# Adaptive Introgression and Maintenance of a Trispecies Hybrid Complex in Range-Edge Populations of *Populus*

Vikram E. Chhatre<sup>1,4</sup>, Luke M. Evans<sup>2</sup>, Stephen P. DiFazio<sup>3</sup>, and Stephen R. Keller<sup>\*1</sup>

<sup>1</sup>Department of Plant Biology, University of Vermont, Burlington VT 05405

<sup>2</sup>Institute of Behavioral Genetics & Department of Ecology and Evolutionary Biology, University of Colorado, Boulder CO 80309

<sup>3</sup>Department of Biology, West Virginia University, Morgantown, WV 25606

<sup>4</sup>Current address: Wyoming INBRE Bioinformatics Core, University of Wyoming, Laramie WY 82071 \*Author for correspondence (srkeller@uvm.edu)

Keywords: rear-edge, admixture, adaptation, hybridization, species-complex, poplar

## Accepted Manuscript MEC-18-0586-R1

This article has been accepted for publication and undergone full peer review but has not been through the copyediting, typesetting, pagination and proofreading process, which may lead to differences between this version and the Version of Record. Please cite this article as doi: 10.1111/mec.14820

### Abstract

In hybrid zones occurring in marginal environments, adaptive introgression from one species into the genomic background of another may constitute a mechanism facilitating adaptation at range limits. Although recent studies have improved our understanding of adaptive introgression in widely distributed tree species, little is known about the dynamics of this process in populations at the margins of species ranges. We investigated the extent of introgression between three species of the genus Populus sect. Tacamahaca (P. balsamifera, P. angustifolia, and P. trichocarpa) at the margins of their distributions in the Rocky Mountain region of the United States and Canada. Using genotyping-by-sequencing (GBS), we analyzed  $\sim 83,000$  single nucleotide polymorphisms genotyped in 296 individuals from 29 allopatric and sympatric populations of the three species. We found a tri-species hybrid complex present throughout the zone of range overlap, including early as well as advanced generation backcross hybrids, indicating recurrent gene flow in this hybrid complex. Using genomic clines analysis, we found evidence of non-neutral patterns of introgression at 23%of loci in hybrids, of which 47% and 8% represented excess ancestry from *P. angustifolia* and *P.* balsamifera, respectively. Gene ontology analysis suggested these genomic regions were enriched for genes associated with photoperiodic regulation, metal ion transport, maintenance of redox homeostasis, and cell wall metabolites involved in regulation of seasonal dormancy. Our study demonstrates the role of adaptive introgression in a multi-species hybrid complex in range edge populations, and has implications for understanding the evolutionary dynamics of adaptation in hybrid zones, especially at the margins of species distributions.

## 1. Introduction

Hybridization is a dynamic evolutionary process resulting from the admixture and recombination of diversity between divergent parental genomes (Barton & Hewitt 1985; Harrison & Larson 2014), with diverse outcomes for genetic variability and fitness (Gompert & Buerkle 2016). Hybrid zones have long been recognized as natural laboratories for understanding the genetics of species divergence and reproductive isolation (Harrison & Larson 2016), including dissolution of species differences between weakly diverged taxa (Currat *et al* 2008) and the reinforcement of reproductive isolation resulting from reduced fitness and fecundity of hybrids (Fishman & Willis 2001). However, hybrid zones may also display increased genetic variance and may be an important source of evolutionary novelty (Rieseberg *et al* 2003).

Plant genomes seem to be especially tolerant of hybridization (Rieseberg 1997; Soltis & Soltis 2009; Whitney *et al* 2010; Suarez-Gonzalez *et al* 2018b), and the increase in trait variability can translate into community-level consequences for plant-herbivore or plant-microbe associations (Lamit *et al* 2011; Evans *et al* 2012; Floate *et al* 2016). Selection may also act on this increased variance to drive introgression of genomic regions between parental species, opening up physiological niches unavailable to either parental species and permitting expansion into new habitats by increasing niche breadth or stress tolerance (Rieseberg *et al* 2007; Whitney *et al* 2010; Goulet *et al* 2017).

In widely distributed plant species such as forest trees, hybrid zones are commonly associated with range shifts during climate change (Excoffier *et al* 2009). For example, among European trees, admixture is prominent along migration pathways representing northward range expansions out of Mediterranean glacial refugia (Petit *et al* 2003; Grivet *et al* 2006). Conversely, hybridization may also occur during the contraction of species ranges. For example, historical periods of glacial advance may have promoted secondary contact between related taxa retreating into shared refugial areas, evident as genomic admixture in areas where the rear edge of a species forms isolated "climate relict" populations (Hampe & Petit 2005; Hampe & Jump 2011; Woolbright *et al* 2014). Further, the effect of hybridization may be magnified in small, peripheral populations where the effects of

gene flow may be large relative to the demographic size of populations (Rieseberg 1997; Meirmans *et al* 2010). However, the role of hybridization in marginal populations is not well studied, and we currently have few examples of how hybridization shapes standing genetic variation in range edges.

In *Populus*, naturally occurring interspecific hybrids are commonplace where ranges overlap between interfertile species (DiFazio et al 2011). Furthermore, Populus hybrids are routinely used in breeding programs due to their abundant heterosis, so interfertility of the species and sections of the genus are well-characterized (Stettler et al 1996; DiFazio et al 2011). However, the consequences of introgression for the genetic architecture of traits and fitness, and the evolutionary forces maintaining species boundaries are less well known (Suarez-Gonzalez et al 2018b). For example, Geraldes et al (2014) found evidence for introgression between the sister species P. balsamifera and P. trichocarpa in northern British Columbia, which has resulted in asymmetric adaptive introgression of disease resistance between P. balsamifera and P. trichocarpa (Suarez-Gonzalez et al 2016, 2018a). A well-studied hybrid zone between P. angustifolia and P. fremontii on the Weber River in Utah shows primarily unidirectional introgression toward *P. angustifolia*, with substantial implications for ecosystem composition and function (Whitham et al 2006; Martinsen et al 2001). Complex hybrid ancestry has also been found between plantations of exotic poplar species and native populations of P. deltoides and P. balsamifera in Quebec (Meirmans et al 2010). In some cases, introgression occurs between multiple *Populus* species within the same region, such as the contact zone between P. trichocarpa, P. balsamifera, and P. angustifolia in the Oldman River drainage in Alberta, where tri-species ancestry proportions vary along a local gradient of elevation (Floate 2004; Floate et al 2016). In other cases, hybrids may only occur among some of the sympatric species at a site such as the San Miguel River in Colorado, where P. fremontii, P. deltoides, and P. angustifolia are sympatric, but bidirectional introgression occurs primarily between the latter two species only (Hersch-Green *et al* 2014).

The ecological, genetic, and genomic factors underlying the diverse outcomes in *Populus* hybrid zones are poorly understood, but some insights on selection and adaptive introgression can be gained by characterizing these zones with genome-wide molecular markers (Goulet *et al* 2017; Suarez-Gonzalez *et al* 2018b). For example, in Europe, population genomic analyses of replicate

hybrid zones between P. alba and P. tremula have shown strong post-zygotic isolation due to genetic incompatibilities between species, but that isolation was not strong enough to prevent introgression driven by selection (Lexer *et al* 2010; Lindtke *et al* 2014). Moreover, whole genome studies based on common garden experiments in these replicate hybrid zones indicate selection against recombinant hybrids despite fertility of F1's (Christe *et al* 2016). This suggests an overall tension may exist in *Populus* hybrid zones between reproductive isolation limiting introgression of genomic diversity as a whole, and positive selection driving the introgression of certain alleles favorable in alternative genomic backgrounds.

Here, we describe a new tri-species *Populus* hybrid zone in the Rocky Mountains of North America, and test for adaptive introgression between three *Populus* species in section *Tacamahaca*: *P. balsamifera*, *P. trichocarpa*, and *P. angustifolia*. For two of these species (*P. balsamifera* and *P. trichocarpa*), this region consists of numerous isolated rear edge populations likely representing climate relicts descended from glacial refugial populations (Keller *et al* 2011; Geraldes *et al* 2014). In allopatric portions of their range, strong local adaptation along gradients of photoperiod and temperature are known for a suite of phenological and physiological traits related to dormancy timing and growth, with corresponding signatures in the genome consistent with local selection on standing genetic variation (Keller *et al* 2011, 2012; Evans *et al* 2014; Zhou *et al* 2014; McKown *et al* 2014; Evans *et al* 2016). Here we use population genomic analysis of genome-wide SNP variation in allopatric populations of all three species and their sympatric hybrid zone to address the following questions:

- 1. What are the patterns of population structure in these species in their zone of range-overlap?
- 2. How extensive is the level of genome introgression in hybrid populations? Is there evidence of advanced generation hybrids?
- 3. Is introgression in genotypes of hybrid ancestry being driven by selection on certain genomic regions?
- 4. Do regions experiencing selective introgression between species correspond to regions that have undergone positive selection in the parental species?

We discuss our results within the specific context of adaptive introgression as a source of adaptive variation in marginal populations, and more broadly in terms of phylogeography of range edges and the importance of hybrid zones in these regions for understanding the evolution of species complexes.

## 2. Materials & Methods

### (a) Sampling

We sampled dormant stem cuttings from 242 individuals across 21 populations of P. balsamifera and *P. angustifolia* (Table 1). We used range maps and herbarium records to identify putative hybrid zones between P. balsamifera and P. angustifolia where their ranges overlap in the Rocky Mountains of the USA and Canada (Figure 1). Within this region, P. balsamifera is reported to form numerous rear edge climate relict populations, geographically separated from the core of its range, while *P. angustifolia* is common but patchily distributed along floodplains of rivers and mountain streams. Eight putative hybrid zone populations were sampled from Arizona, Utah, Wyoming and Alberta where the species ranges are sympatric. Allopatric sites were included for reference, and were sampled from the range core of each species. For *P. balsamifera*, this consisted of nine populations collected from its natural distribution in Saskatchewan, Ontario, Quebec and Manitoba, from areas geographically well separated from any known zones of range overlap and hybridization. These samples are part of a larger study investigating the population genomics of local adaptation in *P. balsamifera* (Chhatre & Keller *et al. in preparation*). Accordingly, trees within the putative hybrid zone were phenotypically selected with a preference for individuals exhibiting P. balsamifera-like traits (bud size, leaf shape), and thus do not represent a random sample of all possible hybrid genotypes. Samples of *P. angustifolia* were from a previous study of microsatellite variation and population genetic structure (Evans et al 2015). Due to the relatively few samples of *P. angustifolia* within certain populations, we aggregated samples from geographically proximal and genetically undifferentiated sites into a single population (UTAH).

