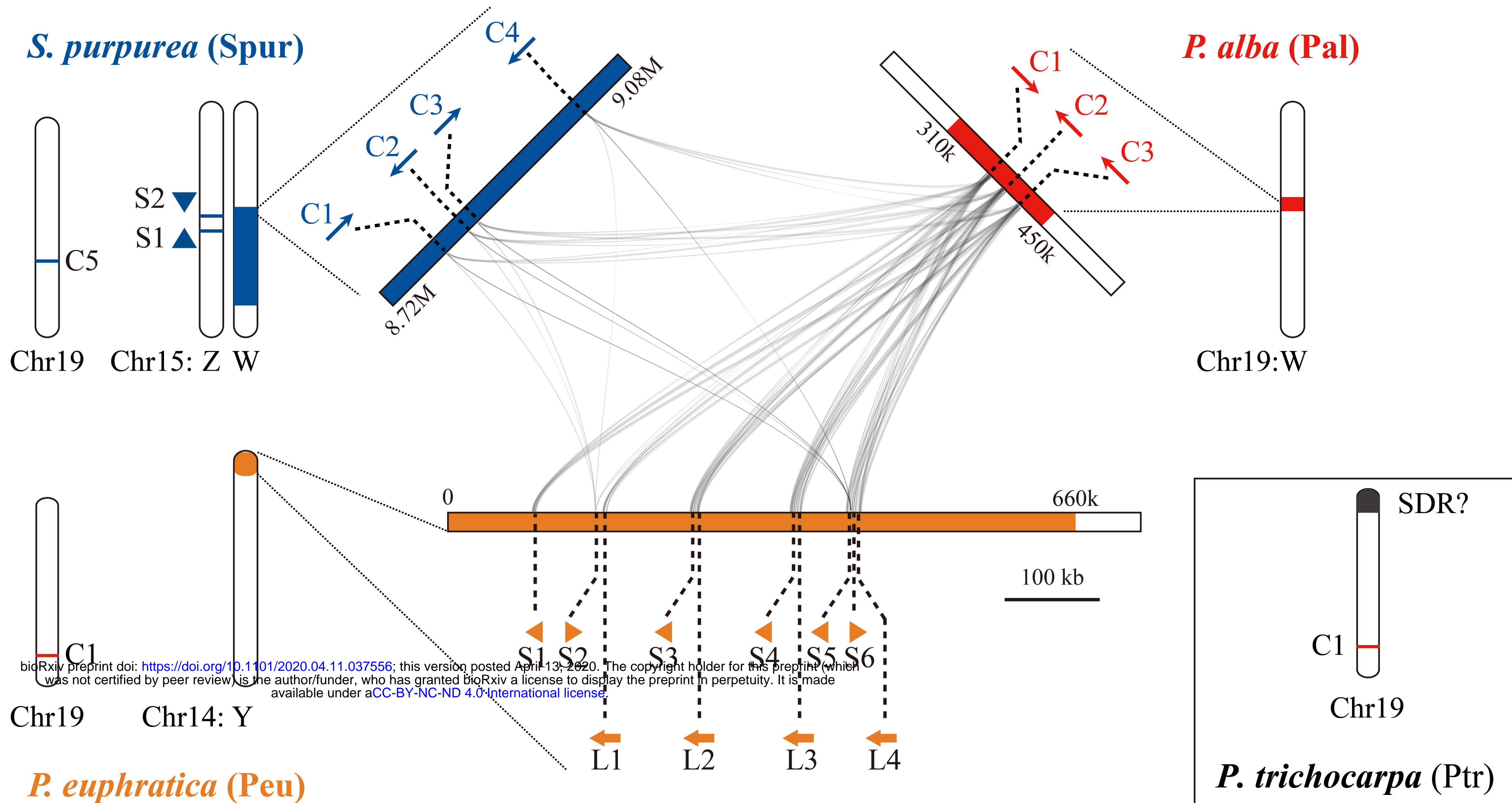


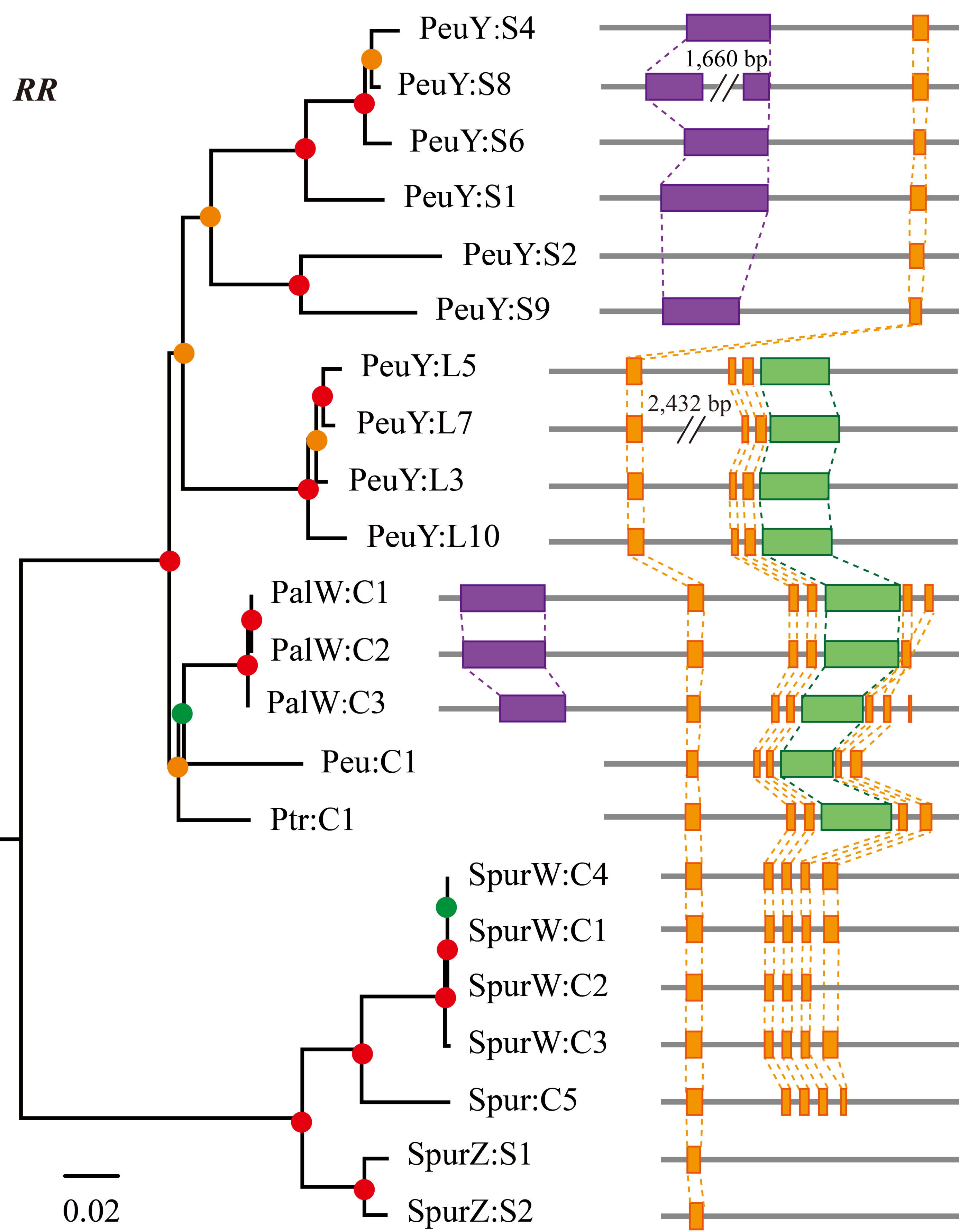

B

| Reference genome | Scaffold ID | Chr ID | Position (bp) Start | Position (bp) End | SNP* | Female Homo(%) | Female Hete(%) | Male Homo(%) | Male Hete(%) |
|---|---|---|---|---|---|---|---|---|---|
| Female | Contig42 | 19 | 317,074 | 440,815 | 48 | 8.44 | 91.56 | 96.61 | 3.39 |
| | Contig111 | 10 | 131,470 | 132,638 | 4 | 6.67 | 93.33 | 96.08 | 3.92 |
| | Contig319 | 1 | 26,758 | 26,805 | 3 | 2.23 | 97.78 | 96.67 | 3.33 |
| | **Total** | **-** | **-** | **-** | **55** | **7.95** | **92.05** | **96.58** | **3.42** |

