Adaptive introgression and standing genetic variation, two facilitators of adaptation to high latitudes in European aspen (*Populus tremula* L.)

Martha Rendón-Anaya[1], Jonathan Wilson[1], Sæmundur Sveinsson[2], Aleksey Fedorkov[3], Joan Cottrell[4], Mark E.S. Bailey[5], Dainis Ruņģis[6], Christian Lexer[7†], Stefan Jansson[8], Kathryn M. Robinson[8], Nathaniel R. Street[8], Pär K. Ingvarsson[1*]

[1] Linnean Centre for Plant Biology, Department of Plant Biology, Uppsala BioCenter, Swedish University of Agricultural Science, Uppsala, Sweden.

[2] Matis Ltd. Vinlandsleid, Reykjavik Iceland.

[3] Institute of Biology, Komi Science Center, Russian Academy of Sciences, Syktyvkar, Russia.

[4] Forest Research, Northern Research Station, Roslin, UK

[5] School of Life Sciences, College of Medical, Veterinary and Life Sciences, University of Glasgow, Glasgow, UK.

[6] Genetic Resource Centre, Latvian State Forest Research Institute “Silava”, LV2169 Salaspils, Latvia

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

[8] Umeå Plant Science Centre, Department of Plant Physiology, Umeå University, Umeå, Sweden

† Deceased.

* Corresponding author: par.ingvarsson@slu.se

Abstract

Understanding local adaptation in plants from a genomic perspective has become a key research area given the ongoing climate challenge and the concomitant requirement to conserve genetic resources. Perennial plants, such as forest trees, are good models to study local adaptation given their wide geographic distribution, largely outcrossing mating systems and demographic histories. We evaluated signatures of local adaptation in European aspen (*Populus tremula*) across Europe by means of whole genome re-sequencing of a collection of 411 individual trees. We dissected admixture patterns between aspen lineages and observed a strong genomic mosaicism in Scandinavian trees, evidencing different colonization trajectories into the peninsula from Russia, Central and Western Europe. As a consequence of the secondary contacts between populations after the last glacial maximum (LGM), we detected an adaptive introgression event in a genome region of ~500kb in chromosome 10, harboring a large-effect locus that has previously been shown to contribute to adaptation to the short growing seasons characteristic of northern Scandinavia. Demographic simulations and ancestry inference suggest an Eastern origin - probably Russian - of the adaptive Nordic allele which nowadays is present in a homozygous state at the north of Scandinavia. The strength of introgression and positive selection signatures in this region is a unique feature in the genome. Furthermore, we detected signals of balancing selection, shared across regional populations, that highlight the importance of standing variation as a primary source of alleles that facilitate local adaptation. Our results therefore emphasize the importance of migration-selection balance underlying the genetic architecture of key adaptive quantitative traits.

Introduction

Local adaptation, the means by which populations of a species genetically adjust to local environments, is a powerful process of evolution. It occurs because multiple environmental factors imposing different selective pressures exist and the strength of each factor varies across habitats. When a population colonizes a new habitat, certain environmental conditions will impose higher selective pressures, while natural selection may be relaxed for other environmental factors. The overall shift in the selection landscape leads to local adaptation and consequent fitness trade-offs [1]. Two fundamental genetic sources for local adaptation, particularly in temperate forest trees that have large effective population size and low nucleotide substitution rates per unit of time, are standing variation and intra/inter-species hybridization. Hybridization occurs when reproductive isolation is not complete between species or when species lineages that are separated geographically meet after a secondary contact. Species capable of hybridization will therefore have access to a larger pool of genetic variation that provides the raw material for selection and accelerated adaptation. At the same time, standing variation can be maintained through balancing selection (BS) within populations. The signatures of BS include increased diversity around the target of selection, differentiation between populations departing from the genome-wide average and increased linkage disequilibrium, among others [2]. When selection varies geographically, it may favor locally adapted alleles in the derived lineages or subpopulations, giving rise to genomic regions with elevated F_{ST} and absolute divergence D_{XY} [3, 4].

Throughout the entire Pleistocene, range expansions and contractions of different species occurred in Europe, influencing the current patterns of diversity in many taxa [5]. Paleoclimate model simulations and pollen and plant macrofossil records have revealed a long-term decline in tree populations, with a threshold at the time of Heinrich Stadial (HS) 2 (24 thousand years ago [ka]), when tree populations decreased dramatically and did not recover until the end of HS1 (15ka). At this point, there were no refugia for temperate trees north of 45°N in Europe, which means that present day populations of temperate trees in northern Europe are essentially young, in contrast to conifers and other boreal tree species [6]. This also implies that temperate trees in northern Europe derive from southern populations, where they survived in glacial refugia until temperature and moisture conditions allowed for the recolonization of higher latitudes [7]. The distribution of glacial refugia across the continent and the successive contact of isolated populations following range expansions often led to hybridization events that played a major role in the evolution of forest trees and other plants [8, 9].

European aspen, *Populus tremula,* is a dioecious temperate angiosperm tree with a very extensive range across both the European and Asian continents. Its role in many ecosystems is that of a primary colonizer for forests, growing quickly and being able to cover large areas across both different latitudes and elevations [10]. Natural variation of *P. tremula* along a north-south gradient on the Scandinavian peninsula has been widely studied. GWAS and population differentiation analyses have identified a region on chromosome 10 harboring the *PtFT*2 locus, a gene known to be involved in controlling seasonal phenology [11], as a key contributor to climate adaptation at high latitudes in Sweden. Other loci harboring genes related to senescence processes or

112 plant growth have also been proposed to have played a role in local adaptation, based
113 on differences in haplotype frequencies identified with the hapFLK statistic [12]
114 between Swedish populations originating at different latitudes [13].
115
116 Introgression has been hypothesized to be a major driver of adaptation in the genus
117 *Populus*, where the lack of reproductive barriers allows for prevalent inter-species
118 hybridizations. One such example is given by the adaptive introgression of a telomeric
119 region on chromosome 15 from *P. balsamifera* to *P. trichocarpa* that has been
120 implicated in climate adaptation in the recipient species [14]. In the case of *P. tremula*,
121 admixture in contact zones along the Scandinavian peninsula has been documented
122 and, even though different genetic lineages of *P. tremula* can be defined across Europe
123 [15], no cases of adaptive introgression have been described so far in this species.
124
125 Using whole-genome re-sequencing data from an extensive set of Eurasian samples
126 of *P. tremula* we describe different aspects of the post-glacial colonization,
127 adaptation and admixture in the species across the Eurasian continent. We dissect
128 the population structure of the species, revealing a complex and intricate genomic
129 mosaicism supporting and extending earlier results from much more limited datasets
130 [15]. Furthermore, we studied different types of selection that underlie local
131 adaptation of aspen. Interestingly, we observed a clear and unique case of adaptive
132 introgression, in which the strong selective sweep previously observed on
133 chromosome 10 that harbors a gene controlling bud set in Northern Scandinavia
134 resulted from a recent introgression event from the Russian gene pool. Overall, our
135 results highlight the importance both of standing genetic variation and of genomic
136 introgression as sources of new alleles for local adaptation.

Materials and methods

Sample collection and sequencing

The Swedish samples used in this paper belong to the SwAsp collection which has been extensively described previously [11, 16]. Leaf samples were collected from the common garden at Sävar [16] for all but two genotypes that were present only at the Ekebo common garden, for which wood tissue was sampled. Sampled tissue was stored in cool conditions until frozen at the laboratory. An additional 296 samples of *P. tremula* were obtained from across Eurasia. Trees selected from the U.K. were cloned from root cuttings and grown in one of two common gardens [17]: a Scottish national aspen collection maintained by Forest Research at Bush, Midlothian, and the Eadha Enterprises clone garden at Lochwinnoch, Renfrewshire. Leaves were sampled from 125 genotypes at Bush and 15 genotypes at Lochwinnoch, representing a range of national seed zones defined by the UK Forestry Commission. Samples were freeze-dried prior to transportation to Sweden. In Norway and in Russia, leaf material was sampled from 24 trees within a region. In Latvia, ten trees were sampled in each of five regions spanning the country. Fewer samples were obtained from Iceland due to the limited availability of natural stands of *P. tremula*. Due to the clonal growth of *P. tremula*, individual samples were collected with enough physical separation (0.3 km) between individuals to minimize the risk of sampling identical genets. Leaf samples were stored in silica gel before shipping to Sweden. Geographical coordinates were recorded for the sampled trees. The

individuals included in the study (after the quality control steps that follow) and their geographic origins are listed (Additional File2: Table S1).

Total genomic DNA was extracted from frozen leaf tissue for all individuals using the DNeasy plant mini prep kit (QIAGEN, Valencia, CA,USA). For the Icelandic samples, library preparation and sequencing were conducted by Génome Québec, Canada. Briefly,1 µg of high-quality DNA was used for paired-end libraries construction. The 12 libraries were sequenced in three lanes of Illumina HiSeqv4 with read length of 125 bp. All other samples were subjected to paired-end sequencing using libraries with an average insert size of either 350bp or 650bp and samples were sequenced by the National Genomics Infrastructure at Science for Life Laboratory, Stockholm, on an Illumina HiSeq 2000 or Illumina HiSeq X platform to a mean per-sample depth of approximately 17X.

Mapping and SNP calling

Re-sequenced accessions were mapped against the reference genome of *P. tremula* v2.0 [18], using BWA (v0.7.17; [19] – mem alignments for paired-end reads using default parameters. Post-mapping filtering removed reads with MQ<20 (using samtools v1.10; [20]. Depth and breadth of coverage were assessed in order to confirm all samples had a minimum coverage of 8X (Additional File2: Table S1). Finally, before the variant calling step, we tagged duplicate reads (using Picard MarkDuplicates v2.10.3; Picard Toolkit." 2019. Broad Institute, GitHub Repository: http://broadinstitute.github.io/picard/) and found that they did not exceed 14% of the sequencing reads in individual libraries (ranging from 3 to 13.8%).

We used GATK v3.8 to call variants [21]. First, we performed a local realignment around indels with RealignerTargetCreator and IndelRealigner (using default parameters). We called per-sample variants using HaplotypeCaller to produce gVCF files (-ERC GVCF). Given the large number of individuals, we produced intermediate gVCF files using CombineGVCFs that were finally used in the joint-call step with GenotypeGVCFs. From the resulting VCF file, we selected SNPs using SelectVariants and filtered them with VariantFiltration (QD < 2.0; FS > 60.0; MQ < 40.0; ReadPosRankSum < -8.0; SOR > 3.0; MQRankSum < -12.5). At this point, the filtered VCF was lifted over to version 2.2 of the genome of *P. tremula*, available at ftp://plantgenie.org/Data/PopGenIE/Populus_tremula/v2.2/, using picard LiftoverVcf, which allowed for better resolution at the chromosome level. Further SNP pruning with vcf/bcftools removed positions with extreme depth values (min-meanDP 10, max-meanDP 25; these thresholds correspond to the average depth ± one standard deviation), absent in more that 30% of the samples, non-biallelic or displaying an excess of heterozygosity (FDR <0.01).

Batch removal on SwAsp individuals.

Given the fact that two sequencing batches from different Illumina platforms (Illumina HiSeq 2000 and HiSeq X) were used to cover the SwAsp collection, we observed noisy signals at the population structure level, presumably as a result of the different sequencing equipment and library preparation methods. By means of principal component analyses (PCA) we identified differences in the grouping of the individuals where differences among samples along PC1 and, to a lesser extent, PC2 were explained by the sequencing platform and not the geographic origin of samples

(Additional File1: Figure S1A; PC1 separates the SwAsp collection in two sets, each corresponding to a different Illumina sequencing platform). To address these batch effects, we removed SNPs associated with both components without losing the geographic structure of the Swedish population. For this purpose, we generated a genotype file with vcftools for each chromosome. We ran independent PCA analyses on each SNP configuration (homozygous reference, heterozygous, homozygous alternative) and realized it was at the heterozygous level where the batch effect was present. Using the libraries “ggfortify”, “factoextra” and “FactoMineR”, we estimated the contribution of each variant to the components affected by the platform effect. We started by assuming a uniform contribution of each variant to each component, and removed those that would deviate from this premise in both PC1 and PC2 using the following rational:

Assuming uniformity in the contribution of the variants, we calculate the expected C value:

C=(1/nrow(SNPmatrix))*100

Threshold (T) of variants contribution both to PC1 and PC2 under uniformity:

T=((C*eig1)+(C*eig2))/(eig1+eig2)

(where eig1 and eig2 are the eigenvalues corresponding to PC1 and PC2)

For each variant, we calculate the contribution value:

var_contrib =((var_contrib[PC1]*eig1)+(var_contrib[PC2]*eig2))/(eig1+eig2)

We kept only those variants that did not contribute more than expected under the uniformity hypothesis (var_contrib <= T). While removing the batch effect using this procedure, we observed loss of more than $4e^6$ SNPs (Additional File1: Figure S1B).

In order to evaluate if we could remove the batch effect without compromising such a large number of variants, we assigned a p-value to the contribution of each SNP to the components using dimdesc (FactoMineR) at the chromosome level. We tested different cut-offs (Additional File1: Figure S2) and observed that the batch effect was removed while maintaining the geographic distribution of samples when using a p-value threshold of 0.05 for PC1 and 0.01 for PC2 (Additional File1: Figure S2B). With these thresholds, ~1.8 e^6 SNPs were filtered out to remove the batch effects but without compromising the overall population structure of the samples.