We also included samples from a closely related third species, P. trichocarpa, which is not historically reported from the area of the Rocky Mountains containing the putative hybrid zone, but

does occur in nearby areas of Montana, Idaho, Utah, and Nevada (Figure 1). Our aim was to identify any contemporary or historical P. trichocarpa ancestry in our samples of P. balsamifera and P. angustifolia, despite apparent absence of this species in the immediate neighborhood of the putative hybrid zone. Sample locations of allopatric P. trichocarpa (N=54 individuals from 8 populations) were specifically chosen from river drainages in Washington and British Columbia, with no known introgression from related poplar species based on a previous whole-genome sequencing study (Evans *et al* 2014). In total, the current study consisted of 296 individuals from 29 populations of the three species.

Cuttings were rooted and grown in a controlled environment and fresh leaves were used for isolation of whole genomic DNA using DNeasy 96 Plant Mini Kits (Qiagen, Valencia, CA, USA). DNA was quantified using the fluorometric Qubit BR assay (Invitrogen) and run on 1% agarose gels to determine purity and presence of high molecular weight DNA.

#### (b) Illumina Sequencing & SNP Genotyping

We used genotyping-by-sequencing (GBS) (Elshire *et al* 2011) in order to obtain high-density SNP genotypes for all newly sequenced individuals of *P. balsamifera* and *P. angustifolia* (*P. trichocarpa* were previously sequenced). Briefly, genomic libraries were prepared by digesting 100 ng of genomic DNA with the restriction endonuclease EcoT22I and ligating 96 unique barcoded adapters (following Elshire *et al* (2011)) to the ends of resulting restriction fragments. The resulting barcoded fragments were pooled in equimolar quantities, purified using the QIAquick PCR purification kit, PCR amplified for 18 cycles to append Illumina sequencing primers, cleaned using the QIAquick PCR purification kit, and the resulting library assessed for fragment size distribution on a Bioanalyzer. Each GBS library was then single-end sequenced (100 bp reads) on an Illumina HiSeq 2000 to 48-plex. GBS library preparation and Illumina sequencing were performed by the Cornell University Genomic Diversity Facility (Ithaca, NY USA).

Raw sequence data was processed with the Tassel GBS Pipeline (Glaubitz *et al* 2014). Sequence reads were filtered for quality (perfect match to expected barcode sequence, presence of the restriction site, no N's), trimmed to a length of 64 bp, and aligned to the *P. trichocarpa* reference genome

version 3.0 (Tuskan *et al* 2006) using BWA (Li & Durbin 2009). Single nucleotide polymorphisms (SNPs) were called based on the reference assembly and duplicate variants were removed. Details of SNP genotype calling for *P. trichocarpa* were reported previously (Evans *et al* 2014). SNP genotype data were converted to Variant Call Format (VCF) files and used for downstream genotype filtering using VCFtools (Danecek *et al* 2011).

We filtered SNPs to keep only biallelic sites, with individual-wise missing data <50%, locus-wise missing data <20%, genotype quality (GQ)  $\ge 90$ , allele depth  $\ge 5$ , and indels removed. We did not set a minimal minor allele frequency threshold to avoid discarding rare alleles that may be under selection. Finally, we tested all SNPs for deviations from Hardy-Weinberg equilibrium. Composite data sets involving two or more species could lead to a deficit of heterozygotes due to Wahlund effect. To avoid this bias, we performed  $\chi^2$  tests of Hardy-Weinberg violations within each species individually and removed sites with heterozygote excess (P < 0.001). After combining samples from all three species, we obtained 83,835 SNPs from all 19 chromosomes that were polymorphic in at least one species. All computations were performed on the Mount Moran high performance computing cluster at the University of Wyoming (ARCC 2012).

### (c) Population Structure, Admixture & Migration Events

We assessed overall levels of population structure across all 83,835 SNPs in 29 populations of the three species. Genetic differentiation  $(F_{\rm ST})$  was measured (1) between species, (2) among allopatric populations within species, (3) among populations in the hybrid zone, and (4) over all populations using VCFtools. We also estimated pairwise  $F_{\rm ST}$  (Weir & Cockerham 1984) between all population pairs using a random subset of 10,000 SNPs using the pairwise.WCfst function from the R library HIERFSTAT (Goudet & Jombart 2015). To assess evidence for mixed genetic ancestry within individuals, we used maximum likelihood genotype clustering implemented in ADMIXTURE (Alexander *et al* 2009). For this analysis, we evaluated models from K = 1 to 5 subpopulations. Inference of K was determined based on the lowest error rate generated using 2000 bootstrap replicates across 10–fold cross-validation iterations. Barplots of cluster membership were visualized at K = 2-4 using a modified DISTRUCT v2.3 Python script (Raj *et al* 2014; Chhatre 2018). To

test for potential introgression from other Populus species, we repeated the ADMIXTURE analysis by combining our study samples with those from 5 additional species (P. deltoides, P. nigra, P.euphratica, P. tremula, and P. fremontii). Due to different data sources, this additional species dataset consisted of 11,794 of the original 83,835 SNPs. We applied the same run conditions to this data set as in the original ADMIXTURE run and tested for up to 10 Hardy-Weinberg genetic clusters (i.e. K subpopulations).

Two of our study species, *P. trichocarpa* and *P. balsamifera*, are recently speciated sister taxa with a shallow divergence time of ~76,000 years (Levsen *et al* 2012), and share a more distant common ancestor with *P. angustifolia* (Hamzeh & Dayanandan 2004; Cervera *et al* 2005). Given the close phylogenetic relationships, and proximate current geographic range limits of all three species (Levsen *et al* 2012), the potential for admixture events arising from secondary contact during glacial advance and retreat may have existed for millennia (Suarez-Gonzalez *et al* 2016, 2018a). To better understand the joint history of these populations and the most likely migration events between them that generated admixture, we used TREEMIX (Pickrell & Pritchard 2012) to analyze a Gaussian model of genetic drift with admixture to estimate a population phylogeny. The phylogenetic tree was rooted with the more distantly related Eastern Cottonwood (*P. deltoides* WV94 v2.1 - available from https://phytozome.jgi.doe.gov). This analysis was based on a subset of SNPs (n=11,794) that were available across all four taxa. We accounted for non-independence due to linkage disequilibrium by analyzing windows of 500 contiguous SNPs. We allowed for gene flow between populations by increasing the total number of migration events until the residual error was minimized (Pickrell & Pritchard 2012).

#### (d) Extent of Recurrent Hybridization

We evaluated evidence that populations in this hybrid zone produce early (F1) or advanced generation (F2 and backcross) genotypic classes using NEWHYBRIDS (Anderson & Thompson 2002). We restricted our analysis to the 375 SNPs showing the highest level of allele frequency divergence between the parental species (e.g. fixed or nearly fixed differences between allopatric populations. We evaluated 12 independent progeny genotypic classes (Table 3) comprising parental, F1, F2 and

advanced generation backcross hybrids. Because there is not a straightforward means to estimate genotypic classes among more than 2 species in NEWHYBRIDS, we performed two separate analyses. First, we combined P. balsamifera and P. trichocarpa individuals together into a single "balsam poplar" group. Allopatric populations of the balsam poplars and P. angustifolia were classified as the two parental species, and the sympatric populations were designated as putative hybrids. For the second analysis, we excluded all P. trichocarpa samples, retaining only P. balsamifera and P. angustifolia and the hybrids. For each data set, we set up three independent runs of NEWHY-BRIDS using unique seed numbers and Jeffrey's prior for Pi and Theta parameters, and explored the MCMC parameter space for a total of 200,000 sweeps, of which we discarded the first 100,000 sweeps as burn-in.

#### (e) Selection & Differential Introgression

Selection in hybrid zones may favor the introgression of genomic regions that provide an adaptive advantage in hybrids, while preventing introgression of other regions that result in loss of local adaptation (i.e., "genomic islands of divergence") or are involved in reproductive isolation (Wolf & Ellegren 2017). We studied the extent to which genomic introgression results in over-representation of locus-specific ancestry in hybrid genotypes using the maximum likelihood based INTROGRESS package in R (Gompert & Buerkle 2010). Because the model implemented in INTROGRESS is limited to two parental species, we focused this analysis on introgression between P. angustifolia (parent 1), P. balsamifera (parent 2) and the putative hybrids (progeny), and omitted P. trichocarpa samples. Informative inference of introgression requires analyzing genomic regions that show a large allele frequency differential between parental species. Therefore, we analyzed a subset of 7,523 SNPs that showed an absolute allele frequency difference of 0.8 or higher between species. We used INTROGRESS to estimate a hybrid index of interspecific ancestry, analogous to cluster membership in genotypic clustering analyses when K = 2. The hybrid index was then used to compare the observed genomic clines of locus-specific ancestry to the expected genotype proportions under a neutral demographic model, based on 1,000 parametric simulations for each locus. Deviations from neutral expectations at each locus were recorded as the relative genotypic contribution of each of the two parent species to the hybrids. We interpreted SNPs violating neutral expectations at P = 0 as evidence of differential introgression. SNPs with significant intermediate p-values (0 > P <= 0.05) were not considered as outliers for the purposes of further interpretation.

The parametric model behind INTROGRESS may be sensitive to neutral demographic effects if the admixed population is very small, time since admixture is large, and/or the hybrid zone is composed of multiple structured populations, leading to the potential for false positives among the outlier loci for adaptive introgression (Gompert & Buerkle 2009, 2011). The *Populus* hybrid zone we sampled is genetically diverse and mostly consists of hybridization that has occurred within the last several generations (see Results). Thus the potential for false positives in our analysis is primarily from genetic structure among populations within the hybrid zone. To guard against this possibility, we randomly dropped 1 of the 8 hybrid populations and re-estimated genomic clines based on the remaining 7 hybrid populations plus the original set of parental populations. We repeated this procedure 100 times and identified robust outliers as those loci that had an INTROGRESS *P*-value = 0 in all 100 iterations.

