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

To understand the processes behind the formation of 'differentiation islands' along the speciation continuum, we summarize four different scenarios according to the divergence models proposed by Han *et al* (2017) and Irwin *et al* (2018) (Fig. 1a). In this context, nucleotide diversity (π) as well as two

measures of divergence, the relative measure F_{ST} and the absolute measure D_{XY}, are instrumental in distinguishing between evolutionary scenarios (Charlesworth 1998). Genomic regions with high F_{ST} may occur in the absence of local gene flow due to ecological or non-ecological isolating barriers, or due to locally reduced *Ne* in regions of low recombination or selective sweeps (Noor and Bennett 2009; Turner and Hahn 2010; Renaut, et al. 2014; Campagna, et al. 2015; Gagnaire, et al. 2018; Henderson and Brelsford 2020). Disentangling the underlying processes using only F_{ST} is impractical, as this is directly influenced by the level of within-groups genetic diversity (e.g., Cruickshank and Hahn, 2014). D_{XY} is a complementary measure that is independent of the levels of diversity within the groups. D_{XY} is sensitive to ancestral variation, with higher values in regions of restricted gene flow and decreased in regions under background selection or selective sweeps (Han, et al. 2017; Irwin, et al. 2018). Scanning the whole genome of pairs of species using π, F_{ST} and D_{XY}, different scenarios can be hypothesized regarding the extent of local gene flow to identify the main drivers of genomic landscapes of differentiation (Fig. 1a). The first scenario is 'divergence with gene flow' where selection at loci which contribute to reproductive isolation restricts gene exchange between populations, locally elevating genomic differentiation (leading to both high F_{ST} and D_{XY}) and reducing genetic diversity. The second scenario is 'allopatric selection' where positive selection acts on distinct regions of the genome after a species split into two populations, leading to lower π and higher F_{ST}. As D_{XY} is sensitive to ancestral polymorphism, its values are expected to remain relatively stable under this model. Besides, selective sweeps in allopatry in one or both populations may also contribute to this divergence pattern. The third scenario is 'recurrent selection' where background selection or selective sweeps at certain genomic regions reduce genetic diversity in both the common ancestor and the two daughter populations, leading to lower D_{XY} and π, but higher F_{ST}. The fourth and last scenario is 'balancing selection' where ancestral polymorphism is maintained at selected sites, resulting in increased differentiation between species due to lineage sorting. When employed in conjunction, the use of genomic estimates of F_{ST}, D_{XY} and π can thus enhance our understanding of the mechanisms at play that shaped the genomic differentiation landscape.

Conserved genomic landscapes of differentiation among independent species pairs could be generated by long-term linked selection in heterogeneous recombination

landscapes. And it has been argued that background selection may more likely lead to shared genomic patterns than genetic hitchhiking because deleterious mutations are far more common than beneficial ones (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.

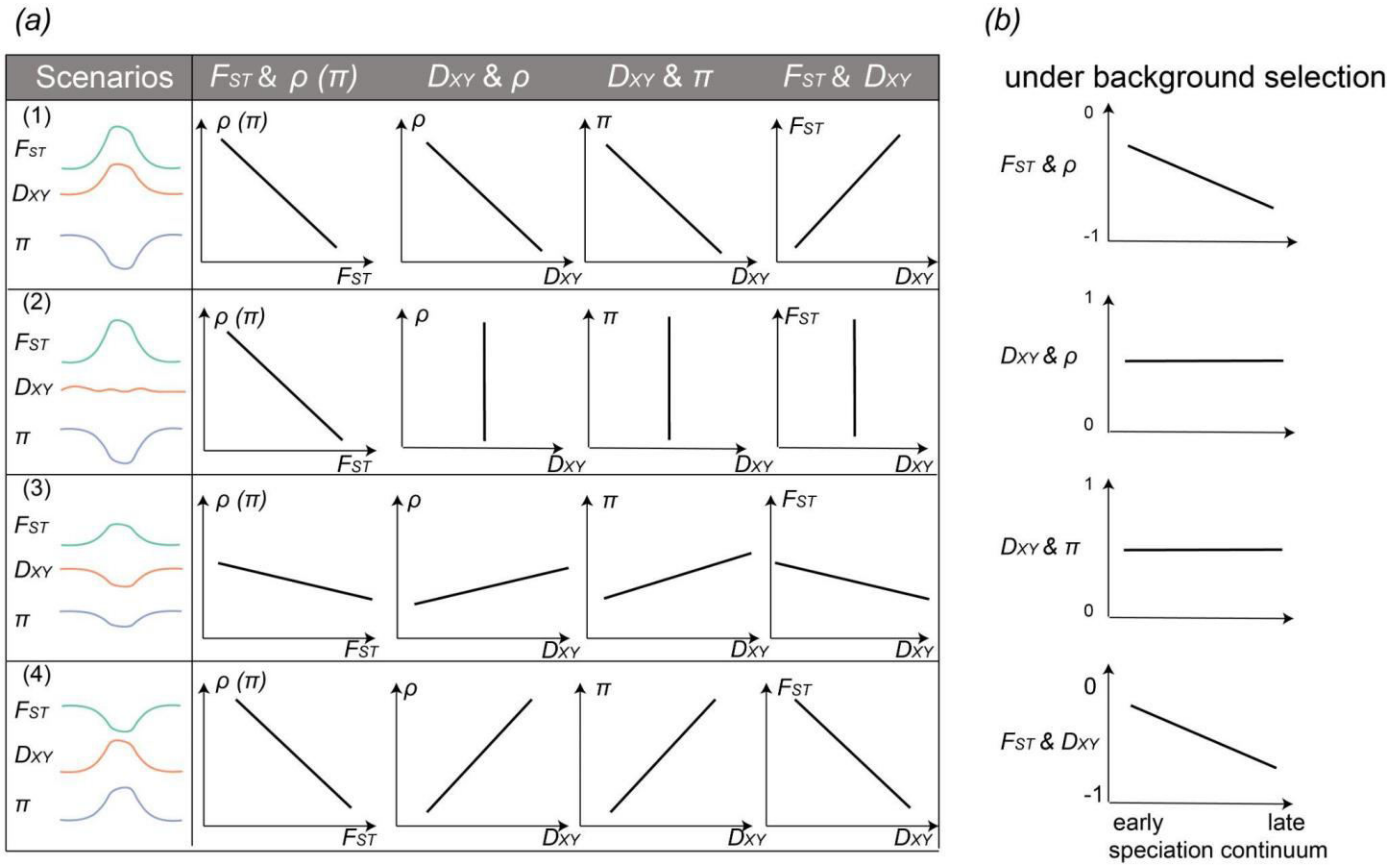

Fig. 1. *Expected patterns of genomic differentiation under different scenarios. (a) Genetic diversity, divergence change along the genome under four scenarios: (1) 'divergence with gene flow', (2) 'allopatric selection', (3) 'background selection/balancing selection', and (4) 'balancing selection'. 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*; Schiffthaler, et al. 2019). Taken together, the !rich genomic resources, small genomes and widespread geographic distribution make the *Populus* system a useful model to address open questions regarding the formation of genomic islands along the speciation continuum.