Population structure and admixture in *P. tremula*

We used plink (v1.90b4.9; [22]) to identify linked and low frequency SNPs (--indep-pairwise 100 10 0.2 –maf 0.05) that we removed with vcftools. We used the resulting set of pruned positions to compute the relatedness between samples from the Eurasian collection by calculating genome-wide estimates of identity by descent (IBD). We removed from our downstream analyses one sample from each pair of closely related individuals, using a threshold value of relatedness of 0.4.

Next, we used vcftools to output the genotype likelihood information contained in the pruned VCF file; the resulting beagle-formatted file was input into NGSAdmix to

estimate individual admixture proportions across a varying number of ancestral populations (K=3 to K=6).

Finally, we assessed individual ancestries using EIGMIX implemented in SNPRelate. This eigen analysis, developed by [23], is computationally efficient for estimating ancestral proportions by making assumptions of surrogate samples for ancestral populations. For this, we chose the Latvian, Scottish and Russian populations as proxies of the ancestral populations (referred to as Central European, Western European and Siberian, respectively) from which admixture was to be estimated, based on earlier estimates of the post-glacial colonization history of *P. tremula* across Europe [15]. We used the LD/MAF pruned SNPs and produced a GDS object for the analysis [24].

Gene flow

We combined several statistics developed to identify introgressed genomic regions. These included the classic ABBA-BABA test in its developed F_d statistic form. For these calculations we used the popgen pipeline developed by Simon Martin, available at https://github.com/simonhmartin/genomics_general. We treated each chromosome independently; the corresponding vcf file was converted to .geno format using parseVCF.py. From these files, we calculated diversity and divergence values for each subpopulation (D_{XY}, F_{ST} and π) in non-overlapping 10 Kb windows, using the script genomics.py (Additional File2: Table S2). Finally, we used the script ABBABABAwindows.py to compute the four-taxon D statistic and f estimators in non-overlapping windows of 10 Kb as well. For these calculations, we tested gene flow from the Russian subpopulation (P3; 23 individuals) into Northern Scandinavia (P2;

56 individuals: 34 from Northern Sweden and 22 from Norway), setting the Chinese (O; 15 individuals) samples as the outgroup, and the Latvian (45 individuals) or Southern Scandinavian (50 from Sweden and 23 from Norway) groups as the other potential receptors (P1). Since we were interested in estimating shared variation between Russian and Northern Scandinavian individuals, we focused on the f_{dM} introgression statistic, described by [25]. f_{dM} gives positive values when introgression occurred between P3 and P2, and negative values if it occurred between P3 and P1. In order to avoid stochastic errors that could produce meaningless values, we only considered windows with at least 100 biallelic SNPs.

We also calculated the basic distance fraction, Bd_f (PopGenome; v2.7.1; [26], which combines both f_d and distances estimations and that is less sensitive to the time of gene-flow. Just as in the previous analysis, we set a four-taxon tree hypothesis to estimate the proportion of introgression from Russia into Northern Scandinavia. We calculated Bd_f in 10 Kb windows treating each chromosome independently. We combined all 19 chromosome estimations and assigned a Z-score and p-value to each Bd_f value using genome-wide results and did an extra FDR correction. We selected those regions that had an FDR<0.05.

Finally, we ran the Efficient local ancestry inference (ELAI) method [27] to confirm the introgression event on chromosome 10. We used two upper-layer clusters and 10 lower-level clusters; 20 Expectation-Maximization steps and 100 generations of admixture between the ancestral populations, that corresponded to the Russian and Latvian aspen populations. The plotted allele dosages correspond to the average

over all the Swedish individuals from the northern population. We ran five independent EM runs.

Positive and balancing selection

We scanned the genome of *P. tremula* for signals of positive selection using a newly developed strategy called “integrated selection of alleles favoured by evolution”, iSAFE [28]. This method first calculates haplotype allele frequency scores based on the presence of derived alleles in a particular haplotype, which is then used to calculate SAFE scores. These SAFE scores are in turn calculated across a region of given size in 50% overlapping windows of 300 bp to culminate in an iSAFE signal. These statistics can be calculated for large regions of phased haplotypes, which we obtained with BEAGLE v.4.1 [29], so chromosomes were divided into 3 Mb windows for each iSAFE iteration. The iSAFE software can be set to run under a case-control mode, with the case populations being, for example, the Northern Scandinavian population, and the control population being all remaining individuals not used in the case population. We ran this screening for Northern and Southern Scandinavia, Russia, and combining all the Nordic individuals from Norway, Sweden and Russia together. As recommended by the authors, we considered iSAFE values significant when they were >0.1.

In addition, we ran betascan [30],to detect possible signals of balancing selection, by dividing the collection in geographic zones: Northern Scandinavia (Sweden and Norway), Central Europe (Latvia and Southern Sweden), Western Europe (Scotland) and Southwestern Scandinavia (Southern Norway and Iceland), and using the following parameters: -fold -m 0.1 -p 20. To avoid any bias due to the

unbalanced number of individuals in each group, we randomly chose 55 individuals from each geographic cluster. In order to avoid spurious signals of balancing selection, we binned SNPs with high beta-scores (FDR<0.01) into 50 Kb windows and searched for genes encoded within bins having at least 100 significant SNPs. Functional enrichments were obtained using topGO v.2.36.0 [31].

Population demographics

We chose StairwayPlot [32] to infer changes in effective population sizes in the past. For this, the folded site frequency spectrum (SFS) for the tested subpopulation was calculated using ANGSD v.0.920 [33] using the LD-pruned SNPs (plink --indep-pairwise 100 10 0.2). We ran independent analyses for Scandinavia, Russia, Latvia and for a set of 50 randomly selected individuals from the entire collection. We generated 100 bootstrap replicates and assumed a mutation rate (μ) of 2.5e-9 and a generation time of 15 years [34].

In addition, we aimed to confirm the directionality of the gene flow events we detected on chromosome 10 between the Russian and Scandinavian individuals. For this, we explored alternative demographic models using the diffusion approximation method of dadi [35] to analyze the site frequency spectra of our aspen subpopulations. We used two chromosomes for this purpose, chromosome 10 and chromosome 8, since there is strong evidence of synteny between both chromosomes deriving from the 40 My old genome duplication that are shared by all member of the genus *Populus* [36]. We ran analyses independently for chromosome 8 and 10 using all LD pruned, biallelic sites including all biallelic sites present in the putatively selective sweep/introgressed region on chromosome 10. We fit 19

360 demographic models (Additional File1: Figure S3) for the Russian and Scandinavian
361 subpopulations using the demographic modelling workflow (dadi_pipeline) from [37].
362 We tested different classes of models: simple models of divergence with and without
363 migration; simple models plus instantaneous size changes, ancient migration or
364 secondary contact, ancient migration plus instantaneous size change, and island
365 models of vicariance and founder events. In total, we tested 19 different demographic
366 scenarios using all polymorphic sites on chromosome 10. First, we ran the general
367 optimization routine (dadi_Run_Optimizations.py), which includes fitting the model
368 using particular settings for a given number of replicates, then using the parameters
369 from the best scoring replicate to seed a subsequent round of model fitting using
370 updated settings. We used four rounds, with 10, 20, 30 and 40 replicates. The
371 arguments controlling the steps of the optimization algorithm and the perturbation of
372 starting parameters were maxiter=[3,5,10,15] and folds=[3,2,2,1]; we defined the
373 extrapolation grid size to pts = [80,90,100] and projection sizes to proj = [15,15]. The
374 optimization routine with four rounds had an important effect in minimizing
375 differences in the likelihoods achieved at the end of the runs (Additional File1: Figure
376 S4).
377
378 In order to generate confidence intervals for the parameters estimated from the
379 demographic models with the highest likelihood, we created 100 bootstrap
380 replicates of the spectra and calculated the standard deviations of the best-fit
381 parameters using the Godambe methods [38].
382
383
384 **Results**

385

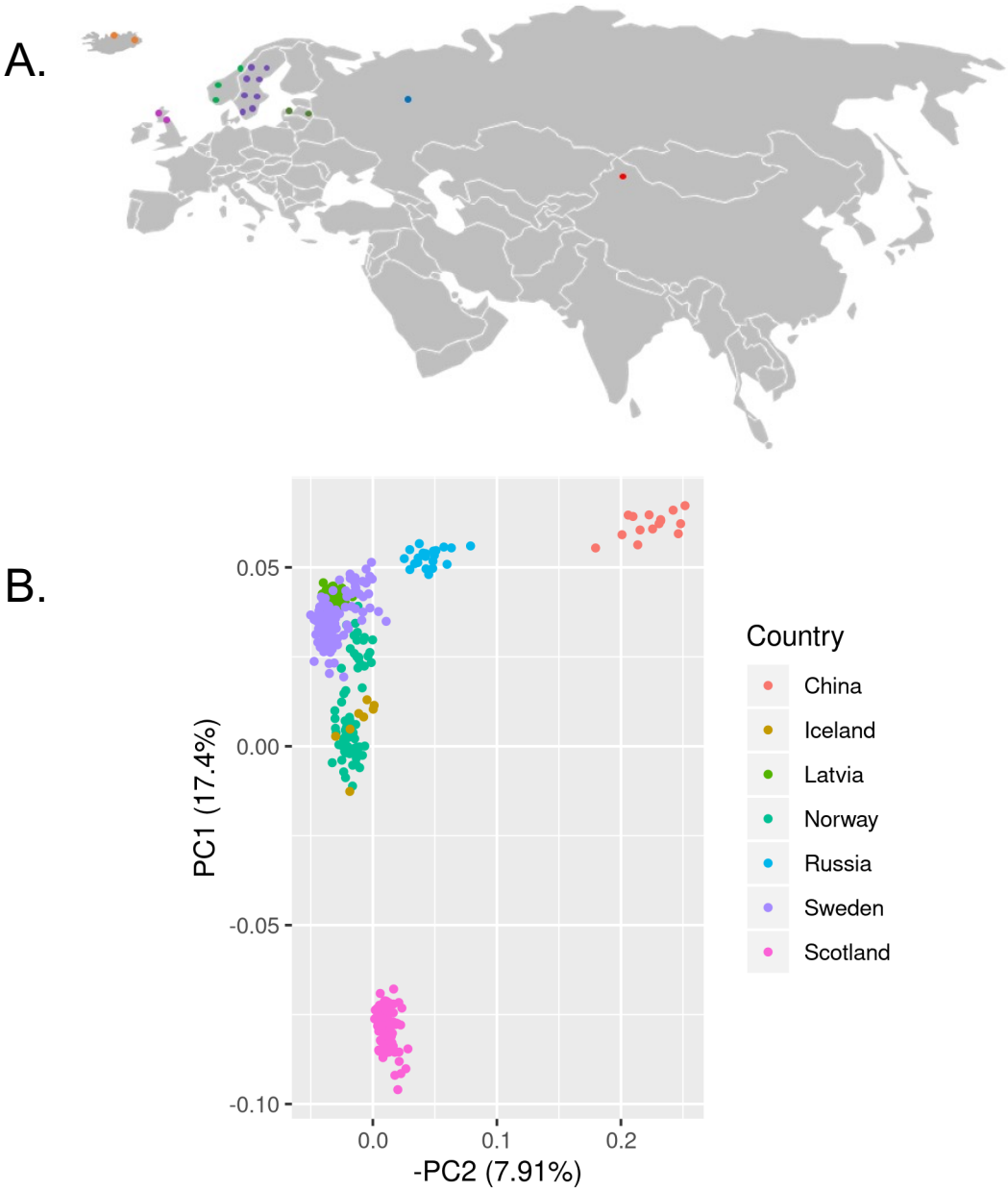
386 **Population structure and admixture**

387

388 In order to elucidate demographic and adaptive events that have accompanied the
389 post-glacial radiation of European aspen, we re-sequenced a large collection of
390 individuals, covering a wide geographic range across the Eurasian continent. Our
391 collection comprises 411 *P. tremula* individuals sampled from China [39], Russia,
392 Latvia, Norway, Sweden, Iceland and Scotland (Figure 1A; Additional File2: Table
393 S1). Sequencing reads for all individuals were mapped against the reference *P.*
394 *tremula* v2.0 genome [18] and obtained ~20.8e^6 SNPs after several filtering steps
395 (depth of coverage, missingness, level of heterozygosity and batch effect caused by
396 the incorporation of samples in the Swedish collection from two different sequencing
397 platforms; Additional File1: Figure S1,2). It should be highlighted that the removal of
398 false heterozygous positions did not alter the results of the downstream analyses
399 (Additional File1: Figure S5). Finally, we removed 48 samples that were possible
400 hybrids or that were highly related samples, yielding a total of 363 individuals for all
401 downstream analyses.

402

403 Due to the broad geographic coverage of *P. tremula*, we first focused on estimating
404 the population structure of the species. As depicted in Fig 1B, a principal component
405 analysis (PCA) using 2.8e^5 pruned, unlinked SNPs clearly separated at least three
406 independent clusters of individuals, one corresponding to samples of Chinese origin,
407 another comprising all samples collected across continental European and a third
408 comprising samples collected on the British Isles (primarily Scotland). A deeper
409 analysis of the composition of the species structure with NGSAdmix revealed five



410

411 **Figure 1.** *P. tremula* collection. A. Sampling sites across the Eurasian continent. B. PCA of

412 Principal component analysis of *P. tremula* individuals, using 283,505 pruned sites across the 19

413 chromosomes.