To assess if selection within the parental species contributed to genomic regions experiencing differential introgression in hybrids, we tested for overlap between introgression outliers and regions under positive selection in the parental populations. Selective sweeps within the pure parental populations of *P. angustifolia*, *P. balsamifera*, and *P. trichocarpa* were tested using the composite likelihood ratio (CLR) test of Nielsen *et al* (2005), as implemented in SweeD (Pavlidis *et al* 2013). We computed the CLR in chromosomal regions of 100kb, and retained regions with CLR in the top 1% within each species for comparisons to introgression outliers. Significance of overlap between introgression outliers and genomic regions under selective sweeps was tested using 1000 permutations of the sweep windows, and then re-calculating the random overlap with the observed INTROGRESS SNP loci. We also tested for evidence of local adaptation within *P. balsamifera* and *P. trichocarpa* (*P. angustifolia* was not included due to small sample size within most populations). Local adaptation in *P. balsamifera* was tested using BAYESCAN (Foll & Gaggiotti 2008). Run conditions were optimized using 20 pilot runs each with a length of 5,000 sweeps. An initial 50,000 sweeps were discarded as burn-in and data were collected for the next 50,000 sweeps. inferred the strength of positive selection from the locus-specific effect  $\alpha$ , with  $\alpha > 1$  indicative of a history of local adaptation. We set a rigorous prior expectation of 1 locus in 10,000 to be under local selection to minimize false positives (Lotterhos & Whitlock 2014). We similarly tested for introgressed regions associated with *P. trichocarpa* loci under local adaptation using previously reported genome scans for local selection in that species (Evans *et al* 2014). The *P. trichocarpa* analysis of local adaptation was based on results from BAYENV (Günther & Coop 2013) using the first and second principal components of multivariate climate variables (see Evans *et al* (2014) for

#### (f) Functional Enrichment

additional details).

We tested for functional enrichment of introgression outliers from INTROGRESS using gene ontology (GO) terms based on the *P. trichocarpa* reference genome annotation v3.0. Statistical support for enrichment of GO terms with outlier SNPs was assessed using the R package SNP2GO (Szkiba et al 2014). We repeated this analysis for different sets of loci: (a) introgression outliers from a single INTROGRESS run with all N=8 hybrid populations, (b) introgression outliers based on the 100 resampled datasets, (c) the subset of (b) outlier loci that also showed significant deviation among the expected genotypic classes, and (d) two subsets of loci in (c) showing preponderance vs deficit of *P. angustifolia* alleles in the hybrids. Given the large expected degree of linkage disequilibrium in early generation hybrids and our relatively modest marker density, we compared candidate vs non-candidate SNPs for GO terms associated with genes within a window of 1 Mb upstream and downstream from each SNP. GO term enrichment was generated by randomly re-sampling SNPs 100,000 times to determine null expectations, and the significance of enrichment determined using a q-value of 0.1. Functional redundancy of enriched GO terms was summarized using REVIGO (Supek et al 2011). We used REVIGO to cluster closely related GO terms based on their degree of semantic similarity (>70% based on the SimRel similarity measure), and retained terms representative of each cluster based on keeping the term with the highest FDR-corrected  $-\log 10P$ value from the enrichment test in SNP2GO. We visualized the reduced REVIGO clusters using multi-dimensional scaling of the pairwise semantic similarity matrix.

### 3. Results

#### (a) Patterns of Population Structure & Admixture

Analysis of population structure outside of the hybrid zone revealed strong genetic differentiation among populations of *P. angustifolia* ( $F_{\rm ST}$ =0.153), confirming previous findings in this species (Evans *et al* 2015), whereas the balsam poplars showed little genetic differentiation among populations for either *P. balsamifera* ( $F_{\rm ST}$ =0.046) or *P. trichocarpa* ( $F_{\rm ST}$ =0.028). Pairwise population divergence based on a subset of randomly chosen 10,000 SNPs ranged from as low as  $F_{\rm ST}$ =0.0021 (MMT *vs* OFR) to 0.754 (AZPW *vs* SKYKOMISH), with the largest pairwise differences between *P. angustifolia* and the balsam poplars, and lower differentiation between *P. balsamifera* and *P. trichocarpa* (see Supplementary Table 1). As expected, the mean differentiation between the two balsam poplars was substantially lower ( $F_{\rm ST}$ =0.037), consistent with their more recent evolutionary divergence, and in line with previous estimates (Keller *et al* 2011; Evans *et al* 2014). Within the hybrid zone, genetic differentiation between *P. angustifolia* and *P. balsamifera* as a whole ( $F_{\rm ST}$ =0.094).

Ancestry analysis using genotypic clustering at K=2 subpopulations showed little to no admixture within the allopatric populations of *P. balsamifera*, *P. trichocarpa* and *P. angustifolia* outside of the region of secondary contact (Figure 2a). The major division of population structure at K=2was between *P. angustifolia* and the two species of balsam poplars, with admixture clearly evident for populations in Colorado, Wyoming, and Alberta (MSG, CYH, SSR, JKH, RMP, RMPANG) as well as some populations in Utah where *P. angustifolia* is the only species present (UTAH & WYSR). At K=3, *P. trichocarpa* separated from *P. balsamifera* (Figure 2b), with substantial admixture evident between these species in the Rocky Mountains of Alberta and Wyoming (CYH, JKH, SSR, MSG). Some individuals in these populations also contained ancestry from *P. angustifolia*, indicating the presence of tri-species hybrids in the regions of range overlap. At K=4 (Figure 2c), additional population substructure was revealed within *P. angustifolia*, but the pattern of interspecific admixture observed at K=2 and K=3 remained unchanged. Combined ADMIXTURE analysis with the five outgroup species did not reveal significant admixture from any of them into our study populations. Cross validation error was the lowest at K=3 where our three main study species clustered separately. At K=4 the five outgroup species formed a single homogenous cluster separate from our original study sample (Figure S5). Further K clustering revealed substructure within two of the three study species (*P. balsamifera* and *P. angustifolia*), but no introgression from the outgroup species.

#### (b) Historical Divergence & Gene Flow

The demographic history of the sampled populations revealed a complex pattern of divergence among the three parental species accompanied by extensive gene flow (Figure 3 & Figure S3). The population tree estimated by TREEMIX confirmed established phylogenetic relationships in section *Tacamahaca*, with a greater proportion of the total genetic drift separating populations of P. angustifolia from the two balsam poplars (Figure 3). Populations showing high levels of interspecific admixture in genotypic clustering (MSG, SSR, and JKH) also occupied basal positions to the main population clusters of P. balsamifera and P. trichocarpa, diverging either prior to the species split (JKH) or prior to the main group of *P. balsamifera* populations (MSG, SSR) (Figure 3). Allowing for gene flow between populations resulted in an estimated 7 migration events across the tree, with the most prominent migration pattern showing extensive bidirectional gene flow between populations of *P. balsamifera* and *P. angustifolia*. Gene flow between these species was strongest between the sympatric populations in Colorado (RMP and RMPANG), but was also extensive elsewhere throughout the hybrid zone (e.g. from RMPANG into JKH, from ABOR into MSG, and from an intermediate ancestor of *P. angustifolia* populations into CYH). Gene flow was detected from the common ancestor of the two balsam poplars into the Wyoming WYSR population of *P. angustifolia* possibly indicating the historic presence of an ancestral balsam poplar gene pool contributing to admixture. There was also evidence for gene flow into RMPANG from outside the hybrid zone involving allopatric P. balsamifera. Increasing the number of migration routes from 7 to 10 improved model fit and reduced the residual variance (Figure S3). At these higher levels of migration, gene flow between the balsam poplars was evident from a population ancestral to P. trichocarpa into the Wyoming population SSR of P. balsamifera (Figure 3c & d): a pattern that was also evident from the ADMIXTURE analysis (Figure 2). Finally, a gene flow event from the Arizona populations of *P. angustifolia* into *P. deltoides* was apparent, potentially reflecting historical hybridization during previous range overlap.

#### (c) Extent of Introgressive Hybridization

The hybrid index showed close correspondence to ancestry proportions from ADMIXTURE, ranging from close to 0 (P. angustifolia-like) to approaching 1 (P. balsamifera-like). A number of hybrids were identified with hybrid index ranging somewhat continuously from 0.5 to 1, indicative of intermediate (F1, F2) and advanced generation hybrids backcrossing towards P. balsamifera (Figure 4). Hybrid individuals tended to show slightly reduced levels of interspecific heterozygosity relative to maximal expectations from their hybrid index, while parental types (hybrid index near 0 or 1) were slightly above the expected values. These probably reflect shared ancestral polymorphisms segregating within pure populations of each species, although heterozygote deficit in hybrids could also arise if one of the two alleles in a hybrid was divergent from the reference genome and thus failed to map properly during variant calling.

Assignment of individuals into hybrid genotype classes between P. angustifolia and the balsam poplars (both with and without P. trichocarpa) showed evidence for F1, F2, and several advanced generation backcrosses between P. angustifolia and the balsam poplars (Table 2). The majority of individuals with ancestry values near 0.5 in admixture analyses consisted of F1 individuals (N=21), although three F2 individuals were also detected. In addition, we uncovered multiple advanced generation backcross hybrids (N=8), with the majority of individuals (7 out of 8) reflecting backcrosses towards the P. balsamifera parent.

#### (d) Selection Driven Introgression

Genomic clines estimated by INTROGRESS showed significant deviation from genotype frequencies expected under a neutral demographic model (P=0) for 3,723 of the 7,523 SNPs. Of the 3,723 significant SNPs, 3,263 showed significant deviation from neutral frequencies for at least one of the inter-specific genotypic classes (Table 3). Of these, 1,592 sites (43%) showed more *P. angustifolia* 

homozygous ancestry than expected, whereas 949 (25%) showed more P. balsamifera homozygous ancestry than expected; interspecific heterozygous excess and deficit occurred in similar proportions (28% and 33%, respectively). Genomic regions experiencing differential introgression were evident as regions of inter-specific ancestry that deviated from the genomic background for a given hybrid index (Figure 5 & Figure S4). Of particular note was the high frequency of genomic regions showing P. angustifolia ancestry in hybrids backcrossed to P. balsamifera, especially ones evident across multiple populations (Figure S4).

