Conserved genomic landscapes of differentiation across *Populus* speciation continuum

Huiying Shang[1,2], Martha Rendón-Anaya[3], Ovidiu Paun[1], David L Field[4], Jaqueline Hess[5], Claus Vogl[6], Jianquan Liu[7], Pär K. Ingvarsson[3,9]*, Christian Lexer[1,9 †], Thibault Leroy[1,8,9]*

[1]Department of Botany and Biodiversity Research, University of Vienna, Vienna, Austria.

[2]Vienna Graduate School of Population Genetics, Vienna, Austria.

[3]Swedish University of Agricultural Sciences (SLU), Uppsala, Sweden.

[4]Edith Cowan University, Perth, Australia.

[5]Helmholtz Centre for Environmental Research, Halle (Saale), Germany.

[6]Department of Biomedical Sciences, Vetmeduni Vienna, Vienna, Austria

[7]Key Laboratory for Bio-resources and Eco-environment, College of Life Science, Sichuan University, Chengdu, People's Republic of China

[8]IRHS-UMR1345, Université d'Angers, INRAE, Institut Agro, SFR 4207 QuaSaV, 49071 Beaucouzé, France

[9]Shared last authorship.

[†]Deceased.

*Corresponding authors:

- Thibault Leroy, Department of Botany and Biodiversity Research, University of Vienna, Rennweg 14, 1030 Vienna, Austria

Email: thibault.leroy@univie.ac.at

- Pär K. Ingvarsson, Swedish University of Agricultural Sciences (SLU), Uppsala, Sweden

Email: par.ingvarsson@slu.se

Abstract

Investigating genome-wide variation patterns along a speciation continuum is of central importance to understand the evolutionary processes contributing to lineage diversification. To identify which forces have shaped the genomic landscapes in *Populus*, we resequenced 201 whole-genomes from eight closely related species, with pairs of species at different stages along the speciation continuum. Using population structure and identity by descent analyses, we show extensive introgression between some species pairs, especially those with parapatric distributions. Inference of historical changes in effective population sizes support species-specific demographic trajectories, including recent population expansions in species characterized by broad present-day distributions. We observe highly conserved genomic landscapes, either focusing on within-species (genetic diversity: π and recombination rate: ρ) or among-species variation (relative divergence: F_{ST} and absolute divergence: D_{XY}). Independent of the stage across the divergence continuum, we recovered positive correlations between the pair π and ρ and the pair D_{XY} and ρ across all species pairs, which is consistent with a substantial contribution of linked selection in shaping these genomic landscapes. However, the positive correlations between π and D_{XY} became weaker as the overall divergence level (d_a) increased, suggesting that background selection is not the only factor at play. Positive correlations between F_{ST} and D_{XY} in all species pairs, regardless of the rate of gene flow, also indicates the high F_{ST} could be due to divergent sorting of ancient polymorphism before speciation. Our study showcases the importance of investigating genomic patterns on multiple species across the speciation continuum to better understand the genomic landscapes of diversity and differentiation.

Keywords: differentiation islands, divergence, introgression, identity-by-descent, linked selection, recombination

Introduction

Understanding the evolutionary forces that shape natural genetic variation is a central goal of evolutionary biology. Over the last decade, population genomic studies have documented highly heterogeneous genomic landscapes of differentiation, identifying both regions of elevated and reduced differentiation between diverging populations (Ellegren, et al. 2012; Martin, et al. 2013; Lamichhaney, et al. 2015; Vijay, et al. 2016; Sendell-Price, et al. 2020). Notwithstanding the specific features of individual groups, the overall drivers of this omnipresent pattern can be many (see below) and their relative contribution is still debated (Wolf and Ellegren 2017).

Hotspots of elevated genetic differentiation relative to genomic background are often referred to as 'differentiation islands' or 'speciation islands' and are assumed to form around loci underlying local adaptation and reproductive isolation. Thus, delineating differentiation islands has recently become a major topic of research in the field of speciation and adaptation genomics (Burri 2017b; Martin and Jiggins 2017; Ravinet, et al. 2018; Tavares, et al. 2018; Stankowski, et al. 2019). Such investigations are facilitated in groups still experiencing interspecific gene flow, i.e., species diverging under an isolation-with-migration or a secondary contact scenario (Harrison and Larson 2016; Roux, et al. 2016; Wolf and Ellegren 2017; Leroy, et al. 2020; Yamasaki, et al. 2020). For such species pairs, effective migration can be reduced around 'speciation islands' due to the effect of selection against hybrids, whereas the remainder of the genome can be homogenized by gene flow (Wu 2001). In *Heliconius* butterflies, for instance, admixture footprints were found to be locally weaker around loci involved in Mullerian mimicry (Martin, et al. 2013). In many plant species, such as monkey-flowers (Martin, et al. 2019), snapdragons (Tavares, et al. 2018) and morning-glories (Rifkin, et al. 2019), differentiation islands have been reported to harbor key adaptive loci controlling floral traits, proving the role of selection and gene flow in shaping a highly heterogeneous genomic landscape of differentiation.

Other empirical studies reported, however, that heterogeneous differentiation landscapes can emerge due to genomic features not causally linked to speciation and adaptation. These features can include background selection, recombination rate variation, biased gene conversion, and genomic characteristics influenced by life history traits (Corbett-Detig, et al. 2015; Wolf and Ellegren 2017). In regions of low recombination, genetic hitchhiking (Smith and Haigh 1974)

85 and background selection (Charlesworth, et al. 1993) may reduce effective population size (*Ne*)
86 and lead to decreased diversity levels and elevated relative differentiation across a larger
87 genomic region around the selected locus (Charlesworth and Campos, 2014). Thus, a positive
88 correlation between recombination and genome-wide diversity or a negative correlation between
89 recombination and genome-wide relative differentiation would be expected (Burri 2017a). The
90 correlations have been reported in *Ficedula* flycatchers (Burri, et al. 2015), *Helianthus*
91 sunflowers (Renaut, et al. 2013), the Pacific cupped oyster (Gagnaire, et al. 2018), warblers
92 (Irwin, et al. 2018) and hummingbirds (Henderson and Brelsford 2020). However, a recent study
93 in a triplet of *Leptidea* butterfly species suggested that divergence landscapes are mainly shaped
94 by directional selection, rather than recombination or introgression (Talla, et al. 2019). Another
95 study in *Boechera stricta* demonstrated that ancestral balanced polymorphism may have
96 contributed to the genomic regions of high divergence (Wang, et al. 2019). These examples
97 demonstrate that debates still exist regarding the relative importance of different evolutionary
98 processes that contribute to the heterogeneous landscape of differentiation. One possible
99 explanation for these different findings is that each study system lies at different time points
100 along the speciation continuum, where the relative importance of genomic features and
101 individual evolutionary processes changes as species pairs move along the speciation continuum.
102 This highlights that disentangling the relative contribution of these different evolutionary
103 processes to interpret the underlying mechanisms driving genomic divergence remains
104 challenging with only a few species pairs (Wolf and Ellegren 2017). Therefore, new insights into
105 the relative contribution of these evolutionary processes will likely come from whole genomic
106 studies that include multiple pairs of independent lineages along a speciation continuum where
107 we can also detect genomic features (i.e., recombination rate), demographic history and the
108 presence and timing of gene flow.

109 To understand the processes behind the formation of 'differentiation islands' along the
110 speciation continuum, we summarize four different scenarios according to the divergence models
111 proposed by Han *et al* (2017) and Irwin *et al* (2018) (Fig. 1a). In this context, nucleotide
112 diversity (π) as well as two measures of divergence, the relative measure F_{ST} and the absolute
113 measure D_{XY}, are instrumental in distinguishing between evolutionary scenarios (Charlesworth
114 1998). Genomic regions with high F_{ST} may occur in the absence of local gene flow due to

115 ecological or non-ecological isolating barriers, or due to locally reduced *Ne* in regions of low
116 recombination or selective sweeps (Noor and Bennett 2009; Turner and Hahn 2010; Renaut, et
117 al. 2014; Campagna, et al. 2015; Gagnaire, et al. 2018; Henderson and Brelsford 2020).
118 Disentangling the underlying processes using only F_{ST} is impractical, as this is directly
119 influenced by the level of within-groups genetic diversity (e.g., Cruickshank and Hahn, 2014).
120 D_{XY} is a complementary measure that is independent of the levels of diversity within the groups.
121 D_{XY} is sensitive to ancestral variation, with higher values in regions of restricted gene flow and
122 decreased in regions under background selection or selective sweeps (Han, et al. 2017; Irwin, et
123 al. 2018). Scanning the whole genome of pairs of species using π, F_{ST} and D_{XY}, different
124 scenarios can be hypothesized regarding the extent of local gene flow to identify the main drivers
125 of genomic landscapes of differentiation (Fig. 1a). The first scenario is 'divergence with gene
126 flow' where selection at loci which contribute to reproductive isolation restricts gene exchange
127 between populations, locally elevating genomic differentiation (leading to both high F_{ST} and D_{XY})
128 and reducing genetic diversity. The second scenario is 'allopatric selection' where positive
129 selection acts on distinct regions of the genome after a species split into two populations, leading
130 to lower π and higher F_{ST}. As D_{XY} is sensitive to ancestral polymorphism, its values are expected
131 to remain relatively stable under this model. Besides, selective sweeps in allopatry in one or both
132 populations may also contribute to this divergence pattern. The third scenario is 'recurrent
133 selection' where background selection or selective sweeps at certain genomic regions reduce
134 genetic diversity in both the common ancestor and the two daughter populations, leading to
135 lower D_{XY} and π, but higher F_{ST}. The fourth and last scenario is 'balancing selection' where
136 ancestral polymorphism is maintained at selected sites, resulting in increased differentiation
137 between species due to lineage sorting. When employed in conjunction, the use of genomic
138 estimates of F_{ST}, D_{XY} and π can thus enhance our understanding of the mechanisms at play that
139 shaped the genomic differentiation landscape.