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

Results and Discussion

Strong interspecific structure despite interspecific introgression

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

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

Neighbor-joining (Fig. 2b) and Admixture (Alexander and Lange 2011) analyses (Fig. 2c) based on all SNPs identified seven genetic groups, which were consistent with previously identified species boundaries based on phylogenomic analysis (Shang et al, 2020). Additionally, Admixture also indicated a potential introgression between the subtropical

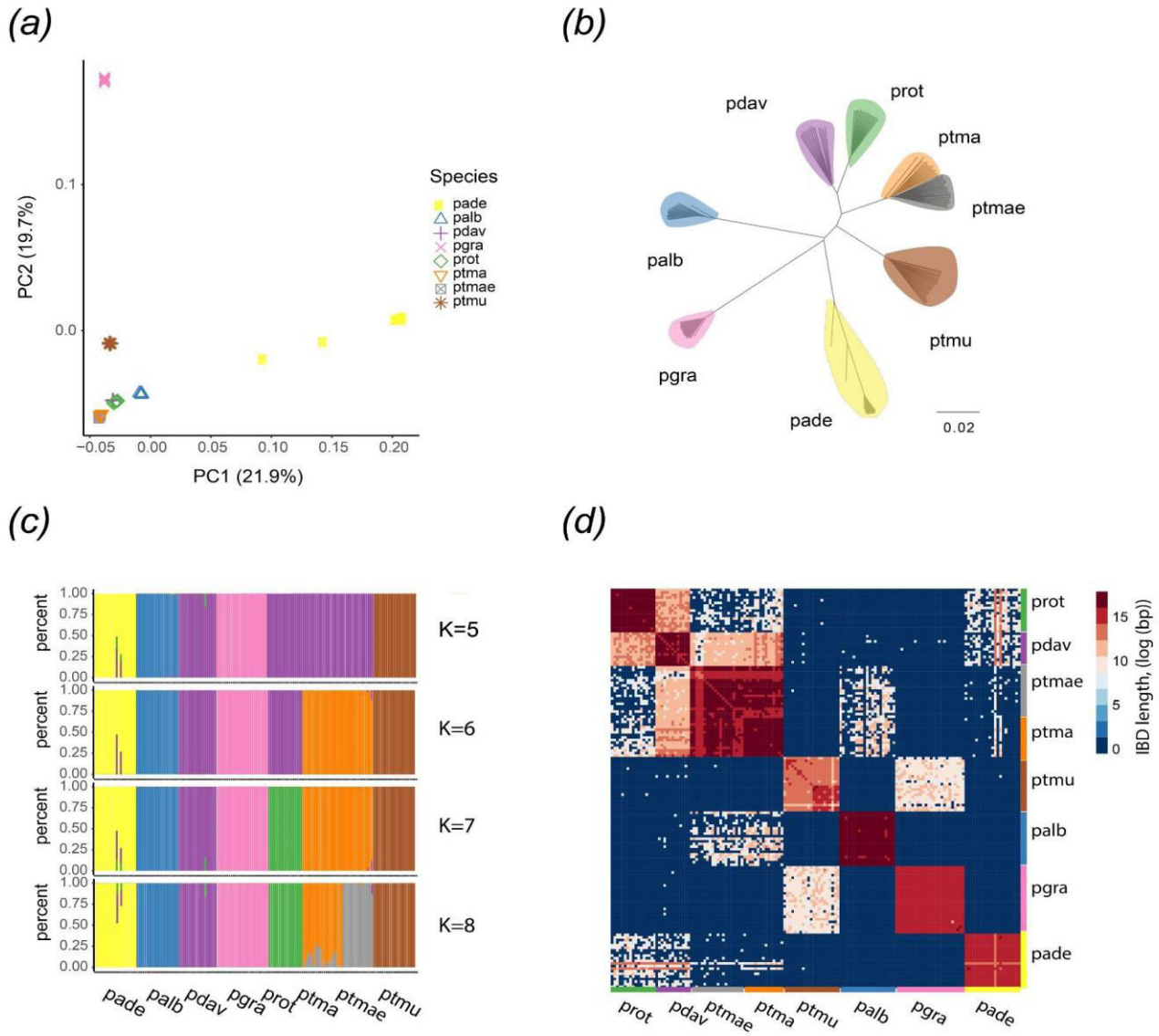

Fig. 2. *Genetic structure within Populus (poplar and aspen) accessions investigated here.* ***(a)*** *Principal Component Analysis (PCA) based on all SNPs for seven Populus species. Colored symbols represent different species according to legend.* ***(b)*** *Unrooted neighbour-joining tree of all SNPs.*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

species *P. adenopoda* and two recently diverged species, *P. davidiana* and *P. rotundifolia* (Fig. 1c and supplementary material, Fig. S5). Identity-by-descent (IBD) analyses (Fig. 2d) also identified seven reliable clusters, corresponding to the same species boundaries as before, but further pinpointed some shared haplotypes among the aspen species *P. davidiana*, *P. rotundifolia* and *P. tremula*, indicating recent introgression among these species. The IBD results also provide support for extensive introgression between species with overlapping distribution, such as *P. alba* and *P. tremula*, or *P. grandidentata* and *P. tremuloides*. These results are consistent with a scenario of divergence with ongoing gene flow, either due to isolation-with-migration or secondary contact, maintained even after substantial divergence times (net divergence d_a: 0.023 for *P. alba* - *P. tremula*; d_a: 0.025 for *P. tremuloides* - *P. grandidentata*). For these two species pairs, net divergence values are indeed larger than the upper boundary for the 'grey zone of speciation' reported by Roux, et al. (2016) for both vertebrate and invertebrate animals (d_a from 0.005 to 0.02). Future investigations on the plant grey zone based on a large number of taxa are needed to get more general insights about plant diversification.