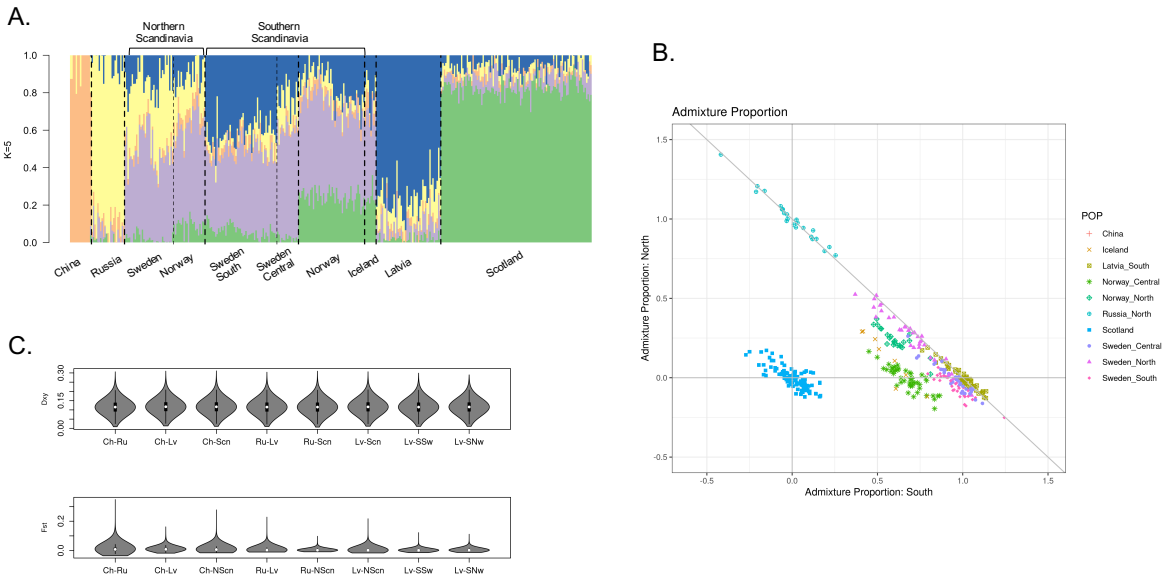
414

415 different ancestral populations and different levels of hybridization in particular

416 geographic regions. Chinese and Russian samples, represented by solid colors in

417 Figure 2A (orange and yellow, respectively), did not show strong signs of admixture

418 with European samples. However, when moving towards Western Europe, admixture
419 events occur and we can recognize three additional populations: a Scandinavian
420 (purple), a Central-European (Latvia, blue) and a Western European (Scottish,
421 green) population. Interestingly, the genomic background of the Scandinavian
422 samples (from Norway and Sweden) is a complex mosaic of up to four populations,
423 confirming earlier results based on microsatellite data [15] and providing additional
424 evidence for several waves of migration that led to the colonization of the
425 Scandinavian peninsula.
426



427

428 **Figure 2.** Population structure and admixture proportion. A. NGSadmix plot (K=5) showing the
429 hybrid genetic background of the European populations. B. Admixture plot (EIGMIX) where the
430 Russian, Scottish and Latvian populations where selected as proxies of the ancestral gene-pools
431 in Europe and thus, are located at the vertices of the plot. C. Diversity and differentiation
432 estimators between aspen subpopulations.
433

434 Given these observations, we took a closer look at the genomic mosaicism of the
435 Scandinavian population. We used EIGMIX, taking Russian, Scottish and Latvian

436 populations as proxies of the ancestral Siberian, Western and Central European
437 populations, respectively (Figure 2B, see also [15]). These populations correspond to
438 the vertices of the graph in Figure 2B, while the remaining samples were positioned
439 according to their level of admixture along the sides of the triangle. This analysis,
440 consistent with the observations derived from the NGSadmix analysis, shows that
441 individuals collected in Northern Scandinavia have on average 26% of Russian origin
442 with a maximum value of 52% in Sweden (SWASP_108 and SWASP_115) and 37%
443 in Norway (NO_MIR20). While the second highest ancestral component ($\bar{x}$=0.66) in
444 Northern Scandinavia corresponds to the Central European population, in the
445 samples from northern Norway there is up to 20% of ancestry from the Western
446 European population. The opposite trend was observed in Southern Scandinavia,
447 where the dominant ancestral component is Central European ($\bar{x}$=0.75), with less
448 than 10% of Russian admixture and up to 18% and 44% of Western European
449 ancestry in samples from Sweden and Norway, respectively. This Western European
450 genomic component (colored in green in Figure 2A), which is strongly represented
451 among the Scottish and Norwegian samples, derives from an additional ancestral
452 population, probably located between the Iberian and Italian peninsulas that we did
453 not cover in our sampling. These observations also fit the results from earlier
454 analyses based on microsatellite data [15]. Icelandic samples largely overlap with
455 Norwegian individuals, suggesting that the population in Iceland has been recently
456 introduced and has a clear Norwegian origin.
457
458 These observations were also supported by F_{ST} estimations (Figure 2C, Table 1), as
459 we obtained a small dispersion of values when calculating pairwise comparisons
460 between Russia and Northern Scandinavia (mean F_{ST}=0.003, maximum at 0.09), or

461 Latvia and Southern Scandinavia (mean F_{ST} between Latvia and Sweden of 0.002,
462 maximum at 0.07 and between Latvia and Norway of 0.004, maximum at 0.1), while
463 comparisons between populations without signs of admixture, such as China and
464 Russia, reached values as high as 0.35, which approach F_{ST} values observed
465 between closely related aspen species [39].
466
467 **Gene flow.**
468
469 We scanned the genome of *P. tremula* for potential introgression events that could
470 have occurred given the admixed background we observed in Scandinavia. We
471
472 **Table 1**. Genome-wide F_{ST} comparison.

		CH	ICE	LV	NOC	NON	NOS	RU	SWC	SWN	SWS	UK
	CH	-										
	ICE	0,014	-									
	LV	0,012	0,003	-								
	NOC	0,016	0,003	0,005	-							
	NON	0,013	0,002	0,004	0,003	-						
F_{ST}	NOS	0,016	0,003	0,004	0,003	0,003	-					
	RU	0,012	0,006	0,005	0,007	0,005	0,007	-				
	SWC	0,016	0,004	0,003	0,004	0,003	0,003	0,006	-			
	SWN	0,012	0,003	0,004	0,004	0,002	0,003	0,004	0,001	-		
	SWS	0,011	0,002	0,002	0,003	0,002	0,002	0,005	0,001	0,002	-	
	UK	0,010	0,002	0,009	0,004	0,005	0,004	0,008	0,004	0,007	0,008	-

CH:China; ICE: Iceland; LV: Latvia; NOC: Central Norway; NOCN: Northern Norway; NOS: Southern Norway; RU: Russia; SWC: Central Sweden; SWN: Northern Sweden; SWS: Southern Sweden; UK: Scotland.

followed a four-taxon approach for these scans, keeping the Chinese population as an outgroup in all our comparisons. We calculated the f_{dM} introgression statistic [25], and the basic distance fraction, Bd_f (PopGenome; v2.7.1), which combines both f_d and distance estimations and that is less sensitive to the timing of gene-flow. Both estimators give positive values when introgression occurs between P3 (donor population: Russia or Latvia) and P2 (Northern Scandinavia), and negative values if it occurred between P3 and P1 (Southern Scandinavia), on a scale from -1 to 1. When evaluating gene flow from Russia into Scandinavia, we obtained f_{dM} and Bd_f mean genomic values of -0.008 and 0.014, respectively (Figure 3A). After FDR correction at a threshold of 0.01, one region on chromosome 10, spanning from 4.5 to 4.9 Mb had a highly significant introgression signal, reaching values of f_{dM} =0.85 (FDR=$3.6e^{-3}$) and Bd_f =0.89 (FDR=$1.33e^{-50}$). As expected, there were differences in D_{XY} and F_{ST} patterns between subpopulations at this region: F_{ST} is remarkably high between Nordic subpopulations and any other geographic location, while D_{XY} decreases when northern Norwegian, Swedish and Russian subpopulations are contrasted (Additional File1: Figure S6,7). Introgression on chromosome 10 was further validated applying ELAI on each individual from the north of Sweden. This method implements a two-layer HMM (hidden Markov model) to infer local ancestry of admixed individuals without prior definition of window sizes or haplotype phasing, returning the most likely

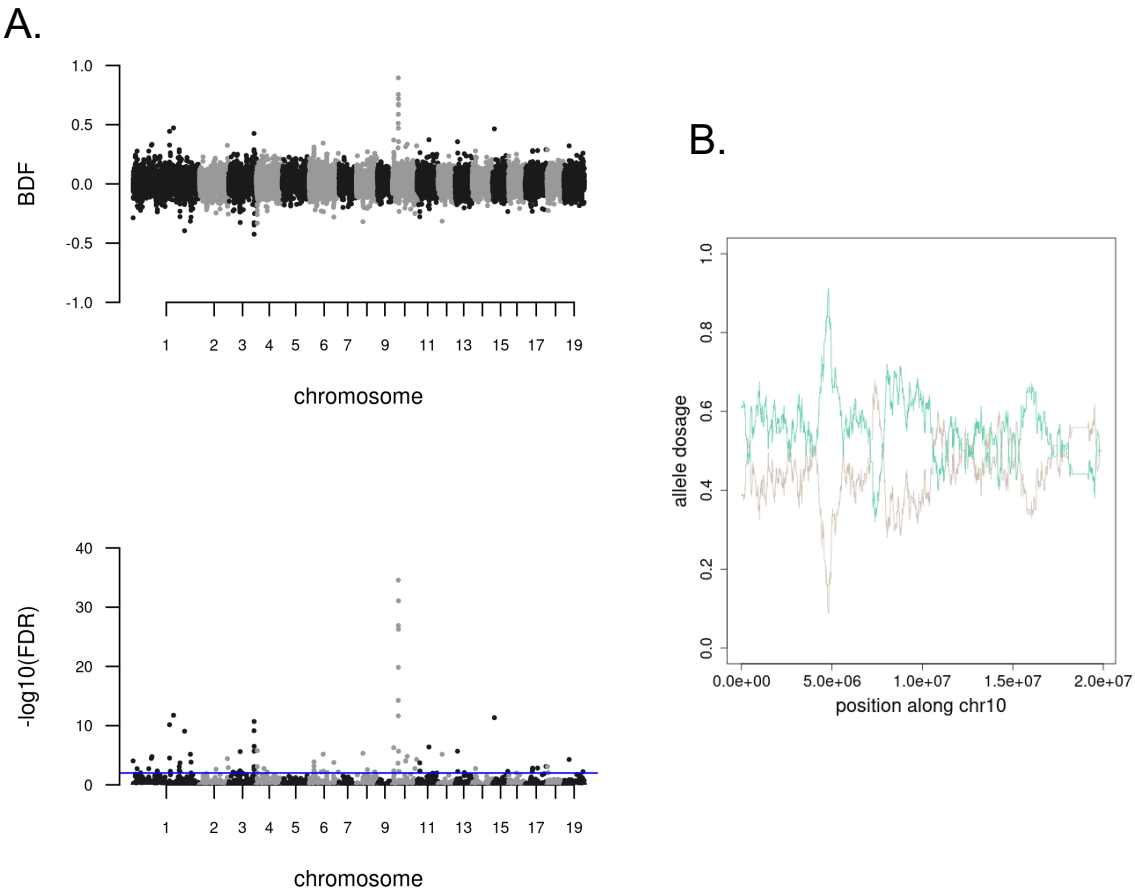


Figure 3. Introgression signals between Scandinavia and its ancestral gene pools. A. Gene flow from Russia using a four-taxon tree hypothesis. B. Allele dosage along chromosome 10 estimated with ELAI. In green, the dosage from the ancestral Russian population; in brown from the Latvian population.

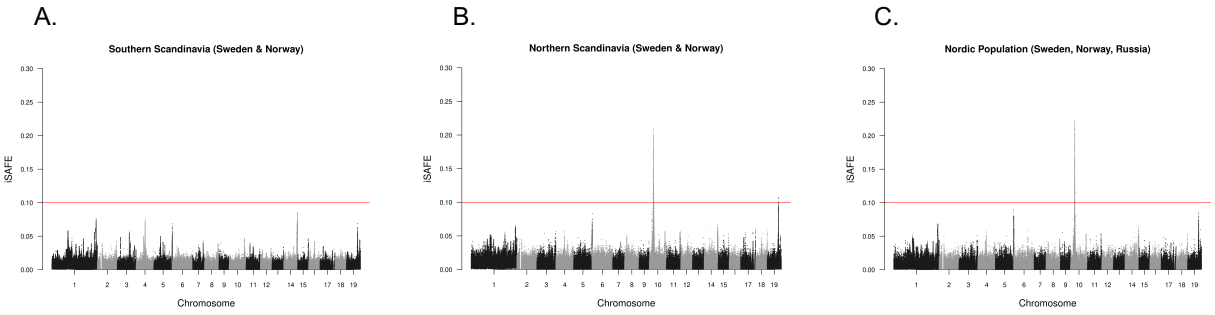
proportions of ancestry at each variable position of the chromosome. ELAI confirmed that, while chromosome 10 shares alleles from both ancestral or training populations and on average, northern individuals have 44% and 56% admixture proportions from Latvian and Russian populations respectively, the region encompassing 4.5-4.9 Mb has a predominant Russian ancestry (Figure 3B; Additional file 1: Figure S8).

In terms of gene content, two relevant loci annotated as Heading-date 3A-like, or Flowering Locus T (PtFT2) genes, are encoded in this introgressed genomic block: Potra2n10c20839 (4772117 - 4776457 bp) and Potra2n10c20842 (4789807 – 4792846 bp) and the region surrounding these two genes is known to have been subject to a selective sweep in the northern Swedish population [11].

Alleles favored by evolution or positive selection.