140 Conserved genomic landscapes of differentiation among independent species pairs could be
141 generated by long-term linked selection in heterogeneous recombination landscapes. And it has
142 been argued that background selection may more likely lead to shared genomic patterns than
143 genetic hitchhiking because deleterious mutations are far more common than beneficial ones
144 (Burri 2017a). However, the conserved differentiation landscapes could also be the consequence

of polygenic adaptation from standing genetic variation. To have a better understanding which forces contribute to the similar patterns, one powerful tool is to parse out the signal of background selection (Burri, 2017a; Stankowski, et al. 2019). Under the influence of background selection, we would expect a significantly negative correlation between F_{ST} and D_{XY}. Conversely, a positive correlation between these two parameters may reflect divergent hitchhiking or ancient polymorphism contributing to the genomic landscape of differentiation (Fig. 1a). In addition, according to the predictions of background selection proposed by Burri 2017a, the impact of background on the genomic landscape of diversity and divergence could be estimated by measuring the dynamic correlations between genomic variation and recombination across speciation continuum. For example, the negative correlations of F_{ST} and ρ become increased with advancing differentiation (Fig. 1b). This is due to the fact that at the early stage of divergence, differentiation starts building up through lineage sorting. With divergence increasing, the effects of background selection will accumulate and the genomic landscape of differentiation will increasingly reflect recombination. The positive correlations between π and ρ should remain highly correlated with each other under background selection. Should these correlations become weaker as divergence time increases, a prediction of background selection would be violated; thus other evolutionary processes may play a role in generating the heterogeneous landscape of differentiation.

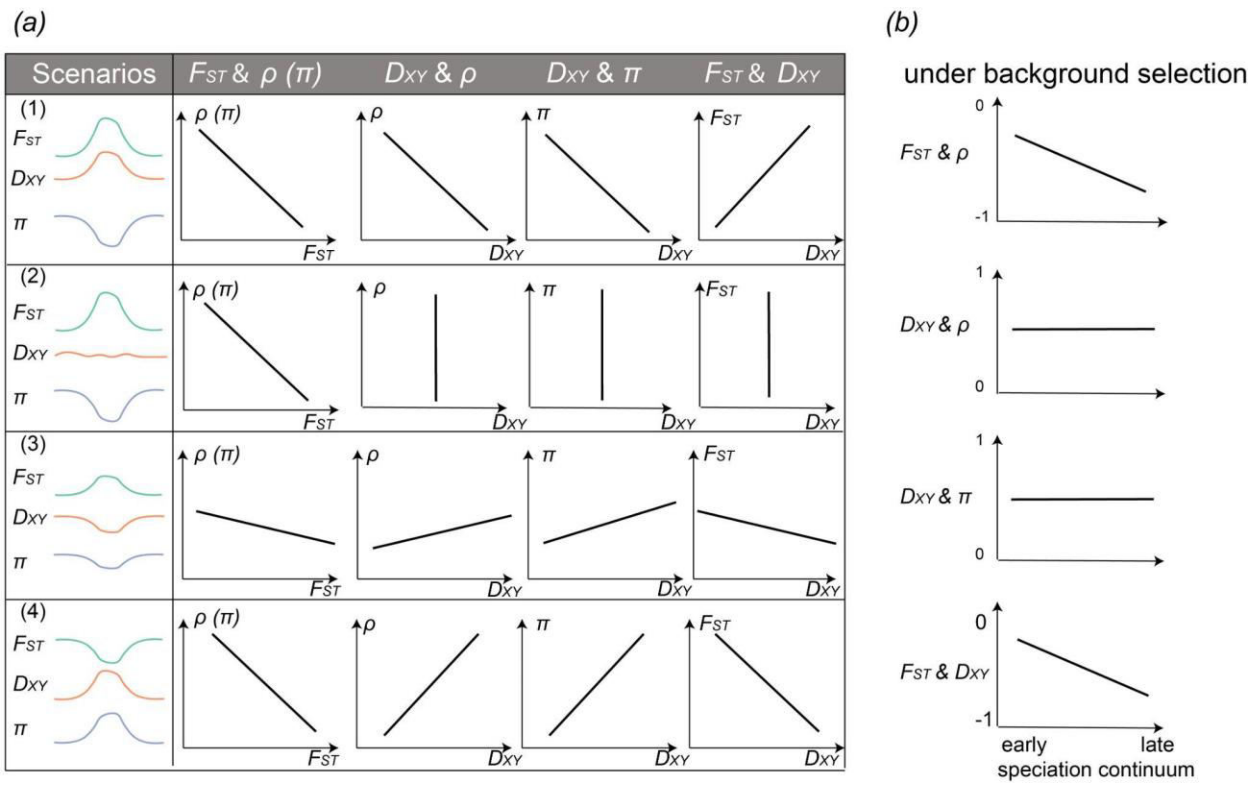


Fig. 1. *Expected patterns of genomic differentiation under different scenarios. (a) Genetic diversity, divergence change along the genome under four scenarios: (1) 'divergence with gene flow', (2) 'allopatric selection', (3) 'background selection/balancing selection', and (4) 'balancing selection'. Besides, expected correlations between different parameters are shown under each scenario. (b) The expected correlations between different parameters change under background selection along the speciation continuum from the early to the late stage.*

Species of the genus *Populus* are perennial woody plants, dioecious, and widely distributed across the Northern Hemisphere (Stettler, et al. 1996). The genus *Populus* comprises six sections containing 29 species, among which ten species form the section *Populus* (Stettler, et al. 1996; Jansson, et al. 2010). The genus *Populus* is well studied in evolutionary biology not only due to the trees' economic and ecological importance, but also due to their small genome sizes (<500Mb), diploidy through the genus ($2n$ = 38), wind pollination, extensive gene flow among species, and sexual and vegetative reproductive strategies (Rajora and Dancik 1992; Martinsen, et al. 2001; Suarez-Gonzalez, et al. 2016). Among all woody perennial angiosperm species, the genome of *P. trichocarpa* was sequenced and published first (Tuskan, et al. 2006). In addition to *Populus trichocarpa*, another well-annotated genome assembly is available (*P. tremula*;

180 Schiffthaler, et al. 2019). Taken together, the rich genomic resources, small genomes and
181 widespread geographic distribution make the *Populus* system a useful model to address open
182 questions regarding the formation of genomic islands along the speciation continuum.

183 In this study, we focused on white poplars and aspens from the section *Populus* which are
184 widely distributed in Eurasia and North America (Supplementary material, Fig. S1 and Table S1)
185 and provide a broad set of species pairs along the speciation continuum. The divergence time
186 among species from *Populus* section *Populus* varies from 1.3 to 4.8 million years ago, as
187 previously reported based on a species tree estimation (Shang, et al. 2020). This variation
188 represents different stages along the speciation continuum and therefore provides an excellent
189 system to investigate the tempo of genomic differentiation and whether the relative contribution
190 of different evolutionary processes and genomic features changes through time. Here, we use
191 whole genome resequencing data from eight *Populus* species (Supplementary material, Fig. S1
192 and Table S1) to address the following questions: (1) How do genomic landscapes of
193 differentiation accumulate along the speciation continuum? (2) Are differentiation patterns
194 across the genomic landscape repeatable among independent lineages? (3) What are the main
195 evolutionary processes driving these heterogeneous landscapes of diversity and differentiation
196 along the speciation continuum? (4) Which scenario is suitable to explain the 'differentiation
197 islands' in each species pair?

198

199 ## Results and Discussions

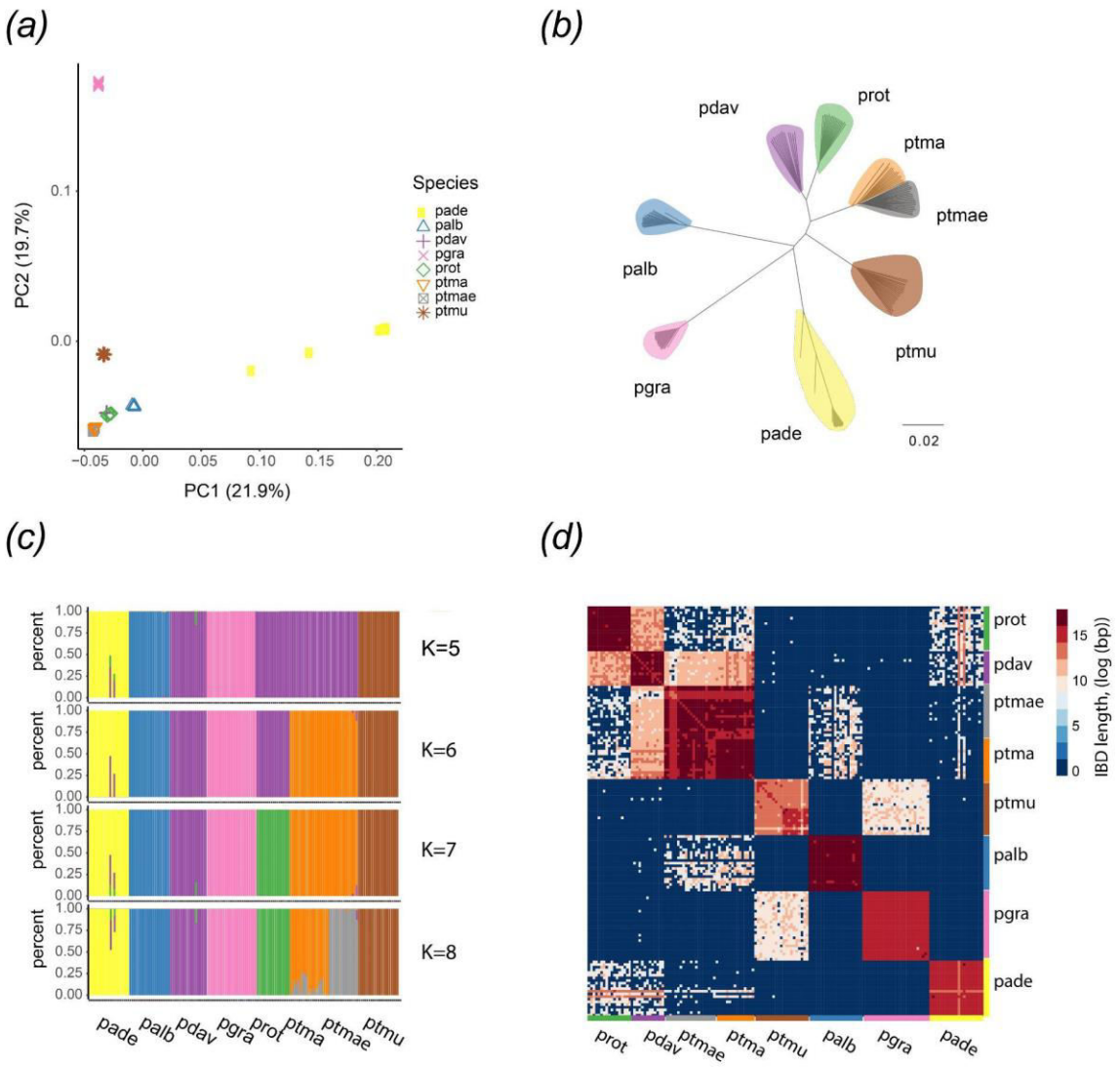
200 ### Strong interspecific structure despite interspecific introgression