Between species variability in demographic trajectories

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

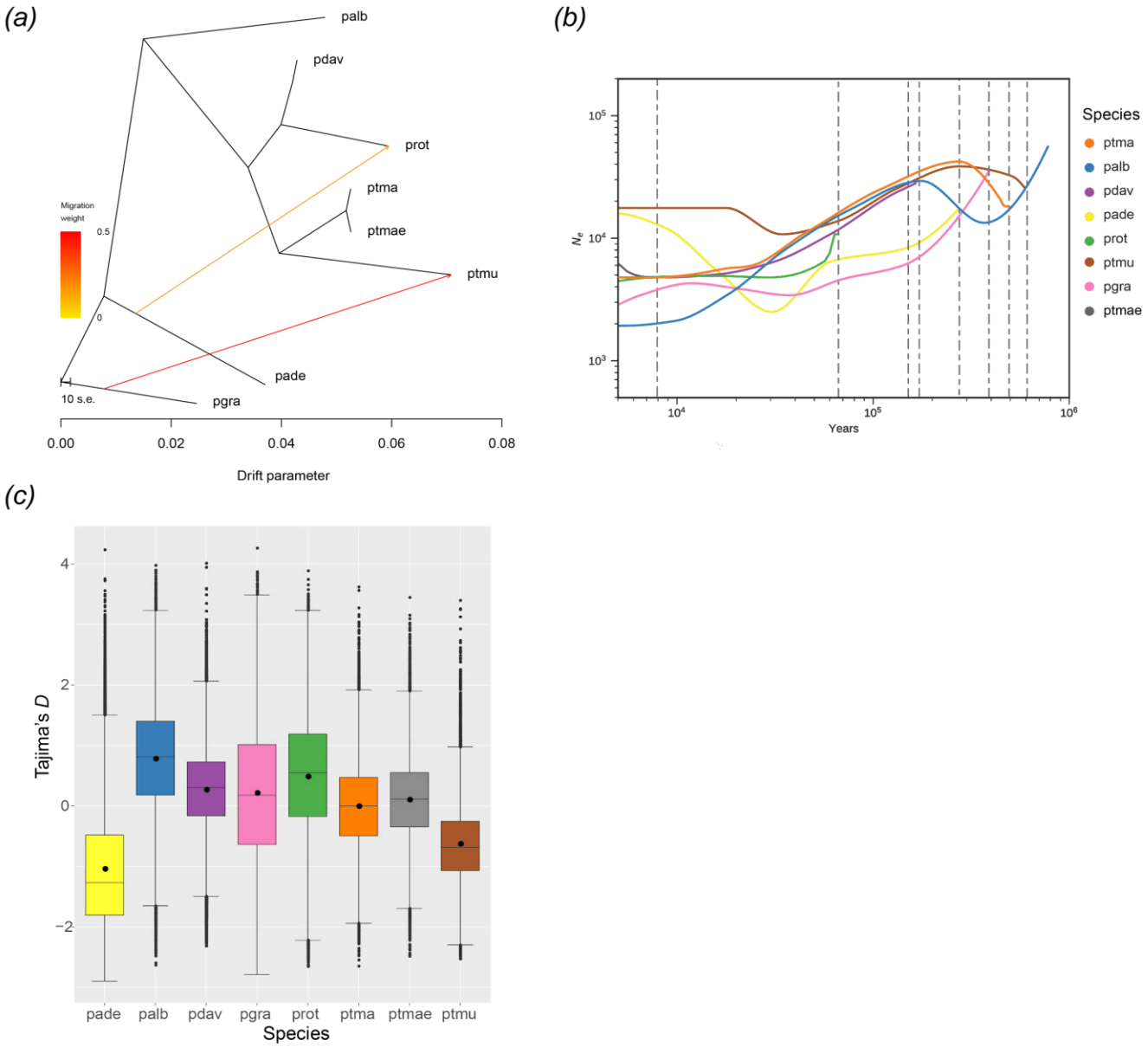

Fig. 3. *Demographic history of Populus species.* ***(a)*** *The maximum likelihood tree inferred by TreeMix under a strictly bifurcating model with two migration events.* ***(b)*** *Changes in effective population size through time inferred by SMC++.* ***(c)*** *The distribution of Tajima's D values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