The resampling analysis to assess the robustness of introgression outliers to exclusion of individual populations identified a set of 1,764 SNP outliers out of 7,523 tested that were significant in all 100 INTROGRESS replicate runs (Figure 6a). Of these 1,764 outliers, 1,016 also deviated significantly in their expected genotypic classes across all 100 runs (Table 3). These outliers fell into three major categories: 832 (47%) showed more *P. angustifolia* homozygous ancestry than expected, while 138 (8%) showed more *P. balsamifera* homozygous ancestry than expected. Interspecific heterozygote excess and deficit were observed in 308 (17%) and 213 (12%) SNPs of the total respectively.

If selection is indeed driving asymmetric introgression of P. angustifolia ancestry into the genomic background of P. balsamifera, then we should expect these genomic regions to become homogenized within the hybrid zone and show less divergence between species than genomic regions acting neutrally or resisting introgression. Consistent with this prediction, we found per-locus  $F_{\rm ST}$ for P. angustifolia-derived alleles in the hybird zone to be ranging from 0.021 to 0.777 for the combined set of neutral and candidate loci; however, the distribution of candidate loci was markedly shifted towards lower  $F_{\rm ST}$  values compared to putatively neutral loci (Figure 6b). This reduction in  $F_{\rm ST}$  among outlier loci is consistent with selection introgressing genomic regions of P. angustifolia ancestry that are adaptive on a P. balsamifera genomic background, while the rest of the genome reflects the neutral process of divergence and admixture between the species.

Of the 1,764 introgression outliers that were significant in all 100 resampling iterations, 51 overlapped with the upper 1% of genomic regions (N=120 windows) that showed evidence of a selective sweep in one of the three parental species (Figure S6). Given a genome-wide number of

100kb windows (N=3,946) tested in 3 species, we would expect the number of outlier SNPs to fall within 120 windows to be  $1764 \times \frac{120}{3946} = 54$ , assuming independence of beneficial mutations among species and linkage equilibrium among loci. Thus, the observed overlap is close to what is expected to occur by chance. In accordance, permutation tests indicated that the observed overlap between the outliers was not significant (Figure S7). Consistent with this finding, while there was evidence of local adaptation among *P. balsamifera* allopatric populations for 10 SNPs (FDR < 0.1), none of these loci were associated with differential introgression in the hybrid zone. Similarly, among 78 genomic regions showing local adaptation in a broader population genomics study in *P. trichocarpa* (Evans *et al* 2014) (BAYENV PC1/PC2), none showed overlap with the 1,764 introgression outlier SNPs from our study.

#### (e) Functional Enrichment of Introgression Outliers

Genomic regions flanking SNPs that showed differential introgression between P. balsamifera and P. angustifolia for the INTROGRESS run with all 8 hybrid populations were significantly enriched for 91 gene ontology (GO) terms (Table S2a). Clustering of GO terms based on similarity scores resulted in 52 minimally redundant groups, including binding and transport of metals and other cations (Zn, Fe, Mg, and Al) across cell membranes (Figure 7; Table S2). Other enriched GO terms with minimal redundancy involved biosynthesis and metabolism of cell wall components such as cellulose and cellular glucan (i.e. apoplast, glucan and xyloglucan metabolism), and maintenance of cell redox homeostasis (Figure 7). Among the 1,764 loci significant across all 100 resampled runs, there was a single enriched term (GO:0009416, photoperiodic regulation). However, among the subset of the 1,764 outliers showing significant deviation from expected proportion of genotypic classes, 11 GO terms were enriched for loci with excess P. angustifolia ancestry (832 outliers), including maintenance of cell redox homeostasis through metal ion binding (Figure 7; Tables S2b & S2c). In contrast, none were enriched for those showing a lower proportion of P. angustifolia or a higher proportion of P. balsamifera ancestries than expected.

### 4. Discussion

Few studies have examined how gene flow between species at the edges of their distributional ranges may contribute to adaptation to peripheral habitats or the maintenance of species boundaries. Populations near the edges of species ranges typically occur at lower abundances and occupy stressful environments near their physiological limits, potentially amplifying the evolutionary consequences that hybridization has for transferring adaptive variants between divergent genomes or driving local gene pools to extinction via repeated directional gene flow. While hybridization is well documented in the genus *Populus* in North America (e.g. Eckenwalder (1984); Talbot *et al* (2012); Geraldes *et al* (2014)), and recent reports have highlighted the importance of adaptive introgression between *P. balsamifera* and *P. trichocarpa* (Suarez-Gonzalez *et al* 2016, 2018a), our study provides the first genome-wide assessment of adaptive introgression between *Populus* species specifically where they overlap at their range limits.

Our results identified a tri-species complex of *Populus* in the Northern Rocky Mountains, along the southern and eastern range edges of the two balsam poplars, *P. balsamifera* and *P. trichocarpa*, and near the northern range edge of *P. angustifolia*. Populations from this genetically diverse zone of overlap contained hybrids with complex patterns of inter-specific ancestry, including both early (F1, F2) and advanced generation backcrosses. There was extensive genome-wide evidence for selection-driven introgression between the parental genomes in the hybrid zone. These signatures of selection in range-edge populations may reflect the importance of hybridization as a source of segregating genetic variation for adaptation to stressful or novel environmental conditions found near the edges of a species range (Rieseberg 1997). We discuss these results in light of the unique population genetic processes occurring in range edge populations, and the role that hybridization plays in their genetic structure and adaptive potential.

#### (a) Recurrent Gene Flow in Range-Edge Populus Hybrid Zones

We found that recurrent interspecific gene flow characterizes the genomic diversity of isolated *Populus* populations growing in the zone of overlap between the range edges of *P. balsamifera*, *P.* 

trichocarpa, and P. angustifolia. Genotypic clustering clearly separates each of these species into relatively pure demes where populations are allopatric, and reveals individuals with a range of mixed ancestry in populations within the zone of range overlap (Figure 2). Most populations contained hybrids between two species, for example the RMP population showed frequent F1 hybrids between P. balsamifera and P. angustifolia, while the SSR population showed hybridization exclusively between P. balsamifera and P. trichocarpa. Interestingly, several individuals in other populations appeared to be tri-hybrids, reminiscent of other *Populus* hybrid zones where tri-hybrids exist between populations of P. balsamifera, P. angustifolia, and P. deltoides (Floate et al 2016). While complex admixture patterns could also result due to ancestry contributions from other *Populus* spp. outside of our sample, we found no evidence of introgression from 5 other related species (Fig. S5). Thus this hybrid zone in the Rocky Mountains appears to be a unique zone of introgression among these 3 lineages where the edges of their ranges overlap.

While many hybrids appear to be F1's or F2's, a range of advanced generation hybrids were also present indicating repeated backcrossing to P. balsamifera. The prevalence of backcross hybrids with predominantly P. balsamifera ancestry likely reflects our sampling of individuals with more balsamifera-like traits, but we cannot rule out a bias in the direction of hybrid compatibility between parental genomes. Such asymmetric introgression is known in the P. angustifolia and P. fremontii hybrid zone in Utah where the backcrosses have been observed only with the narrowleaf cottonwood (Keim *et al* 1989; Whitham *et al* 2006), and is also reported between P. trichocarpa and P. balsamifera hybrid zones, in which introgression is biased towards P. balsamifera (Suarez-Gonzalez *et al* 2018a). Future sampling should target a broader range of individuals within this hybrid zone to determine if hybrid viability varies asymmetrically between the genomes of the parental species. Nevertheless, our current results suggest that although most hybrid individuals that we sampled may be of relatively recent origin, introgression involving all three species can occur in zones of secondary contact and leads to viable progeny and the formation of advanced generation hybrids.

The evolutionary history of population splitting and mixing further suggests that interspecific hybridization at range edges may have been recurrent in the history of these three species. Migration events appear to be frequent and bidirectional between *P. balsamifera* and *P. angustifolia* in this

region (Figure 3 & Figure S3). Additional evidence for a longer-term history of hybridization between these two species comes from the inference of migration events involving current allopatric populations of *P. angustifolia* in Wyoming (WYSR) and the common ancestor of contemporary *P. balsamifera* and *P. trichocarpa* populations. Thus, hybrid genomes in the area of sympatry in the Rocky Mountains probably contain vestiges of hybridization events that happened many generations ago, potentially even pre-dating the divergence and range expansion of species into their current distributions (i.e., occurring during previous secondary contact in glacial refugia).

#### (b) Selection Driven Introgression in Hybrid Populus Genomes

One of the mechanisms for maintaining species integrity in the face of recurrent gene flow between species is a reduction in fitness of recombinant hybrids due to intrinsic incompatibilities or dilution of local adaptation. Christe *et al* (2016) documented that hybrid zones between *P. tremula* and *P. alba* in Europe showed selection against recombinant hybrids, leading to overrepresentation of F1 hybrids in the progeny and an apparent lack of advanced generation hybrids. In our study, we found little evidence of selection acting against recombinants. F2 and advanced generation hybrids were relatively common in hybrid zone populations, and genomic clines analysis showed no consistent excess or deficit of interspecific heterozygosity among introgressed regions (Figure 4).

Interspecific hybridization between divergent lineages can create large blocks of loci in linkage disequilibrium, creating enhanced potential for hitchhiking selection in hybrid zones. Genomic clines analysis with resampling revealed a large portion of the genome (23% of tested SNPs) that showed patterns of excess ancestry in hybrids beyond neutral demographic expectations based on the overall degree of admixture (i.e., hybrid index). Other studies of introgression in hybrid zones have also found a high ratio of candidate to neutral loci, such as in spruce (52%; De La Torre et al (2014); 76%; Hamilton et al (2013)), house mice (91%; Teeter et al (2010)), and European poplars (35%; Lexer et al (2010)). It is prudent to be cautious about interpreting deviations from neutral expectations during genomic clines analysis as evidence of selection given that the INTROGRESS method (Gompert & Buerkle 2009) can be sensitive to effects of genetic drift. For example, if reference (parental) population frequencies are not representative of the species where the hybrid

zones form, this could lead to a bias in generating the neutral expectation (personal comm. A. Buerkle). This could arise due to high genetic drift between core and edge populations. However, the high reproducibility of outlier loci under our resampling approach (Figure 6a), and the excess of *P. angustifolia*-derived alleles in the hybrids as evidenced by the shift of outlier loci towards lower  $F_{\rm ST}$  compared to non-outlier loci (Figure 6b), point to introgression of *P. angustifolia* ancestry into the *P. balsamifera* genomic background at levels that exceed demographic expectation.