201 As a first step, we estimated the relatedness between individuals using the KING toolset
202 (available at http://people.virginia.edu/~wc9c/KING/) to trace potential clone mates produced by
203 vegetative reproduction. This analysis identified 13 duplicated genotypes out of a total of 32
204 samples from the Korean population of *P. davidiana*. In addition, all individuals of *P.*
205 *qiongdaoensis* were identified as clone mates (supplementary material, Fig. S2). Therefore, these
206 two populations were eliminated from subsequent analyses. After filtering and quality control
207 (see Methods), the depth of coverage was relatively homogeneous (supplementary material, Fig.
208 S3) and varied from 21× to 32×, resulting in a dataset of 30,539,136 high-quality SNPs.

209 We then explored population genetic structure across all seven *Populus* species. Genetic
210 clustering as indicated by the first two principal components of the PCA (Fig. 2a) is consistent
211 with expected genetic divisions, identifying the species previously described as the most
212 divergent (Shang, et al. 2020) such as *P. adenopoda* or *P. grandidentata*, but not the most
213 recently diverged aspens (*P. davidiana* and *P. rotundifolia*). More specifically, the first PC
214 explains 21.9% of the total variation and separates most species, with a particularly strong
215 separation of *P. adenopoda* on one side, and all other *Populus* species on the other. The second
216 PC explains 19.7% of the variance and separates several species, especially *P. grandidentata*
217 from the rest (Fig. 2a). We performed an additional PCA analysis with the recently diverged
218 aspens only (supplementary material, Fig. S4). There, the first principal component (26.2%
219 variance explained) separates *P. tremuloides* from the other three species; the second principal
220 component (21.7% variance explained) separates *P. tremula* from *P. davidiana* and *P.*
221 *rotundifolia*.

222 Neighbor-joining (Fig. 2b) and Admixture (Alexander and Lange 2011) analyses (Fig. 2c)
223 based on all SNPs identified seven genetic groups, which were consistent with previously
224 identified species boundaries based on phylogenomic analysis (Shang et al, 2020). Additionally,
225 Admixture also indicated a potential introgression between the subtropical species *P. adenopoda*
226 and two recently diverged species, *P. davidiana* and *P. rotundifolia* (Fig. 1c and supplementary
227 material, Fig. S5). Identity-by-descent (IBD) analyses (Fig. 2d) also identified seven reliable
228 clusters, corresponding to the same species boundaries as before, but further pinpointed some
229 shared haplotypes among the aspen species *P. davidiana*, *P. rotundifolia* and *P. tremula*,
230 indicating recent introgression among these species. The IBD results also provide support for
231 extensive introgression between species with overlapping distribution, such as *P. alba* and *P.*
232 *tremula*, or *P. grandidentata* and *P. tremuloides*. These results are consistent with a scenario of
233 divergence with ongoing gene flow, either due to isolation-with-migration or secondary contact,
234 maintained even after substantial divergence times (net divergence d_a: 0.023 for *P. alba* - *P.*
235 *tremula*; d_a: 0.025 for *P. tremuloides* - *P. grandidentata*). For these two species pairs, net
236 divergence values are indeed larger than the upper boundary for the 'grey zone of speciation'
237 reported by Roux, et al. (2016) for both vertebrate and invertebrate animals (d_a from 0.005 to

238 0.02). Future investigations on the plant grey zone based on a large number of taxa are needed to
239 get more general insights about plant diversification.



240

241 ***Fig. 2.*** *Genetic structure within Populus (poplar and aspen) accessions investigated.* ***(a)*** *Principal*
242 *Component Analysis (PCA) based on all SNPs for seven Populus species. Colored symbols represent*
243 *different species according to legend.* ***(b)*** *Unrooted neighbour-joining tree of all SNPs.* ***(c)*** *Estimated*
244 *membership of each individual's genome for K = 5 to K = 8 as estimated by Admixture (best K = 7).* ***(d)***
245 *Identity by descent (IBD) analysis for seven Populus species. Heatmap colours represent the shared*
246 *haplotype length between species. Species abbreviations: pade, P. adenopoda; palb, P. alba; prot, P.*
247 *rotundifolia; pgra, P. grandidentata; ptmu, P. tremuloides; ptma, P. tremula collected from China;*
248 *ptmae, P. tremula collected from Europe; pdav, P. davidiana collected from China.*

249 **Between species variability in demographic trajectories**

250 We then used TreeMix (Pickrell and Pritchard 2012) to recover the phylogenetic
251 relationship among species and infer putative migration events in *Populus* (Fig. 3a). The tree
252 topology was consistent with phylogenetic relationships found in the previous study (Shang, et
253 al. 2020). A drift-only model of divergence (i.e., without migration edges) already explained
254 95.8% of the total variance. Adding one single migration edge allowed us to account for 98.9%
255 of the total variance (supplementary material, Fig. S6). This event was inferred from *P.*
256 *grandidentata* to *P. tremuloides* and is consistent with previous reports of extensive
257 hybridization and introgression between these two species (Deacon, et al. 2019). Adding a
258 second migration edge from *P. adenopoda* to *P. rotundifolia* allowed us to account for 99.6% of
259 the total variance. By adding more migration edges, the variance explained increases by less than
260 0.1%, which was considered as too marginal (supplementary material, Fig. S6). Therefore, we
261 considered the bifurcating tree with two migration events as the best scenario explaining the
262 historical relationships among these *Populus* species.

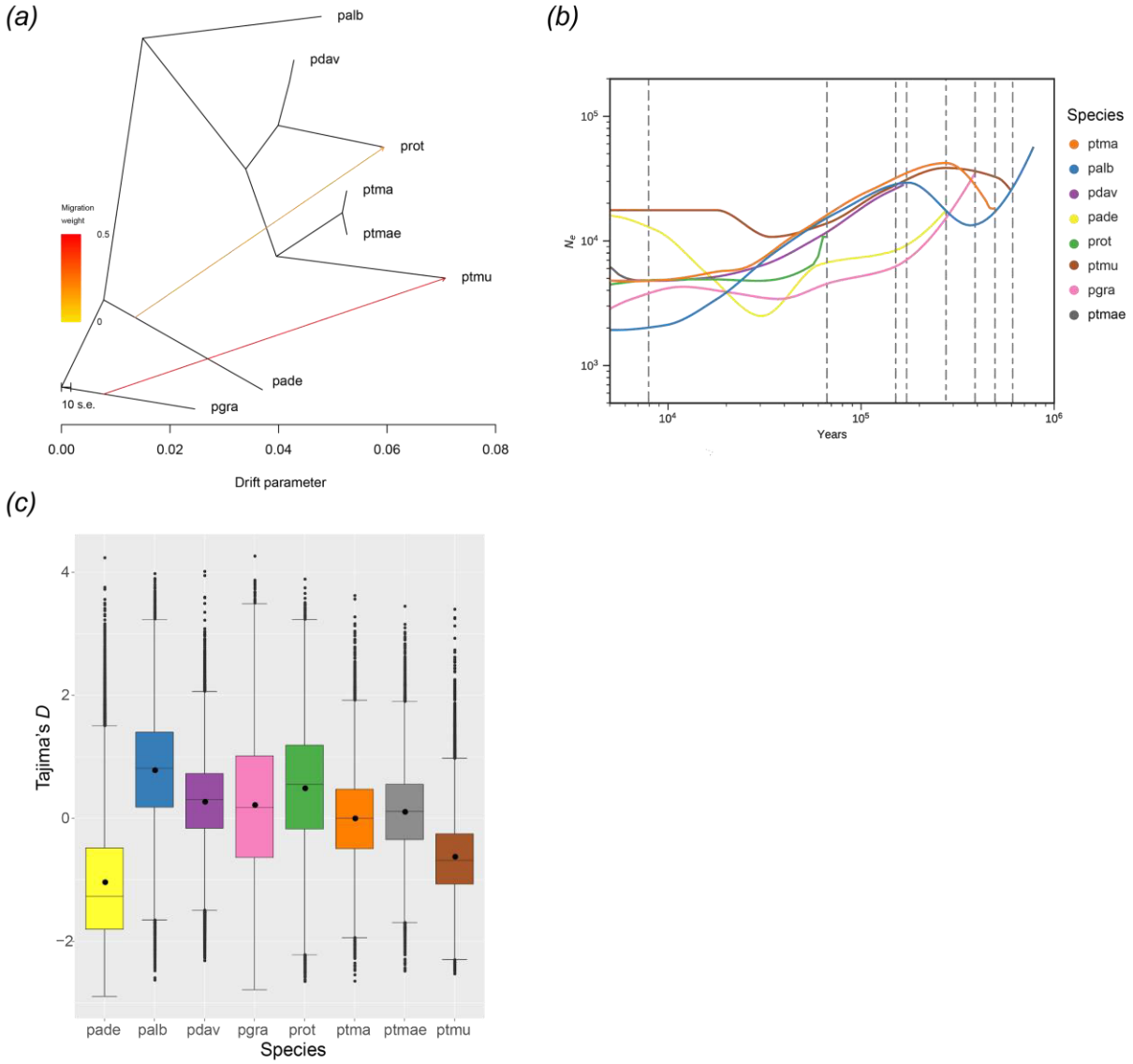


Fig. 3. *Demographic history of Populus species.* ***(a)*** *The maximum likelihood tree inferred by TreeMix under a strictly bifurcating model with two migration events.* ***(b)*** *Changes in effective population size through time inferred by SMC++.* ***(c)*** *The distribution of Tajima's D values calculated over all 10-kb windows.*