A good way to understand local adaptation without necessarily having phenotypic data is to look for signs of positive selection in different populations across a species range. Previous evidence of selective sweeps has shown a particularly strong signal on chromosome 10, in a region encompassing the two Flowering Locus T homologs [11]. This region is strongly implicated in the capacity of Nordic populations to adapt their seasonal phenology to the more variable day lengths and shorter growth seasons experienced at norther latitudes. We scanned the genome of *P. tremula* using a novel method to detect mutations favored by selection, iSAFE (integrated selection of alleles favored by evolution; [28]. We calculated iSAFE scores for each European population, and also contrasted Northern vs. Southern populations in Scandinavia in a case/control analysis mode. The highest iSAFE signal (iSAFE=0.22) was located on chromosome 10, in the region spanning 4.5 to 4.9 Mb where iSAFE values reached the recommended threshold of 0.1. The high iSAFE values were reached not only in Swedish individuals, but also when we combined samples originating in Northern Norway, Northern Sweden and Russia (Figure 4), implying that the sweep is shared among these populations derived from more

532 northern latitudes in the Eurasian continent. As mentioned before, this region is
533 centered around two loci annotated as Heading-date 3A-like, or *PtFT*2, in
534 agreement with previous analyses that have associated *PtFT*2 with climate
535 adaptation *in P. tremula* [11].
536



537

538 **Figure 4**. Selective sweep in chromosome 10 in Nordic populations. A. iSAFE scan of individuals
539 from the southern part of Scandinavia. B. iSAFE scan of individuals from the northern part of
540 Scandinavia. C. iSAFE scan of chromosome 10 grouping Swedish, Norwegian and Russian
541 individuals.
542
543

544 **Balancing selection and functional enrichments.**

545

546 A closer look at the iSAFE patterns in the different subpopulations revealed several
547 strong signals that a) did not reach the accepted significance threshold to be
548 confidently classified as regions targeted by positive selection or b) were not
549 geographically limited, i.e., they are present in at least two distant subpopulations
550 (Additional File1: Figure S9). This suggests that another selective force may be
551 acting to maintain alleles at high frequencies, for instance balancing selection (BS).
552 In order to evaluate this possibility, we ran betascan [30] on all 19 chromosomes,

553 dividing the collection in geographic zones: Northern Scandinavia (Northern Norway
554 and Sweden), Central Europe (Latvia and Southern Sweden), Western Europe
555 (Scotland) and Southwestern Scandinavia (Southern Norway and Iceland). We
556 hypothesized that some genomic tracts with a high beta-score would match regions
557 that we previously classified as iSAFE candidates for positive selection across
558 multiple populations. This pattern was evident on chromosomes 1 (23.4-23.5 and
559 38.2-38.35 Mb), 4 (9.9-10Mb), 5 (8.35-8.4 Mb), 7 (9.1-9.15 Mb),14 (14,7-14.8 Mb),17
560 (8.85-8.9 Mb), 18 (9.4-9.45 Mb) and 19 (5.6-5.65 Mb) (Figure 5). A further validation
561 of the effect of BS on these regions comes from the comparison of π and D_{XY}
562 patterns. While the average of genome wide, pairwise D_{XY} was around 0.12, the
563 average value obtained when considering only the chromosome tracts affected by
564 BS was close to 0.14 (t-test, p<0.01; Additional File2: Table S3), while π values also
565 increased in BS targeted regions compared to genome-wide average (Additional
566 File2: Table S3). Average F_{ST} values ded not differ between BS affected regions and
567 genome wide estimations, however.
568
569 Furthermore, we also observed significant bins that were unique to a specific
570 population and we therefore ran independent functional enrichment tests for
571 biological processes affected by balancing selection for each of these regions.
572 Interestingly, we found a few common terms with significant p-values (Fisher test,
573 p<0.05), such as ethylene metabolic and biosynthetic processes (p_{Nordic}=1,4e^{-3};
574 $p_{CentralEurope}$=9.7e^{-4}), jasmonic acid mediated signaling pathways
575 ($p_{SouthWestScandinavia}$=1.8e^{-2}; $p_{CentralEurope}$ =4.7e^{-3}; $p_{Scotland}$=1.5e^{-2}), and intramembrane
576

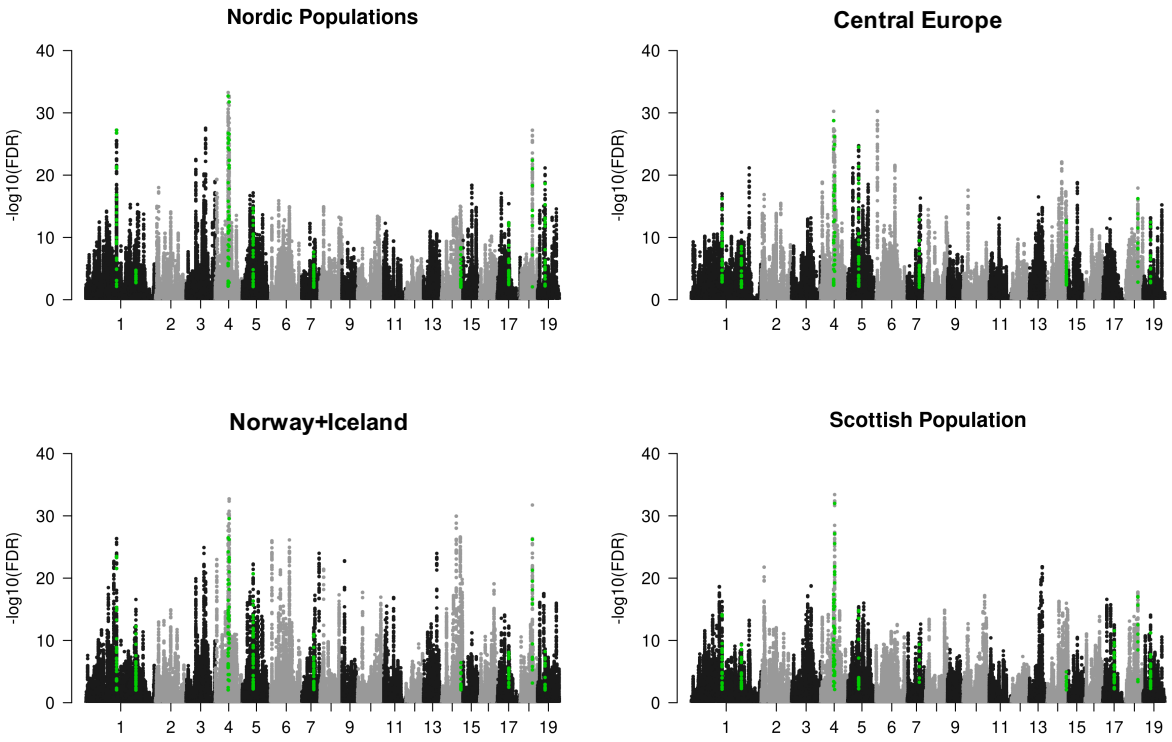


Figure 5. Balancing selection. Green dots highlight shared signals between subpopulations.

Golgi-vacuolar transport (p_{Nordic}=2e^{-3}; $p_{SouthWestScandinavia}$=4.3e^{-3}; $p_{Scotland}$=1.8e^{-2}). Other relevant processes particular to a specific subpopulation include alkene biosynthesis (p_{Nordic} =1,4e^{-3}; $p_{CentralEurope}$=9.7e^{-4}), water transport ($p_{SouthWestScandinavia}$ =6.8e^{-3}) or secondary shoot formation ($p_{CentralEurope}$=3e^{-3}).

Demography and secondary contacts.

The expansion of *P. tremula* across the Eurasian continent provides a good scenario to test for changes in population sizes across time. We evaluated population size changes through time using the stairway plot method, which is based on the composite likelihood of the SFS of the species and/or selected subpopulations. We used 50 random samples from our collection to generate the overall behavior of the

592 species, as well as for specific collection sites (Russia and Northern Scandinavia).
593 We observed the strongest reduction in N_e around ~700-800 ka this sharp decrease
594 in N_e was observed across all subpopulations, which means it was a bottleneck that
595 affected the entire species and that predated its dispersal in the Eurasian continent.
596 A second bottleneck occurred approximately ~150-170 ka in the populations from
597 Central Europe and Scandinavia, and a third, mild bottleneck spanned from ~10-35
598 ka, a period representing the last glacial era in the Northern hemisphere. These two
599 bottlenecks were not present in the Russian subpopulation and instead, we only
600 observed a second decrease of N_e around 60-80 ka in this population (Figure 6).
601
602 Given the admixed nature of the Scandinavian individuals, we also evaluated 19
603 different demographic models including simple models of divergence with and
604 without migration, models with instantaneous size changes, ancient migration or
605 secondary contact, ancient migration plus instantaneous size change, and island
606 models of vicariance and founder events. We used two levels of resolution for this
607 analysis: we ran the 19 models at the chromosome level, using unlinked sites from
608 chromosomes 8 and 10, and also, at a region-targeted level, using the SNPs derived
609 from the segment of chromosome 10 that we hypothesize is the result of adaptive
610 introgression in the Northern Scandinavian population. For these analyses, we
611 considered two populations from our collection, Scandinavia and Russia, in order to
612 identify the most probable site of origin of the selective sweep on chromosome 10.
613 While using all unlinked polymorphic sites on chromosomes 8 and 10 (Figure 7;

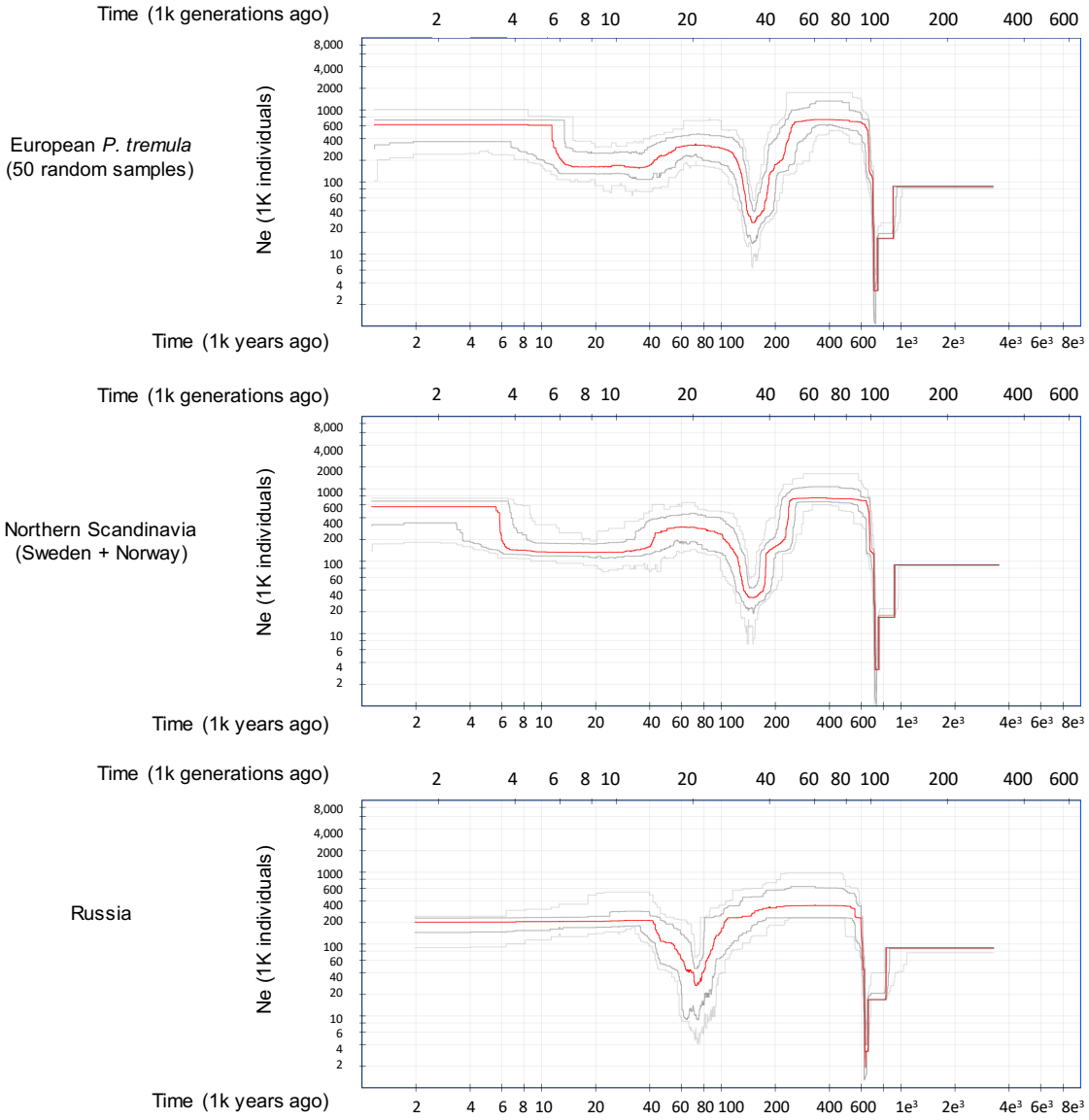
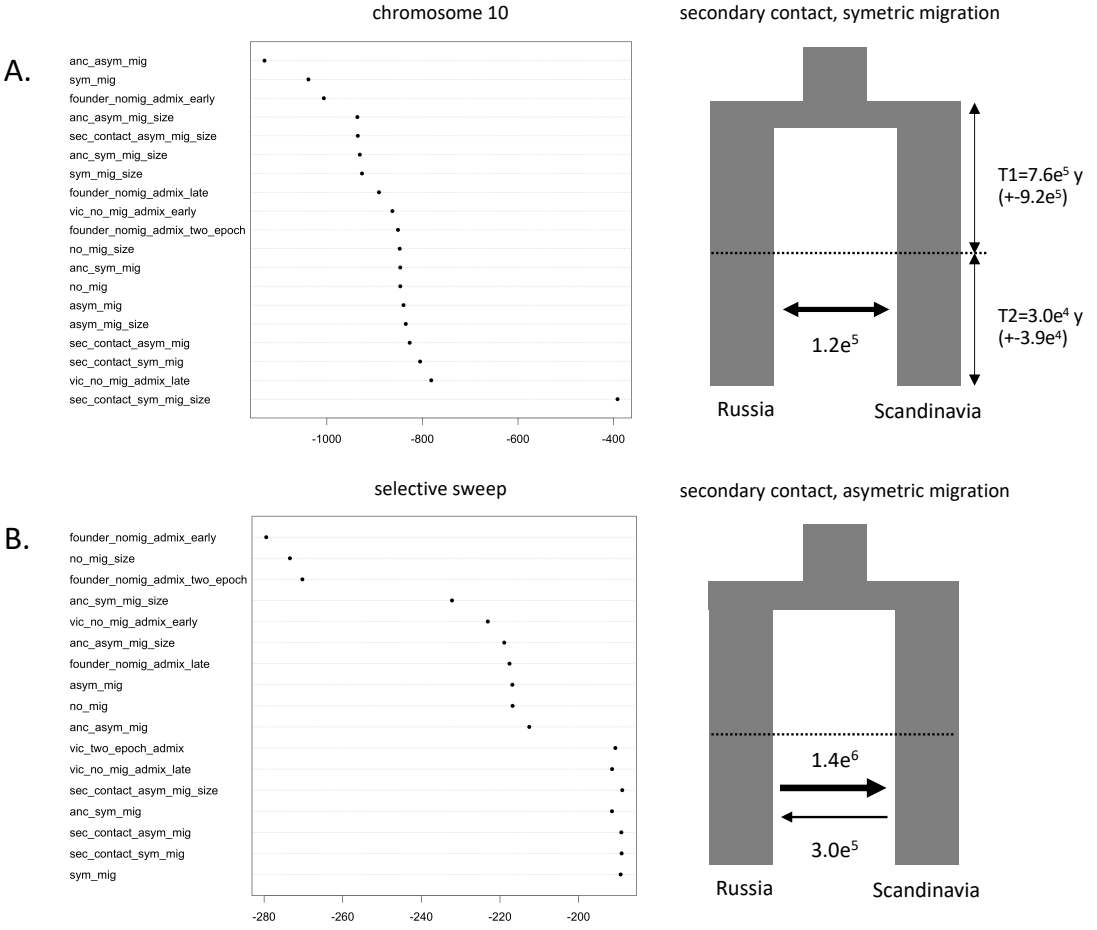