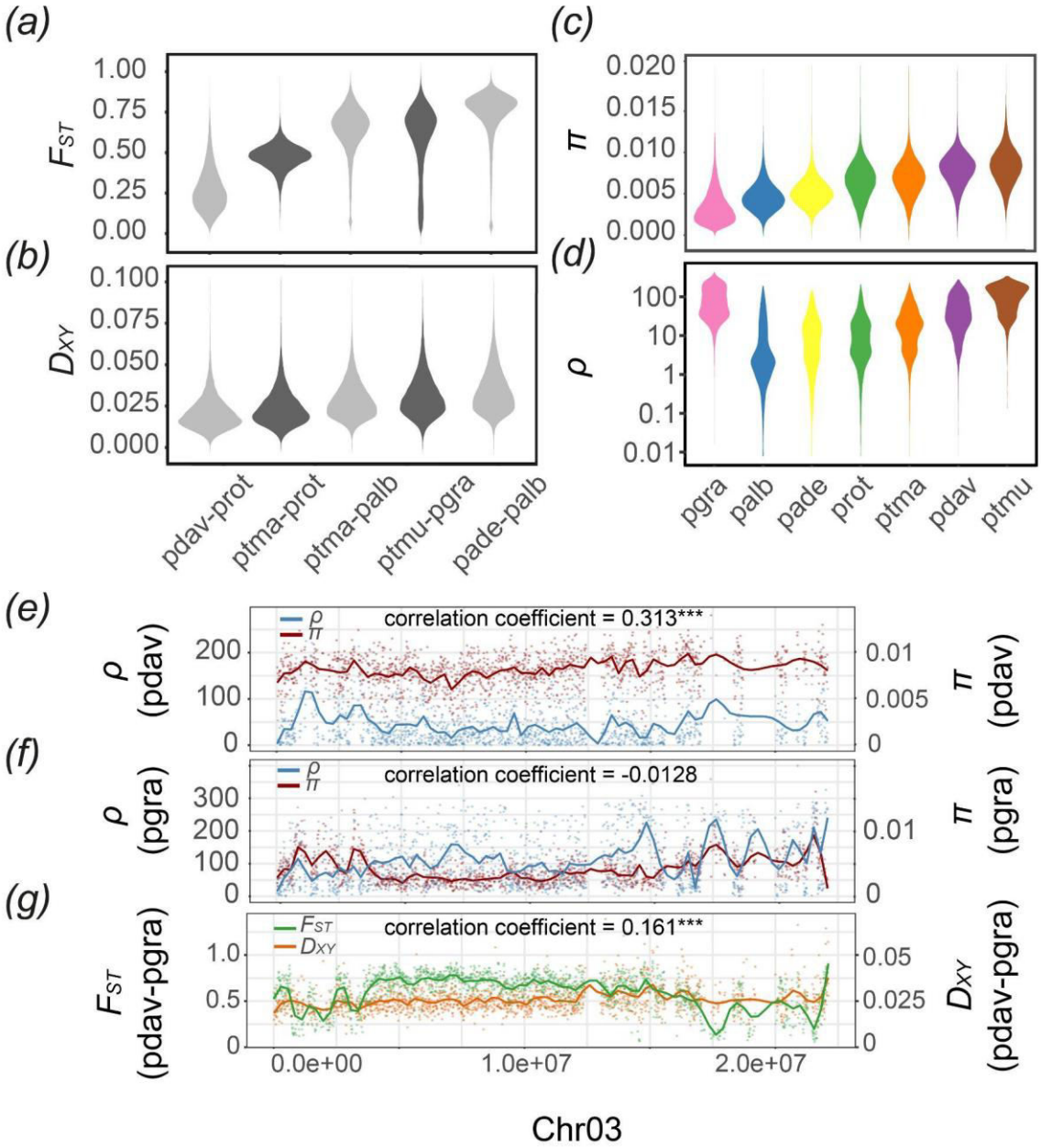

Fig. 4. *Summary violin plots of F_{ST}, D_{XY}, π and ρ calculated across 10kb windows.* ***(a-b)*** *Violin plots of F_{ST} and D_{XY} for five species pairs as representatives of the different stages across the speciation continuum investigated in this study.* ***(c-d)*** *Violin plots of π and ρ 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.*

Conserved genomic landscapes across the continuum of divergence

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

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

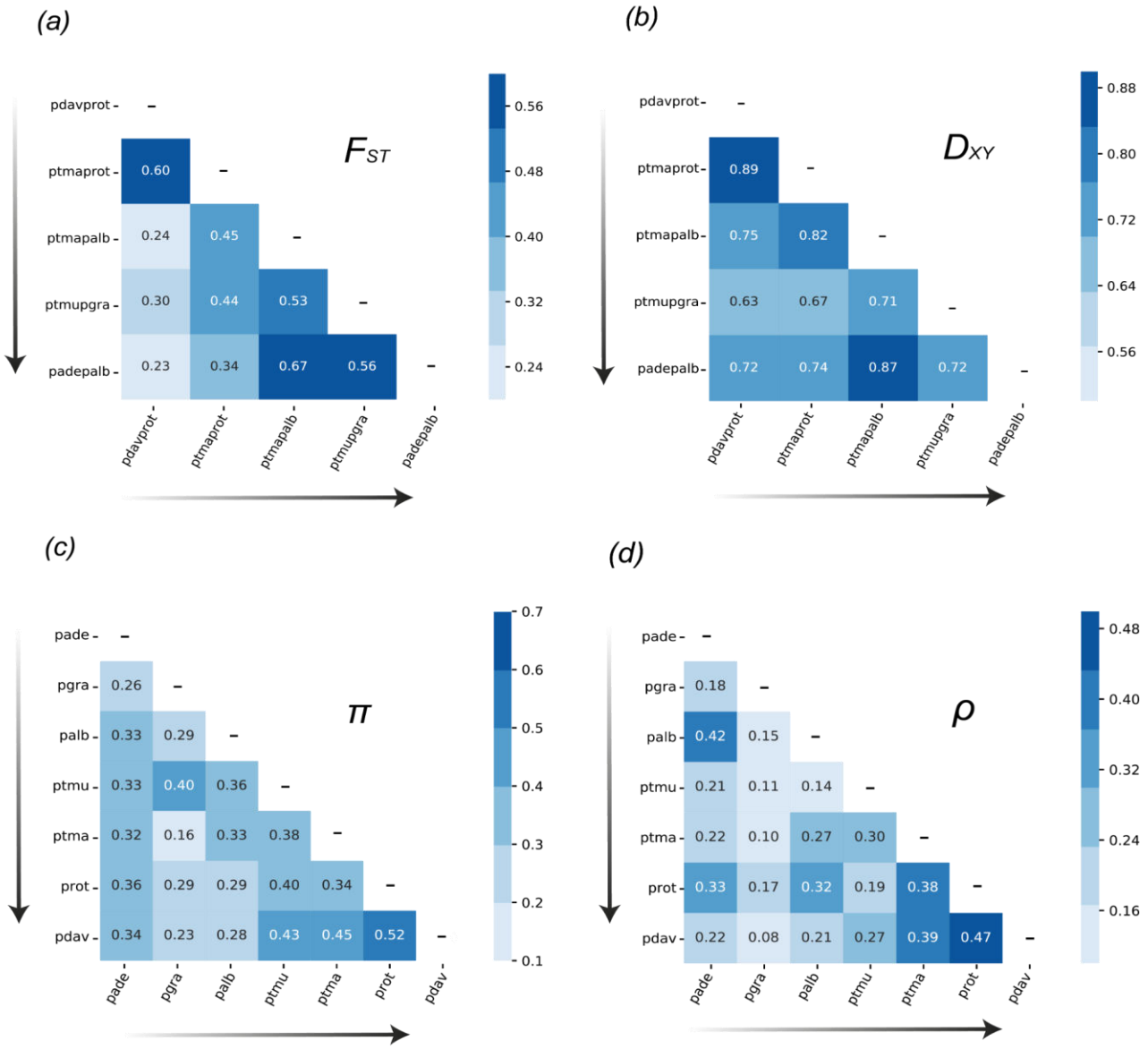

Fig. 5. *Correlations analyses of within-species diversity or among-species divergence landscapes* ***(a-b)*** *Correlation coefficients of F_{ST} or D_{XY} between species pairs. The species pairs are ordered across the speciation continuum (arrows).* ***(c-d)*** *Correlation coefficients of π or ρ 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.*

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.