To infer historical changes in effective population sizes for the investigated *Populus* species we used SMC++ (Terhorst, et al. 2017), a method which only requires unphased sequence data. These inferences (Fig. 3b) support that all species have experienced population size reductions at least over the last 100,000 years, except for the subtropical species *P. adenopoda* and the North

American species *P. tremuloides*. These latter two species have undergone population size expansions from about 30,000 years ago, which seem consistent with their wide present-day distribution areas in South China and North America, respectively (Eckenwalder 1996). Negative mean Tajima's *D* over all non-overlapping 10kb sliding windows spanning the whole genome are consistent with these recent expansions (*P. adenopoda*: -0.97, *P. tremuloides*: -0.61; Fig. 3c). For these two species, the folded-SFS also showed a strong excess of rare variants (*P. adenopoda*: 2.4%, *P. tremuloides*: 5.2%; supplementary material, Fig. S7), which explains the negative Tajima's *D* values (Fig. 3c). Further evidence for different demographic histories among *Populus* species was provided by estimates of genome-wide π and population-scaled recombination rate (ρ) based on 10kb windows. Both π and ρ were found to vary greatly between species (Fig. 4c-f). Since both π and ρ scale with *Ne,* the high π and ρ values of *P. tremuloides* (Fig. 4c-f) are consistent with larger effective population sizes previously inferred for this species (Fig. 3b). Interestingly, for *P. grandidentata* we found relatively low π but high ρ (Fig. 4c-d) compared to other species and there is no significant correlation between these two parameters (Fig. 4f). However, a significantly negative correlation between gene density and π was evident in this species. Besides, we also detected significantly negative correlations between gene density and π in all other *Populus* species (supplementary material, Fig. S8). This result provides evidence for the important role of background selection in reducing genetic diversity and indicates gene density is a suitable indicator of selection.

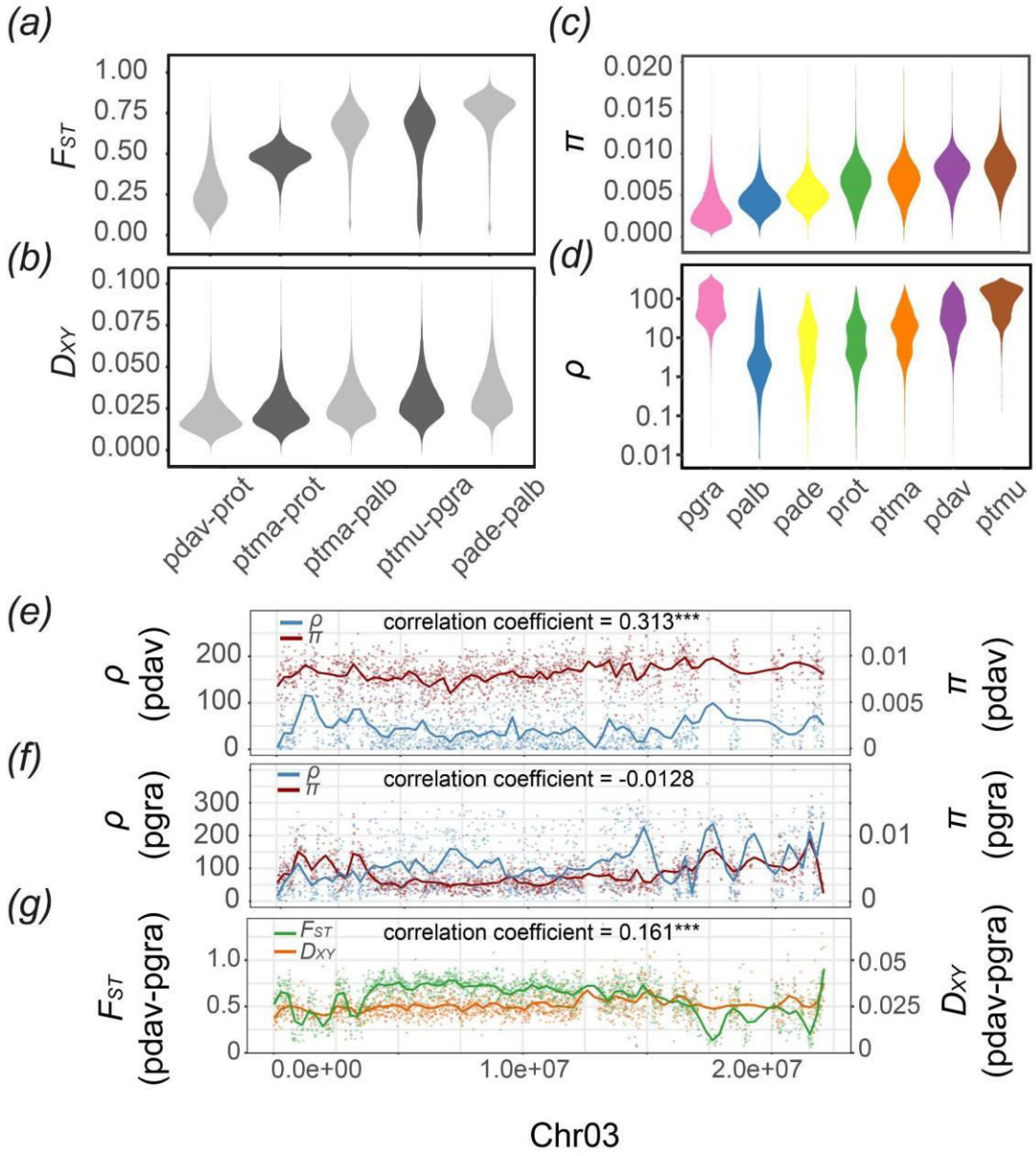


Fig. 4. *Summary violin plots of F_{ST}, D_{XY}, π and ρ calculated across 10kb windows.* ***(a-b)*** *Violin plots of F_{ST} and D_{XY} for five species pairs as representatives of the different stages across the speciation continuum investigated in this study.* ***(c-d)*** *Violin plots of π and ρ for seven Populus species. Different colours represent different species (see Fig. 2 for details). Note that the unit of ρ is 4Ner and that ρ is log-scaled in the* ***(d)*** *panel.* ***(e-g)*** *Landscapes of π, ρ, F_{ST} and D_{XY} in P. davidiana and P. grandidentata on chromosome 3. Correlation analysis for π and ρ in P.davidiana show significantly positive correlation on chromosome 3, while no correlations in P. grandidentata. P values less than 0.001 are summarized with three asterisks.*

301 **Conserved genomic landscapes across the continuum of divergence**

302 To have a knowledge of genome-wide variation in *Populus*, we calculated genome-wide
303 patterns of divergence, nucleotide diversity, gene density and recombination for all seven
304 *Populus* species and 21 possible species pairs across non-overlapping 10kb windows spanning
305 the whole genome. We then selected and violin plotted genomic variation for five species pairs
306 as representatives of the different stages across the speciation continuum (Fig. 4). Mean F_{ST}
307 varied from 0.27 between *P. davidiana* - *P. rotundifolia* to 0.73 between *P. adenopoda* - *P. alba*
308 (Fig. 4a) and D_{XY} ranged from 0.021 (*P. davidiana* - *P. rotundifolia*) to 0.035 (*P. adenopoda* - *P.*
309 *alba*) (Fig. 4b). The average π varied from 0.0035 in *P. grandidentata* to 0.0084 in *P.*
310 *tremuloides* (Fig. 4c). We inferred lower population-scaled recombination rates in *P. alba* and *P.*
311 *adenopoda* than in other species, which is consistent with the slower linkage disequilibrium (LD)
312 decays and lower genetic diversity in these two species (Fig. 4d and supplementary material, Fig.
313 S9).

314 The degree of correlation of both the relative and absolute divergence landscapes between
315 pairs of species supports a highly conserved pattern among the five investigated species pairs
316 (Fig. 5a-b). The correlations of F_{ST} between independent species pairs become stronger with
317 advancing differentiation. This may be the case that the effect of linked selection accumulates
318 with differentiation advance. The correlation coefficients for the pairwise comparisons of the
319 landscape of the nucleotide diversity π vary substantially, from 0.16 (*P. tremula* versus *P.*
320 *grandidentata*) to 0.52 (*P. rotundifolia* versus *P. davidiana*), generally with the phylogenetic
321 distance, with strongest correlation coefficients for the phylogenetically closest pair of species:
322 *P. rotundifolia* and *P. davidiana* (Fig. 5c). Pairwise comparisons of the local recombination rates
323 inferred independently for all species also revealed only positive correlations (Fig. 5d), with the
324 highest positive correlation coefficient of ρ again observed between the two closest related
325 species, *P. davidiana* and *P. rotundifolia* (0.47), while the lowest correlation was observed for *P.*
326 *davidiana* and *P. grandidentata* (0.08). Most of the lower values (correlation coefficients < 0.2)
327 were found when comparing *P. grandidentata* with the other species, suggesting an increased
328 uniqueness in the recombination landscape in this species. Interestingly, correlations of π were
329 higher than those of ρ, indicating that not only the recombination rate variation shapes nucleotide
330 diversity. Overall, landscapes of genetic diversity, divergence and recombination rate remain

relatively stable across different species or species pairs (Fig. 5), which implies relatively conserved genomic features across all species. This phenomenon has also been observed in some other plant and animal models (Nosil and Feder 2012; Renaut, et al. 2014; Burri, et al. 2015; Wang, et al. 2020). The result is consistent with the important role of linked selection, and therefore the recombination landscape, in shaping genome-wide genetic diversity levels.