Figure 6. Demographic changes in P*. tremula*.

Additional File1: Figure S10), we consistently observed that the most strongly supported model included divergence in isolation with continuous symmetrical secondary contact and an instantaneous size change (sec_contact_sum_mig_size). Interestingly, at the sweep level, several models were equally supported, with divergence in isolation with continuous asymmetrical secondary contact (with or without instantaneous size change) being the most significant models. Not only was

623 asymmetrical migration detected, but it was stronger from Russia into Scandinavia
624 than in the other direction (Additional File1: Figure S11). After extrapolating the
625 results of these simulations to time units, the secondary contact between Russian
626 and Scandinavian populations was dated to around 30 ka (±3.9e^4), while the
627 divergence of both lineages occurred 760 ka, consistent with the first bottleneck seen
628 at the species level in the stairway plot analyses.
629



630

631 **Figure 7.** Likelihood of demographic models fitting the SNP data. A. SNPs observed in
632 chromosome 10 are evaluated. B. SNPs comprised within the selective sweep in chromosome 10
633 are tested.
634
635

Discussion

Aspen populations: a mosaic of ancestral gene pools

The large number of aspen individuals included in our collection allowed us to dissect the genetic mosaicism of the species across a large geographic scale. Intra-species admixture in *Populus tremula* is not a novel idea, in fact, it has previously been documented between divergent lineages in Europe using data from microsatellite loci [15]. These results, first described in 2010, are, to a certain extent, corroborated by our current analyses but we have also been able to study the hybridization and introgression patterns at much finer scales across the genome of *P. tremula*. The NGSAdmix analysis identified five genetic components that explain the population structure along the East-West geographic axis in Eurasia, corresponding to Eastern Asia (China), Russia, Scandinavia (Sweden and Norway), Eastern European (Latvia) and what we suggest may correspond to Central and Western Europe (Scotland and Southern Norway + Iceland). While Chinese and Russian individuals appear to be partly isolated from Western populations (genome wide F_{ST} values between China and Western Europe ~0.01, an order of magnitude higher than any other intra-species pairwise comparison), the genomic mosaicism is evident particularly in Scandinavia, where the population can be divided into at least two different groups that follow the latitudinal gradient along the peninsula. Individuals from Northern Scandinavia have a strong Russian ancestral component, whereas those from Southern Scandinavia have strong influence from both Eastern and Central/Western European population. The observations of two independent routes of colonization of Scandinavia, one from the south and another one from the northeast along the ice-

661 free Norwegian Atlantic coast, agree with findings in other organisms, such as brown
662 bears (reviewed by [40]) and humans [41].
663
664 de Carvalho and collaborators [15] considered Scottish and Russian subpopulations
665 as proxies for the ancestral populations contributing to the admixed Swedish
666 population. In our analyses we extended this idea by using the Russian, Latvian and
667 Scottish populations in a triple model of hybridization that gave us a higher resolution
668 for identifying admixture routes of the Scandinavian population. In the south of
669 Scandinavia, we could differentiate two groups of individuals: one in Sweden with a
670 clear Latvian component, and one in Norway, with Latvian but also with a Scottish
671 component. This could likely be explained by independent waves of colonization from
672 the south of the peninsula, where Swedish and Norwegian populations were kept
673 partially isolated by the Scandinavian Mountains, or Scandes, that run along the
674 entire peninsula. In line with previous observations, the individuals from the Scottish
675 populations are largely independent from individuals derived from Central Europe.
676 Unfortunately, the lack of samples from Central Europe and the Iberian Peninsula in
677 our collection make it difficult to identify the origin of this population, although de
678 Carvalho et al (2010) identified that trees from Spain constitute a different lineage
679 themselves, possibly due to human-mediated disturbance. Finally, it is clear that the
680 relatively few aspen clones that occur in Iceland are all recent introductions to the
681 island of individuals with a Norwegian origin.

682 Many very interesting discussions have been carried out in recent years regarding
683 paleographic migration trajectories of trees in Europe [42, 43]. By tracking plant
684 macrofossil and pollen records, it has been suggested that a long-term decline in tree
685 populations occurred around 24 ka followed by a recovery in population sizes

686 approximately 15 ka [7]. The demographic changes that have affected *P. tremula*
687 across the Eurasian continent vary along the East-West axis. The populations from
688 Western Europe have undergone two episodes of bottleneck and subsequent
689 recovery, one around 150 ka and another one during the Middle Pleniglacial period,
690 from which it recovered to its current effective population size between 12-15 ka
691 (Figure 6). On the other hand, the Russian population experienced a single severe
692 bottleneck around 80 ka, from which it recovered before the LGM (Figure 6). The
693 asymmetry in the distributions of tree species along the West to East axis in Europe
694 has been also highlighted by paleobotanical evidence and climate simulations.
695 Western Europe generally lacked tree species north of 46N while higher growing-
696 season warmth in Eastern Europe resulted in increased permafrost thaw and higher
697 water availability so that small populations of boreal trees were able to survive up to
698 approximately 49.8N. The lack of a second bottleneck during the LGM in Russia
699 suggests that the split of the European lineages of aspen predated the glacial
700 maximum and suggests different colonization routes of Russia from southern aspen
701 populations located in far eastern Europe and central Asia.

702

703 **Gene flow from the East facilitated adaptation to extreme latitudes.**

704

705 Evidence of inter-species hybridization in poplars has been extensively documented
706 in recent years [14, 44] as well as intra-species gene flow in *P. tremula* [45]. Under
707 these circumstances, we can think of two opposite outcomes of gene flow: it can
708 either homogenize populations, and thus interfere with processes of local adaptation,
709 or it can introduce genetic variation at higher rates than mutation would in the same
710 time frame, providing a source of novel alleles. If any of those foreign alleles are

711 permanently incorporated in a population as a result of gene flow and successive
712 backcrossing, a process defined as introgression, this can increase the fitness of the
713 recipient population and thus, can be referred to as adaptive introgression.
714 Compared to neutral introgression, where alleles can be lost by drift, adaptively
715 introgressed alleles are maintained by selection that can lead to fixation in the
716 recipient population [46]. The role of introgression for species adaptation and
717 evolution has been recognized in animals (reviewed by [47], humans [48], and plants
718 [49, 50]. A recent clear example of adaptive introgression was reported between two
719 species of cypress in the mountainous region of the eastern Qinghai-Tibet Plateau,
720 where loci from *Cupressus gigantea* introgressed into *C. duclouxiana* and thereby
721 facilitated adaptation in *C. duclouxiana* to cooler and drier conditions at higher
722 latitudes and elevations [51].
723
724 As several of the populations in our aspen collection are hybrids, we evaluated the
725 occurrence of adaptive introgression events between lineages. One noteworthy
726 genomic region was identified as it displayed clear patterns of introgression between
727 Northern Scandinavia and Russia that were detected using different estimators of
728 introgression (Bd_f, f_{dM}, and local ancestry; Figure 3) and that we could validate using
729 patterns of genetic diversity and differentiation (Additional File1: Figure S6,7). This
730 genomic tract is located on chromosome 10 and comprises ~500 Kb surrounding two
731 FT homologs that have already been implicated as a key component for seasonal
732 adaptation of phenology in aspens to high latitudes in Sweden [11]. Not only did the
733 genomic tract display a clear signal of positive selection in Sweden, corroborating
734 earlier studies [11], it also showed strong signs of selection in other high-latitude
735 populations, including northern Norway and Russia (iSAFE>0.2; Figure 4).

In order to determine the population of origin of the adaptive allele in this region on chromosome 10, we used demographic modeling to test for different patterns of contact between ancestral populations and directionality of the hypothesized gene flow. First, using a set of unlinked SNPs derived from two independent aspen chromosomes (chr10 and chr8), we observed that Russian and Northern Scandinavian populations have experienced symmetric migration following a secondary contact, making it possible for backcrossing to occur and thereby promoting introgression between the two populations. Second, using information only from the adaptively introgressed region on chromosome 10, we confirmed that, indeed, the secondary contact models are the most strongly supported (together with vicariance and late admixture) and that the migration in this region was stronger from Russia into Scandinavia than in the opposite direction (Additional File1: Figure S10,11). These analyses strongly suggest that regions surrounding the two FT homologs have been adaptively introgressed from the Russian population into Northern Scandinavia in the recent past, thereby facilitating adaptation in *P. tremula* to grow under high latitude conditions.

At the phenotypic level, there is convincing evidence that the *FT*2 allele present in northern Scandinavia displays partial dominance [11]. Trees in the northern part of Scandinavia carrying the northern *FT*2 allele initiate growth cessation about 30 days earlier than trees carrying the southern allele. Trees carrying the northern allele cannot grow in Southern Scandinavia as their critical photoperiod for growth is never reached [52]. This implies that the Nordic allele could have emerged and been kept in the Russian population in a heterozygous form during the last glaciation. The degree of dominance influences the evolutionary dynamics of alleles in diploid populations as the fixation probability of a newly arise (partially) dominant beneficial

761 allele is higher than for a recessive allele in static or slowly changing environments, a
762 process known as Haldane's sieve [53]. Furthermore, recent modelling suggests that
763 the relationship between dominance and selection coefficients arose as a natural
764 outcome of the functional importance of genes (i.e., degree of connectivity in protein
765 networks) and their optimal expression levels [54]. Given the central role of *FT*2 in
766 regulating the plant phenology signaling pathways and the fine tuning of its
767 expression along plant development, we can speculate that the effect on fitness of
768 the Nordic allele allowed a rapid increase in frequency after the colonization of the
769 Scandinavian peninsula, where it was strongly selected for, allowing a rapid local
770 adaptation to high latitudes. It is worth noting that despite the high recombination rate
771 in *P. tremula* [55], the introgression signal has not been broken and is detectable in a
772 long tract of ~500 Kb, reinforcing the idea of recent contacts and hybridizations
773 between Nordic populations.

774 **Other paths to local adaptation**.

775

776 A major genetic process contributing to local adaptation in many populations is that of
777 selective sweeps. Selective sweeps result in the loss of genetic variation in the
778 neighboring regions of a newly adaptive allele or mutation as selection drives this allele
779 to fixation. Selective sweeps generally fall into two categories, hard sweeps and soft
780 sweeps, depending on the origin of the adaptive allele. Hard sweeps are the result of
781 positive selection acting upon a newly arisen beneficial mutation before divergence of
782 lineages, thus the same variation in neighboring haplotypes is driven to high
783 frequencies alongside the adaptive allele. Conversely, a soft sweep occurs from
784 standing variation so that haplotypes surrounding a future beneficial allele have had
785 time to diverge before the onset of positive selection, meaning that the haplotypes that

786 hitchhike together with the beneficial allele may differ between lineages or populations
787 [56, 57].

