Scale and direction of adaptive introgression between black cottonwood (*Populus trichocarpa*) and balsam poplar (*P. balsamifera*)

Adriana Suarez-Gonzalez¹, Charles A. Hefer^{1,2}, Christian Lexer³, Quentin C. B. Cronk^{1*} and Carl J. Douglas[†]

¹Department of Botany, University of British Columbia, Vancouver, Canada

²Current address: Biotechnology Platform, Agricultural Research Council, Private Bag X05, Onderstepoort, 0110, South Africa

³Department of Botany and Biodiversity Research, University of Vienna, Austria

[†] Deceased 25 July 2016

KEYWORDS:

Adaptive introgression, local adaptation, admixture, selection, subtelomeric regions

*Corresponding author:

Quentin Cronk

Email: quentin.cronk@ubc.ca

Running head: Scale and direction of adaptive introgression

ABSTRACT

Introgression can introduce novel genetic variation at a faster rate than mutation alone, and result in adaptive introgression when adaptive alleles are maintained in the recipient genome over time by natural selection. A previous study from our group demonstrated adaptive introgression from *Populus balsamifera* into *P. trichocarpa* in a target genomic region. Here we expand our local ancestry analysis to the whole genome of both parents to provide a comprehensive view of introgression patterns and to identify additional candidate regions for adaptive introgression genome-wide. *Populus trichocarpa* is a large, fast-growing tree of mild coastal regions of the Pacific Northwest, whereas *P. balsamifera* is a smaller stature tree of continental and boreal regions with intense winter-cold. The species hybridize where they are parapatric. We detected asymmetric patterns of introgression across the whole genome of these two poplar species adapted to contrasting environments, with stronger introgression from *P. balsamifera* to *P. trichocarpa* than vice versa. Admixed *P. trichocarpa* individuals contained more genomic regions with unusually high levels of introgression (19 regions) and also the largest introgressed genome fragment (1.02 Mb) compared with admixed *P. balsamifera* (9 regions). Our analysis also revealed numerous candidate regions for adaptive introgression with strong signals of selection, notably related to disease resistance, and enriched for genes that may play crucial roles in survival and adaptation. Furthermore, we detected a potential overrepresentation of subtelomeric regions in *P. balsamifera* introgressed into *P. trichocarpa* and possible protection of sex-determining regions from interspecific gene flow.

INTRODUCTION

Hybridization, the interbreeding between individuals of different varieties or species, is a widespread natural phenomenon in plants (Ellstrand *et al.* 1999; Stace *et al.* 2015) and animals (Dowling & Secor 1997) that has shaped the genomes of many lineages. In natural hybrid zones, back-crossing of early-generation hybrids over subsequent generations can result in introgression, where genetic material from one parental species is incorporated into another species. The permeability of species boundaries to introgression depends on several factors including

recombination rate and the genomic distribution and number of loci associated with reproductive isolation (Barton & Hewitt 1985; Harrison 1990).

Introgression can introduce novel genetic variation, including new gene combinations, at a faster rate than through mutation alone and result in adaptive introgression when adaptive alleles are maintained over time by natural selection (Twyford & Ennos 2012; Slarkin 1985; Rieseberg *et al.* 2003; Abbott *et al.* 2013; Schmickl *et al.* 2017). Examples of adaptive introgression in plants and animals include increased herbivore resistance in sunflower (Whitney *et al.* 2006), flood tolerance in *Iris* (Martin *et al.* 2006), mimetic wing patterns in butterflies (Pardo-Diaz *et al.* 2012), and altitude adaptation in Tibetans (Huerta-Sanchez *et al.* 2014). Adaptive introgression may be particularly important in rapidly changing environments, where standing genetic variation and mutation alone may only offer limited potential for adaptation (Hedrick 2013; Hamilton & Miller 2015).

Natural hybrid zones provide evolutionary laboratories to identify chromosomal regions that introgress significantly more frequently than expected (Lexer *et al.* 2004; Buerkle & Lexer 2008; Abbott 2017). Forest tree species with extensive natural hybrid zones, large ranges across geographical and climatic clines as well as substantial trait variation (Savolainen *et al.* 2007; Soolanayakanahally *et al.* 2009; Keller *et al.* 2011; McKown *et al.* 2013) are attractive for the study of adaptive introgression. *Populus* spp. in particular, have emerged as models for population genomic studies of adaptation (Weigel & Nordborg 2015) due to porous species barriers and a wealth of genomic resources available (Cronk 2005; Jansson & Douglas 2007; Tuskan *et al.* 2006). *Populus trichocarpa* and *P. balsamifera* are sibling poplar species of the section *Tacamahaca* that diverged in allopatry rather recently and hybridize freely where they overlap (Viereck & Foote 1970). Despite their recent divergence [~ 76 Ka (Levsen *et al.* 2012); but see (Ismail *et al.* 2012)] and morphological similarity, these species are ecologically divergent and adapted to strongly contrasting environments. *Populus balsamifera* is a boreal species distributed from Alaska to Newfoundland, with high frost tolerance and adapted to a very large temperature range (-62°C to 44°C) as well as to moderate annual precipitation. *Populus trichocarpa*, on the other hand, is distributed throughout the western US and Canada, from northern California to southern Alaska, and is adapted to relatively humid, moist (riparian or higher precipitation), and milder conditions (Richardson *et al.* 2014; Geraldès *et al.* 2014).

In a previous study, we explored fine-scale introgression patterns within and around target genes in three chromosomes of *P. trichocarpa* and *P. balsamifera* to detect candidate regions for adaptive introgression (Suarez-Gonzalez *et al.* 2016). Our local ancestry analysis revealed a subtelomeric region in chromosome 15 with *P. balsamifera* ancestry, containing genes associated with higher chlorophyll and leaf nitrogen content. Admixed *P. trichocarpa* individuals with these particular *P. balsamifera* alleles may grow and acquire carbon faster than pure *P. trichocarpa* individuals. This may serve to counteract shorter growth seasons at higher latitudes as has been shown to be physiologically important in poplar (Soolanayakanahally *et al.* 2009; McKown *et al.* 2013) and spruce (Oleksyn *et al.* 1998). A more recent study using admixture mapping and phenotypic analyses provided further support for the hypothesis that hybridization between *P. trichocarpa* and *P. balsamifera* is an important source of adaptive phenotypic variation that may allow improved survival in new environments (Suarez-Gonzalez *et al.* 2018). These results, based on functional and phenotypic tests, showed the potential of local ancestry analyses to identify strong candidates for adaptive introgression (Suarez-Gonzalez *et al.* 2016).

Our previous work, together with other documented empirical cases (Martin *et al.* 2006; Pardo-Diaz *et al.* 2012; Huerta-Sanchez *et al.* 2014), indicate the importance of adaptive introgression. However, little is known about the extent and direction of introgression across the entire genome, and the genomic architecture of introgression at a fine genomic scale. Here we expand our local ancestry analysis to the whole genomes of *P. trichocarpa* and *P. balsamifera* using a population-wide resequencing approach and include additional admixed individuals to address the following questions: (i) are introgression patterns asymmetric between these two poplar species? (ii) is there enrichment for certain biological terms in the introgressed regions, in particular, those related to disease resistance, given that *P. trichocarpa* populations may have been under intense selection pressures for adaptation to leaf rust (La Mantia *et al.* 2013)? and (iii) are certain genomic regions more prone to (e.g. subtelomeric regions) or protected from (e.g. sex-determining regions) interspecific introgression?