Correlated patterns of genome-wide variation across speciation continuum

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

In our analysis, we found that the negative relationships between F_{ST} and π or ρ become stronger as d_a increases, which is consistent with the expectation under background selection (Fig. 6a, d). Similar investigations for π and D_{XY} showed significantly positive correlations while the trend became weaker as divergence increases (Fig. 6b). The trend is not consistent with the general hypothesis that correlations should still be highly correlated as divergence increases. We also recovered a strong positive correlation between π and ρ (Fig. 6c), and a similar trend was found as for the investigation of π and D_{XY}. Correlations between D_{XY} and ρ were significantly positive across the entire speciation continuum, and we do observe that these correlations tend to become stronger as divergence increases across the speciation continuum (Fig. 6e). Positive correlations between F_{ST} and D_{XY} were found across all species pairs, even though the relationships

become weaker as divergence increases (Fig. 6f). Thus the high F_{ST} may be due to elevated D_{XY}, which is not consistent with the expectations of lower D_{XY} under background selection. The observed patterns (Fig. 6b, f) differ from expectations under a scenario with background selection as the sole factor shaping the heterogeneous landscape of differentiation, indicating that there should be additional evolutionary factors that contribute to it (Burri 2017a). Consistent with our findings, studies in monkey flowers, threespine stickleback, and avian species also confirmed that background selection alone is too subtle to make a strong contribution to the conserved genomic patterns (Irwin, et al. 2018; Stankowski, et al. 2019; Rennison, et al. 2020). Indeed, these studies suggest either adaptive introgression or shared standing genetic variation play a major role in generating the similar patterns of genomic differentiation. In our study, the negative correlations between π and F_{ST} (Fig. 6a) and the positive correlations between F_{ST} and D_{XY} (Fig. 6f) in all 21 species pairs indicates high F_{ST} could be caused by the high D_{XY}. As D_{XY} is sensitive to ancestral polymorphism, the high D_{XY} most likely due to the presence of diverged haplotypes in ancestral populations before speciation or resulted from limited gene flow in regions under selection, which can be caused by ecological adaptation.

Scenarios of genomic patterns of differentiation

Based on F_{ST}, D_{XY}, and π, we summarized four scenarios to help explain the genomic divergence patterns observed (supplementary material, Fig. S10) following the strategy developed by Han et al. (2017) and Irwin et al. (2018). The major differences of these scenarios are the role and extent of gene flow, and the type of selection acting to shape genome-wide genetic differentiation. Our results show a heterogeneous distribution of the four scenarios along the genome for all five species pairs (supplementary material, Fig. S11-S15 and Table S2). A great number of regions (74.3%-78.7%) in all five species pairs fit a scenario of "allopatric selection" (scenario b in supplementary material, Fig. S11-S15 and Table S2), i.e., regions exhibiting an elevated F_{ST} (tail of the distribution) but a moderate D_{XY} (in the middle of the distribution). Such a signature is expected to be consistent with recent footprints of positive or background selection on genomic differentiation.

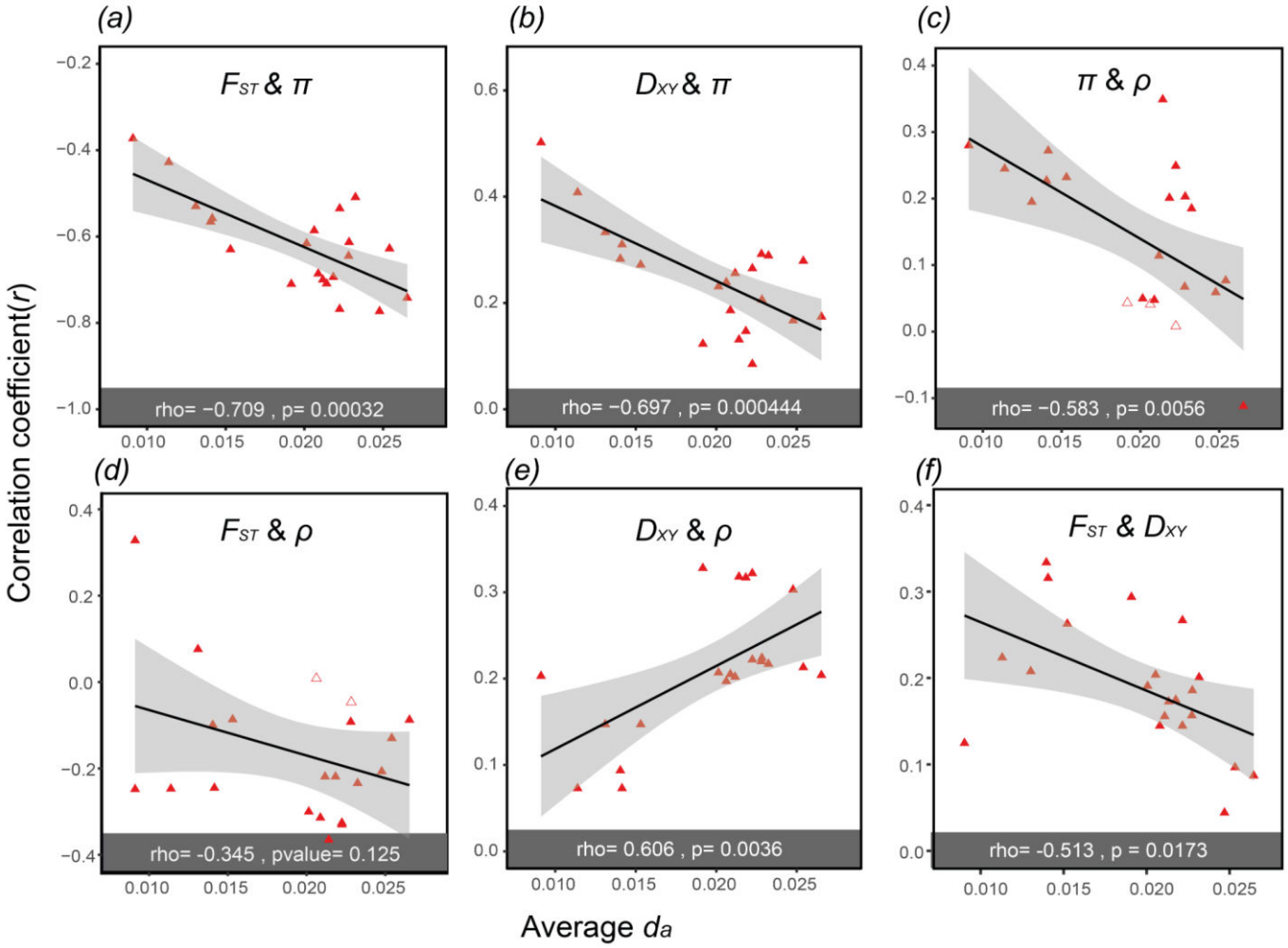

Fig. 6. *Correlations between variables for all species comparisons (red filled and unfilled triangles) plotted against the average d_a, used here as a measure of divergence time. The red filled triangles indicate the correlation coefficients are significant ($p < 0.01$) in each panel. The upper panels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