788 Our scans for positive selection using the iSAFE algorithm revealed a very strong
789 signal of a selective sweep located on chromosome 10. While we also observed
790 other regions across the aspen genome that seemly have experienced selective
791 sweeps, the calculated iSAFE values in these regions were generally too weak to
792 pass the significance threshold and did not display a clear geographic pattern, which
793 precluded them from being interpreted as selective sweeps involved in conferring
794 local adaptation. A natural conclusion from the lack of non-neutral outliers in our
795 analyses is that local adaptation in aspens is primarily polygenic and/or driven by
796 natural selection acting on standing variation rather than from new mutations that
797 would induce hard sweeps, as is the case for the hard sweep identified on
798 chromosome 10.

799 The role of standing variation in adaptive evolution has been widely debated. As
800 highlighted by [15], beneficial alleles that are present as standing variation are
801 generally older than new mutations, implying they could have undergone a selective
802 filter during past environmental conditions or in different parts of a species range.
803 These polymorphisms can be maintained for a long time due to balancing selection
804 that persists for many generations and minimizes the effect of drift. When we think of
805 temperate tree species, such as aspens, that have experienced several population
806 contraction and expansion cycles throughout the last millennia, such “pre-selected”
807 standing variation could represent a very useful pool for adaptive alleles that can
808 rapidly be brought together to mediate local adaptation to novel environmental
809 conditions encountered during range expansion. For instance, one climatic

consequence for plants during the LGM was water stress due to low CO_2 concentrations and the presence of permafrost. Even if boreal trees can grow on continuous permafrost, other factors such as soil texture or timing of the spring-summer thaw determine the amount of water available for growth, thus influencing the species distribution [7].

We evaluated the importance of standing variation for local adaptation in aspen by scanning the genome for signals of balancing selection (BS). The combination of significant beta-scores and the increased genetic diversity (π) and pairwise absolute divergence (D_{XY}) in several regions across multiple chromosomes, shared by at least two aspen lineages, indicated that BS has maintained ancestral polymorphisms in the species. Indeed, excess diversity around selected loci may be due to the retention of ancestral polymorphisms or due to the accumulation of derived polymorphisms, which happens faster than expected under neutrality. The former is expected if the same allelic lineages have been evolving under BS for a long time, with drift and mutation determining polymorphism within each lineage. The latter would occur if new, more recent instances of BS appear, facilitating the establishment of new neutral polymorphisms [2]; given that our sampling belongs to one single species across the Eurasian continent, we can speculate it is the second scenario which has produced the excess of diversity. Even though gene flow is an important process that could mimic trans-lineage polymorphisms and shared haplotypes between subpopulations, our scans of genomic introgression showed no evidence of such pervasive events, except for the Scandinavian and Siberian lineages.

As reported in other species, the BS regions identified in aspens harbor potentially interesting genes for mediating local adaptation. Of particular relevance was the observation of an enrichment of genes in these genomic regions related to ethylene metabolic and biosynthetic processes. This is of note given that this phytohormone plays a pivotal role in defense responses, plant growth and senescence [58] and, most importantly, in xylem development and the formation of tension wood in aspens [59, 60]. Other functional categories such as jasmonic acid mediated signaling pathways and endo-membrane trafficking suggest differential responses to abiotic stresses that have been favored by the maintenance of advantageous alleles across the species range. In addition, the observation of an enrichment of alkene biosynthesis related genes is intriguing, as a significant association between the presence of alkenes in cuticular waxes and tree growth and resistance to leaf spot have been reported in *P. trichocarpa* [61]. Thus, we suggest that standing variation and balancing selection have shaped the capacity of aspen populations to adapt to a wide range of environmental conditions during the post-glacial colonization of Europe.

Conclusion

Our large panel of re-sequenced aspen individuals allowed us to unravel key genomic aspects behind local adaptation across the Eurasian continent. It is clear from our results that intra-species hybridization has played a major role homogenizing the genomic background of the species and has promoted the movement of adaptive alleles between populations. Of major relevance is the observation of a recent adaptive introgression event between Nordic populations

around the *Flowering locus T*, that has facilitated the survival of aspens in high latitudes. This event is however the only such event that can be detected in the species, showing that the emergence of advantageous alleles and their propagation is rather rare. Standing variation and its conservation through balancing selection across lineages seems to be a more efficient way of keeping advantageous alleles and maintaining high levels of diversity. This combination of evolutionary scenarios suggests that aspens may have the capacity to adapt rapidly to new challenging environments, and this augurs well for the survival of the species under a range of possible future climate conditions.

Acknowledgments

The SwAsp common garden was maintained by Skogforsk. The Scottish clone collection was commissioned by Forest Research and NatureScot, and planted and maintained by Forest Research Technical Services Unit, with funding from the Forestry Commission. We thank Peter Livingston and Eadha Enterprises for the samples from the EADHA clone collection. We are grateful to the Bioinformatics facility at Umeå Plant Science Centre for curation of sequencing data. SJ, KMR and NRS are supported by the Trees for the Future (T4F) project. PKI, SJ and KMR are supported by the Knut and Alince Wallenberg Foundation (KAW). The research was also supported by grants from the Swedish Research Council (VR) to PKI and to SJ, and by Formas to SJ. Part of the data generated for the study was supported by Science for Life Laboratory and the National Genomics Infrastructure (NGI, as a SciLifeLab National project) which provided access to massive parallel sequencing. All analyses were performed on resources provided by the Swedish National

Infrastructure for Computing (SNIC) at Uppsala Multidisciplinary Centre for Advanced Computational Science (UPPMAX) under compute projects 2017/7-219, 2018/3-552, 2019/3-597, 2020/5-621 and storage projects sllstore2017050, sllstore2017059.

Data accessibility

All raw read data have been uploaded to NCBI/ENA and can be found under Bioproject IDs PRJNA297202 and PRJEB42846.

Author contributions

PKI, NRS, KR, SJ and MR-A planned and designed the research. MR-A, JW and PKI performed experiments and analyzed data. SS, AF, JC, MESB, CL, DR, KR and NRS provided biological material and performed field work. MR-A and PKI wrote the manuscript with input from NRS and KR. All authors read and approved of the final version of the manuscript.

References

1. VanWallendael A, Soltani A, Emery NC, Peixoto MM, Olsen J, Lowry DB: **A Molecular View of Plant Local Adaptation: Incorporating Stress-Response Networks.** *Annu Rev Plant Biol* 2019, **70:**559-583.
2. Fijarczyk A, Babik W: **Detecting balancing selection in genomes: limits and prospects.** *Mol Ecol* 2015, **24:**3529-3545.
3. Guerrero RF, Hahn MW: **Speciation as a sieve for ancestral polymorphism.** *Mol Ecol* 2017, **26:**5362-5368.
4. Wang B, Mojica JP, Perera N, Lee CR, Lovell JT, Sharma A, Adam C, Lipzen A, Barry K, Rokhsar DS, et al: **Ancient polymorphisms contribute to genome-wide variation by long-term balancing selection and divergent sorting in Boechera stricta.** *Genome Biol* 2019, **20:**126.

5. Widmer A, Lexer C: **Glacial refugia: sanctuaries for allelic richness, but not for gene diversity.** *Trends in Ecology & Evolution* 2001, **16:**267-269.
6. Svenning J-C, Normand S, Kageyama M: **Glacial refugia of temperate trees in Europe: insights from species distribution modelling.** *Journal of Ecology* 2008, **96:**1117-1127.
7. Tzedakis PC, Emerson BC, Hewitt GM: **Cryptic or mystic? Glacial tree refugia in northern Europe.** *Trends Ecol Evol* 2013, **28:**696-704.
8. Canestrelli D, Porretta D, Lowe WH, Bisconti R, Carere C, Nascetti G: **The Tangled Evolutionary Legacies of Range Expansion and Hybridization.** *Trends Ecol Evol* 2016, **31:**677-688.
9. Hewitt GM: **Genetic consequences of climatic oscillations in the Quaternary.** *Philos Trans R Soc Lond B Biol Sci* 2004, **359:**183-195; discussion 195.
10. WORRELL R: **European aspen (Populus tremula L.): a review with particular reference to Scotland II. Values, silviculture and utilization.** *Forestry: An International Journal of Forest Research* 1995, **68:**231-244.
11. Wang J, Ding J, Tan B, Robinson KM, Michelson IH, Johansson A, Nystedt B, Scofield DG, Nilsson O, Jansson S, et al: **A major locus controls local adaptation and adaptive life history variation in a perennial plant.** *Genome Biol* 2018, **19:**72.
12. Fariello MI, Boitard S, Naya H, SanCristobal M, Servin B: **Detecting signatures of selection through haplotype differentiation among hierarchically structured populations.** *Genetics* 2013, **193:**929-941.
13. Ingvarsson PK, Bernhardsson C: **Genome-wide signatures of environmental adaptation in European aspen (Populus tremula) under current and future climate conditions.** *Evol Appl* 2020, **13:**132-142.
14. Suarez-Gonzalez A, Hefer CA, Christe C, Corea O, Lexer C, Cronk QCB, Douglas CJ: **Genomic and functional approaches reveal a case of adaptive introgression from Populus balsamifera (balsam poplar) in P. trichocarpa (black cottonwood).** *Molecular Ecology* 2016, **25:**2427-2442.
15. de Carvalho D, Ingvarsson PK, Joseph J, Suter L, Sedivy C, Macaya-Sanz D, Cottrell J, Heinze B, Schanzer I, Lexer C: **Admixture facilitates adaptation from standing variation in the European aspen (Populus tremula L.), a widespread forest tree.** *Mol Ecol* 2010, **19:**1638-1650.
16. Luquez V, Hall D, Albrectsen BR, Karlsson J, Ingvarsson P, Jansson S: **Natural phenological variation in aspen (Populus tremula): the SwAsp collection.** *Tree Genetics & Genomes* 2008, **4:**279-292.
17. Harrison A: **Aspen growth trials: showing the species' potential in Scotland. .** In *The biodiversity and management of aspen woodlands.* Edited by Cosgrove PaA, Andy. UK: The Cairngorms Local Biodiversity Action Plan, Grantown-on-Spey; 2009: 49-51
18. Schiffthaler B, Delhomme N, Bernhardsson C, Jenkins J, Jansson S, Ingvarsson P, Schmutz J, Street N: **An Improved Genome Assembly of the European Aspen Populus tremula.** *bioRxiv* 2019:805614.
19. Li H, Durbin R: **Fast and accurate short read alignment with Burrows-Wheeler transform.** *Bioinformatics* 2009, **25:**1754-1760.
20. Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, Marth G, Abecasis G, Durbin R, Genome Project Data Processing S: **The Sequence Alignment/Map format and SAMtools.** *Bioinformatics* 2009, **25:**2078-2079.
21. DePristo MA, Banks E, Poplin R, Garimella KV, Maguire JR, Hartl C, Philippakis AA, del Angel G, Rivas MA, Hanna M, et al: **A framework for variation discovery and**

genotyping using next-generation DNA sequencing data. *Nature Genetics* 2011, **43:**491-498.

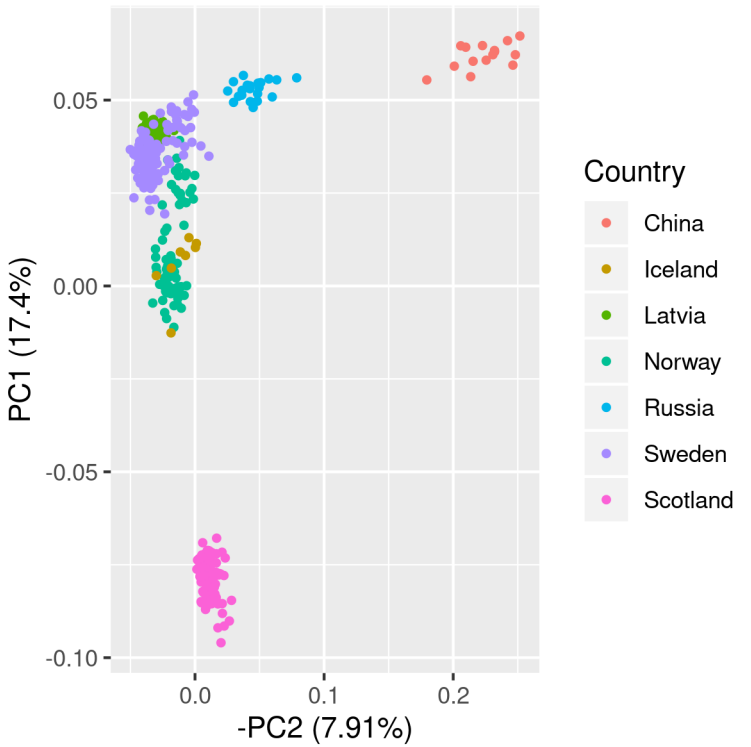
22. Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira MA, Bender D, Maller J, Sklar P, de Bakker PI, Daly MJ, Sham PC: **PLINK: a tool set for whole-genome association and population-based linkage analyses.** *Am J Hum Genet* 2007, **81:**559-575.
23. Zheng X, Weir BS: **Eigenanalysis of SNP data with an identity by descent interpretation.** *Theor Popul Biol* 2016, **107:**65-76.
24. Zheng X, Levine D, Shen J, Gogarten SM, Laurie C, Weir BS: **A high-performance computing toolset for relatedness and principal component analysis of SNP data.** *Bioinformatics* 2012, **28:**3326-3328.
25. Malinsky M, Challis RJ, Tyers AM, Schiffels S, Terai Y, Ngatunga BP, Miska EA, Durbin R, Genner MJ, Turner GF: **Genomic islands of speciation separate cichlid ecomorphs in an East African crater lake.** *Science* 2015, **350:**1493-1498.
26. Pfeifer B, Wittelsburger U, Ramos-Onsins SE, Lercher MJ: **PopGenome: an efficient Swiss army knife for population genomic analyses in R.** *Mol Biol Evol* 2014, **31:**1929-1936.
27. Guan Y: **Detecting structure of haplotypes and local ancestry.** *Genetics* 2014, **196:**625-642.
28. Akbari A, Vitti JJ, Iranmehr A, Bakhtiari M, Sabeti PC, Mirarab S, Bafna V: **Identifying the favored mutation in a positive selective sweep.** *Nat Methods* 2018, **15:**279-282.
29. Browning SR, Browning BL: **Rapid and accurate haplotype phasing and missing-data inference for whole-genome association studies by use of localized haplotype clustering.** *Am J Hum Genet* 2007, **81:**1084-1097.
30. Siewert KM, Voight BF: **Detecting Long-Term Balancing Selection Using Allele Frequency Correlation.** *Mol Biol Evol* 2017, **34:**2996-3005.
31. Alexa AR, Jorg: **topGO: Enrichment Analysis for Gene Ontology.** R package version 2.36.0. edition; 2019.
32. Liu X, Fu YX: **Exploring population size changes using SNP frequency spectra.** *Nat Genet* 2015, **47:**555-559.
33. Korneliussen TS, Albrechtsen A, Nielsen R: **ANGSD: Analysis of Next Generation Sequencing Data.** *BMC Bioinformatics* 2014, **15:**356.
34. Ingvarsson PK: **Multilocus patterns of nucleotide polymorphism and the demographic history of Populus tremula.** *Genetics* 2008, **180:**329-340.
35. Gutenkunst RN, Hernandez RD, Williamson SH, Bustamante CD: **Inferring the joint demographic history of multiple populations from multidimensional SNP frequency data.** *PLoS Genet* 2009, **5:**e1000695.
36. Tuskan GA, Difazio S, Jansson S, Bohlmann J, Grigoriev I, Hellsten U, Putnam N, Ralph S, Rombauts S, Salamov A, et al: **The genome of black cottonwood, Populus trichocarpa (Torr. & Gray).** *Science* 2006, **313:**1596-1604.
37. Portik DM, Leaché AD, Rivera D, Barej MF, Burger M, Hirschfeld M, Rödel M-O, Blackburn DC, Fujita MK: **Evaluating mechanisms of diversification in a Guineo-Congolian tropical forest frog using demographic model selection.** *Molecular Ecology* 2017, **26:**5245-5263.
38. Coffman AJ, Hsieh PH, Gravel S, Gutenkunst RN: **Computationally Efficient Composite Likelihood Statistics for Demographic Inference.** *Mol Biol Evol* 2016, **33:**591-593.