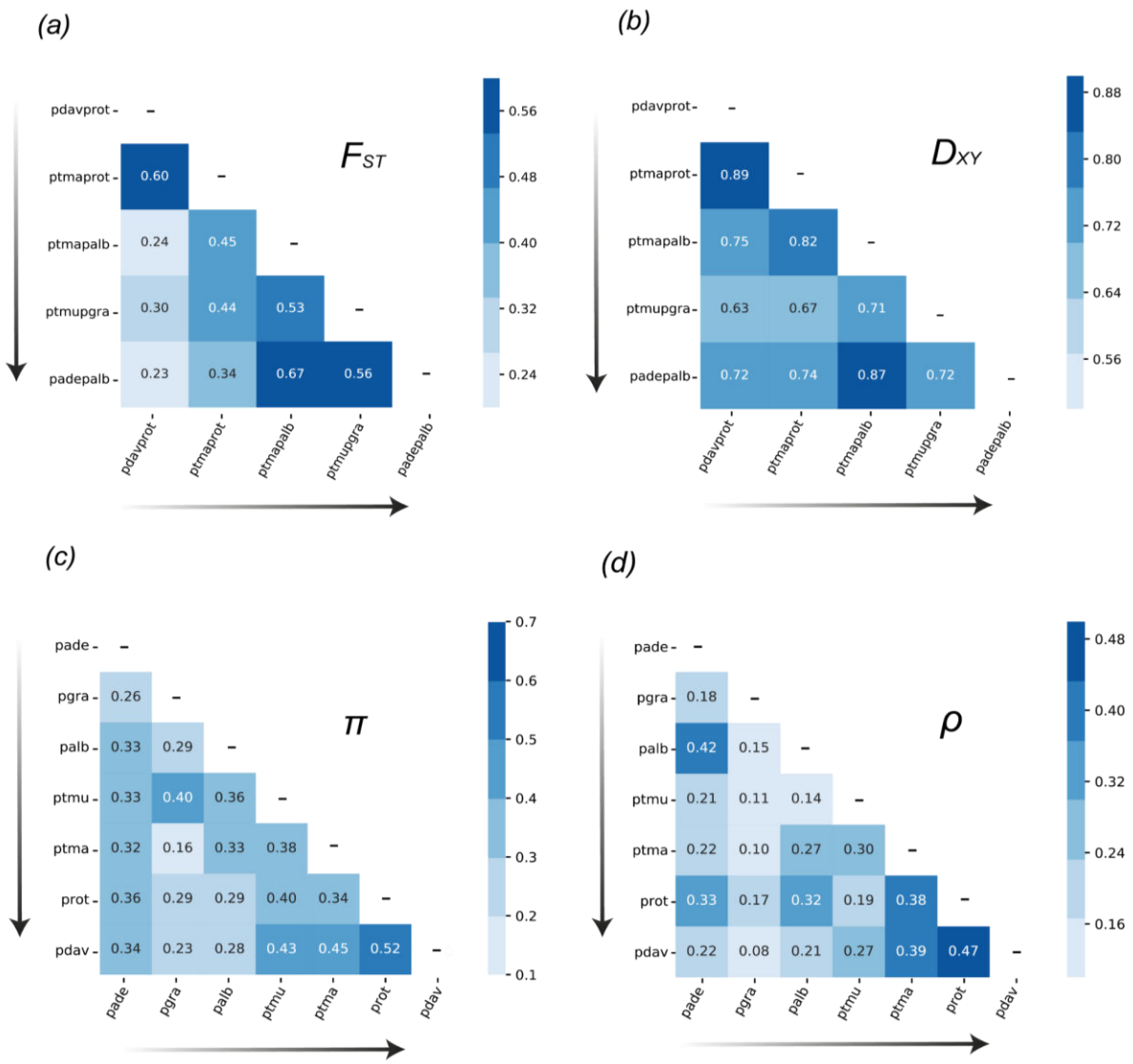


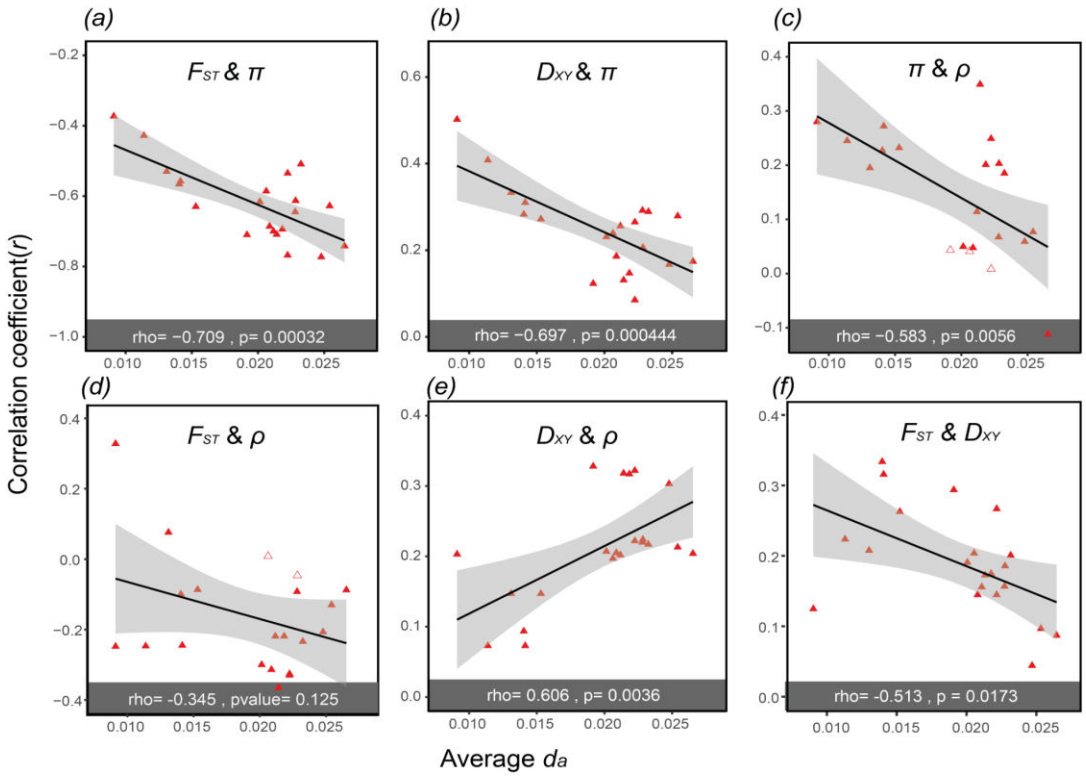
Fig. 5. *Correlations analyses of within-species diversity or among-species divergence landscapes* ***(a-b)*** *Correlation coefficients of F_{ST} or D_{XY} between species pairs. The species pairs are ordered across the speciation continuum (arrows).* ***(c-d)*** *Correlation coefficients of π or ρ between species. The order of the species is based on the order of species divergence from the root. All the values are significantly positively correlated ($p < 0.001$), respectively.*

342 **Correlated patterns of genome-wide variation across speciation continuum**

343 We also investigated the correlations between within-species nucleotide diversity and relative
344 divergence levels, or absolute divergence and recombination rate. Under the influence of linked
345 selection, a lower genetic diversity at high gene density or across low recombination regions is
346 expected. Consistent with this expectation, significantly positive correlations between π and ρ or
347 D_{XY} (Fig. 6b, c, e) and negative correlations between F_{ST} and π (Fig. 6a) were observed across the
348 continuum of divergence represented by all the 21 species pairs. We then used the level of
349 genetic distance between each species pair (d_a: D_{XY} – mean π) as a proxy for the divergence time
350 and estimated how the correlations between genome-wide diversity or divergence and
351 recombination change across the speciation continuum. Under the predictions of background
352 selection, we would expect negative correlations between F_{ST} and π or ρ and the relationships
353 become stronger as d_a increases. Besides, the positive correlations between D_{XY} and π should be
354 highly correlated, because D_{XY} is sensitive to ancestral polymorphism. The correlations between
355 F_{ST} and D_{XY} are expected to be negatively correlated under the influence of recurrent background
356 selection in the common ancestor.

357 In our analysis, we found that the negative relationships between F_{ST} and π or ρ become
358 stronger as d_a increases, which is consistent with the expectation under background selection
359 (Fig. 6a, d). Similar investigations for π and D_{XY} showed significantly positive correlations while
360 the trend became weaker as divergence increases (Fig. 6b). The trend is not consistent with the
361 general hypothesis that correlations should still be highly correlated as divergence increases. We
362 also recovered a strong positive correlation between π and ρ (Fig. 6c), and a similar trend was
363 found as for the investigation of π and D_{XY}. Correlations between D_{XY} and ρ were significantly
364 positive across the entire speciation continuum, and we do observe that these correlations tend to
365 become stronger as divergence increases across the speciation continuum (Fig. 6e). Positive
366 correlations between F_{ST} and D_{XY} were found across all species pairs, even though the
367 relationships become weaker as divergence increases (Fig. 6f). Thus the high F_{ST} may be due to
368 elevated D_{XY}, which is not consistent with the expectations of lower D_{XY} under background
369 selection. The observed patterns (Fig. 6b, f) differ from expectations under a scenario with
370 background selection as the sole factor shaping the heterogeneous landscape of differentiation,
371 indicating that there should be additional evolutionary factors that contribute to it (Burri 2017a).

372 Consistent with our findings, studies in monkey flowers, threespine stickleback, and avian
373 species also confirmed that background selection alone is too subtle to make a strong
374 contribution to the conserved genomic patterns (Irwin, et al. 2018; Stankowski, et al. 2019;
375 Rennison, et al. 2020). Indeed, these studies suggest either adaptive introgression or shared
376 standing variation play a major role in generating the similar patterns of genomic differentiation.