Genomic regions fitting the scenario of 'balancing selection' (scenario d in supplementary material, Fig. S10) are the second most frequent for all investigated species pairs. This scenario is characterized by an elevated D_{XY} but a low F_{ST} implying the action of balancing selection in shaping the heterogeneous landscape of divergence. In addition, we found support for reproductive barriers (scenario a in supplementary material, Fig. S11-S15 and Table S2) in all five species pairs, suggesting that genomic heterogeneity in the levels of gene flow due to the species barriers play a role in shaping genomic differentiation landscapes. Interestingly, this result holds true for all species pairs we investigated, *i.e.,* regardless of the level of gene flow or the stage along the *Populus* speciation continuum. Indeed, the regions with high F_{ST} and D_{XY} found in the highly diverged species pair (*P. adenopoda* - *P. alba*) could not be caused by adaptive introgression but may be due to the shared ancestral polymorphisms, as no gene flow was found between these two species. In contrast, at the early stage of divergence, barriers to gene flow may play an important role in genomic heterogeneous divergence, as the influence of background selection is still very weak at this stage. However, it is too naive to use the four scenarios to explain the empirical data, as the interplay of demographic history, selection, gene flow and recombination may obscure the signature of genomic patterns generated by one evolutionary factor.

Conclusions

In this study, we reconstructed the demographic history and investigated the evolution of the genomic landscape of diversity and divergence across a speciation continuum of eight closely related species of *Populus* section *Populus*. By investigating evolution of diversity and differentiation landscapes across this speciation continuum, we provided, to our knowledge, one of the most ambitious case studies in terms of the number of species pairs analyzed (see also Stankowski, et al. 2019). Our correlation analyses support this role of linked selection, recombination rate and gene density in shaping the empirical patterns of genomic differentiation, but the contribution of linked selection alone is not sufficient in generating these genomic landscapes. Instead, the positive correlations between F_{ST} and D_{XY} in all species pairs indicates diverged ancient polymorphism must play a role. The study also confirmed the importance of gene flow in this system, through extensive introgression among species with parapatric distributions, despite a high level of divergence among the most divergent hybridizing species (d_a = 0.025). Overall, this

work provides a prime example of the strategic importance of investigating these genomic patterns on a large number of species to better access the temporal evolution of the genomic landscapes of diversity and differentiation. In the future, more studies on multiple species pairs across the speciation continuum are needed to discuss the interactions of several evolutionary factors on the heterogeneous genomic landscapes of differentiation.

Materials and methods

Sampling, sequencing and read processing

Two hundred and one samples were collected from eight species of *Populus* section *Populus* in Eurasia and North America (supplemental material, Fig. S1 and Table S1). The leaves were dried in silica gel first and were then used for genomic DNA extraction with Plant DNeasy Mini Kit (Qiagen, Germany). To increase the purity of total DNA, we used the NucleoSpin gDNA Clean-up kit (Macherey-Nagel, Germany). Whole genome resequencing was performed with 2 x 150bp paired-end sequencing technology on Illumina HiSeq 3000 sequencer at the Institute of Genetics, University of Bern, Switzerland.

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

Family relatedness and population structure analysis

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

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

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

Demographic trajectory reconstruction

To reconstruct the demographic history of *Populus* species, we first inferred the history of species splits and mixture based on genome

wide allele frequency data using TreeMix v1.13 (Pickrell and Pritchard 2012). We removed the sites with missing data and performed linkage pruning. We then ran TreeMix implementing a default bootstrap and a block size of 500 SNPs (-k=500). The best migration edge was evaluated according to the greatest increase of total variation explained. The plotting R functions of the Treemix suite were then used to visualize the results. In addition, we used SMC++ to estimate historical changes in effective population size (Terhorst, et al. 2017). The split time for species pairs was estimated based on the joint frequency spectrum for both species. Due to widespread vegetative reproduction in some poplar species, we adjusted the generation time to 20 years rather than the generally thought 15 years (Wang, et al. 2016); a mutation rate of 2.5×10^{-9} per site per year was also assumed in order to scale results to real time (Macaya-Sanz, et al. 2012).

Nucleotide diversity and divergence estimates

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

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

Population-scale recombination rate and linkage disequilibrium

We estimated population scaled recombination rate with FastEPRR (Gao, et al. 2016) for each species separately. To eliminate the effect of sample size on the estimation of recombination rate, we downsampled to 13 randomly selected individuals for each species, corresponding to the number of individuals available for *Populus davidiana* (pdav). First, we filtered all missing and non-biallelic sites with VCFtools and then phased the data with the parameters "*impute=true nthreads=20 window=10,000 overlap=1,000 gprobs=false*" in Beagle v5.1 (Browning and Browning 2013). Finally, we ran FastEPRR v2.0 (Gao, et al. 2016) with a window size of 10kb. After getting the results, we estimated the correlation between recombination rate of one species to another. To evaluate LD decay, we used PLINK (Purcell, et al. 2007) to obtain LD statistics for each species. Parameters were set as follows: '*--maf 0.1 --r^2 gz --ld-window-kb 500 --ld-window 99999 --ld-window-r^2 0*'. LD decay was finally plotted in R.

Divergent regions of exceptional differentiation

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

Acknowledgements

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

Data accessibility

All raw read data was uploaded to NCBI and can be found under Bioproject ID PRJNA299390, PRJNA612655, PRJNA720790, and PRJNA297202. All the scripts used for the analysis are available on https://github.com/Huiying123/Populus_speciation.

Author contributions

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

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.

ORCID

Huiying Shang:
https://orcid.org/0000-0002-1302-8008
Ovidiu Paun:
https://orcid.org/0000-0002-8295-4937
Thibault Leroy:
https://orcid.org/0000-0003-2259-9723
Jaqueline Hess :
https://orcid.org/0000-0003-3281-5434
David Field:
https://orcid.org/0000-0002-4014-8478
Pär K Ingvarsson:
https://orcid.org/0000-0001-9225-7521

References

Alexander DH, Lange K. 2011. Enhancements to the ADMIXTURE algorithm for individual ancestry estimation. BMC Bioinformatics 12:246.

Browning BL, Browning SR. 2013. Improving the accuracy and efficiency of identity-by-descent detection in population data. Genetics 194:459-471.