The candidate SNPs for adaptive introgression in our study represented genomic blocks preferentially inherited from one of the two parental genomes, most often from P. angustifolia into a P. balsamifera genomic background (47% of introgression outliers). Recurrent backcrossing may present an opportunity to capture adaptive genetic variation through introgression that provides a selective advantage in suboptimal range-edge environments. Gene ontology terms associated with these SNPs were significantly overrepresented for photoperiodic regulation and metabolic functions such as metal ion transport across membranes and maintenance of cell redox homeostasis (Figure

7).

Several enriched GO terms among the candidate loci point to potential adaptive processes that may underlie the signal of differential introgression. *P. balsamifera* occupies a wide geographic range and shows substantial local adaptation in growth and phenology to latitude driven gradients of climate (Soolanayakanahally *et al* 2009; Keller *et al* 2011, 2012). In particular, bud phenology is closely tied to length of photoperiod (Olson *et al* 2013) and chilling requirement (Guy 2014). In a recent range-wide study of flowering time genes, Keller *et al* (2017) demonstrated that rearedge populations of *P. balsamifera* differ significantly in their genomic signatures of locally adaptive responses to photoperiod, temperature and elevation. In the present study, GO term enrichment for introgression outliers involved photoperiodic regulation as well as glucan metabolism which affects the cellular matrix deposited in plant cell walls, and plays diverse physiological roles including dormancy cycling. For example Rinne *et al* (2011) demonstrate a role for 1,3- $\beta$  glucan in the release of seasonal endodormancy in *Populus* (see also Brunner *et al* (2014)). Given the extensive introgression between *P. angustifolia* and *P. balsamifera* in this region, and the enrichment of genomic regions involved in dormancy, one possibility is that admixed populations of *P. balsamifera*  along its rear range edge draw upon introgression of adaptive variants from P. angustifolia in order to adapt their dormancy cycling to the range of environmental cues present in marginal environments.

A second signal of functional enrichment that was consistent across sets of introgression outliers involved cation and metal transport across cell membranes and the maintenance of redox homeostasis. Absorption of metals by cells can lead to generation of excess amounts of reactive oxygen species (ROS), contributing to oxidative stress (Gill & Tuteja 2010). ROS toxicity can disrupt a number of important metabolic processes, although ROS can also be important signaling molecules to initiate plant defense mechanisms (Sytar et al 2013). Our study populations growing in the metal rich soils of the Rocky Mountains are likely exposed to metal ion toxicity, where the integrity of redox homeostasis may be essential for survival. Cadmium exposure, for example, is known to cause disruption of antioxidative enzymes in the *Populus* x canescens complex (a hybrid of *P. trem*ula and P. alba) (Schützendübel et al 2002). Furthermore, there is substantial heritable variation for Cd tolerance in hybrids of P. trichocarpa and P. deltoides (Induri et al 2012), which supports the contention that these traits may be under selection in naturally-occurring hybrid zones. Exposure to Zinc can also similarly induce oxidative stress in poplars due to decrease in the levels of Glutathione, a crucial component of the  $H_2O_2$  scavenging mechanism (Tyystjärvi *et al* 1999). Thus, given the complexity of molecular and biochemical response mechanisms to counteract metal ion stress, selection may favor maintenance in the hybrids of allelic variation involved in redox homeostasis.

While we found overlap between the genomic regions with evidence of selective sweeps and the introgression outliers, those associations were not significant. Similarly, loci under local adaptation (in *P. balsamifera* and *P. trichocarpa*) did not exhibit genomic islands of differentiation that inhibited introgression. Significant overlap between local adaptation outliers in pure species and introgression selection candidates has been reported in the *P. trichocarpa* and *P. balsamifera* hybridization zone in British Columbia, Alberta and Manitoba (Suarez-Gonzalez *et al* 2018a). Lack of such association in our study in *P. balsamifera* could reflect different selection regimes in allopatric *vs* sympatric regions or the presence of conditional neutrality (e.g. loci involved in local adaptation

elsewhere in the species range do not experience selection in the hybrid zone).

Interestingly, 54 of our 1,764 introgression outliers are located in the 20 genomic regions found to be introgressing from *P. balsamifera* into admixed *P. trichocarpa* (13 regions) or the opposite (7 regions) as reported by Suarez-Gonzalez *et al* (2018a). This raises an intriguing possibility that genomic regions that are targets of adaptive introgression might be shared in geographically distant hybrid zones, even when a different hybridizing partner is involved. In the *P. balsamifera* x *P. trichocarpa* hybrid zone studied by Suarez-Gonzalez *et al* (2018a), introgressed genomic regions are enriched for disease resistance genes as well as those involved in response to biotic and abiotic stress (including oxidative stress, as in our study).

Still, caution must be exercised when interpreting the adaptive significance of candidate outliers as alternative biological mechanisms may also generate the genomic patterns we observed. For example, genomic regions undergoing high rates of recombination may favor introgression (Martin & Jiggins 2017). Recombination rates are variable across the genome and regions with high recombination may generate patterns that mimic adaptive introgression. On the other hand, these regions could also facilitate adaptive introgression if the involved loci are truly under natural selection. Little is known about genome-wide variability in recombination rates in *P. balsamifera* and P. angustifolia, but Wang et al (2015) report low variation in recombination rates across the genome of P. trichocarpa. Interestingly Suarez-Gonzalez et al (2018a) report that introgression was elevated in subtelomeric regions which are known for typically exhibiting high recombination rates. Thus the candidates from our study that overlap with Suarez-Gonzalez et al (2018a) could result due to the shared genomic feature of high recombination rate rather than adaptive introgression per se. However, we find this possibility unlikely for several reasons: first, P. trichocarpa does not appear to show high levels of recombination in subtelomeric regions compared to the rest of the genome (Wang et al 2015); second, Suarez-Gonzalez et al (2018a) found only a weak correlation between introgression and proximity to telomeres for *P. balsamifera* introgressed ancestry and no correlation for introgression in opposite direction; finally, not all of our shared candidate outliers lie in subtelomeric regions (Figure S6).

Range-edges present unique opportunities for the creation and maintenance of hybrid zones. Marginal populations at the rear-edge of a species' distribution may contain unique variants pre-adapted to warmer climates (Hampe & Petit 2005), and many likely served as refugia during glaciation (Hampe et al 2003; Hewitt 2004; Provan & Bennett 2008; Feliner 2011). Hybrid zones at the rear-edge therefore may have unique genetic material available for evolution, than at other zones of contact.

Due to their expected small effective size, marginal populations may also be more susceptible to hybridization than elsewhere in the range (Ellstrand & Elam 1993), and may be centers of hybrid speciation (Rieseberg 1997).

Information on range-edge hybrid zones remains scarce in plants and particularly in forest tree species. We surveyed over 200 studies of hybrid zones in prominent forest tree species in the world published since 2000 and could not find any that explicitly discussed the dynamics of hybridization in populations at the rear-edge of the species range, although a few studies report rearedge hybridization in other angiosperms. For example, potential extinction of rear-edge populations through hybridization and genetic assimilation with a more abundant congener has been noted in the wintergreen Pyrola spp. (Beatty et al 2010). In another study, coalescent models of two Rubus species in Japan where leading populations of one species meet the rear-edge of the other revealed repeated introgression during climatic oscillations that produced a hybrid cline, likely helping the northward expansion of one of the species (Mimura *et al* 2014).

In *P. balsamifera*, the rear range-edge exists as numerous disjunct populations along the Rocky Mountains in the United States and Canada that likely inhabit environments near the species? physiological limits. Climate modeling suggests this region may have been the site of glacial refugia for this species (Levsen et al 2012), and population genetic structure as well as local adaptation in phenology genes in rear-edge populations are distinct from the rest of the range (Keller *et al* 2017). Thus, hybrid zones along the rear range-edge of a species' distribution likely present unique opportunities and challenges for population persistence under suboptimal conditions, and interspecific hybridization may provide a key source of adaptive genomic diversity.

## 5. Conclusions

Our results demonstrate the presence of a tri-species hybrid complex in *Populus* throughout the sympatric zone of overlap in the Rockies. Hybridization appears to be commonplace between these lineages, including both early and advanced generation hybrids. *Populus* in sect. *Tacamahaca* would thus seem to represent a species complex with permeable genomes, evidenced by incomplete/weak reproductive barriers between species and extensive interspecific hybridization. Formation of hybrid complexes in range-edge positions may facilitate population persistence through the formation of novel genotypic combinations not otherwise occurring in pure parental genomes. Signatures of adaptive introgression indicate selection has been active in shaping patterns of introgression in these range-edge populations. Overlap of some of the introgression outliers with introgressed genomic regions in hybrid individuals in other studies provides a historical context for selectiondriven introgression. While many of the hybrid genotypes appear to result from relatively recent admixture (F1, F2, and recent backcrosses), *Populus* are known to form long-lived vegetative clones that may persist for millennia, and there was support for gene flow events involving ancestral populations (e.g., Figure 3). Thus, it is interesting to speculate whether some of the introgression uncovered here may be a relict of more ancient secondary contact between these species. If so, the evolutionary history of highly outbreeding forest trees such as *Populus* may be characterized by repeated episodes of lineage divergence followed by introgression upon secondary contact, with important implications for their adaptive evolution.

## Acknowledgements

This study was supported by National Science Foundation Award #1461868. Comments from two anonymous reviewers and C. Lexer helped improve the manuscript. We would like to remember and recognize our late colleague and friend, Carl J. Douglas of University of British Columbia for his interest in *Populus* introgressive hybridization and helpful discussions about this study.