MATERIALS AND METHODS

Samples

Populus trichocarpa accessions used in this study were collected by the British Columbia Ministry of Forests, Lands and Natural Resource Operations (MFLNRO) (Xie *et al.* 2009), and planted in a common garden at the University of British Columbia (McKown *et al.* 2013). These individuals were from 28 “drainages” (i.e., topographic units separated by watershed barriers) spanning 14° in latitude (45.6°–59.6°) from throughout the species range. For *P. balsamifera*, we used accessions from 46 provenances throughout the species range obtained from the Agriculture and Agri-Food Canada AgCanBaP collection (Soolanayakanahally *et al.* 2009). For local ancestry analysis in RASPBerry (Wegmann *et al.* 2011), 50 reference individuals (25 *P. balsamifera* and 25 *P. trichocarpa*) and 161 admixed individuals (129 *P. trichocarpa* individuals with *P. balsamifera* admixture, and 32 *P. balsamifera* individuals with *P. trichocarpa* admixture) were selected from the sympatric zone between *P. trichocarpa* and *P. balsamifera* as well as from allopatric populations (Figure 1). On the basis of preliminary results, the 129 admixed *P. trichocarpa* individuals were reduced to 118 for all analyses, as 11 individuals exhibited only trace amounts of admixture (see results). These 211 individuals were selected from a collection of 435 *P. trichocarpa* and 448 *P. balsamifera* genotypes, using a previous genome-wide admixture analysis based on a genome-wide dataset of 971k SNPs (Figure S1; Table S1, Supporting Information) (Gerald *et al.*, in preparation). Additional sample information and data generated in this study are given in Table S1, Supporting Information.

Sequencing, read mapping and variant calling

Sampled leaves were flash frozen in liquid nitrogen, or silica gel dried and lyophilized. Between 50 and 100 mg of leaf tissue were ground in a Precellys 24 tissue homogenizer (Bertin Technologies, Montigny-le Bretonneux, France), and DNA was extracted using a Qiagen DNeasy Plant Mini Kit (Qiagen, Valencia, CA). Protocol modifications include the use of 12 µL of RNase A applied to each sample, and an additional centrifugation for 2 minutes at 20,000g performed after cell lysis to remove tissue debris. All solution volumes were increased by 1.5x in each step of the manufacturer’s protocol. DNA quantification and purity assessment were performed using a NanoDrop spectrophotometer (Thermo Fisher Scientific, Wilmington, DE) and a Qubit fluorometer (Invitrogen, Carlsbad, CA). High quality purified genomic DNA was used to

prepare multiplexed sequencing libraries according to the Standard Operating Protocols at the Genome Sciences Centre, Vancouver, BC. Sequences (100bp paired-end reads) were obtained at estimated 30X or 15X coverage of the Nisqually-1 reference genome on an Illumina HiSeq at the Genome Sciences Centre.

For local ancestry analysis, SNPs were called across the whole genomes for 211 individuals allowing us to generate a dataset of unprecedented size for a study of this sort. Short reads from the sequencing libraries were independently aligned to the *P. trichocarpa* version 3 (v3.0) genome using BWA (version 0.6.1-r104) with default parameters. As the two poplar species are very closely related (Levsen, Tiffin & Olson, 2012; Geraldès et al. 2013) it is possible to call SNPs in both species relative to the same reference with high accuracy (Geraldès et al., 2015). We corrected mate pair metadata and marked duplicate molecules using the FixMateInformation and MarkDuplicates methods in the Picard package (<http://picard.sourceforge.net>). Reads present in areas surrounding InDels were re-aligned using the IndelRealigner method from GATK (version v1.5-25-gf46f7d0). Next, we called SNPs and small indels independently using the UnifiedGenotyper method from GATK. SNPs were then filtered to exclude variants within 3bp of any identified variants, having a mapping quality less than 5, and a variant quality less than 30. Each SNP was annotated using SNPeff (Cingolani *et al.* 2012) with version 3 of the *P. trichocarpa* genome. All raw sequencing data and alignment information have been deposited in SRA (PRJNA276056).

Inference of local ancestry

We estimated the probabilities for each of the possible ancestral configurations (*P. balsamifera*, *P. trichocarpa* or mixed ancestry) in every SNP across the whole genome of 161 admixed individuals using RASPBerry, a software that implements a reliable Hidden Markov model (HMM) for admixture (Wegmann *et al.* 2011), following the same pipeline and parameters as in our previous study on adaptive introgression (Suarez-Gonzalez *et al.* 2016). Briefly, SNPs with missing data in the parental genotypes were removed using Plink 1.07 (Purcell *et al.* 2007), the parental genotypes (50) were then phased with fastphase (Scheet & Stephens 2006) by creating the input files with FCGENE (Roshyara & Scholz 2014), and ancestries for each admixed individual were estimated using ADMIXTURE (Alexander *et al.* 2009).

To determine the ancestral configurations of each SNP in one of the three categories, ancestries were considered for probabilities >95%. The proportion of introgressed ancestry in admixed individuals was calculated, for each SNP, by counting sites with homospecific ancestry as two and sites with mixed (i.e. heterospecific) ancestry as one, as in Suarez-Gonzalez *et al.* (2016). We detected regions with unusually high levels of introgression (relative to general genome-wide background levels) in admixed individuals using a sliding window approach and a significance cut-off of three standard deviations (SD) from the weighted mean across all the chromosomes based on SNP density per window. We used windows of 100-kb with steps of 20-kb since this window size provided the best compromise between large numbers of windows with missing data (50-kb windows) and low resolution (1-Mb windows) (Suarez-Gonzalez *et al.* 2016).

To determine if levels of introgression were correlated with certain structural features of the genome, we performed a Pearson correlation (calculating coefficient r) between admixed ancestry values and distance to telomeres. We also explored levels of introgression in predicted centromeric regions (Pinosio *et al.* 2016).

Association between climate and introgression

To determine if introgression was associated with climate, we compiled 23 climate variables from ClimateNA (Wang 2012) based on 1971–2000 and performed two principal component analyses (PCAs), one for temperature and one for moisture, using the function `prcomp` in R. Eight variables were associated with moisture and 15 were associated with temperature (Table S2A, Supporting information). As these two sets of variables behave very differently (moisture variables broadly correlated with longitude and temperature with latitude), they were analysed separately for ease of interpretation. Hypothesis testing was performed using Pearson correlations.

Enrichment analysis

To detect overrepresented biological terms in introgressed regions, in particular those related to disease resistance, we performed enrichment tests for various terms including Gene Ontology (GO) and Protein Family (Pfam) using Popgenie accessed in November 2016 (Sjödín *et al.* 2009). Since NBS-LRRs (nucleotide binding sites and leucine-rich repeat) are the largest class of disease resistance genes and play critical roles in plant defence (Dangl & Jones 2001), we searched for protein families (Pfam) associated with NBS-LRR domains using the keyword search function in

the Pfam database (Finn *et al.* 2016). We compared these Pfam codes associated with NBS-LRRs domains with the enriched Pfam codes in introgressed regions. In addition, we performed an enrichment analysis in a *P. balsamifera* introgressed region showing signals of positive selection.

Tests of selection in the introgressed regions from the pure species individuals

We estimated Tajima's D values across all introgressed regions for the 50 pure species individuals, using 20-kb windows in VCFtools v0.1.12b (Danecek *et al.* 2011). Average Tajima's D was estimated within each introgressed region and compared with the average from 20-kb windows without introgressed peaks using ANOVA. To detect windows with Tajima's D values significantly different from neutral expectations, we estimated the top and bottom 5% of the distribution across the whole genome. The average proportion of amino acid substitutions driven by directional selection (α ; values range between $-\infty$ and 1) was also estimated in pure species individuals for all introgressed regions using Smith and Eyre-Walker's α (Smith & Eyre-Walker 2002), and generating confidence intervals by randomly selecting genes with replacement (bootstrapping) 1000 times in R v. 3.2.1 (R Development Core Team, <http://www.r-project.org>).

Levels of introgression in sex-determining regions

To determine if sex-associated regions are protected from introgression, we plotted the location of SNPs significantly associated with sex in *Populus* (Geraldes *et al.* 2015) with the levels of admixed ancestry from RASPBerry across six chromosomes (4, 5, 9, 14, 18, 19). Geraldes *et al.* (2015) mapped the sex-linked region to these six chromosomes and two unmapped scaffolds in version 3.0, likely due to genome misassembly.