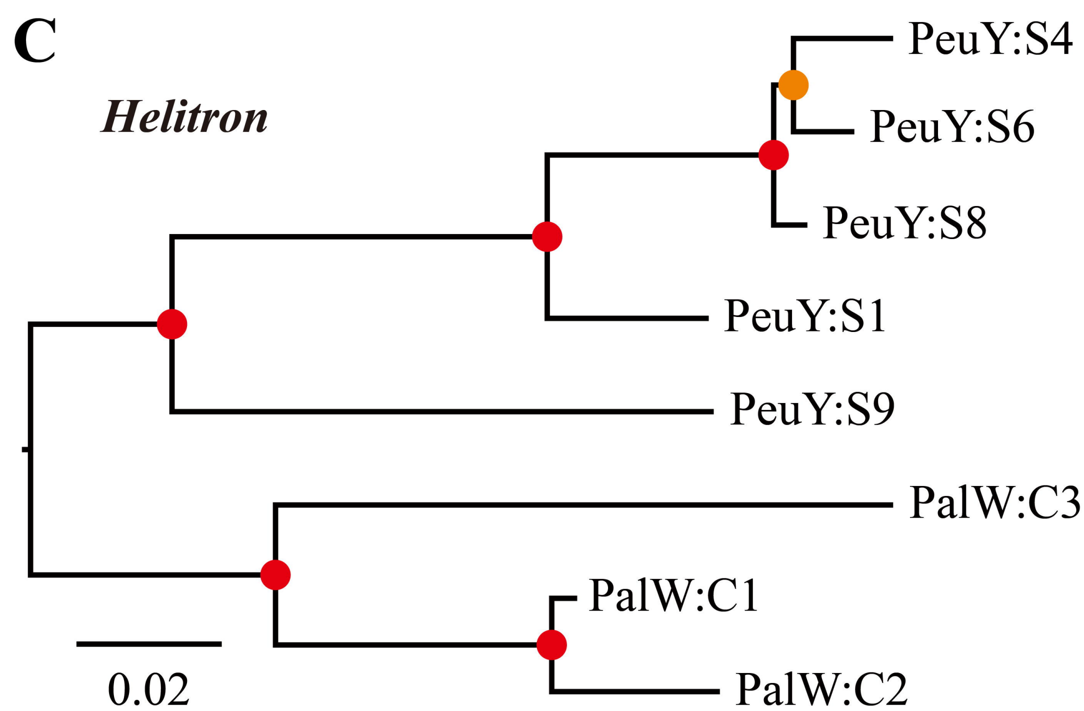


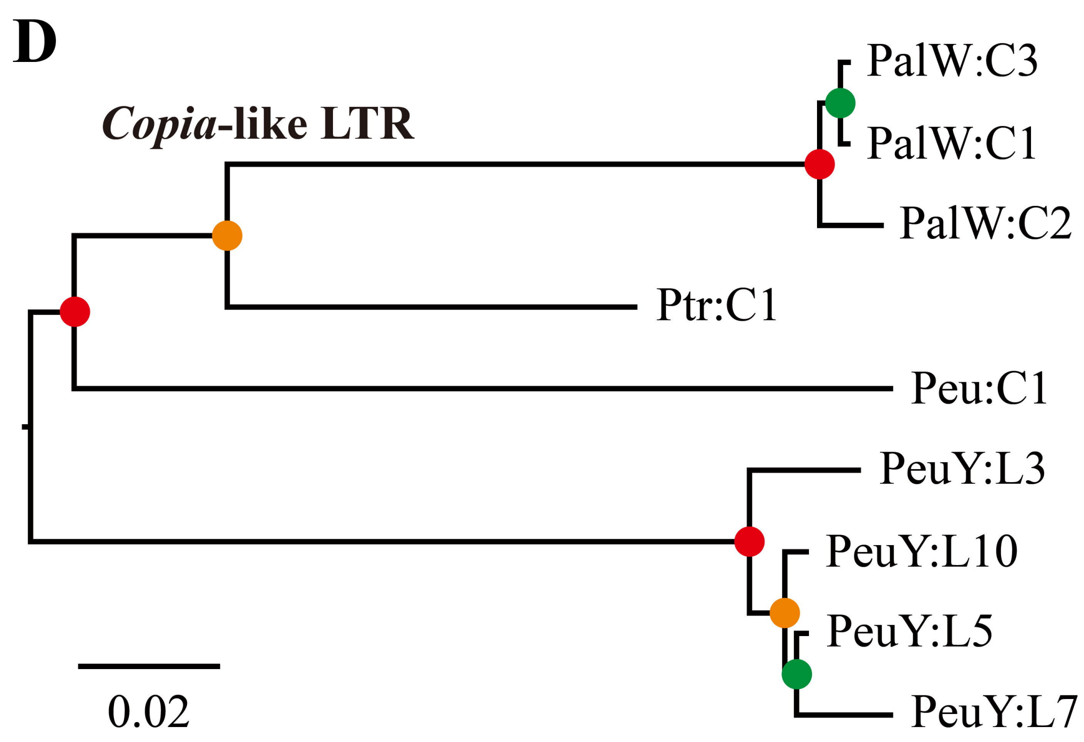


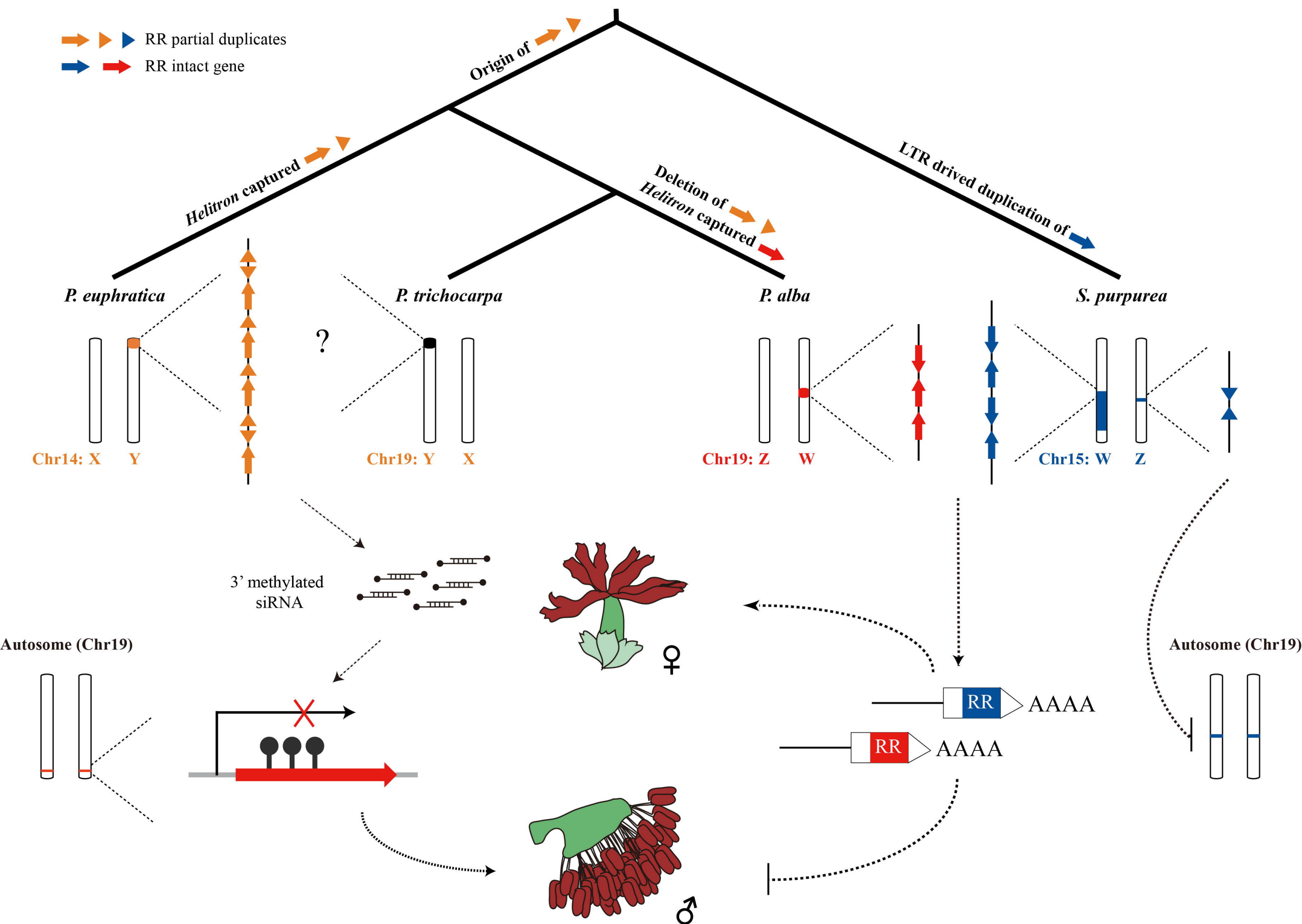

bioRxiv preprint doi: https://doi.org/10.1101/2020.04.11.037556; this version posted April 13, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.






RR partial duplicates
RR intact gene
Origin of
Helitron captured
Deletion of
Helitron captured
LTR drived duplication of
P. euphratica
P. trichocarpa
P. alba
S. purpurea
?
Chr14: X
Y
Chr19: Y
X
Chr19: Z
W
Chr15: W
Z
3’ methylated
siRNA
Autosome (Chr19)
Autosome (Chr19)
RR
AAAA
RR
AAAA
♀
♂