## 6. References

- Alexander DH, Novembre J, Lange K (2009) Fast model-based estimation of ancestry in unrelated individuals. Genome research, 19, 1655–1664.
- Anderson E, Thompson E (2002) A model-based method for identifying species hybrids using multilocus genetic data. *Genetics*, **160**, 1217–1229.
- ARCC (2012) Mount Moran: IBM System X Cluster, Advanced Research Computing Center, University of Wyoming, Laramie, WY. doi: http://n2t.net/ark:/85786/m4159c.
- Barton NH, Hewitt GM (1985) Analysis of hybrid zones. Annual review of Ecology and Systematics, pp. 113–148.
- Beatty GE, Philipp M, Provan J (2010) Unidirectional hybridization at a species range boundary: implications for habitat tracking. *Diversity and Distributions*, **16**, 1–9.
- Brunner AM, Evans LM, Hsu CY, Sheng X (2014) Vernalization and the chilling requirement to exit bud dormancy: shared or separate regulation? *Frontiers in plant science*, **5**, 732.
- Cervera MT, Storme V, Soto A, et al (2005) Intraspecific and interspecific genetic and phylogenetic relationships in the genus populus based on aflp markers. Theoretical and Applied Genetics, **111**, 1440–1456.
- Chhatre VE (2018) Distruct 2.3: A modified admixture plotting script. Available from http:// distruct2.popgen.org.
- Christe C, Stölting KN, Bresadola L, *et al* (2016) Selection against recombinant hybrids maintains reproductive isolation in hybridizing populus species despite f1 fertility and recurrent gene flow. *Molecular ecology*, **25**, 2482–2498.
- Currat M, Ruedi M, Petit RJ, Excoffier L (2008) The hidden side of invasions: massive introgression by local genes. *Evolution*, **62**, 1908–1920.
- Danecek P, Auton A, Abecasis G, et al (2011) The variant call format and vcftools. *Bioinformatics*, **27**, 2156–2158.
- De La Torre AR, Wang T, Jaquish B, Aitken SN (2014) Adaptation and exogenous selection in a picea glauca× picea engelmannii hybrid zone: implications for forest management under climate change. New Phytologist, **201**, 687–699.
- DiFazio SP, Slavov GT, Joshi CP, et al (2011) Populus: a premier pioneer system for plant genomics. Joshi C, Difazio SP, Kole C, editors. Genetics, genomics and breeding of poplar. Enfield, NH: Science Publishers, pp. 1–28.
- Eckenwalder JE (1984) Natural intersectional hybridization between north american species of populus (salicaceae) in sections aigeiros and tacamahaca. ii. taxonomy. *Canadian Journal of Botany*, **62**, 325–335.

- Ellstrand NC, Elam DR (1993) Population genetic consequences of small population size: implications for plant conservation. Annual review of Ecology and Systematics, 24, 217–242.
- Elshire RJ, Glaubitz JC, Sun Q, et al (2011) A robust, simple genotyping-by-sequencing (gbs) approach for high diversity species. PloS one, 6, e19379.
- Evans LM, Allan GJ, DiFazio SP, et al (2015) Geographical barriers and climate influence demographic history in narrowleaf cottonwoods. *Heredity*, **114**, 387–396.
- Evans LM, Allan GJ, Whitham TG (2012) Populus hybrid hosts drive divergence in the herbivorous mite, aceria parapopuli: implications for conservation of plant hybrid zones as essential habitat. *Conservation Genetics*, **13**, 1601–1609.
- Evans LM, Kaluthota S, Pearce DW, et al (2016) Bud phenology and growth are subject to divergent selection across a latitudinal gradient in populus angustifolia and impact adaptation across the distributional range and associated arthropods. *Ecology and Evolution*.
- Evans LM, Slavov GT, Rodgers-Melnick E, et al (2014) Population genomics of populus trichocarpa identifies signatures of selection and adaptive trait associations. Nature genetics, 46, 1089–1096.
- Excoffier L, Foll M, Petit RJ (2009) Genetic consequences of range expansions. Annual Review of Ecology, Evolution, and Systematics, 40, 481–501.
- Feliner GN (2011) Southern european glacial refugia: a tale of tales. Taxon, 60, 365–372.
- Fishman L, Willis JH (2001) Evidence for dobzhansky-muller incompatibilities contributing to the sterility of hybrids between mimulus guttatus and m. nasutus. *Evolution*, **55**, 1932–1942.
- Floate KD (2004) Extent and patterns of hybridization among the three species of populus that constitute the riparian forest of southern alberta, canada. *Canadian Journal of Botany*, 82, 253–264.
- Floate KD, Godbout J, Lau MK, Isabel N, Whitham TG (2016) Plant-herbivore interactions in a trispecific hybrid swarm of populus: assessing support for hypotheses of hybrid bridges, evolutionary novelty and genetic similarity. *New Phytologist*, **209**, 832–844.
- Foll M, Gaggiotti O (2008) A genome-scan method to identify selected loci appropriate for both dominant and codominant markers: a bayesian perspective. *Genetics*, **180**, 977–993.
- Geraldes A, Farzaneh N, Grassa CJ, et al (2014) Landscape genomics of populus trichocarpa: the role of hybridization, limited gene flow, and natural selection in shaping patterns of population structure. Evolution, 68, 3260–3280.
- Gill SS, Tuteja N (2010) Reactive oxygen species and antioxidant machinery in abiotic stress tolerance in crop plants. *Plant physiology and biochemistry*, **48**, 909–930.
- Glaubitz JC, Casstevens TM, Lu F, et al (2014) Tassel-gbs: a high capacity genotyping by sequencing analysis pipeline. PLoS One, 9, e90346.
- Gompert Z, Buerkle AC (2009) A powerful regression-based method for admixture mapping of isolation across the genome of hybrids. *Molecular Ecology*, 18, 1207–1224.

- Gompert Z, Buerkle AC (2010) Introgress: a software package for mapping components of isolation in hybrids. *Molecular Ecology Resources*, **10**, 378–384.
- Gompert Z, Buerkle AC (2011) Bayesian estimation of genomic clines. *Molecular Ecology*, **20**, 2111–2127.
- Gompert Z, Buerkle CA (2016) What, if anything, are hybrids: enduring truths and challenges associated with population structure and gene flow. *Evolutionary Applications*.
- Goulet BE, Roda F, Hopkins R (2017) Hybridization in plants: old ideas, new techniques. *Plant Physiology*, **173**, 65–78.
- Grivet D, DEGUILLOUX M, Petit RJ, Sork VL (2006) Contrasting patterns of historical colonization in white oaks (quercus spp.) in california and europe. *Molecular Ecology*, **15**, 4085–4093.
- Günther T, Coop G (2013) Robust identification of local adaptation from allele frequencies. *Genetics*, **195**, 205–220.
- Guy RD (2014) The early bud gets to warm. New Phytologist, 202, 7–9.
- Hamilton JA, Lexer C, Aitken SN (2013) Differential introgression reveals candidate genes for selection across a spruce (picea sitchensis× p. glauca) hybrid zone. New Phytologist, **197**, 927–938.
- Hampe A, Arroyo J, Jordano P, Petit RJ (2003) Rangewide phylogeography of a bird-dispersed eurasian shrub: contrasting mediterranean and temperate glacial refugia. *Molecular Ecology*, **12**, 3415–3426.
- Hampe A, Jump AS (2011) Climate relicts: past, present, future. Annual Review of Ecology, Evolution, and Systematics, 42, 313–333.
- Hampe A, Petit RJ (2005) Conserving biodiversity under climate change: the rear edge matters. Ecology letters, 8, 461–467.
- Hamzeh M, Dayanandan S (2004) Phylogeny of populus (salicaceae) based on nucleotide sequences of chloroplast trnt-trnf region and nuclear rdna. *American journal of botany*, **91**, 1398–1408.
- Harrison RG, Larson EL (2014) Hybridization, introgression, and the nature of species boundaries. Journal of Heredity, **105**, 795–809.
- Harrison RG, Larson EL (2016) Heterogeneous genome divergence, differential introgression, and the origin and structure of hybrid zones. *Molecular ecology*.
- Hersch-Green EI, Allan GJ, Whitham TG (2014) Genetic analysis of admixture and patterns of introgression in foundation cottonwood trees (salicaceae) in southwestern colorado, usa. *Tree Genetics & Genomes*, **10**, 527–539.
- Hewitt G (2004) Genetic consequences of climatic oscillations in the quaternary. Philosophical Transactions of the Royal Society of London B: Biological Sciences, 359, 183–195.

- Induri BR, Ellis DR, Slavov GT, et al (2012) Identification of quantitative trait loci and candidate genes for cadmium tolerance in populus. Tree physiology, **32**, 626–638.
- Keim P, Paige K, Whitham T, Lark K (1989) Genetic analysis of an interspecific hybrid swarm of populus: occurrence of unidirectional introgression. *Genetics*, **123**, 557–565.
- Keller SR, Chhatre VE, Fitzpatrick MC (2017) Influence of range position on locally adaptive gene–environment associations in populus flowering time genes. *Journal of Heredity*, **109**, 47–58.
- Keller SR, Levsen N, Olson MS, Tiffin P (2012) Local adaptation in the flowering-time gene network of balsam poplar, populus balsamifera l. *Molecular Biology and Evolution*, 29, 3143–3152.
- Keller SR, Soolanayakanahally RY, Guy RD, Silim SN, Olson MS, Tiffin P (2011) Climatedriven local adaptation of ecophysiology and phenology in balsam poplar, populus balsamifera l.(salicaceae). American Journal of Botany, 98, 99–108.
- Lamit L, Bowker MA, Holeski L, et al (2011) Genetically-based trait variation within a foundation tree species influences a dominant bark lichen. Fungal Ecology, 4, 103–109.
- Levsen ND, Tiffin P, Olson MS (2012) Pleistocene speciation in the genus populus (salicaceae). Systematic biology, **61**, 401–412.
- Lexer C, Joseph JA, van Loo M, et al (2010) Genomic admixture analysis in european populus spp. reveals unexpected patterns of reproductive isolation and mating. *Genetics*, **186**, 699–712.
- Li H, Durbin R (2009) Fast and accurate short read alignment with burrows-wheeler transform. *Bioinformatics*, **25**, 1754–1760.
- Lindtke D, Gompert Z, Lexer C, Buerkle CA (2014) Unexpected ancestry of populus seedlings from a hybrid zone implies a large role for postzygotic selection in the maintenance of species. *Molecular ecology*, 23, 4316–4330.
- Lotterhos KE, Whitlock MC (2014) Evaluation of demographic history and neutral parameterization on the performance of fst outlier tests. *Molecular ecology*, 23, 2178–2192.
- Martin SH, Jiggins CD (2017) Interpreting the genomic landscape of introgression. Current opinion in genetics & development, 47, 69–74.
- Martinsen GD, Whitham TG, Turek RJ, Keim P (2001) Hybrid populations selectively filter gene introgression between species. *Evolution*, **55**, 1325–1335.
- McKown AD, Klápště J, Guy RD, *et al* (2014) Genome-wide association implicates numerous genes underlying ecological trait variation in natural populations of populus trichocarpa. *New Phytologist*, **203**, 535–553.
- Meirmans PG, Lamothe M, Gros-Louis MC, et al (2010) Complex patterns of hybridization between exotic and native north american poplar species. American Journal of Botany, 97, 1688–1697.
- Mimura M, Mishima M, Lascoux M, Yahara T (2014) Range shift and introgression of the rear and leading populations in two ecologically distinct rubus species. *BMC evolutionary biology*, 14, 209.