RESULTS

Local ancestry in RASPBerry vs previous admixture analysis

Eleven of the 129 individuals initially considered as admixed *P. trichocarpa* did not show any *P. balsamifera* ancestry with RASPBerry analysis and were therefore removed from all downstream analyses. These 11 individuals could have been incorrectly classified previously as admixed or have so little *P. balsamifera* ancestry that it was undetectable using RASPBerry (Figure S1; Q values for *P. balsamifera* ancestry = 0.011-0.003). The remaining 118 admixed *P. trichocarpa* contained 0.049% to 15.184% of *P. balsamifera* alleles across their genomes according to RASPBerry analysis. These

estimates were in agreement with previous admixture analysis, though RASpberry estimates (mean: 0.0559, SD 0.0399) were, for the most part, lower than previous estimates of Q (mean 0.0682, SD: 0.0562) based on a different model-based analysis (Alexander et al. 2009). Admixture levels above 1% were geographically limited to interior and northern populations of *P. trichocarpa* (Figure 1). Using introgressed individuals exclusively (*P. trichocarpa* admixture levels between 6.161% and 18.164%) allowed us to detect introgressed blocks against genome-wide expectations for admixture.

Across the genomes (1,168,955 SNPs examined) of 118 admixed *P. trichocarpa* individuals, we detected 19 regions, which included 1107 genes, with unusually high levels of *P. balsamifera* ancestry (Figure 2, Table 1). These regions included candidate genes for adaptive introgression previously identified by our previous targeted study (Suarez-Gonzalez *et al.* 2016). The 19 regions, occurring across 11 chromosomes, showed *P. balsamifera* ancestry peaks with heights (i.e. percentage of *P. balsamifera* ancestry) ranging from 0.1778 to 0.2733 and widths ranging from 140 kb to 1.02 Mb.

Of the 32 individuals initially considered as admixed *P. balsamifera*, 11 were early generation hybrids (*P. balsamifera* ancestry estimates above 20%) and were not included in further downstream calculations (Figure S1; Supporting Information). Local ancestry analyses in the remaining 21 *P. balsamifera* admixed individuals revealed nine regions, containing 545 genes, with unusually high levels of *P. trichocarpa* ancestry, across five chromosomes (Figure 2, Table 1, S1; Supporting information). These introgressed regions showed *P. trichocarpa* ancestry peaks with heights ranging from 0.3262 to 0.4273 and widths ranging from 160 kb to 880 kb.

Introgression patterns are strongly asymmetric

In admixed *P. trichocarpa*, 2.3% of the genome showed unusually high levels of introgression from *P. balsamifera* (total size of introgressed peaks: 9Mb; genome size: 394.5 Mb; Table 1), while in admixed *P. balsamifera*, only 1.04% of the genome showed unusually high levels of introgression from *P. trichocarpa* (total size of introgressed peaks: 4.1Mb; Table 1). Overall levels of genomic admixture were higher in admixed *P. balsamifera* individuals (mean *P. trichocarpa* ancestry: 0.1157, SD 0.0666, weighted by SNP density), compared to admixed *P. trichocarpa* individuals (mean *P. balsamifera* ancestry 0.0559, SD 0.0399), suggesting the former are earlier generation hybrids.

The average width of regions with unusually high levels of introgression was slightly larger, though not significantly ($p > 0.05$), in blocks with *P. balsamifera* ancestry within admixed *P. trichocarpa* individuals (mean 473.68 kb, SD: 252.371) compared to those with *P. trichocarpa* ancestry within *P. balsamifera* individuals (mean 455.56 kb; SD: 227.547 (Figure S2; Supporting Information).

There was no overlap between regions showing unusually high levels of introgression in *P. balsamifera* admixed individuals (9) and those in *P. trichocarpa* admixed individuals (19). Furthermore, the level of introgressed ancestry across the 19 chromosomes differed between admixed *P. trichocarpa* and admixed *P. balsamifera* individuals. In the former, the three chromosomes with the highest levels of *P. balsamifera* ancestry were chromosomes 17, 13 and 9. In the latter, the top three chromosomes with the highest levels of *P. trichocarpa* ancestry were 12, 5 and 16 (Figure S3; Supporting Information).

PCA analysis revealed strong associations between levels of *P. balsamifera* introgression into *P. trichocarpa* and the first principal component of variables associated with climate (temperature PC1 explained 60% of the variance; Pearson's $r = 0.51$, $p < 0.00001$; moisture PC1 explained 63% of the variance; Pearson's $r = 0.23$, $p < 0.01$); with increased levels of introgression occurring in colder and drier environments (Figure 3; Table S2B, Supporting information). *P. trichocarpa* introgression into *P. balsamifera* also increased in colder climates (Pearson's $r = 0.55$, $p < 0.01$), but there was no association with changes in moisture.

Subtelomeric enrichment of introgressed regions

Overall, regions with unusually high levels of introgression were found at various chromosomal locations, although a weak but highly significant negative correlation was evident between levels of *P. balsamifera* introgressed ancestry and distance to telomeres (Pearson $r^2 = -0.074$, $p\text{-value} < 10^{-15}$, Figure 4). In contrast, there was no correlation between levels of *P. trichocarpa* introgressed ancestry and distance to the telomeres in admixed *P. balsamifera* individuals ($r^2 = -0.002$, $p\text{-value} = 0.734$).

Levels of introgression in predicted centromeric regions (Pinosio *et al.* 2016) were well below the significance threshold (i.e. 3SD from the whole genome average) and similar to the

genomic average (*P. trichocarpa* ancestry in admixed *P. balsamifera* individuals: 0.1293; *P. balsamifera* ancestry in *P. trichocarpa* individuals: 0.0492; Figure 2).

Introgressed regions show enrichment for disease resistance genes

Enrichment analyses of biological terms revealed a different set of overrepresented GO and Pfam terms in regions exhibiting unusually high levels of introgression in *P. trichocarpa* admixed individuals (14 GO and 31 Pfam terms) compared to those in *P. balsamifera* admixed individuals (14 GO and 27 Pfam terms; Table S3; Supporting Information). Of the 58 enriched Pfam terms, two were associated with disease resistance genes: The Toll/interleukin-1 receptor (*TIR*, PF01582) was enriched in subtelomeric *P. balsamifera* introgressed regions (6 genes in chromosome 5 and 4 genes in chromosome 7), and the Leucine-rich repeat protein family (*LRR_1*, PF00560) was enriched in *P. trichocarpa* introgressed genes (1 gene on chromosome 5, 1 gene on chromosome 8, and 4 genes on chromosome 16) (Figure 5, Table S4; Supporting Information). In chromosome 5, six *TIR* genes occurred in tandem. In chromosome 7, four *TIR* genes were interspersed with another type of disease resistance genes encoding NB-ARC domains (PF00931) and showed signals of purifying selection (Figure 6). Genes encoding *LRR_1* proteins were located on chromosome 5 (1 gene), 8 (1 gene) and chromosome 16 (4 genes). In chromosome 16, the group of *LRR_1* genes was interspersed with genes coding macrophage migration inhibitory factors (MIF, PF01187).

Only a few common KEGG (1 term) and micro RNA (10 terms) terms were enriched in both *P. balsamifera* and *P. trichocarpa* introgressed regions (Table S3; Supporting Information). The *P. balsamifera* introgressed region in chromosome 3 was enriched for genes that may play crucial roles during response to biotic and abiotic stresses, such as those related to response to oxidative stress (GO:0006979) and peroxidase activity (GO:0004601) (Table S3; Supporting Information).