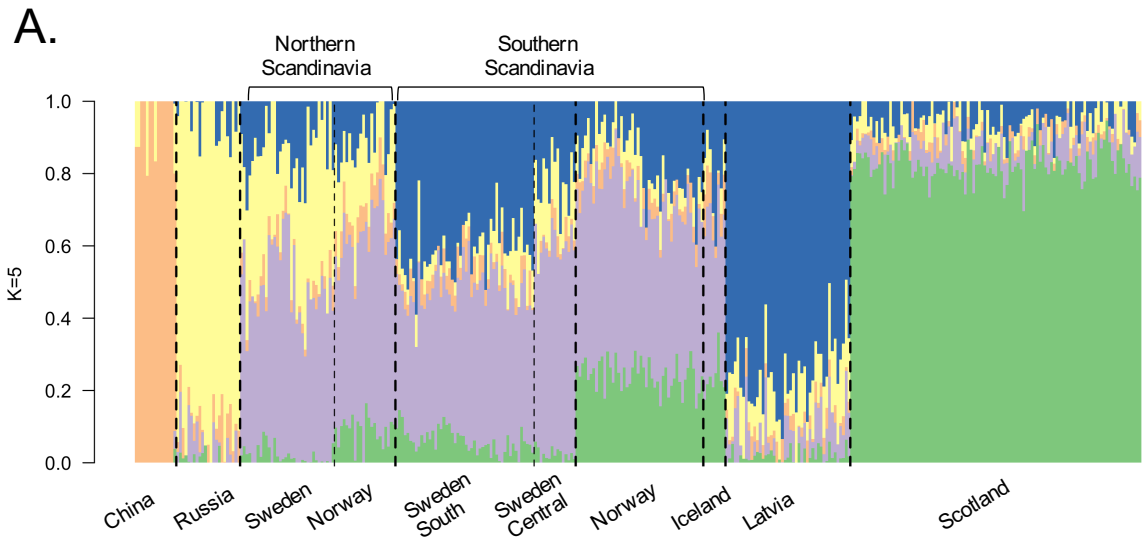
39. Shang HR-A, M.; Paun, O.; Field, D.; Hess, J.; Vogl, C.; Ingvarsson, P.K.; Leroy, T.; Christian Lexer, C. : **Conserved genomic landscapes of differentiation over the Populus speciation continuum** *in prep* 2021.
40. Rosengren E: **The Colonisation History of the Scandinavian Fauna Presented through Subfossil Finds along One of Its Major Immigration Routes; Scania, Southern Sweden.** *Environment and Ecology Research* 2015, **3:**143-149.
41. Gunther T, Malmstrom H, Svensson EM, Omrak A, Sanchez-Quinto F, Kilinc GM, Krzewinska M, Eriksson G, Fraser M, Edlund H, et al: **Population genomics of Mesolithic Scandinavia: Investigating early postglacial migration routes and high-latitude adaptation.** *PLoS Biol* 2018, **16:**e2003703.
42. Wetterich S, Rudaya N, Tumskoy V, Andreev AA, Opel T, Schirrmeister L, Meyer H: **Last Glacial Maximum records in permafrost of the East Siberian Arctic.** *Quaternary Science Reviews* 2011, **30:**3139-3151.
43. Westergaard KB, Zemp N, Bruederle LP, Stenoien HK, Widmer A, Fior S: **Population genomic evidence for plant glacial survival in Scandinavia.** *Mol Ecol* 2019, **28:**818-832.
44. Chhatre VE, Evans LM, DiFazio SP, Keller SR: **Adaptive introgression and maintenance of a trispecies hybrid complex in range-edge populations of Populus.** *Molecular Ecology* 2018, **27:**4820-4838.
45. Fussi B, Lexer C, Heinze B: **Phylogeography of Populus alba (L.) and Populus tremula (L.) in Central Europe: secondary contact and hybridisation during recolonisation from disconnected refugia.** *Tree Genetics & Genomes* 2010, **6:**439-450.
46. Burgarella C, Barnaud A, Kane NA, Jankowski F, Scarcelli N, Billot C, Vigouroux Y, Berthouly-Salazar C: **Adaptive Introgression: An Untapped Evolutionary Mechanism for Crop Adaptation.** *Front Plant Sci* 2019, **10:**4.
47. Hedrick PW: **Adaptive introgression in animals: examples and comparison to new mutation and standing variation as sources of adaptive variation.** *Mol Ecol* 2013, **22:**4606-4618.
48. Gouy A, Excoffier L: **Polygenic Patterns of Adaptive Introgression in Modern Humans Are Mainly Shaped by Response to Pathogens.** *Mol Biol Evol* 2020, **37:**1420-1433.
49. Bechsgaard J, Jorgensen TH, Schierup MH: **Evidence for Adaptive Introgression of Disease Resistance Genes Among Closely Related Arabidopsis Species.** *G3 (Bethesda)* 2017, **7:**2677-2683.
50. Hohmann N, Koch MA: **An Arabidopsis introgression zone studied at high spatio-temporal resolution: interglacial and multiple genetic contact exemplified using whole nuclear and plastid genomes.** *BMC Genomics* 2017, **18:**810.
51. Ma Y, Wang J, Hu Q, Li J, Sun Y, Zhang L, Abbott RJ, Liu J, Mao K: **Ancient introgression drives adaptation to cooler and drier mountain habitats in a cypress species complex.** *Commun Biol* 2019, **2:**213.
52. Michelson IH, Ingvarsson PK, Robinson KM, Edlund E, Eriksson ME, Nilsson O, Jansson S: **Autumn senescence in aspen is not triggered by day length.** *Physiol Plant* 2018, **162:**123-134.
53. Devi A, Jain K: **The Impact of Dominance on Adaptation in Changing Environments.** *Genetics* 2020, **216:**227-240.
54. Huber CD, Durvasula A, Hancock AM, Lohmueller KE: **Gene expression drives the evolution of dominance.** *Nat Commun* 2018, **9:**2750.

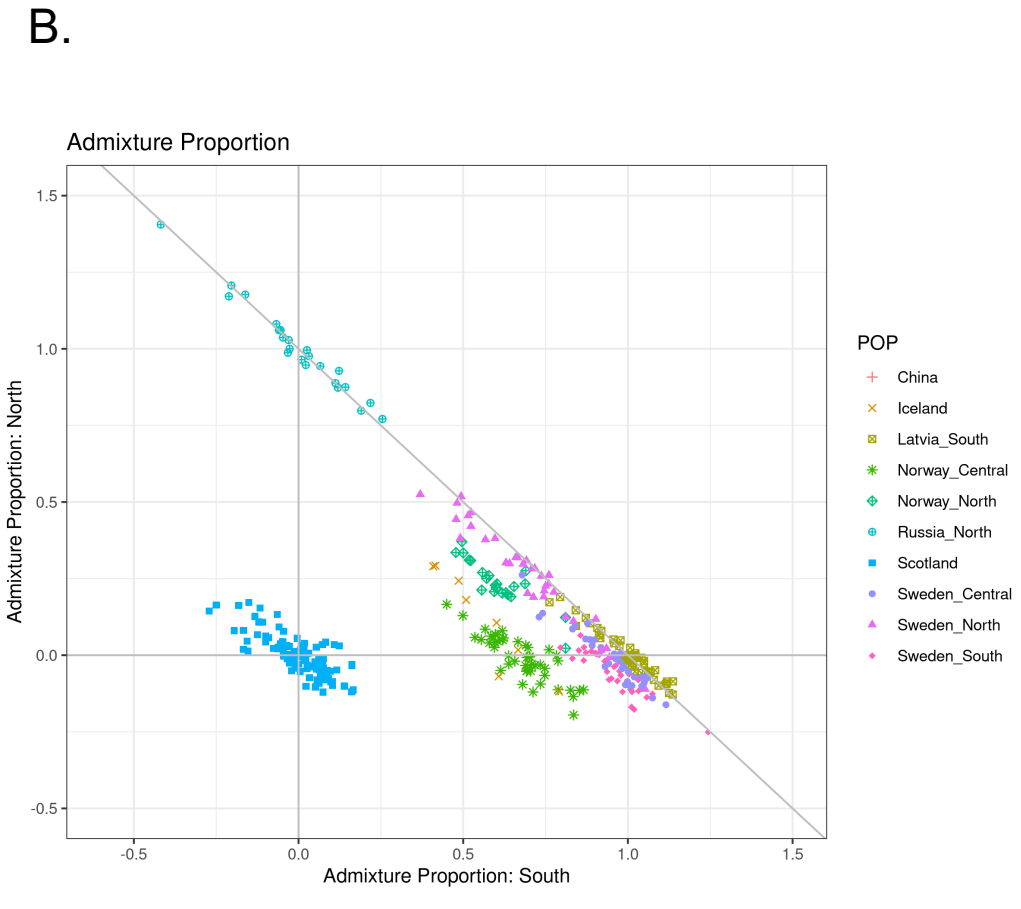
55. Apuli RP, Bernhardsson C, Schiffthaler B, Robinson KM, Jansson S, Street NR, Ingvarsson PK: **Inferring the Genomic Landscape of Recombination Rate Variation in European Aspen (Populus tremula).** *G3 (Bethesda)* 2020, **10:**299-309.
56. Messer PW, Petrov DA: **Population genomics of rapid adaptation by soft selective sweeps.** *Trends in Ecology & Evolution* 2013, **28:**659-669.
57. Hermisson J, Pennings PS, Kelley J: **Soft sweeps and beyond: understanding the patterns and probabilities of selection footprints under rapid adaptation.** *Methods in Ecology and Evolution* 2017, **8:**700-716.
58. Iqbal N, Khan NA, Ferrante A, Trivellini A, Francini A, Khan MIR: **Ethylene Role in Plant Growth, Development and Senescence: Interaction with Other Phytohormones.** *Front Plant Sci* 2017, **8:**475.
59. Felten J, Vahala J, Love J, Gorzsás A, Gerber L, Kumar M, Kangasjärvi J, Sundberg B: **Ethylene signaling via Ethylene Response Factors (ERFs) modifies wood development in hybrid aspen.** *BMC Proceedings* 2011, **5**.
60. Love J, Bjorklund S, Vahala J, Hertzberg M, Kangasjarvi J, Sundberg B: **Ethylene is an endogenous stimulator of cell division in the cambial meristem of Populus.** *Proc Natl Acad Sci U S A* 2009, **106:**5984-5989.
61. Gonzales-Vigil E, Hefer CA, von Loessl ME, La Mantia J, Mansfield SD: **Exploiting Natural Variation to Uncover an Alkene Biosynthetic Enzyme in Poplar.** *Plant Cell* 2017, **29:**2000-2015.