Burri R. 2017a. Dissecting differentiation landscapes: a linked selection's perspective. J Evol Biol 30:1501-1505.

Burri R. 2017b. Linked selection, demography and the evolution of correlated genomic landscapes in birds and beyond. Mol Ecol 26:3853-3856.

Burri R, Nater A, Kawakami T, Mugal CF, Olason PI, Smeds L, Suh A, Dutoit L, Bureš S, Garamszegi LZ, et al. 2015. Linked selection and recombination rate variation drive the evolution of the genomic landscape of differentiation across the speciation continuum of *Ficedula* flycatchers. Genome Res 25:1656-1665.

Campagna L, Gronau I, Silveira LF, Siepel A, Lovette IJ. 2015. Distinguishing noise from signal in patterns of genomic divergence in a highly polymorphic avian radiation. Mol Ecol 24:4238-4251.

Charlesworth B. 1998. Measures of divergence between populations and the effect of forces that reduce variability. Mol Biol Evol 15:538-543.

Charlesworth B, Campos JL. 2014. The relations between recombination rate and patterns of molecular variation and evolution in *Drosophila*. Annu Rev Genet 48:383-403.

Charlesworth B, Morgan MT, Charlesworth D. 1993. The effect of deleterious mutations on neutral molecular variation. Genetics 134:1289-1303.

Corbett-Detig RB, Hartl DL, Sackton TB. 2015. Natural selection constrains neutral diversity across a wide range of species. PLoS Biol 13:e1002112.

Cruickshank TE, Hahn MW. 2014. Reanalysis suggests that genomic islands of speciation are due to reduced diversity, not reduced gene flow. Mol Ecol 23:3133-3157.

Deacon NJ, Grossman JJ, Cavender-Bares J. 2019. Drought and freezing vulnerability of the isolated hybrid aspen *Populus x smithii* relative to its parental species, *P. tremuloides* and *P. grandidentata*. Ecol Evol 9:8062-8074.

DePristo MA, Banks E, Poplin R, Garimella KV, Maguire JR, Hartl C, Philippakis AA, del Angel G, Rivas MA, Hanna M, et al. 2011. A framework for variation discovery and genotyping using next-generation DNA sequencing data. Nat Genet 43:491-498.

Eckenwalder J. 1996. Systematics and evolution of *Populus*. U: Stettler, RF, Bradshaw, HD, Heilman, PE, Hinckley, TM, eds.(1996): Biology of *Populus* and Its Implications for Management and Conservation. In: NRC Research Press, Ottawa, Ontario, Canada.

Ellegren H, Smeds L, Burri R, Olason PI, Backström N, Kawakami T, Künstner A, Mäkinen H, Nadachowska-Brzyska K, Qvarnström A, et al. 2012. The genomic landscape of species divergence in *Ficedula* flycatchers. Nature 491:756-760.

Gagnaire PA, Lamy JB, Cornette F, Heurtebise S, Dégremont L, Flahauw E, Boudry P, Bierne N, Lapègue S. 2018. Analysis of Genome-Wide Differentiation between Native and Introduced Populations of the Cupped Oysters *Crassostrea gigas* and

Crassostrea angulata. Genome Biol Evol 10:2518-2534.

Gao F, Ming C, Hu W, Li H. 2016. New Software for the Fast Estimation of Population Recombination Rates (FastEPRR) in the Genomic Era. G3 (Bethesda) 6:1563-1571.

Han F, Lamichhaney S, Grant BR, Grant PR, Andersson L, Webster MT. 2017. Gene flow, ancient polymorphism, and ecological adaptation shape the genomic landscape of divergence among Darwin's finches. Genome Res 27:1004-1015.

Harrison RG, Larson EL. 2016. Heterogeneous genome divergence, differential introgression, and the origin and structure of hybrid zones. Mol Ecol 25:2454-2466.

Henderson EC, Brelsford A. 2020. Genomic differentiation across the speciation continuum in three hummingbird species pairs. BMC Evol Biol 20:113.

Irwin DE, Milá B, Toews DPL, Brelsford A, Kenyon HL, Porter AN, Grossen C, Delmore KE, Alcaide M, Irwin JH. 2018. A comparison of genomic islands of differentiation across three young avian species pairs. Mol Ecol 27:4839-4855.

Jansson S, Bhalerao R, Groover A. 2010. Genetics and genomics of *Populus*: Springer.

Lamichhaney S, Berglund J, Almén MS, Maqbool K, Grabherr M, Martinez-Barrio A, Promerová M, Rubin CJ, Wang C, Zamani N, et al. 2015. Evolution of Darwin's finches and their beaks revealed by genome sequencing. Nature 518:371-375.

Leroy T, Rougemont Q, Dupouey JL, Bodénès C, Lalanne C, Belser C, Labadie K, Le Provost G, Aury JM, Kremer A, et al. 2020. Massive postglacial gene flow between European white oaks uncovered genes underlying species barriers. New Phytol 226:1183-1197.

Li H. 2013. Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. arXiv preprint arXiv:1303.3997.

Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, Marth G, Abecasis G, Durbin R. 2009. The Sequence Alignment/Map format and SAMtools. Bioinformatics 25:2078-2079.

Macaya-Sanz D, Heuertz M, López-de-Heredia U, De-Lucas AI, Hidalgo E, Maestro C, Prada A, Alía R, González-Martínez SC. 2012. The Atlantic-Mediterranean watershed, river basins and glacial history shape the genetic structure of Iberian poplars. Mol Ecol 21:3593-3609.

Martin SH, Dasmahapatra KK, Nadeau NJ, Salazar C, Walters JR, Simpson F, Blaxter M, Manica A, Mallet J, Jiggins CD. 2013. Genome-wide evidence for speciation with gene flow in *Heliconius* butterflies. Genome Res 23:1817-1828.

Martin SH, Davey JW, Salazar C, Jiggins CD. 2019. Recombination rate variation shapes barriers to introgression across butterfly genomes. PLoS Biol 17:e2006288.

Martin SH, Jiggins CD. 2017. Interpreting the genomic landscape of introgression. Curr Opin Genet Dev 47:69-74.

Martinsen GD, Whitham TG, Turek RJ, Keim P. 2001. Hybrid populations selectively filter gene introgression between species. Evolution 55:1325-1335.

Noor MA, Bennett SM. 2009. Islands of speciation or mirages in the desert? Examining the role of restricted recombination in maintaining species. Heredity (Edinb) 103:439-444.

Nosil P, Feder JL. 2012. Widespread yet heterogeneous genomic divergence. Mol Ecol 21:2829-2832.