377

378 ***Fig. 6.*** *Correlations between variables for all species comparisons (red triangles) plotted against the*
379 *average d_a, used here as a measure of divergence time. The filled triangles indicate when the correlation*
380 *coefficients are significant ($p < 0.01$). The upper panels show how the relationships between average π*
381 *and* ***(a)*** *F_{ST}, or* ***(b)*** *D_{XY} vary for pairs of species with increasing divergence time; and between average π*
382 *and ρ for all species pairs investigated. The lower panels* ***(d)*** *and* ***(e)*** *show the relationships between ρ*
383 *and F_{ST} or D_{XY}, respectively. The last panel* ***(f)*** *shows the relationships between F_{ST} and D_{XY}.*

384

385 In our study, the negative correlations between π and F_{ST} (Fig. 6a) and the positive correlations
386 between F_{ST} and D_{XY} (Fig. 6f) in all 21 species pairs indicates high F_{ST} could be caused by the high
387 D_{XY}. As D_{XY} is sensitive to ancestral polymorphism, the high D_{XY} most likely due to the presence

388 of diverged haplotypes in ancestral populations before speciation or resulted from limited gene
389 flow in regions under selection, which can be caused by ecological adaptation.

390

391 **Scenarios of genomic patterns of differentiation**

392 Based on F_{ST}, D_{XY}, and π, we summarized four scenarios to help explain the genomic divergence
393 patterns observed (supplementary material, Fig. S10) following the strategy developed by Han et
394 al. (2017) and Irwin et al. (2018). The major differences of these scenarios are the role and extent
395 of gene flow, and the type of selection acting to shape genome-wide genetic differentiation. Our
396 results show a heterogeneous distribution of the four scenarios along the genome for all five
397 species pairs (supplementary material, Fig. S11-S15 and Table S2). A great number of regions
398 (74.3%-78.7%) in all five species pairs fit a scenario of "allopatric selection" (scenario b in
399 supplementary material, Fig. S11-S15 and Table S2), i.e., regions exhibiting an elevated F_{ST} (tail
400 of the distribution) but a moderate D_{XY} (in the middle of the distribution). Such a signature is
401 expected to be consistent with recent footprints of positive or background selection on genomic
402 differentiation. Genomic regions fitting the scenario of 'balancing selection' (scenario d in
403 supplementary material, Fig. S10) are the second most frequent for all investigated species pairs.
404 This scenario is characterized by an elevated D_{XY} but a low F_{ST} implying the action of balancing
405 selection in shaping the heterogeneous landscape of divergence. In addition, we found support
406 for reproductive barriers (scenario a in supplementary material, Fig. S11-S15 and Table S2) in all
407 five species pairs, suggesting that genomic heterogeneity in the levels of gene flow due to the
408 species barriers play a role in shaping genomic differentiation landscapes. Interestingly, this
409 result holds true for all species pairs we investigated, *i.e.,* regardless of the level of gene flow or
410 the stage along the *Populus* speciation continuum. Indeed, the regions with high F_{ST} and D_{XY}
411 found in the highly diverged species pair (*P. adenopoda* - *P. alba*) could not be caused by
412 adaptive introgression but may be due to the shared ancestral polymorphisms, as no gene flow
413 was found between these two species. In contrast, at the early stage of divergence, barriers to
414 gene flow may play an important role in genomic heterogeneous divergence, as the influence of
415 background selection is still very weak at this stage. However, it is too naive to use the four
416 scenarios to explain the empirical data, as the interplay of demographic history, selection, gene

417 flow and recombination may obscure the signature of genomic patterns generated by one
418 evolutionary factor.

419

420 Conclusions

421 In this study, we reconstructed the demographic history and investigated the evolution of the
422 genomic landscape of diversity and divergence across a speciation continuum of eight closely
423 related species of *Populus* section *Populus*. By investigating evolution of diversity and
424 differentiation landscapes across this speciation continuum, we provided, to our knowledge, one
425 of the most ambitious case studies in terms of the number of species pairs analyzed (see also
426 Stankowski, et al. 2019). Our correlation analyses support this role of linked selection,
427 recombination rate and gene density in shaping the empirical patterns of genomic differentiation,
428 but the contribution of linked selection alone is not sufficient in generating these genomic
429 landscapes. Instead, the positive correlations between F_{ST} and D_{XY} in all species pairs indicates
430 diverged ancient polymorphism must play a role. The study also confirmed the importance of
431 gene flow in this system, through extensive introgression among species with parapatric
432 distributions, despite a high level of divergence among the most divergent hybridizing species
433 (d_a = 0.025). Overall, this work provides a prime example of the strategic importance of
434 investigating these genomic patterns on a large number of species to better access the temporal
435 evolution of the genomic landscapes of diversity and differentiation. In the future, more studies
436 on multiple species pairs across the speciation continuum are needed to discuss the interactions
437 of several evolutionary factors on the heterogeneous genomic landscapes of differentiation.

438

439 Materials and method

440 Sampling, sequencing and read processing

441 Two hundred and one samples were collected from eight species of *Populus* section *Populus* in
442 Eurasia and North America (supplemental material, Fig. S1 and Table S1). The leaves were dried
443 in silica gel first and were then used for genomic DNA extraction with Plant DNeasy Mini Kit

444 (Qiagen, Germany). To increase the purity of total DNA, we used the NucleoSpin gDNA Clean-
445 up kit (Macherey-Nagel, Germany). Whole genome resequencing was performed with 2 x 150bp
446 paired-end sequencing technology on Illumina HiSeq 3000 sequencer at the Institute of Genetics,
447 University of Bern, Switzerland.

448 All raw sequencing reads were mapped to the *P. tremula* 2.0 reference genome (Schiffthaler,
449 et al. 2019) using BWA-MEM, as implemented in bwa v0.7.10 (Li 2013). Samtools v1.3.1 was
450 used to remove alignments with mapping quality below 20 (Li, et al. 2019). Read-group
451 information including library, lane, sample identity and duplicates were recorded using Picard
452 v2.5 (http://broadinstitute.github.io/picard/). Sequencing reads around insertions and deletions
453 (i.e., indels) were realigned using RealignerTargetCreator and IndelRealigner in the Genome
454 Analysis Toolkit (GATK v3.6) (DePristo, et al. 2011). We used the GATK HaplotypeCaller and
455 then GenotypeGVCFs for individual SNP calling and for joint genotyping, respectively, among
456 all samples using default parameters. Finally, we performed several filtering steps using GATK
457 to retain only high-quality SNPs: (1) 'QD' < 2.0; (2) 'FS > 60.0'; (3) 'MQ < 40.0'; (4)
458 'ReadPosRankSum < -8.0'; (5) 'SOR > 4.0'; (6) 'MQRankSum < -12.5'.Moreover, we also
459 excluded loci with missing data of more than 30% and discarded two individuals with very low
460 depth of coverage (< 10), as calculated using VCFtools v0.1.15
461 (http://vcftools.sourceforge.net/man_latest.html). The scripts for SNP calling are available at
462 https://github.com/Huiying123/Populus_speciation/tree/2fff35298813b6be7a0ef69d1558c6cd91
463 80ef5c/population_snp_calling.

464 **Family relatedness and population structure analysis**

465 To avoid pseudoreplication due to the inclusion of clone mates, we estimated kinship coefficients
466 using the KING toolset for family relationship inference based on pairwise comparisons of SNP
467 data (http://people.virginia.edu/~wc9c/KING/manual.html). The software classifies pairwise
468 relationships into four categories according to the estimated kinship coefficient: a negative
469 kinship coefficient estimation indicates the lack of a close relationship. Estimated kinship
470 coefficients higher than >0.354 correspond to duplicates, while coefficients ranging from [0.177,
471 0.354], [0.0884, 0.177] and [0.0442, 0.0884] correspond to 1st-degree, 2nd-degree, and 3rd-degree
472 relationships, respectively.

473 After discarding individuals with low depth and high inbreeding coefficient (F > 0.9, *P.*
474 *qiongdaoensis*) as well as clones identified with the KING toolset, we used VCFtools v0.1.15
475 (http://vcftools.sourceforge.net/man_latest.html) to calculate the mean depth of coverage and
476 heterozygosity for each individual. Based on the results of these analyses, we concluded that all
477 *P. qiongdaoensis* individuals are probably clone mates. As a consequence, only seven species
478 were used for the subsequent analyses.

479 We used PLINK (Purcell, et al. 2007) to generate a variance-standardized relationship matrix
480 for principal components analysis (PCA) and a distance matrix to build a neighbor joining tree
481 (NJ-tree). The NJ tree was constructed using PHYLIP v.3.696 (https://evolution.genetics
482 .washington.edu/phylip.html). Both PCA and NJ-tree analyses were performed based on the full
483 set of SNPs. In addition, we used ADMIXTURE v1.3 for the maximum-likelihood estimation of
484 individual ancestries (Alexander and Lange 2011). First, we generated the input file from a VCF
485 containing unlinked SNPs. Besides, sites with missing data more than 30% have been filtered
486 out. This analysis was run for *K* from 1 to 10, and the estimated parameter standard errors were
487 generated using 200 bootstrap replicates. The best *K* is the one with the lowest cross-validation
488 error. We also performed an IBD blocks analysis using BEAGLE v5.1 (Browning and Browning
489 2013) to detect identity-by-descent segments between pairs of species. The parameters we used
490 are: window=100,000; overlap=10,000; ibdtrim=100; ibdlod=10.

491 **Demographic trajectory reconstruction**

492 To reconstruct the demographic history of *Populus* species, we first inferred the history of
493 species splits and mixture based on genome wide allele frequency data using TreeMix v1.13
494 (Pickrell and Pritchard 2012). We removed the sites with missing data and performed linkage
495 pruning. We then ran TreeMix implementing a default bootstrap and a block size of 500 SNPs (-
496 k=500). The best migration edge was evaluated according to the greatest increase of total
497 variation explained. The plotting R functions of the Treemix suite were then used to visualize the
498 results. In addition, we used SMC++ to estimate historical changes in effective population size
499 (Terhorst, et al. 2017). The split time for species pairs was estimated based on the joint
500 frequency spectrum for both species. Due to widespread vegetative reproduction in some poplar
501 species, we adjusted the generation time to 20 years rather than the generally thought 15 years

502 (Wang, et al. 2016); a mutation rate of 2.5×10^{-9} per site per year was also assumed in order to
503 scale results to real time (Macaya-Sanz, et al. 2012).