Introgressed regions in pure *P. balsamifera* and *P. trichocarpa* are under selection

All introgressed regions except for two in chromosome 5 and 16 (Table S5; Supporting Information) contained windows with Tajima's D values either above or below 5% of the whole genome distribution. Significantly positive Tajima's D values (i.e., indicating an excess of intermediate frequency alleles that may result from balancing selection) were present in pure *P. trichocarpa* within three focal regions of special interest to this study (two *P. trichocarpa* regions identified as introgressed in admixed *P. balsamifera* and one *P. balsamifera* region identified as

introgressed in admixed *P. trichocarpa*) and in one region in pure *P. balsamifera* (a region of putative *P. trichocarpa* origin in admixed *P. balsamifera*). Significantly negative Tajima's D values (indicating an excess of low-frequency polymorphisms relative to expectation possibly resulting from purifying selection) were evident in pure *P. trichocarpa* in two focal genome regions (two *P. balsamifera* regions identified as introgressed in admixed *P. trichocarpa*) and two focal regions in pure *P. balsamifera* (two *P. balsamifera* regions in admixed *P. trichocarpa*) (Table 1, Table S5, Supporting information, Figure 6). Two of these *P. balsamifera* regions in *P. trichocarpa* (within chromosomes 7 and 11) and the region showing positive Tajima's D values (within chromosome 10) were telomeric. We note that unusual Tajima's D values in focal regions in pure species does not *per se* allow us to predict balancing or purifying selection in admixed individuals, as any fitness and/or phenotypic effects of introgressed alleles in such individuals is also dependent on their genomic background (Suarez-Gonzalez et al. 2018).

In addition, Alpha - a measure of selection at the amino acid level - revealed signals of positive selection in the *P. balsamifera* introgressed region in chromosome 3 in about 60% of all amino acid substitutions in both *P. trichocarpa* and *P. balsamifera* pure individuals (Table S6, Supporting Information)..

Are sex-determining regions protected from interspecific introgression?

The sex-linked region did not show signals of introgression above the significance threshold (i.e. 3SD from the genomic average). Most SNPs (80%) significantly associated with sex mapped to chromosomes 18 and 19 in the *P. trichocarpa* genome assembly version 3.0, where no introgressed regions were present (Figure S4). Sex-determining regions in chromosomes 4 and 9 showed admixed ancestry levels below 10%, and in chromosome 5 and chromosome 14, the regions that mapped to sex-determination (5 kb in chromosome 5 and 3.5 kb in chromosome 14) were in close proximity to introgressed regions but did not show admixture levels above the significance threshold (Figure S4).

DISCUSSION

Direction and scale of introgression into *P. trichocarpa* in relation to environment

Our whole genome analysis revealed asymmetric patterns of introgression, with stronger signatures of adaptive introgression from *P. balsamifera* into *P. trichocarpa* than in the reverse

direction. Introgression evident in admixed *P. trichocarpa* individuals was geographically limited to northwestern Canada and the Canadian Rockies, where *P. balsamifera* genes associated with adaptation to higher latitudes might allow admixed *P. trichocarpa* to colonize colder environments as indicated in our previous work (Suarez-Gonzalez *et al.* 2016). Here, a PCA analysis revealed increased levels of introgression in both colder and drier environments, and a previous landscape genomic analysis identified 21 SNPs in eight regions with unusually high levels of introgression in *P. trichocarpa* that were strongly correlated with several geoclimate variables including latitude, mean annual temperature and mean coldest temperature (Geraldes *et al.* 2014). Some correlation of this sort is to be expected because of the difference in the parental niches but it should be noted that admixed individuals make up the majority (or entirety) of northern populations in our *P. trichocarpa* collection (Geraldes *et al.* in preparation) and an admixture mapping study revealed strong associations between introgressed regions and adaptive characters that may allow improved survival in new environments (Suarez-Gonzalez *et al.* 2018). Currat *et al.* (2008) have argued for an important effect of demography in promoting the introgression of local alleles to invading populations particularly during active invasions and range expansions. However, our species are largely parapatric and separated over much of their range by natural barriers such as the Rocky Mountains. We therefore do not think that an invasion model matches well with our results. Instead, we appear to have gene flow occurring through a hybrid zone, rather than dispersal of pure species deep inside the range of another. Also, the studied *Populus* species are wind-pollinated and wind-dispersed obligate outcrossers, thus gene flow *within* each species is expected to be high. This makes it even less likely that introgression between these species is facilitated by demographic processes described for invading species (Currat *et al.* 2008; Petit & Excoffier 2009).

Although introgression from *P. balsamifera* to *P. trichocarpa* can be interpreted in adaptive terms, the reverse is less obvious. The type of environments these species are currently occupying does not intuitively suggest that introgression from *P. trichocarpa* into *P. balsamifera* is likely to be adaptive. In fact, introgression from *P. trichocarpa*, a fast-growing tree adapted to mild and coastal climates, to *P. balsamifera*, a slower-growing stress-tolerant species occupying northern latitudes with extreme temperature ranges, might be maladaptive as it could decrease the stress tolerance of *P. balsamifera* (Richardson *et al.* 2014; Geraldes *et al.* 2014).

We uncovered several candidate regions related to adaptive introgression with signals of selection, including subtelomeric regions occupied by disease response genes (see below). These signatures of selection suggest that introgressed regions may be underlying adaptive processes in pure individuals and potentially in hybrids, supporting the hypothesis of adaptive introgression as has been shown in our previous studies (Suarez-Gonzalez et al. 2016; Suarez-Gonzalez et al. 2017). Alternative explanations for these patterns are recent introgression, unidirectional incompatibilities, or increased levels of recombination. Future functional and phenotypic studies could be used to further test if variants acquired from *P. balsamifera* have indeed accelerated adaptation in *P. trichocarpa* and resulted in improved survival in colder environments.

Evidence for the role of selection in introgression

We obtained several indications that introgression in this system is not neutral. While none of these indications is conclusive on its own, taken overall they are strongly suggestive of selection. First, certain genome regions are hotspots for introgression. One explanation for this, and perhaps the simplest one in light of our local ancestry results (Fig. 2), is that these regions are under selection and therefore tend to introgress faster. The term “comet alleles” has been used to describe similar rapidly introgressed haplotypes in mice (Staubach et al. 2012).

Second, we have detected significant enrichment for certain gene ontology terms, which suggests that the strongly introgressed regions are not a random sampling of the genome as might be expected under neutrality. In addition, it is striking that some of the functional categories enriched contain genes that may play crucial roles in survival and adaptation, such as those related to disease resistance (see below).

Third, we found signatures of purifying and balancing selection in pure individuals in genomic regions identified as introgressed in admixed samples. Purifying selection in the donor species *P. balsamifera* suggests that variation at such loci is functionally relevant. The functional importance of such genes in the donor species is exemplified by one protein family associated with disease resistance (TIR, PF01582), that is over-represented in a *P. balsamifera* introgressed region in chromosome 7, showing signals of purifying selection. Although we cannot be certain that the signals of balancing selection in pure *P. trichocarpa* also apply to admixed individuals, this finding is consistent with the maintenance of allelic variation (including potentially introgressed variants) by

selection. As noted elsewhere “... natural selection should favor introgression for alleles at genes evolving under multi-allelic balancing selection, such as the MHC in vertebrates, disease resistance, or self-incompatibility genes in plants” (Castric et al., 2008). This is unsurprising, as diseases can be important agents of tree mortality (Holdenrieder et al., 2004). Finally, we should note the findings of a previous study that many introgressed regions are in fact associated with adaptively-relevant phenotypes (Suarez-Gonzalez et al., 2018).

Implications for adaptation to future climates

Changing climate conditions are expected to have a far-reaching impact on forest species (IPCC et al. 2014; Sturrock et al. 2011), especially if warmer temperatures and increased precipitation increase the probability of tree diseases in northern locations. Southern *P. trichocarpa* populations may have been under intense selection pressures for disease resistance, as suggested by the negative correlation between latitude and leaf rust severity across *P. trichocarpa* populations from the Pacific Northwest (La Mantia et al. 2013). If changes in climate result in more favorable conditions for pathogen reproduction and survival at higher latitudes (Hicke et al. 2012; IPCC et al. 2014; Helfer 2014), pathogen aggressiveness may become increasingly important for *P. balsamifera*.