- Nielsen R, Williamson S, Kim Y, Hubisz MJ, Clark AG, Bustamante C (2005) Genomic scans for selective sweeps using snp data. *Genome research*, **15**, 1566–1575.
- Olson MS, Levsen N, Soolanayakanahally RY, *et al* (2013) The adaptive potential of populus balsamifera l. to phenology requirements in a warmer global climate. *Molecular Ecology*, **22**, 1214–1230.
- Pavlidis P, Živković D, Stamatakis A, Alachiotis N (2013) Sweed: likelihood-based detection of selective sweeps in thousands of genomes. *Molecular biology and evolution*, **30**, 2224–2234.
- Petit RJ, Aguinagalde I, de Beaulieu JL, et al (2003) Glacial refugia: hotspots but not melting pots of genetic diversity. science, **300**, 1563–1565.
- Pickrell JK, Pritchard JK (2012) Inference of population splits and mixtures from genome-wide allele frequency data. *PLoS Genet*, 8, e1002967.
- Provan J, Bennett K (2008) Phylogeographic insights into cryptic glacial refugia. Trends in ecology & evolution, 23, 564–571.
- Raj A, Stephens M, Pritchard JK (2014) faststructure: variational inference of population structure in large snp data sets. *Genetics*, **197**, 573–589.
- Rieseberg LH (1997) Hybrid origins of plant species. Annual review of Ecology and Systematics, 28, 359–389.
- Rieseberg LH, Kim SC, Randell RA, et al (2007) Hybridization and the colonization of novel habitats by annual sunflowers. *Genetica*, **129**, 149–165.
- Rieseberg LH, Raymond O, Rosenthal DM, et al (2003) Major ecological transitions in wild sunflowers facilitated by hybridization. Science, **301**, 1211–1216.
- Rinne PL, Welling A, Vahala J, et al (2011) Chilling of dormant buds hyperinduces flowering locus t and recruits ga-inducible 1, 3- $\beta$ -glucanases to reopen signal conduits and release dormancy in populus. The Plant Cell, 23, 130–146.
- Schützendübel A, Nikolova P, Rudolf C, Polle A (2002) Cadmium and  $h_{2}o_{2}$ -induced oxidative stress in populus× canescens roots. *Plant Physiology and Biochemistry*, **40**, 577–584.
- Soltis PS, Soltis DE (2009) The role of hybridization in plant speciation. Annual review of plant biology, 60, 561–588.
- Soolanayakanahally RY, Guy RD, Silim SN, Drewes EC, Schroeder WR (2009) Enhanced assimilation rate and water use efficiency with latitude through increased photosynthetic capacity and internal conductance in balsam poplar (populus balsamifera l.). *Plant, Cell & Environment*, **32**, 1821–1832.
- Stettler R, Zsuffa L, Wu R (1996) The role of hybridization in the genetic manipulation of populus. Biology of Populus and its implications for management and conservation, pp. 87–112.

- Suarez-Gonzalez A, Hefer CA, Christe C, *et al* (2016) Genomic and functional approaches reveal a case of adaptive introgression from *populus balsamifera* (balsam poplar) in *p. trichocarpa* (black cottonwood). *Molecular ecology*.
- Suarez-Gonzalez A, Hefer CA, Lexer C, Cronk QC, Douglas CJ (2018a) Scale and direction of adaptive introgression between black cottonwood (*Populus trichocarpa*) and balsam poplar (*p. balsamifera*). Molecular ecology, 27, 1667–1680.
- Suarez-Gonzalez A, Lexer C, Cronk QC (2018b) adaptive introgression: a plant perspective. *Biology* Letters, 14.
- Supek F, Bošnjak M, Škunca N, Šmuc T (2011) Revigo summarizes and visualizes long lists of gene ontology terms. *PloS one*, 6, e21800.
- Sytar O, Kumar A, Latowski D, Kuczynska P, Strzałka K, Prasad M (2013) Heavy metal-induced oxidative damage, defense reactions, and detoxification mechanisms in plants. Acta physiologiae plantarum, 35, 985–999.
- Szkiba D, Kapun M, von Haeseler A, Gallach M (2014) Snp2go: functional analysis of genome-wide association studies. *Genetics*, **197**, 285–289.
- Talbot P, Schroeder WR, Bousquet J, Isabel N (2012) When exotic poplars and native populus balsamifera l. meet on the canadian prairies: Spontaneous hybridization and establishment of interspecific hybrids. *Forest ecology and management*, **285**, 142–152.
- Teeter KC, Thibodeau LM, Gompert Z, Buerkle CA, Nachman MW, Tucker PK (2010) The variable genomic architecture of isolation between hybridizing species of house mice. *Evolution*, 64, 472– 485.
- Tuskan GA, Difazio S, Jansson S, et al (2006) The genome of black cottonwood, populus trichocarpa (torr. & gray). science, **313**, 1596–1604.
- Tyystjärvi E, Riikonen M, Arisi ACM, Kettunen R, Jouanin L, Foyer CH (1999) Photoinhibition of photosystem ii in tobacco plants overexpressing glutathione reductase and poplars overexpressing superoxide dismutase. *Physiologia Plantarum*, **105**, 409–416.
- Wang J, Street NR, Scofield DG, Ingvarsson PK (2015) Natural selection and recombination rate variation shape nucleotide polymorphism across the genomes of three related populus species. *Genetics*, pp. genetics–115.
- Weir BS, Cockerham CC (1984) Estimating f-statistics for the analysis of population structure. Evolution, **38**, 1358–1370.
- Whitham TG, Bailey JK, Schweitzer JA, et al (2006) A framework for community and ecosystem genetics: from genes to ecosystems. Nature Reviews Genetics, 7, 510–523.
- Whitney KD, Ahern JR, Campbell LG, Albert LP, King MS (2010) Patterns of hybridization in plants. Perspectives in Plant Ecology, Evolution and Systematics, 12, 175–182.
- Wolf JB, Ellegren H (2017) Making sense of genomic islands of differentiation in light of speciation. Nature Reviews Genetics, 18, 87.

- Woolbright SA, Whitham TG, Gehring CA, Allan GJ, Bailey JK (2014) Climate relicts and their associated communities as natural ecology and evolution laboratories. *Trends in ecology & evolution*, **29**, 406–416.
- Zhou L, Bawa R, Holliday J (2014) Exome resequencing reveals signatures of demographic and adaptive processes across the genome and range of black cottonwood (populus trichocarpa). *Molecular ecology*, **23**, 2486–2499.

## Data Archiving

All raw sequence data is available through NCBI SRA accession number SRP070954. All input files have been archived in the github repository located at https://cryptic0.github.io/HybridPoplar.

## Author Contributions

VEC and SRK conceived the study, performed data analysis and wrote the manuscript. LME and SDF wrote parts of manuscript and provided extensive feedback on earlier drafts. LME contributed *P. angustifolia* samples. SDF and LME provided genomic data on five additional *Populus* species. All authors read the final manuscript.

## 7. Tables

# List of Tables

1	Sampling localities for <i>P. angustifolia</i> , <i>P. balsamifera</i> , <i>P. trichocarpa</i> , and putative	
	hybrids	35
2	NEWHYBRIDS analysis of genotype assignment. Values are the numbers of indi-	
	viduals assigned to F1, F2 and advanced generation hybrids in populations from	
	sympatric zones inferred from 375 SNP markers. Results are based on two separate	
	analyses, with first combining the balsam poplars into a single parental population,	
	and second excluding <i>P. trichoarpa</i> from the analysis, both of which yielded highly	
	similar probabilities of genotype assignments.	36
3	Distribution of genotypic classes at candidate SNPs for differential introgression in-	
	ferred from a single run of INTROGRESS $vs$ 100 bootstrapped runs. Key: Genotypic	
	proportions per expectation (=), excess (+), and deficit (-). $\ldots$	37

(1)	Geog. zone	Species	Popcode	Location	N. Ind.	Latitude	Longitude
	Allopatric	PANGUST	AZBR	Blue River, AZ	2	33.730	-109.052
	Allopatric	PANGUST	AZPW	Pumphouse Wash, AZ	2	35.013	-111.736
	Allopatric	PANGUST	UTAH	Oak Creek, UT	3	38.713	-111.630
	Allopatric	PANGUST	UTWR	Weber River, UT	31	40.848	-111.380
	Allopatric	PBALSAM	$\operatorname{CPL}$	Chapleau, ON	14	47.516	-83.262
	Allopatric	PBALSAM	MMT	Moose Mtn. Area, SK	14	49.876	-102.593
	Allopatric	PBALSAM	OFR	Over Flowing River, MB	15	53.145	-101.101
	Allopatric	PBALSAM	RAD	Radisson, QC	14	53.429	-77.641
	Allopatric	PBALSAM	SKN	Saskatoon, SK	14	52.337	-106.613
	Allopatric	PBALSAM	TIM	Timmins, ON	14	48.363	-81.633
	Allopatric	PBALSAM	TUR	Turtleford, SK	13	53.203	-108.348
	Allopatric	PBALSAM	UMI	Umiujaq, QC	4	56.335	-76.255
	Allopatric	PBALSAM	WTR	White River, ON	12	48.587	-85.276
	Allopatric	PTRICHO	NISQUALLY	Nisqually River, WA	9	47.104	-122.636
	Allopatric	PTRICHO	NOOKSACK	Nooksack, WA	8	48.804	-122.173
	Allopatric	PTRICHO	OLYMPIC	Olympic Peninsula, WA	3	47.589	-122.900
	Allopatric	PTRICHO	PUYALLUP	Puyallup, WA	8	47.086	-122.203
	Allopatric	PTRICHO	SKAGIT	Skagit, WA	8	48.511	-122.079
	Allopatric	PTRICHO	SKYKOMISH	Skykomish, WA	8	47.832	-121.837
	Allopatric	PTRICHO	THIRTEEN	Thirteen River, WA	2	54.150	-128.600
	Allopatric	PTRICHO	THIRTYONE	Thirty One River, WA	8	49.706	-125.065
	Sympatric	PANGUST	ABOR	Oldman River, AB	5	49.791	-113.099
	Sympatric	PANGUST	RMPANG	Rocky Mtn. NP, CO	6	40.401	-105.627
	Sympatric	PANGUST	WYSR	Snake River, WY	5	43.242	-110.869
	Sympatric	PBALSAM	SSR	Shoeshone River, WY	16	44.462	-109.610
	Sympatric	PBALSAM	CYH	Cypress Hill PP, SK	13	49.636	-109.806
	Sympatric	PBALSAM	JKH	Jackson Hole, WY	14	43.825	-110.477
	Sympatric	PBALSAM	MSG	Mosier Gulch, WY	15	44.323	-106.873
	Sympatric	PBALSAM	RMP	Rocky Mtn. NP, CO	16	40.252	-105.563

Table	1	Sampling	localities f	or	Ρ.	angustifolia,	P.	balsamifera,	Ρ.	tricho carpa,	and	putative l	hybrie	ds
-------	---	----------	--------------	----	----	---------------	----	--------------	----	---------------	-----	------------	--------	----

PANGUST: P. angustifolia; PBALSAM: P. balsamifera; PTRICHO: P. trichocarpa

, e

Table 2 NEWHYBRIDS analysis of genotype assignment. Values are the numbers of individuals assigned to F1, F2 and advanced generation hybrids in populations from sympatric zones inferred from 375 SNP markers. Results are based on two separate analyses, with first combining the balsam populars into a single parental population, and second excluding P. trichoarpa from the analysis, both of which yielded highly similar probabilities of genotype assignments.