504 **Nucleotide diversity and divergence estimates**

505 Nucleotide diversity, as well as relative and absolute divergence estimates were calculated based
506 on genotype likelihoods. We used ANGSD v0.93
507 (http://www.popgen.dk/angsd/index.php/ANGSD) to estimate statistical parameters from the
508 BAM files for all *Populus* species. First, we used '*dosaf 1*' to calculate site allele frequency
509 likelihood and then used '*realSFS*' to estimate folded site frequency spectra (SFS). Genome-
510 wide diversity and Tajima's *D* were calculated with the parameter '*-doThetas 1*' in ANGSD
511 based on the folded SFS of each species. We selected two population genomic statistics to
512 estimate divergence F_{ST} and D_{XY}. We estimated SFS for each population separately and then used
513 it as a prior to generate a 2D-SFS for each species pair. F_{ST} of each species pair were estimated
514 with the parameters '*realSFS fst*' based on the 2D-SFS. Finally, we averaged the F_{ST} value of
515 sites over 10kb windows. To estimate D_{XY}, we used ANGSD to calculate minor allele
516 frequencies with the parameters '*-GL 1 -doMaf 1 -only_proper_pairs 1 -uniqueOnly 1 -*
517 *remove_bads 1 -C 50 -minMapQ 30 -minQ 20 -minInd 4 -SNP_pval 1e-3 -skipTriallelic 1 -*
518 *doMajorMinor 5*' and then computed D_{XY} as follows: $D_{XY} = A_1*B_2+A_2*B_1$, with A and B being
519 the allele frequencies of A and B, and 1 and 2 being the two populations. We averaged D_{XY}
520 across 10kb windows.

521 To examine the relationships among diversity, differentiation, and recombination landscapes,
522 we estimated Pearson's correlation coefficient between pairs of these statistics. These tests were
523 performed across genomic windows for the 21 possible *Populus* species pairs. Finally, we used
524 d_a (D_{XY} – mean π) as a measure of divergence time.

525 **Population-scale recombination rate and linkage disequilibrium**

526 We estimated population scaled recombination rate with FastEPRR (Gao, et al. 2016) for each
527 species separately. To eliminate the effect of sample size on the estimation of recombination rate,
528 we downsampled to 13 randomly selected individuals for each species, corresponding to the
529 number of individuals available for *Populus davidiana* (pdav). First, we filtered all missing and

530 non-biallelic sites with VCFtools and then phased the data with the parameters "*impute=true*
531 *nthreads=20 window=10,000 overlap=1,000 gprobs=false*" in Beagle v5.1 (Browning and
532 Browning 2013). Finally, we ran FastEPRR v2.0 (Gao, et al. 2016) with a window size of 10kb.
533 After getting the results, we estimated the correlation between recombination rate of one species
534 to another. To evaluate LD decay, we used PLINK (Purcell, et al. 2007) to obtain LD statistics
535 for each species. Parameters were set as follows: '*--maf 0.1 --r^2 gz --ld-window-kb 500 --ld-*
536 *window 99999 --ld-window-r^2 0*'. LD decay was finally plotted in R.

537 **Divergent regions of exceptional differentiation**

538 We further investigated genomic differentiation landscapes across multiple species pairs along
539 the speciation continuum and identified which evolutionary factors contribute to genomic
540 differentiation. We reported genomic regions showing elevated or decreased values of F_{ST}, D_{XY}
541 and π across 10kb windows. Windows falling above the top 5% or below the bottom 5% of F_{ST}
542 and D_{XY} were considered. For these specific windows, we then classified them following the four
543 models of divergence suggested by Irwin *et al*. 2018 and Han *et al* 2017. These four models
544 differ in the role of gene flow (with or without), or the type of selection (selective sweep,
545 background selection or balancing selection).

546

547

548 ## Acknowledgements

549 We thank members of Jianquan Liu's lab for collecting samples and to Chris Cole for providing
550 samples from *P. grandidentata*. Michael Barfuss and Elfi Grasserbauer provided assistance in
551 the laboratory. Sequencing was performed at the Next Generation Sequencing Platform of the
552 University of Berne and at the National Genomics Infrastructure (NGI) of Science for Life
553 Laboratory, Stockholm. The Vienna Scientific Cluster (VSC) and the Swedish National
554 Infrastructure for Computing (SNIC) at Uppsala Multidisciplinary Centre for Advanced
555 Computational Science (UPPMAX) provided access to computational resources. We are also

grateful to members of the PopGen Vienna graduate school for helpful discussions. This manuscript is dedicated to the memory of our friend and colleague Prof. Christian Lexer.

Funding

This work was supported by a fellowship from the China Scholarship Council (CSC) to Huiying Shang, Swiss National Science Foundation (SNF) grant no.31003A_149306 to Christian Lexer, doctoral programme grant W1225-B20 to a faculty team including Christian Lexer, and the University of Vienna.

Conflict of interest disclosure

None of the authors have a conflict of interest to declare regarding the publication of this manuscript.

Data, script and code availability

The raw read data have been deposited with links to BioProject accession numbers PRJNA299390, PRJNA612655, PRJNA720790, and PRJNA297202 in the NCBI BioProject database (https://www.ncbi.nlm.nih.gov/bioproject/). All the scripts used for the analysis are available on: https://github.com/Huiying123/Populus_speciation/tree/2fff35298813b6be7a0ef69d1558c6cd9180ef5c/population_snp_calling.

Supplementary information

Supplementary information is available in the "Supplementary material" section of the bioRxiv page of the article, https://doi.org/10.1101/2021.08.26.457771.

580 ## Author contributions

581 Study conceived and designed by Christian Lexer and Huiying Shang. Laboratory work
582 conducted by Huiying Shang. Population genomic data analysis by Huiying Shang.
583 Interpretation of the results was undertaken by Huiying Shang, Christian Lexer, Martha Rendón-
584 Anaya, Pär K. Ingvarsson, Thibault Leroy, Jaqueline Hess, David Field, Claus Vogl and Ovidiu
585 Paun. The manuscript was drafted by Huiying Shang, with help from Thibault Leroy and Ovidiu
586 Paun.

587

588 ## References

589 Alexander DH, Lange K. 2011. Enhancements to the ADMIXTURE algorithm for individual
590 ancestry estimation. BMC Bioinformatics 12:246.
591 Browning BL, Browning SR. 2013. Improving the accuracy and efficiency of identity-by-descent
592 detection in population data. Genetics 194:459-471.
593 Burri R. 2017a. Dissecting differentiation landscapes: a linked selection's perspective. J Evol
594 Biol 30:1501-1505.
595 Burri R. 2017b. Linked selection, demography and the evolution of correlated genomic
596 landscapes in birds and beyond. Mol Ecol 26:3853-3856.
597 Burri R, Nater A, Kawakami T, Mugal CF, Olason PI, Smeds L, Suh A, Dutoit L, Bureš S,
598 Garamszegi LZ, et al. 2015. Linked selection and recombination rate variation drive the
599 evolution of the genomic landscape of differentiation across the speciation continuum of
600 *Ficedula* flycatchers. Genome Res 25:1656-1665.
601 Campagna L, Gronau I, Silveira LF, Siepel A, Lovette IJ. 2015. Distinguishing noise from signal
602 in patterns of genomic divergence in a highly polymorphic avian radiation. Mol Ecol
603 24:4238-4251.
604 Charlesworth B. 1998. Measures of divergence between populations and the effect of forces that
605 reduce variability. Mol Biol Evol 15:538-543.
606 Charlesworth B, Campos JL. 2014. The relations between recombination rate and patterns of
607 molecular variation and evolution in *Drosophila*. Annu Rev Genet 48:383-403.

608 Charlesworth B, Morgan MT, Charlesworth D. 1993. The effect of deleterious mutations on
609 neutral molecular variation. Genetics 134:1289-1303.
610 Corbett-Detig RB, Hartl DL, Sackton TB. 2015. Natural selection constrains neutral diversity
611 across a wide range of species. PLoS Biol 13:e1002112.
612 Cruickshank TE, Hahn MW. 2014. Reanalysis suggests that genomic islands of speciation are
613 due to reduced diversity, not reduced gene flow. Mol Ecol 23:3133-3157.
614 Deacon NJ, Grossman JJ, Cavender-Bares J. 2019. Drought and freezing vulnerability of the
615 isolated hybrid aspen *Populus x smithii* relative to its parental species, *P. tremuloides* and *P.*
616 *grandidentata*. Ecol Evol 9:8062-8074.
617 DePristo MA, Banks E, Poplin R, Garimella KV, Maguire JR, Hartl C, Philippakis AA, del
618 Angel G, Rivas MA, Hanna M, et al. 2011. A framework for variation discovery and
619 genotyping using next-generation DNA sequencing data. Nat Genet 43:491-498.
620 Eckenwalder J. 1996. Systematics and evolution of *Populus*. U: Stettler, RF, Bradshaw, HD,
621 Heilman, PE, Hinckley, TM, eds.(1996): Biology of *Populus* and Its Implications for
622 Management and Conservation. In: NRC Research Press, Ottawa, Ontario, Canada.
623 Ellegren H, Smeds L, Burri R, Olason PI, Backström N, Kawakami T, Künstner A, Mäkinen H,
624 Nadachowska-Brzyska K, Qvarnström A, et al. 2012. The genomic landscape of species
625 divergence in *Ficedula* flycatchers. Nature 491:756-760.
626 Gagnaire PA, Lamy JB, Cornette F, Heurtebise S, Dégremont L, Flahauw E, Boudry P, Bierne
627 N, Lapègue S. 2018. Analysis of Genome-Wide Differentiation between Native and
628 Introduced Populations of the Cupped Oysters *Crassostrea gigas* and *Crassostrea angulata*.
629 Genome Biol Evol 10:2518-2534.
630 Gao F, Ming C, Hu W, Li H. 2016. New Software for the Fast Estimation of Population
631 Recombination Rates (FastEPRR) in the Genomic Era. G3 (Bethesda) 6:1563-1571.
632 Han F, Lamichhaney S, Grant BR, Grant PR, Andersson L, Webster MT. 2017. Gene flow,
633 ancient polymorphism, and ecological adaptation shape the genomic landscape of divergence
634 among Darwin's finches. Genome Res 27:1004-1015.
635 Harrison RG, Larson EL. 2016. Heterogeneous genome divergence, differential introgression,
636 and the origin and structure of hybrid zones. Mol Ecol 25:2454-2466.
637 Henderson EC, Brelsford A. 2020. Genomic differentiation across the speciation continuum in
638 three hummingbird species pairs. BMC Evol Biol 20:113.