Here we show that genes introgressed from *P. trichocarpa* into *P. balsamifera* were enriched for the leucine-rich repeat protein family (LRR_1, PF00560), which has been associated with innate immunity and sensing of pathogen-associated molecular patterns (Ng & Xavier 2011). Furthermore, these introgressed *LRR_1* genes were interspersed by genes coding for macrophage migration inhibitory factors (MIF, PF01187), which in mammals regulate adaptive and innate immunity (Ståldal et al. 2012). It has recently been suggested that *MIF* genes may have a parallel function in plants, regulating the stress response including innate immunity (Panstruga et al. 2015).

In addition, our results suggest that introgression from *P. trichocarpa* into *P. balsamifera* may be more recent than introgression in the opposite direction due to the higher levels of admixture found, implying earlier generation hybrids with less backcrossing (despite the similar sizes of introgressed genomic blocks). However, this introgression is more evident in cold environments, contrary to our expectations. This raises the intriguing possibility of climate change reversing the historical direction of introgression. If *P. trichocarpa* can indeed contribute beneficial disease

resistance genes, adaptive introgression into *P. balsamifera* could play crucial roles under changing climatic conditions.

Disease resistance genes and structural properties may explain subtelomeric enrichment of introgressed regions

The potential sub-telomeric enrichment of introgressed regions in *P. trichocarpa* is an interesting finding of our study and could be explained by (1) the type of gene families found in the subtelomeres and (2) the unstable structural properties of these regions. In yeast, subtelomeric families are enriched for specific functional categories including response to stress and toxins, metabolism of a broad spectrum of compounds, as well as transporters (Brown *et al.* 2010). In plants, including maize, barley, and poplar, subtelomeres are often occupied by large clusters of plant resistance genes (Geffroy *et al.* 2000; Wei *et al.* 1999; Duplessis *et al.* 2009). Our study revealed one protein family associated with disease resistance (*TIR*, PF01582) overrepresented in *P. balsamifera* introgressed regions and, interestingly, all of the *TIR* introgressed genes were subtelomeric and arranged in clusters that showed signals of purifying selection (i.e. negative values of Tajima's D lower than the neutral expectation). Taken together, these results show that subtelomeres are occupied by, and may be enriched for, gene families associated with local adaptation, where interspecific introgression may transfer important adaptive traits.

Subtelomeres, often the most variable region of eukaryotic genomes (Winzeler *et al.* 2003; Brown *et al.* 2010), are composed of various repeated elements and characterized by increased recombination, duplication, and mutation which may allow evolutionary adaptation and innovation (Rudd *et al.* 2007; Linardopoulou *et al.* 2005; Barton *et al.* 2008). While increased duplication and mutation rates also imply the possibility of falsely calling genotypes as heterozygous in bioinformatic pipelines that rely on mapping sequence reads to a reference genome, this is less likely to be an issue in our study, as our results are based on local ancestries rather than on raw genotype data.

In poplar, there is a weak negative correlation between distance to telomeres and levels of recombination (Slavov *et al.* 2012). A computational analysis in yeast revealed that subtelomeric families are evolving and expanding much faster than families that do not contain subtelomeric genes. Furthermore, non-subtelomeric genes that belong to the same functional categories showed lower variability compared with their subtelomeric counterparts (Brown *et al.* 2010). These results

suggest that the high rates of sequence evolution and rapid gene turnover are inherent properties of subtelomeric regions rather than of the functional categories of genes. In *Plasmodium falciparum*, the subtelomeric location of *var* genes may favour the increased diversity in antigenicity, an effective system to evade immune systems, due to higher levels of recombination in subtelomeric regions (Rubio *et al.* 1996). In humans, subtelomere dynamics including recent duplications between the subtelomeres of different chromosomes, have resulted in a striking variation of subtelomerically located *OR* genes and may contribute to the diversity in olfactory perception (Mefford & Trask 2002; Trask *et al.* 1998).

Overall, these studies show that subtelomeres harbor fast-evolving gene families and are hot spots of recombination and duplications. Our analysis suggests adaptive introgression as an important potential corollary of the extensive sequence variation found in subtelomeric genes, and more fine-scale genomic studies of introgression are needed to evaluate the relative importance of the subtelomeric properties in shaping the architecture of adaptive introgression.

Possible protection of the sex-determining regions from interspecific gene flow

Sex-determining regions have previously been shown to be protected from interspecific gene flow (Hu & Filatov 2016). Our results are consistent with this, as shown here by the lack of significant signals of introgression in the poplar sex-determining region (SDR) (Geraldes *et al.* 2015). Previous studies in *Populus* have consistently mapped the gender determining locus to the proximal telomeric region of chromosome 19 (Gaudet *et al.* 2008; Yin *et al.* 2008; Geraldes *et al.* 2015), which is also in close proximity to the largest NBS disease resistance gene cluster in poplar (Kohler *et al.* 2008). Although subtelomeric regions and genes associated with adaptation could be more prone to be introgressed, our analysis suggests that sexually antagonistic mutations may be impeding introgression in the SDR of chromosome 19. In fact, it has been suggested that suppressed recombination associated with resistance genes may have co-evolved with sexual differentiation and allowed the emergence of an incipient sex chromosome in poplar (Tuskan *et al.* 2012; Caseys *et al.* 2015). Geraldes *et al.* (2015), the first study to fully characterize the SDR in *P. trichocarpa*, found no evidence of an incipient sex chromosome. Instead, the non-recombining region has remained remarkably small (13 genes over c. 100kb), although it does contain several putative resistance genes (Geraldes *et al.* 2015).

Geraldes et al. (2015) also found SNPs associated with sex in five additional chromosomes, likely due to a very poor genome assembly (in both v2.2 and v3.0) of the subtelomeric SDR of chromosome 19. Since the sex locus is fragmented through the genome by poor assembly in the current version of the genome (used in this study), further refinement of the assembly is needed to make firm inferences about introgression patterns near the SDR in *Populus*.

Taken on balance, the results detailed here provide strong support for the hypothesis that hybridization between species is an important source of adaptively significant variation. The asymmetry of this effect between *P. balsamifera* and *P. trichocarpa* further indicates that the effects of introgression on adaptation depends strongly on ecological contexts.

ACKNOWLEDGEMENTS

This work was supported by the Genome Canada Large-Scale Applied Research Program (POPCAN, project 168BIO), funds to QCBC and CJD, by a Natural Sciences and Engineering Research Council of Canada Discovery Grant to CJD (RGPIN 36485-12), and by grants from the Swiss National Science Foundation (SNF) to CL. We thank Michael Friedmann (UBC) for assistance with project management; and Daniel Wegmann (U Fribourg) and Camille Christe for assistance and discussions. We also thank Richard Abbott, and three anonymous referees for comments that greatly improved the manuscript.

REFERENCES

- Abbott RJ (2017) Plant speciation across environmental gradients and the occurrence and nature of hybrid zones. *Journal of Systematics & Evolution* **55**, 238-258.
- Abbott RJ, Albach D, Ansell S et al. (2013) Hybridization and speciation. *Journal of Evolutionary Biology* **26**: 229–246.
- Alexander DH, Novembre J, Lange K (2009) Fast model-based estimation of ancestry in unrelated individuals. *Genome Research*, **19**, 1655-1664.

512 Barton AB, Pekosz MR, Kurvathi RS, Kaback DB (2008) Meiotic recombination at the ends of
 513 chromosomes in *Saccharomyces cerevisiae*. *Genetics*, **179**, 1221-1235.

514 Barton NH, Hewitt GM (1985) Analysis of Hybrid Zones. *Annual Review of Ecology and Systematics*,
 515 **16**, 113-148.

516 Brown CA, Murray AW, Verstrepen KJ (2010) Rapid expansion and functional divergence of
 517 subtelomeric gene families in yeasts. *Current Biology*, **20**, 895-903.