			Exp. Ancestry Proportions				
Genotype Class	Cross	AA	Aa	aA	aa	observed	
ANG	P. angustifolia (A)	1	0	0	0	9	
BAL	P. balsamifera (a)	0	0	0	1	48	
F1	P. angustifola x P. balsamifera	0	0.5	0.5	0	22	
F2	F1 x F1	0.25	0.25	0.25	0.25	3	
BC-F1xANG	F1 x P. angustifolia	0.5	0.25	0.25	0	0	
BC-F1xBAL	F1 x P. balsamifera	0	0.25	0.25	0.5	3	
BC-F2xANG	F2 x P. angustifolia	0.5	0.125	0.125	0.25	0	
BC-F2xBAL	F2 x P. balsamifera	0.25	0.125	0.125	0.5	0	
BC-ANGx(F1xANG)	P. angustifolia x (F1 x P. angustifolia)	0.75	0.125	0.125	0	1	
BC-BALx(F1xBAL)	P. balsamifera x (F1 x P. balsamifera)	0	0.125	0.125	0.75	3	
BC-ANGx(F2xANG)	P. angustifolia x (F2 x P. angustifolia)	0.625	0.125	0.125	0.125	0	
BC-BALx (F2xBAL)	$P. \ balsamifera \ge (F2 \ge P. \ balsamifera)$	0.125	0.125	0.125	0.625	1	
			Total N	90			

\*For the second analysis where *P. trichocarpa* samples were excluded, results could not be obtained for 2 hybrid individuals from the population RMP.

Arti

Dte

Genotype Coo
AA= Aa= aa
AA+ Aa+ aa
AA- Aa- aa
AA+ Aa= aa
AA+ Aa- aa
AA+ Aa= aa
AA+ Aa+ aa
AA+ Aa- aa
AA+ Aa= aa
AA+ Aa+ aa
AA+ Aa- aa
AA- Aa= aa
AA- Aa+ aa
AA- Aa- aa
AA- Aa+ aa
AA- Aa= aa
AA- Aa= aa
AA- Aa+ aa
AA- Aa- aa
AA= Aa= aa
AA= Aa+ aa
AA= Aa- aa
AA= Aa= aa
AA= Aa+ aa
AA= Aa- aa
AA= Aa- aa
AA= Aa+ aa

**Table 3** Distribution of genotypic classes at candidate SNPs for differential introgression inferred from a single run of INTROGRESS vs 100 bootstrapped runs. Key: Genotypic proportions per expectation (=), excess (+), and deficit (-).

Genotype Code	Explanation	Num sites 1 run		Num sites 100 runs	
		Of 3723	Percent	Of 1764	Percent
AA= Aa= aa=	All per expectation	459	12.33	2	0.11
AA+ Aa+ aa+	All more than expected	0	0	0	0.00
AA- Aa- aa-	All less than expected	0	0	0	0.00
AA+ Aa= aa+	More ANG, exp Hets, More BAL	0	0	0	0.00
AA+ Aa- aa+	More ANG, Less Hets, More BAL	15	0.4	0	0.00
AA+ Aa= aa-	More ANG, Exp Hets, Less BAL	641	17.22	242	13.72
AA+ Aa+ aa-	More ANG, More Hets, Less BAL	658	17.67	289	16.38
AA+ Aa- aa-	More ANG, Less Hets , Less BAL	134	3.6	19	1.08
AA+ Aa= aa=	More ANG, exp Hets, exp BAL	21	0.56	247	14.00
AA+ Aa+ aa=	More ANG, More HETS, Exp BAL	0	0	0	0.00
AA+ Aa- aa=	More ANG, Less Hets, Exp BAL	123	3.3	35	1.98
AA- Aa= aa+	Less ANG, exp Hets, More BAL	104	2.79	1	0.06
AA- Aa+ aa+	Less ANG, More Hets, More BAL	0	0	0	0.00
AA- Aa- aa+	Less ANG, Less Hets, More BAL	502	13.48	117	6.63
AA- Aa+ aa-	Less ANG, More Hets, Less BAL	11	0.3	0	0.00
AA- Aa= aa-	Less ANG, exp Hets, less BAL	0	0	0	0.00
AA- Aa= aa=	Less ANG, exp Hets, exp BAL	454	12.19	3	0.17
AA- Aa+ aa=	Less ANG, More Hets, Exp BAL	112	3.01	1	0.06
AA- Aa- aa=	Less ANG, Less Hets, Exp BAL	0	0	0	0.00
AA= Aa= aa+	Exp ANG, exp Hets, More BAL	24	0.64	0	0.00
AA= Aa+ aa+	Exp ANG, More Hets, More BAL	0	0	0	0.00
AA= Aa- aa+	Exp ANG, Less Hets, More BAL	304	8.17	20	1.13
AA= Aa= aa-	Exp ANG, exp Hets, Less BAL	9	0.24	0	0.00
AA= Aa+ aa-	Exp ANG, More Hets, Less BAL	125	3.36	9	0.51
AA= Aa- aa-	Exp ANG, Less Hets, Less BAL	0	0	0	0.00
AA= Aa- aa=	Exp ANG, Less Hets, Exp BAL	6	0.16	22	1.25
AA= Aa+ aa=	Exp ANG, More Hets, Exp BAL	21	0.56	9	0.51
	Total	3723	100	1016	57.60

## 8. Figures

## List of Figures

1	Map of sampling localities showing species ranges and area of overlap in the Rocky Mountain sympatric zone	39
2	Genotypic clustering from ADMIXTURE. Colors correspond to ancestry contribu- tion from parental species <i>P. balsamifera</i> (dark green), <i>P. angustifolia</i> (orange, blue), and <i>P. trichocarpa</i> (light green). Population label colors represent the three zones, one per pure species (darkgreen, lightgreen and orange labels), and the hybrid zone	00
0	(blue labels)	40
3	TREEMIX analysis of population splitting and migration events. Colored arrows represent migration events from population $i$ to population $j$ with color ramp in-	
	dicating the magnitude of gene flow. (a) through (d): 7-10 migration routes. For	/1
4	Triangle plot of interspecific heterozygosity between $P$ analyticalia (hybrid index =	41
1	0) and P. balsamifera (hybrid index = 1).	42
5	Ancestry plot showing probability of species-specific genotypic ancestry for the 19	
	poplar chromosomes across a range of individuals varying in hybrid index. Orange	
	regions represent $P.$ angustifolia-like genotypes $(AA)$ , darkgreen regions represent $P.$	
	balsamifera-like genotypes ( $aa$ ) and blue represents interspecific heterozygotes ( $Aa$ ).	43
6	(a) Reproducibility of INTROGRESS results based on 100 bootstrap resamples. Or-	
	ange bar depicts loci that were candidates for adaptive introgression in all bootstrap	
	runs. (b) Density distribution of $F_{\rm ST}$ in <i>P. angustifolia</i> derived alleles from 1764	
7	candidate loci (orange) $vs$ 1912 neural loci (gray)	44
1	among introgression outliers. Points show minimally redundant terms based on	
	clustering semantically similar terms (warmer colors reflect a greater uniqueness of	
	a given GO term in the similarity matrix). Circle size is proportional to the $-\log 10P$	
	value of the test for functional enrichment of SNPs showing differential introgression.	45



Figure 1 Map of sampling localities showing species ranges and area of overlap in the Rocky Mountain sympatric zone







**Figure 2** Genotypic clustering from ADMIXTURE. Colors correspond to ancestry contribution from parental species *P. balsamifera* (dark green), *P. angustifolia* (orange, blue), and *P. trichocarpa* (light green). Population label colors represent the three zones, one per pure species (darkgreen, lightgreen and orange labels), and the hybrid zone (blue labels)



Figure 3 TREEMIX analysis of population splitting and migration events. Colored arrows represent migration events from population i to population j with color ramp indicating the magnitude of gene flow. (a) through (d): 7-10 migration routes. For residuals, see Supplementary Figure S3.







Figure 5 Ancestry plot showing probability of species-specific genotypic ancestry for the 19 poplar chromosomes across a range of individuals varying in hybrid index. Orange regions represent *P. angustifolia*-like genotypes (AA), darkgreen regions represent *P. balsamifera*-like genotypes (aa) and blue represents interspecific heterozygotes (Aa).



CHHATRE et

al

Figure 6 (a) Reproducibility of INTROGRESS results based on 100 bootstrap resamples. Orange bar depicts loci that were candidates for adaptive introgression in all bootstrap runs. (b) Density distribution of  $F_{\rm ST}$  in *P. angustifolia* derived alleles from 1764 candidate loci (orange) vs 1912 neural loci (gray).



**Figure 7** REVIGO multi-dimensional scaling plot of Gene Ontology (GO) terms enriched among introgression outliers. Points show minimally redundant terms based on clustering semantically similar terms (warmer colors reflect a greater uniqueness of a given GO term in the similarity matrix). Circle size is proportional to the -log10P value of the test for functional enrichment of SNPs showing differential introgression.