639 Irwin DE, Milá B, Toews DPL, Brelsford A, Kenyon HL, Porter AN, Grossen C, Delmore KE,
640 Alcaide M, Irwin JH. 2018. A comparison of genomic islands of differentiation across three
641 young avian species pairs. Mol Ecol 27:4839-4855.
642 Jansson S, Bhalerao R, Groover A. 2010. Genetics and genomics of *Populus*: Springer.
643 Lamichhaney S, Berglund J, Almén MS, Maqbool K, Grabherr M, Martinez-Barrio A,
644 Promerová M, Rubin CJ, Wang C, Zamani N, et al. 2015. Evolution of Darwin's finches and
645 their beaks revealed by genome sequencing. Nature 518:371-375.
646 Leroy T, Rougemont Q, Dupouey JL, Bodénès C, Lalanne C, Belser C, Labadie K, Le Provost
647 G, Aury JM, Kremer A, et al. 2020. Massive postglacial gene flow between European white
648 oaks uncovered genes underlying species barriers. New Phytol 226:1183-1197.
649 Li H. 2013. Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM.
650 arXiv preprint arXiv:1303.3997.
651 Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, Marth G, Abecasis G, Durbin R.
652 2009. The Sequence Alignment/Map format and SAMtools. Bioinformatics 25:2078-2079.
653 Macaya-Sanz D, Heuertz M, López-de-Heredia U, De-Lucas AI, Hidalgo E, Maestro C, Prada A,
654 Alía R, González-Martínez SC. 2012. The Atlantic-Mediterranean watershed, river basins
655 and glacial history shape the genetic structure of Iberian poplars. Mol Ecol 21:3593-3609.
656 Martin SH, Dasmahapatra KK, Nadeau NJ, Salazar C, Walters JR, Simpson F, Blaxter M,
657 Manica A, Mallet J, Jiggins CD. 2013. Genome-wide evidence for speciation with gene flow
658 in *Heliconius* butterflies. Genome Res 23:1817-1828.
659 Martin SH, Davey JW, Salazar C, Jiggins CD. 2019. Recombination rate variation shapes
660 barriers to introgression across butterfly genomes. PLoS Biol 17:e2006288.
661 Martin SH, Jiggins CD. 2017. Interpreting the genomic landscape of introgression. Curr Opin
662 Genet Dev 47:69-74.
663 Martinsen GD, Whitham TG, Turek RJ, Keim P. 2001. Hybrid populations selectively filter gene
664 introgression between species. Evolution 55:1325-1335.
665 Noor MA, Bennett SM. 2009. Islands of speciation or mirages in the desert? Examining the role
666 of restricted recombination in maintaining species. Heredity (Edinb) 103:439-444.
667 Nosil P, Feder JL. 2012. Widespread yet heterogeneous genomic divergence. Mol Ecol 21:2829-
668 2832.

669 Pickrell JK, Pritchard JK. 2012. Inference of population splits and mixtures from genome-wide
670 allele frequency data. PLoS Genet 8:e1002967.
671 Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira MA, Bender D, Maller J, Sklar P, de
672 Bakker PI, Daly MJ, et al. 2007. PLINK: a tool set for whole-genome association and
673 population-based linkage analyses. Am J Hum Genet 81:559-575.
674 Rajora OP, Dancik BP. 1992. Genetic characterization and relationships of *Populus alba*, *P.*
675 *tremula*, and *P. x canescens*, and their clones. Theor Appl Genet 84:291-298.
676 Ravinet M, Yoshida K, Shigenobu S, Toyoda A, Fujiyama A, Kitano J. 2018. The genomic
677 landscape at a late stage of stickleback speciation: High genomic divergence interspersed by
678 small localized regions of introgression. PLoS Genet 14:e1007358.
679 Renaut S, Grassa CJ, Yeaman S, Moyers BT, Lai Z, Kane NC, Bowers JE, Burke JM, Rieseberg
680 LH. 2013. Genomic islands of divergence are not affected by geography of speciation in
681 sunflowers. Nat Commun 4:1827.
682 Renaut S, Owens GL, Rieseberg LH. 2014. Shared selective pressure and local genomic
683 landscape lead to repeatable patterns of genomic divergence in sunflowers. Mol Ecol 23:311-
684 324.
685 Rennison DJ, Delmore KE, Samuk K, Owens GL, Miller SE. 2020. Shared patterns of genome-
686 wide differentiation are more strongly predicted by geography than by ecology. Am Nat
687 195:192-200.
688 Rifkin JL, Castillo AS, Liao IT, Rausher MD. 2019. Gene flow, divergent selection and
689 resistance to introgression in two species of morning glories (*Ipomoea*). Mol Ecol 28:1709-
690 1729.
691 Roux C, Fraïsse C, Romiguier J, Anciaux Y, Galtier N, Bierne N. 2016. Shedding light on the
692 grey zone of speciation along a continuum of genomic divergence. PLoS Biol 14:e2000234.
693 Schiffthaler B, Delhomme N, Bernhardsson C, Jenkins J, Jansson S, Ingvarsson P, Schmutz J,
694 Street N. 2019. An improved genome assembly of the European aspen *Populus tremula*.
695 bioRxiv:805614.
696 Sendell-Price AT, Ruegg KC, Anderson EC, Quilodrán CS, Van Doren BM, Underwood VL,
697 Coulson T, Clegg SM. 2020. The Genomic Landscape of Divergence Across the Speciation
698 Continuum in Island-Colonising Silvereyes (*Zosterops lateralis*). G3 (Bethesda) 10:3147-
699 3163.

700 Shang H, Hess J, Pickup M, Field DL, Ingvarsson PK, Liu J, Lexer C. 2020. Evolution of strong
701 reproductive isolation in plants: broad-scale patterns and lessons from a perennial model
702 group. Philos Trans R Soc Lond B Biol Sci 375:20190544.
703 Smith JM, Haigh J. 1974. The hitch-hiking effect of a favourable gene. Genet Res 23:23-35.
704 Stankowski S, Chase MA, Fuiten AM, Rodrigues MF, Ralph PL, Streisfeld MA. 2019.
705 Widespread selection and gene flow shape the genomic landscape during a radiation of
706 monkeyflowers. PLoS Biol 17:e3000391.
707 Stettler R, Bradshaw T, Heilman P, Hinckley T. 1996. Biology of *Populus* and its implications
708 for management and conservation: NRC Research Press.
709 Suarez-Gonzalez A, Hefer CA, Christe C, Corea O, Lexer C, Cronk QC, Douglas CJ. 2016.
710 Genomic and functional approaches reveal a case of adaptive introgression from *Populus*
711 *balsamifera* (balsam poplar) in *P. trichocarpa* (black cottonwood). Mol Ecol 25:2427-2442.
712 Talla V, Johansson A, Dincă V, Vila R, Friberg M, Wiklund C, Backström N. 2019. Lack of
713 gene flow: Narrow and dispersed differentiation islands in a triplet of *Leptidea* butterfly
714 species. Mol Ecol 28:3756-3770.
715 Tavares H, Whibley A, Field DL, Bradley D, Couchman M, Copsey L, Elleouet J, Burrus M,
716 Andalo C, Li M, et al. 2018. Selection and gene flow shape genomic islands that control
717 floral guides. Proc Natl Acad Sci U S A 115:11006-11011.
718 Terhorst J, Kamm JA, Song YS. 2017. Robust and scalable inference of population history from
719 hundreds of unphased whole genomes. Nat Genet 49:303-309.
720 Turner TL, Hahn MW. 2010. Genomic islands of speciation or genomic islands and speciation?
721 Mol Ecol 19:848-850.
722 Tuskan GA, Difazio S, Jansson S, Bohlmann J, Grigoriev I, Hellsten U, Putnam N, Ralph S,
723 Rombauts S, Salamov A, et al. 2006. The genome of black cottonwood, *Populus trichocarpa*
724 (Torr. & Gray). Science 313:1596-1604.
725 Vijay N, Bossu CM, Poelstra JW, Weissensteiner MH, Suh A, Kryukov AP, Wolf JB. 2016.
726 Evolution of heterogeneous genome differentiation across multiple contact zones in a crow
727 species complex. Nat Commun 7:13195.
728 Wang B, Mojica JP, Perera N, Lee CR, Lovell JT, Sharma A, Adam C, Lipzen A, Barry K,
729 Rokhsar DS, et al. 2019. Ancient polymorphisms contribute to genome-wide variation by

730 long-term balancing selection and divergent sorting in *Boechera stricta*. Genome Biol
731 20:126.
732 Wang J, Street NR, Scofield DG, Ingvarsson PK. 2016. Natural Selection and Recombination
733 Rate Variation Shape Nucleotide Polymorphism Across the Genomes of Three Related
734 *Populus* Species. Genetics 202:1185-1200.
735 Wang M, Zhang L, Zhang Z, Li M, Wang D, Zhang X, Xi Z, Keefover-Ring K, Smart LB,
736 DiFazio SP, et al. 2020. Phylogenomics of the genus *Populus* reveals extensive interspecific
737 gene flow and balancing selection. New Phytol 225:1370-1382.
738 Wu CI. 2001. The genic view of the process of speciation. Journal of evolutionary biology
739 14:851-865.
740 Wolf JB, Ellegren H. 2017. Making sense of genomic islands of differentiation in light of
741 speciation. Nat Rev Genet 18:87-100.
742 Yamasaki YY, Kakioka R, Takahashi H, Toyoda A, Nagano AJ, Machida Y, Moller PR, Kitano
743 J. 2020. Genome-wide patterns of divergence and introgression after secondary contact
744 between *Pungitius* sticklebacks. Philos Trans R Soc Lond B Biol Sci 375:20190548.
745