518 Buerkle CA, Lexer C (2008) Admixture as the basis for genetic mapping. *Trends in Ecology &*
 519 *Evolution*, **23**, 686-694.

520 Caseys C, Stritt C, Glauser G, Blanchard T, Lexer C (2015) Effects of hybridization and evolutionary
 521 constraints on secondary metabolites: The genetic architecture of phenylpropanoids in
 522 european *Populus* species. *PLoS ONE*, **10**, e0128200.

523 **Castric, V., Bechsgaard, J., Schierup, M. H., & Vekemans, X. (2008). Repeated adaptive introgression**
 524 **at a gene under multiallelic balancing selection. PLoS Genetics, 4(8), e1000168.**

525 Cingolani P, Platts A, Wang LL, et al (2012) A program for annotating and predicting the effects of
 526 single nucleotide polymorphisms, SnpEff. *Fly*, **6**, 80-92.

527 Cronk QCB (2005) Plant eco-devo: the potential of poplar as a model organism. *New Phytologist*,
 528 **166**, 39-48.

529 Currat M, Ruedi M, Petit R, Excoffier L (2008) The hidden side of invasions: Massive introgression by
 530 local genes. *Evolution*, **62**, 1908-1920.

531 Danecek P, Auton A, Abecasis G, et al (2011) The variant call format and VCFtools. *Bioinformatics*,
 532 **27**, 2156-2158.

533 Dangl JL, Jones JDG (2001) Plant pathogens and integrated defence responses to infection. *Nature*,
 534 **411**, 826-833.

535 Dowling TE, Secor CL (1997) The role of hybridization and introgression in the diversification of
 536 animals. *Annual Review of Ecology and Systematics*, **28**, 593-619.

537 Duplessis S, Major I, Martin F, Seguin A (2009) Poplar and pathogen interactions: Insights from
 538 Populus genome-wide analyses of resistance and defense gene families and gene expression
 539 profiling. *Critical Reviews in Plant Sciences*, **28**, 309-334.

540 Ellstrand NC, Prentice HC, Hancock JF (1999) Gene flow and introgression from domesticated plants
 541 into their wild relatives. *Annual Review of Ecology and Systematics*, **30**, 539-563.

542 Finn RD, Coghill P, Eberhardt RY, et al (2016) The Pfam protein families database: towards a more
 543 sustainable future. *Nucleic Acids Research*, **44**, 279-285.

544 Gaudet M, Jorge V, Paolucci I, Beritognolo I, Mugnozza GS, Sabatti M (2008) Genetic linkage maps of
 545 *Populus nigra* L. including AFLPs, SSRs, SNPs, and sex trait. *Tree Genetics & Genomes*, **4**, 25-36.

546 Geffroy V, Seignac M, De Oliveira, Julio C. F., et al (2000) Inheritance of partial resistance against
 547 *Colletotrichum lindemuthianum* in *Phaseolus vulgaris* and co-localization of quantitative trait loci
 548 with genes involved in specific resistance. *Molecular Plant-Microbe Interactions*, **13**, 287-296.

549 Geraldès A., Difazio, S.P., Slavov, G.T. et al. (2013). A 34K SNP genotyping array for *Populus*
 550 *trichocarpa*: design, application to the study of natural populations and transferability to other
 551 *Populus* species. *Molecular Ecology Resources*, **13**(2), 306-323.

552 Geraldès A, Hefer CA, Capron A, et al (2015) Recent Y chromosome divergence despite ancient origin
 553 of dioecy in poplars (*Populus*). *Molecular Ecology*, **24**, 3243-3256.

554 Geraldès A, Farzaneh N, Grassa CJ, et al (2014) Landscape genomics of *Populus trichocarpa*: the role
 555 of hybridization, limited gene flow, and natural selection in shaping patterns of population
 556 structure. *Evolution*, **68**, 3260-3280.

557 Hamilton JA, Miller JM (2015) Adaptive introgression as a resource for management and genetic
558 conservation in a changing climate. *Conservation Biology*, **30**, 33-41.

559 Harrison RG (1990) Hybrid zones: windows on evolutionary proces. In: *Oxford Surveys in*
560 *Evolutionary Biology* (eds. Futuyma D, Antonovics J) pp. 69-128 Oxford University Press, New York.

561 Hedrick PW (2013) Adaptive introgression in animals: examples and comparison to new mutation
562 and standing variation as sources of adaptive variation. *Molecular Ecology*, **22**, 4606-4618.

563 Helfer S (2014) Rust fungi and global change. *New Phytologist*, **201**, 770-780.

564 Hicke JA, Allen CD, Desai AR, et al (2012) Effects of biotic disturbances on forest carbon cycling in the
565 United States and Canada. *Global Change Biology*, **18**, 7-34.

566 Holdenrieder, O., Pautasso, M., Weisberg, P. J., & Lonsdale, D. (2004). Tree diseases and landscape
567 processes: the challenge of landscape pathology. *Trends in Ecology & Evolution*, **19**(8), 446-452.

568 Hu X, Filatov DA (2016) The large-X effect in plants: increased species divergence and reduced gene
569 flow on the *Silene* X-chromosome. *Molecular Ecology*, **25**, 2609-2619.

570 Huerta-Sanchez E, Jin X, Asan, et al (2014) Altitude adaptation in Tibetans caused by introgression of
571 Denisovan-like DNA. *Nature*, **512**, 194-197.

572 IPCC, Pachauri RK, Meyer LA, eds (2014) *Climate Change 2014: Synthesis Report. Contribution of*
573 *Working Groups I, II and III to the Fifth Assessment Report of the Intergovernmental Panel on Climate*
574 *Change* IPCC, Geneva.

575 Ismail M, Soolanayakanahally R, Ingvarsson P, et al (2012) Comparative nucleotide diversity across
576 North American and European *Populus* species. *Journal of Molecular Evolution*, **74**, 257-272.

577 Jansson S, Douglas CJ (2007) *Populus*: A model system for plant biology. *Annual Review of Plant*
578 *Biology*, **58**, 435-458.

579 Keller SR, Soolanayakanahally RY, Guy RD, Silim SN, Olson MS, Tiffin P (2011) Climate-driven local
 580 adaptation of ecophysiology and phenology in balsam poplar, *Populus balsamifera* L. (Salicaceae).
 581 *American Journal of Botany*, **98**, 99-108.

582 Kohler A, Rinaldi C, Duplessis S, et al (2008) Genome-wide identification of NBS resistance genes in
 583 *Populus trichocarpa*. *Plant Molecular Biology*, **66**, 619-636.

584 La Mantia J, Klápště J, El-Kassaby Y, et al (2013) Association analysis
 585 identifies *Melampsora* × *columbiana* Poplar leaf rust resistance SNPs. *PLoS ONE*, **8**, e78423.

586 Levensen ND, Tiffin P, Olson MS (2012) Pleistocene speciation in the genus *Populus* (Salicaceae).
 587 *Systematic Biology*, **61**, 401-412.

588 Lexer C, Heinze B, Alia R, Rieseberg LH (2004) Hybrid zones as a tool for identifying adaptive genetic
 589 variation in outbreeding forest trees: lessons from wild annual sunflowers (*Helianthus* spp.). *Forest*
 590 *Ecology and Management*, **197**, 49-64.

591 Linardopoulou EV, Williams EM, Fan Y, Friedman C, Young JM, Trask BJ (2005) Human subtelomeres
 592 are hot spots of interchromosomal recombination and segmental duplication. *Nature*, **437**, 94-100.

593 Little EL (1971) Conifers and important hardwoods. In: *Atlas of United States trees* (ed. U.S.
 594 Department of Agriculture Miscellaneous Publication) pp. 1146-1155 U.S. Department of Agriculture
 595 Miscellaneous Publication, .

596 Martin N, Bouck A, Arnold M (2006) Detecting adaptive trait introgression between *Iris fulva* and *I.*
 597 *brevicaulis* in highly selective field conditions. *Genetics*, **172**, 2481-2489.