Pickrell JK, Pritchard JK. 2012. Inference of population splits and mixtures from genome-wide allele frequency data. PLoS Genet 8:e1002967.

Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira MA, Bender D, Maller J, Sklar P, de Bakker PI, Daly MJ, et al. 2007. PLINK: a tool set for whole-genome association and population-based linkage analyses. Am J Hum Genet 81:559-575.

Rajora OP, Dancik BP. 1992. Genetic characterization and relationships of *Populus alba*, *P. tremula*, and *P. x canescens*, and their clones. Theor Appl Genet 84:291-298.

Ravinet M, Yoshida K, Shigenobu S, Toyoda A, Fujiyama A, Kitano J. 2018. The genomic landscape at a late stage of stickleback speciation: High genomic divergence interspersed by small localized regions of introgression. PLoS Genet 14:e1007358.

Renaut S, Grassa CJ, Yeaman S, Moyers BT, Lai Z, Kane NC, Bowers JE, Burke JM, Rieseberg LH. 2013. Genomic islands of divergence are not affected by geography of speciation in sunflowers. Nat Commun 4:1827.

Renaut S, Owens GL, Rieseberg LH. 2014. Shared selective pressure and local genomic landscape lead to repeatable patterns of genomic divergence in sunflowers. Mol Ecol 23:311-324.

Rennison DJ, Delmore KE, Samuk K, Owens GL, Miller SE. 2020. Shared patterns of genome-wide differentiation are more strongly predicted by geography than by ecology. Am Nat 195:192-200.

Rifkin JL, Castillo AS, Liao IT, Rausher MD. 2019. Gene flow, divergent selection and resistance to introgression in two species of morning glories (*Ipomoea*). Mol Ecol 28:1709-1729.

Roux C, Fraïsse C, Romiguier J, Anciaux Y, Galtier N, Bierne N. 2016. Shedding light on the grey zone of speciation along a continuum of genomic divergence. PLoS Biol 14:e2000234.

Schiffthaler B, Delhomme N, Bernhardsson C, Jenkins J, Jansson S, Ingvarsson P, Schmutz J, Street N. 2019. An improved genome assembly of the European aspen *Populus tremula*. bioRxiv:805614.

Sendell-Price AT, Ruegg KC, Anderson EC, Quilodrán CS, Van Doren BM, Underwood VL, Coulson T, Clegg SM. 2020. The Genomic Landscape of Divergence Across the Speciation Continuum in Island-Colonising Silvereyes (*Zosterops lateralis*). G3 (Bethesda) 10:3147-3163.

Shang H, Hess J, Pickup M, Field DL, Ingvarsson PK, Liu J, Lexer C. 2020. Evolution of strong reproductive isolation in plants: broad-scale patterns and lessons from a perennial model group. Philos Trans R Soc Lond B Biol Sci 375:20190544.

Smith JM, Haigh J. 1974. The hitch-hiking effect of a favourable gene. Genet Res 23:23-35.

Stankowski S, Chase MA, Fuiten AM, Rodrigues MF, Ralph PL, Streisfeld MA. 2019. Widespread selection and gene flow shape the genomic landscape during a radiation of monkeyflowers. PLoS Biol 17:e3000391.

Stettler R, Bradshaw T, Heilman P, Hinckley T. 1996. Biology of *Populus* and its

implications for management and conservation: NRC Research Press.

Suarez-Gonzalez A, Hefer CA, Christe C, Corea O, Lexer C, Cronk QC, Douglas CJ. 2016. Genomic and functional approaches reveal a case of adaptive introgression from *Populus balsamifera* (balsam poplar) in *P. trichocarpa* (black cottonwood). Mol Ecol 25:2427-2442.

Talla V, Johansson A, Dincă V, Vila R, Friberg M, Wiklund C, Backström N. 2019. Lack of gene flow: Narrow and dispersed differentiation islands in a triplet of *Leptidea* butterfly species. Mol Ecol 28:3756-3770.

Tavares H, Whibley A, Field DL, Bradley D, Couchman M, Copsey L, Elleouet J, Burrus M, Andalo C, Li M, et al. 2018. Selection and gene flow shape genomic islands that control floral guides. Proc Natl Acad Sci U S A 115:11006-11011.

Terhorst J, Kamm JA, Song YS. 2017. Robust and scalable inference of population history from hundreds of unphased whole genomes. Nat Genet 49:303-309.

Turner TL, Hahn MW. 2010. Genomic islands of speciation or genomic islands and speciation? Mol Ecol 19:848-850.

Tuskan GA, Difazio S, Jansson S, Bohlmann J, Grigoriev I, Hellsten U, Putnam N, Ralph S, Rombauts S, Salamov A, et al. 2006. The genome of black cottonwood, *Populus trichocarpa* (Torr. & Gray). Science 313:1596-1604.

Vijay N, Bossu CM, Poelstra JW, Weissensteiner MH, Suh A, Kryukov AP, Wolf JB. 2016. Evolution of heterogeneous genome differentiation across multiple contact zones in a crow species complex. Nat Commun 7:13195.

Wang B, Mojica JP, Perera N, Lee CR, Lovell JT, Sharma A, Adam C, Lipzen A, Barry K, Rokhsar DS, et al. 2019. Ancient polymorphisms contribute to genome-wide variation by long-term balancing selection and divergent sorting in *Boechera stricta*. Genome Biol 20:126.

Wang J, Street NR, Scofield DG, Ingvarsson PK. 2016. Natural Selection and Recombination Rate Variation Shape Nucleotide Polymorphism Across the Genomes of Three Related *Populus* Species. Genetics 202:1185-1200.

Wang M, Zhang L, Zhang Z, Li M, Wang D, Zhang X, Xi Z, Keefover-Ring K, Smart LB, DiFazio SP, et al. 2020. Phylogenomics of the genus *Populus* reveals extensive interspecific gene flow and balancing selection. New Phytol 225:1370-1382.

Wu CI. 2001. The genic view of the process of speciation. Journal of evolutionary biology 14:851-865.

Wolf JB, Ellegren H. 2017. Making sense of genomic islands of differentiation in light of speciation. Nat Rev Genet 18:87-100.

Yamasaki YY, Kakioka R, Takahashi H, Toyoda A, Nagano AJ, Machida Y, Moller PR, Kitano J. 2020. Genome-wide patterns of divergence and introgression after secondary contact between *Pungitius* sticklebacks. Philos Trans R Soc Lond B Biol Sci 375:20190548.