598 McKown AD, Guy RD, Klápště J, et al (2013) Geographical and environmental gradients shape
 599 phenotypic trait variation and genetic structure in *Populus trichocarpa*. *New Phytologist*, **201**, 1263-
 600 1276.

601 Mefford HC, Trask BJ (2002) The complex structure and dynamic evolution of human subtelomeres.
602 *Nature reviews. Genetics*, **3**, 91-102.

603 Ng A, Xavier RJ (2011) Leucine-rich repeat (LRR) proteins. *Autophagy*, **7**, 1082-1084.

604 Nielsen R (2005) Molecular signatures of natural selection. *Annual Reviews in Genetics*, **39**, 197-218.

605 Oleksyn J, Modrzyński J, Tjoelker MG, Zaytkowiak R, Reich PB, Karolewski P (1998) Growth and
606 physiology of *Picea abies* populations from elevational transects: common garden evidence for
607 altitudinal ecotypes and cold adaptation. *Functional Ecology*, **12**, 573-590.

608 Panstruga R, Baumgarten K, Bernhagen J (2015) Phylogeny and evolution of plant macrophage
609 migration inhibitory factor/D-dopachrome tautomerase-like proteins. *BMC Evolutionary Biology*, **15**,
610 1-16.

611 Pardo-Diaz C, Salazar C, Baxter SW, Merot C, Figueiredo-Ready W, et al. (2012) Adaptive
612 introgression across species boundaries in *Heliconius* butterflies. *PLoS Genetics*, **8**, e1002752.

613 Petit RJ, Excoffier L (2009) Gene flow and species delimitation. *Trends in Ecology and Evolution*, **24**,
614 386-393.

615 Pinosio S, Giacomello S, Faivre-Rampant P, et al (2016) Characterization of the Poplar Pan-Genome
616 by Genome-Wide Identification of Structural Variation. *Molecular Biology and Evolution*, **33**, 2706-
617 2719.

618 Purcell S, Neale B, Todd-Brown K, et al (2007) PLINK: a toolset for whole-genome association and
619 population-based linkage analysis. *American Journal of Human Genetics*, **81**, 559-575.

620 Richardson J, Isebrands JG, Ball JB (2014) Ecology and physiology of poplars and willows. In: *Poplars*
621 *and Willows: Trees for Society and the Environment* (eds. Isebrands JG, Richardson J) pp. 92-123
622 CABI, Oxfordshire, UK.

623 Rieseberg LH, Raymond O, Rosenthal DM, et al (2003) Major Ecological Transitions in Wild
 624 Sunflowers Facilitated by Hybridization. *Science*, **301**, 1211.

625 Roshyara NR, Scholz M (2014) fcGENE: A versatile tool for processing and transforming SNP
 626 datasets. *PLoS ONE*, **9**, e97589.

627 Rubio JP, Thompson JK, Cowman AF (1996) The var genes of *Plasmodium falciparum* are located in
 628 the subtelomeric region of most chromosomes. *The EMBO journal*, **15**, 4069-4077.

629 Rudd MK, Friedman C, Parghi SS, Linardopoulou EV, Hsu L, Trask BJ (2007) Elevated Rates of Sister
 630 Chromatid Exchange at Chromosome Ends. *PLoS Genetics*, **3**, e32.

631 Savolainen O, Pyhäjärvi T, Knürr T (2007) Gene Flow and Local Adaptation in Trees. *Annual Review of*
 632 *Ecology, Evolution, and Systematics*, **38**, 595-619.

633 Scheet P, Stephens M (2006) A fast and flexible statistical model for large-scale population genotype
 634 data: Applications to inferring missing genotypes and haplotypic phase. *American Journal of Human*
 635 *Genetics*, **78**, 629-644.

636 Schmickl R, Marburger S, Bray S, Yant L. 2017. Hybrids and horizontal transfer: introgression allows
 637 adaptive allele discovery. *Journal of Experimental Botany* **68**, 5453-5470.

638 Sjödin A, Street NR, Sandberg G, Gustafsson P, Jansson S (2009) PopGenIE: The *Populus* genome
 639 integrative explorer. A new tool for exploring the *Populus* genome. *New Phytologist*, **182**, 1013-
 640 1025.

641 Slarkin M (1985) Gene Flow in Natural Populations. *Annual Review of Ecology and Systematics*, **16**,
 642 393-430.

643 Slavov GT, DiFazio SP, Martin J, et al (2012) Genome resequencing reveals multiscale geographic
 644 structure and extensive linkage disequilibrium in the forest tree *Populus trichocarpa*. *New*
 645 *Phytologist*, **196**, 713-725.

646 Smith NGC, Eyre-Walker A (2002) Adaptive protein evolution in *Drosophila*. *Nature*, **415**, 1022-1024.

647 Soolanayakanahally RY, Guy RD, Silim SN, Drewes EC, Schroeder WR (2009) Enhanced assimilation
648 rate and water use efficiency with latitude through increased photosynthetic capacity and internal
649 conductance in balsam poplar (*Populus balsamifera* L.). *Plant, Cell & Environment*, **32**, 1821-1832.

650 Stace CA, Preston CD, Pearman DA (2015) *Hybrid flora of the British Isles*. Bristol: Botanical Society of
651 Britain and Ireland.

652 Ståldal V, Cierlik I, Chen S, et al (2012) The *Arabidopsis thaliana* transcriptional activator STYLISH1
653 regulates genes affecting stamen development, cell expansion and timing of flowering. *Plant*
654 *Molecular Biology*, **78**, 545-559.

655 Sturrock RN, Frankel SJ, Brown AV, et al (2011) Climate change and forest diseases. *Plant Pathology*,
656 **60**, 133-149.

657 Suarez-Gonzalez A, Hefer CA, Christe C, et al (2016) Genomic and functional approaches reveal a
658 case of adaptive introgression from *Populus balsamifera* (balsam poplar) in *P. trichocarpa* (black
659 cottonwood). *Molecular ecology*, **25**, 2427-2442.

660 Suarez-Gonzalez, A., Hefer, C. A., Lexer, C., Douglas, C. J., & Cronk, Q. C. B. (2018). Introgression
661 from *Populus balsamifera* underlies adaptively significant variation and range boundaries in *P.*
662 *trichocarpa*. *New Phytologist* 217, 416-427.

663

664 Staubach, F., Lorenc, A., Messer, P. W., Tang, K., Petrov, D. A., & Tautz, D. (2012). Genome patterns
665 of selection and introgression of haplotypes in natural populations of the house mouse (*Mus*
666 *musculus*). *PLoS Genetics*, 8(8), e1002891.

667 Trask BJ, Friedman C, Martin-Gallardo A, et al (1998) Members of the olfactory receptor gene family
668 are contained in large blocks of DNA duplicated polymorphically near the ends of human
669 chromosomes. *Human Molecular Genetics*, **7**, 13-26.

670 Tuskan GA, DiFazio S, Jansson S, et al (2006) The genome of black cottonwood, *Populus trichocarpa*
671 (Torr. & Gray). *Science*, **313**, 1596-1604.

672 Tuskan GA, DiFazio S, Faivre-Rampant P, et al (2012) The obscure events contributing to the
673 evolution of an incipient sex chromosome in *Populus*: a retrospective working hypothesis. *Tree*
674 *Genetics & Genomes*, **8**, 559-571.

675 Twyford AD, Ennos RA (2012) Next-generation hybridization and introgression. *Heredity*, **108**, 179-
676 189.

677 Viereck LA, Foote JM (1970) The status of *Populus balsamifera* and *P. trichocarpa* in Alaska.
678 *Canadian Field-Naturalist*, **84**, 169-173.

679 Wegmann D, Kessner DE, Veeramah KR, et al (2011) Recombination rates in admixed individuals
680 identified by ancestry-based inference. *Nature Genetics*, **43**, 847-853.

681 Wei F, Gobelman-Werner K, Morroll SM, et al (1999) The *Mla* (Powdery Mildew) resistance cluster is
682 associated with three NBS-LRR gene families and suppressed recombination within a 240-kb DNA
683 interval on chromosome 5S (1HS) of barley. *Genetics*, **153**, 1929.

684 Weigel D, Nordborg M (2015) Population genomics for understanding adaptation in wild plant
685 species. *Annual Review of Genetics*, **49**, 315-338.

686 Whitney K, Randell RA, Rieseberg L (2006) Adaptive introgression of herbivore resistance traits in
687 the weedy sunflower *Helianthus annuus*. *The American Naturalist*, **167**, 794-807.

688 Winzeler EA, Castillo-Davis C, Oshiro G, et al (2003) Genetic Diversity in Yeast Assessed With Whole-
689 Genome Oligonucleotide Arrays. *Genetics*, **163**, 79.

690 Xie C, Ying CC, Yanchuk AD, Holowachuk DL (2009) Ecotypic mode of regional differentiation caused
691 by restricted gene migration: a case in black cottonwood (*Populus trichocarpa*) along the Pacific
692 Northwest coast. *Canadian Journal of Forest Research*, **39**, 519-525.

693 Yin T, DiFazio SP, Gunter LE, et al (2008) Genome structure and emerging evidence of an incipient
694 sex chromosome in *Populus*. *Genome Research*, **18**, 422-430.

695

696 DATA ACCESSIBILITY

697 All the raw sequencing data and alignment information have been deposited on SRA
698 (PRJNA276056). The list and annotation of all the introgressed genes, and the results from Tajima's
699 D estimations have been archived in dryad (doi:10.5061/dryad.p82f2).

700

701 AUTHOR CONTRIBUTIONS

702 A.S.-G., C.J.D., Q.C.B.C, and C.L. designed the study. A.S.-G performed the research, analysed data,
703 and wrote the manuscript. C.H. performed S.N.P. calling. C.J.D., Q.C.B.C. and C.L. provided funding.
704 Q.C.B.C, and C.L edited the manuscript and all the authors approved the final version.

705

706 TABLES

707 **Table 1.** Summary of the *Populus balsamifera* introgressed regions in admixed *P. trichocarpa*
708 individuals (italics) and *P. trichocarpa* introgressed regions in admixed *P. balsamifera* individuals
709 (bold). The size of the introgressed peaks is represented in kilobases (kb) and number of genes (#).
710 The height of the introgressed peaks is given in terms of the maximum units of standard deviations
711 from the mean within each peak ('Max. balsa' & 'Max. tricho'). Average Tajima's D (D) was
712 calculated in each introgressed region and compared with the average from windows without
713 introgressed peaks using ANOVA. Average Tajima's Ds (D) noted were significantly different from
714 the average in windows without introgressed regions. A measure of selection at the amino acid
715 levels is given using alpha. The value is given for those regions where a significant value was
716 recovered.

Genomic coordinates			Size		Height		Selection			
Chro	Start	End	kb	#Genes	Max ^a balsa	Max ^a tricho	D ^b balsa	D ^b tricho	Alpha balsa	Alpha tricho
Ch01	47,120,000	47,540,000	420	43	3.88					

	48,000,000	48,380,000	380	23	3.51		-0.72	
Ch03	7,380,000	7,740,000	360	22		3.27	0.36	
	21,040,000	21,620,000	580	95	3.87			0.62 0.59
Ch05	140,000	380,000	240	43	3.64			
	4,060,000	4,640,000	580	93		3.73	0.39	
	4,980,000	5,440,000	460	50		3.75		
	11,580,000	11,820,000	240	17	3.28		-0.92	
Ch06	3,540,000	3,700,000	160	22	3.09			
Ch07	14,680,000	15,700,000	1,020	145	5.51		-0.33	
Ch08	740,000	1,620,000	880	162		4.47	0.47	
Ch09	1,380,000	1,760,000	380	22	3.53			
Ch10	22,180,000	22,660,000	480	74	3.63		0.44	
Ch11	1,940,000	2,120,000	180	25	3.24			
	17,580,000	17,780,000	200	16	3.68			
	17,800,000	18,600,000	800	97	4.11		-0.57	
Ch12	1,380,000	1,680,000	300	43		3.25		
Ch14	12,720,000	13,300,000	580	63	4.33			
	13,340,000	13,480,000	140	39	3.46			
Ch15	0	580,000	580	87	3.72			
	13,480,000	13,920,000	440	71	3.48			
Ch16	11,200,000	11,420,000	220	24		3.31		
	12,000,000	12,160,000	160	16		3.16		
	12,320,000	12,980,000	660	93		4.28		
	13,120,000	13,600,000	480	42		4.68		
Ch17	3,080,000	3,900,000	820	98	4.03			
	9,720,000	10,420,000	700	42	5.17			
	12,720,000	13,380,000	660	85	4.39			

^aThe height of the introgressed peaks in terms of units of standard deviations from the mean. 'Max. balsa' and 'Max. tricho' represent the highest value within the corresponding introgressed peak in *P. balsamifera* and *P. trichocarpa* respectively.

^b Average Tajima's D (D) significantly different from the average in windows without introgressed regions in pure *P. balsamifera* (balsa) and *P. trichocarpa* (tricho).

FIGURES

Figure 1. Geographic distribution of admixed (squares) and reference (triangles) individuals used in local ancestry analysis across the contact zones between *P. trichocarpa* and *P. balsamifera*. Blue represents pure *P. balsamifera*. Red represents pure *P. trichocarpa*. Ranges of *P. trichocarpa* and *P. balsamifera* are shown in red and blue, respectively (Little 1971).

Figure 2. Proportion of admixed ancestry across 19 chromosomes showing unusually high levels of introgression (sliding window analysis: size: 100-kb, step: 20-kb) in some regions. Top: Proportion of *P. balsamifera* ancestry in admixed *P. trichocarpa* individuals. Bottom: Proportion of *P. trichocarpa* ancestry in admixed *P. balsamifera* individuals. Regions with unusually high levels of introgression – peaks above broken line - have ancestry higher than 3 standard deviations from the weighted mean across the whole genome based on SNP density per window (broken line). Chromosomes in gray did not show unusually high levels of introgression. Putative centromeres are shown by solid gray lines (Pinosio et al 2016). Blue color marks chromosomes with significant introgression from *P. balsamifera* and red marks chromosomes with introgression from *P. trichocarpa*

Figure 3. Relationship between levels of admixture and environmental variables. Top: variables associated with moisture. Bottom: variables associated with temperature. Left: *P. balsamifera* (Pb) levels in admixed *P. trichocarpa* individuals. Right: *P. trichocarpa* (Pt) levels in admixed *P. balsamifera* individuals.

Figure 4. Relationship between the levels of *P. balsamifera* introgressed (=admixed) ancestry and distance to the telomeres in admixed *P. trichocarpa* individuals (Pearson $r^2 = -0.074$, $p\text{-value} < 10^{-15}$).

Figure 5. Chromosome diagrams depicting introgressed regions (arrows) enriched for Pfam terms associated with R proteins. Purple: All genes encoding NBS-LRR domains across four chromosomes. Blue: *P. balsamifera* introgression into admixed *P. trichocarpa* individuals. Red: *P. trichocarpa* introgression into admixed *P. balsamifera* individuals.

Figure 6. Average Tajima's D within each introgressed region (average of 20-kb windows within each of the introgressed blocks) compared with all non-introgressed windows combined (far left, error bars too small to be visible) in pure *P. trichocarpa* and *P. balsamifera*. Red bars are *P. trichocarpa* introgressed regions in *P. balsamifera* and blue bars are *P. balsamifera* introgressed regions in *P. trichocarpa* that showed average Tajima's D values significantly different from those in windows

761 without introgressed regions. Gray bars show introgressed regions where the average Tajima's D
762 values were not significantly different from those in windows without introgressed regions.

763

Accepted author version