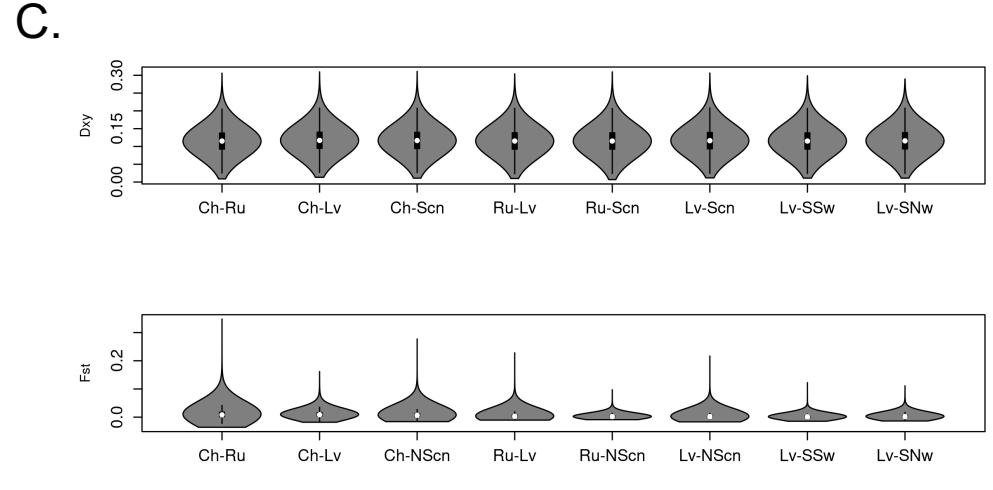
A.

B.

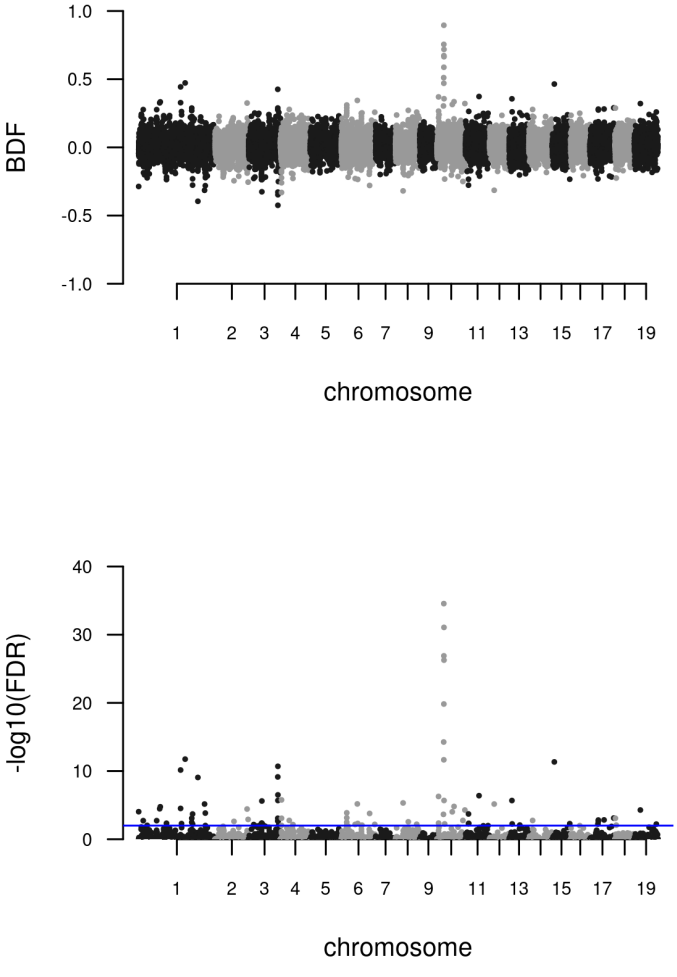

0.05
0.00
-0.05
-0.10
PC1 (17.4%)
0.0
0.1
0.2
-PC2 (7.91%)
Country
China
Iceland
Latvia
Norway
Russia
Sweden
Scotland


A.
Northern Scandinavia
Southern Scandinavia
K=5
1.0
0.8
0.6
0.4
0.2
0.0
China
Russia
Sweden
Norway
Sweden South
Sweden Central
Norway
Iceland
Latvia
Scotland

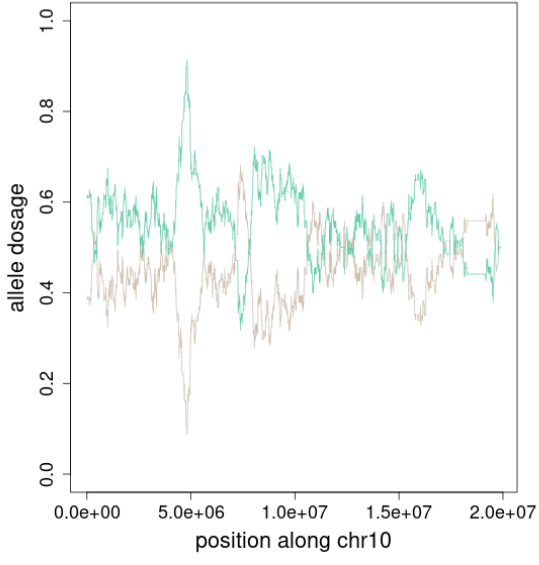

B.
Admixture Proportion
Admixture Proportion: North
Admixture Proportion: South
1.5
1.0
0.5
0.0
-0.5
POP
China
Iceland
Latvia_South
Norway_Central
Norway_North
Russia_North
Scotland
Sweden_Central
Sweden_North
Sweden_South


C.
Dxy
0.00
0.15
0.30
Ch-Ru
Ch-Lv
Ch-Scn
Ru-Lv
Ru-Scn
Lv-Scn
Lv-SSw
Lv-SNw
Fst
0.0
0.2
Ch-Ru
Ch-Lv
Ch-NScn
Ru-Lv
Ru-NScn
Lv-NScn
Lv-SSw
Lv-SNw

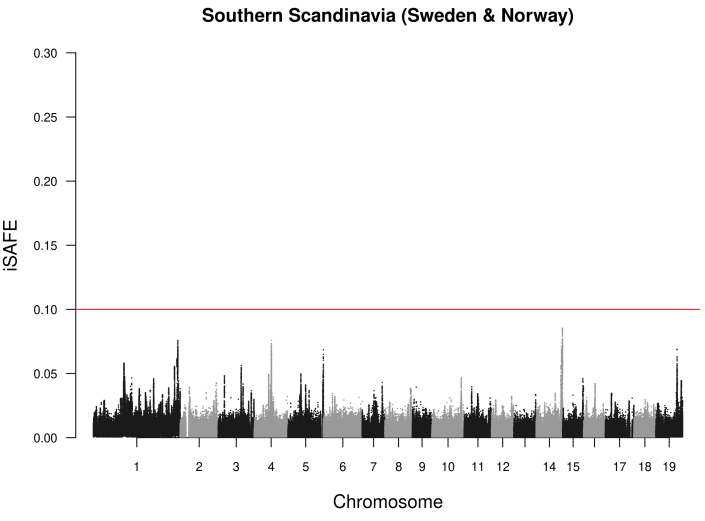
A.


BDF
1.0
0.5
0.0
-0.5
-1.0
1
2
3
4
5
6
7
9
11
13
15
17
19
chromosome
-log10(FDR)
40
30
20
10
0
1
2
3
4
5
6
7
9
11
13
15
17
19
chromosome

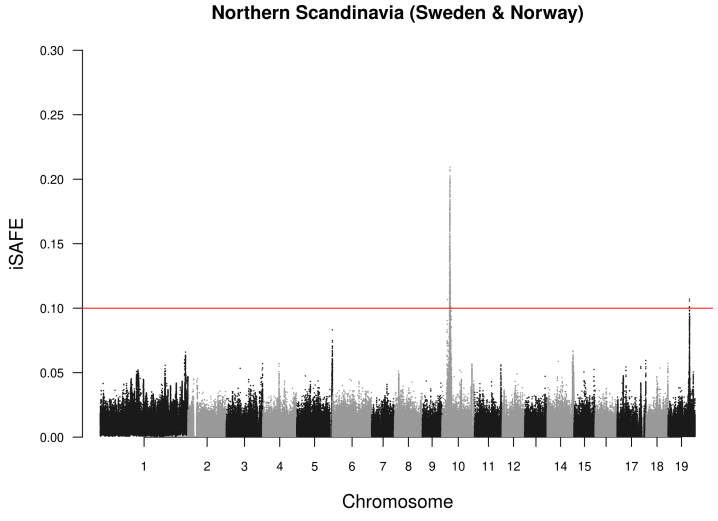
B.


allele dosage
1.0
0.8
0.6
0.4
0.2
0.0
0.0e+00
5.0e+06
1.0e+07
1.5e+07
2.0e+07
position along chr10

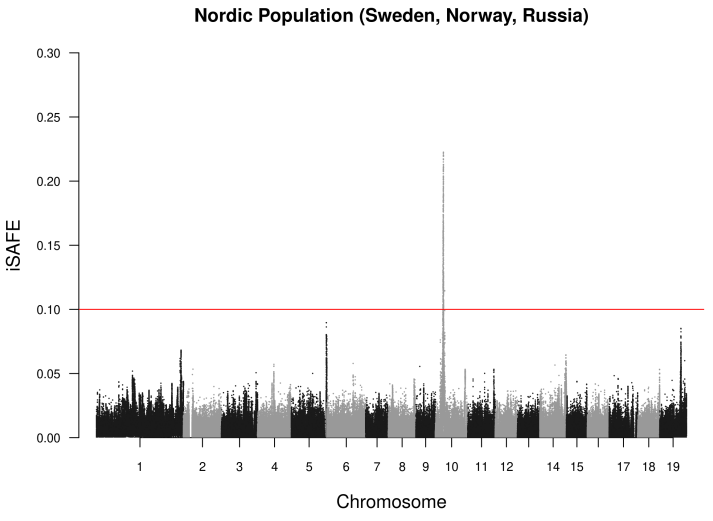
A.

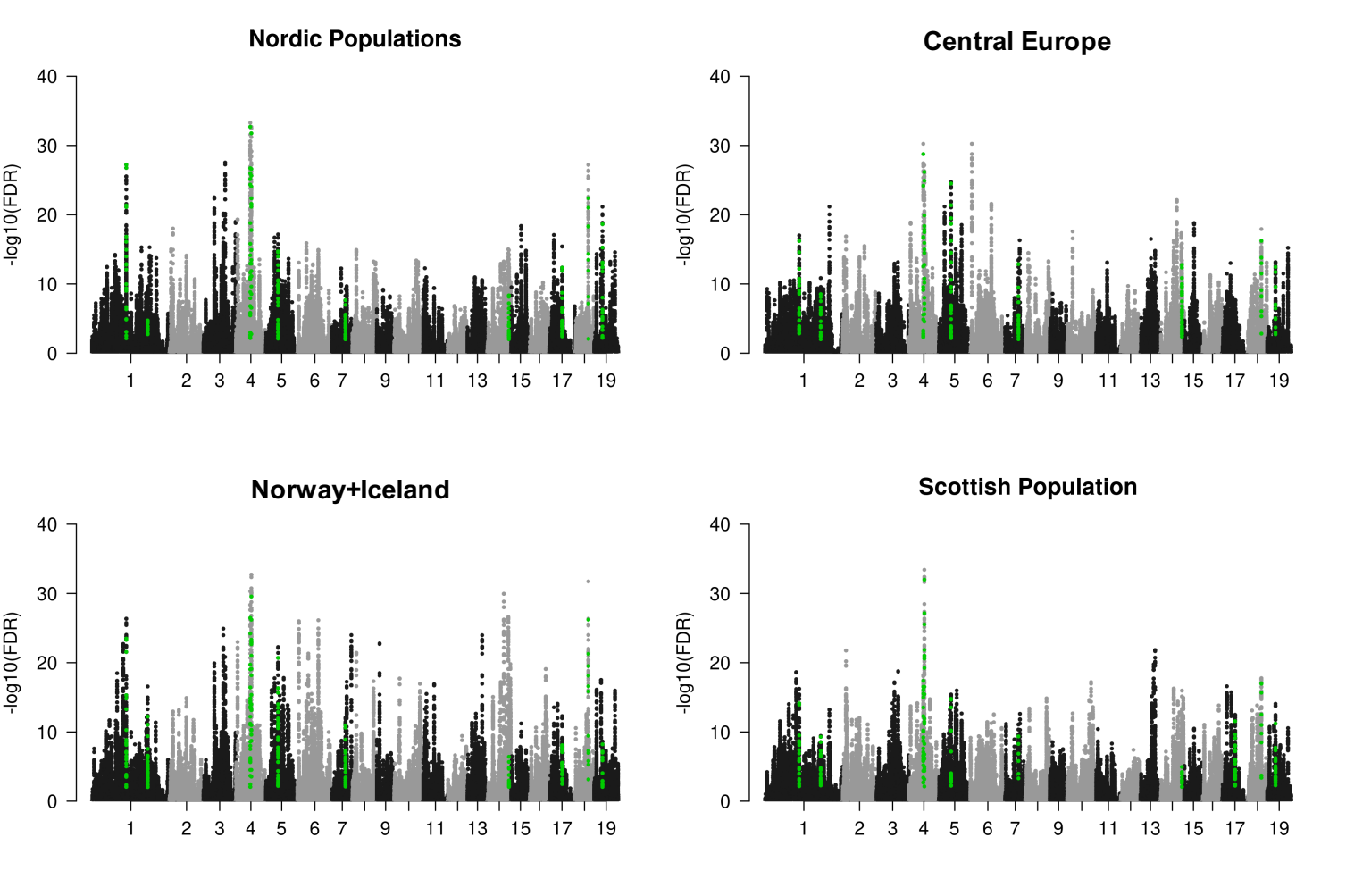

Southern Scandinavia (Sweden & Norway)
iSAFE
0.30
0.25
0.20
0.15
0.10
0.05
0.00
1 2 3 4 5 6 7 8 9 10 11 12 14 15 17 18 19
Chromosome

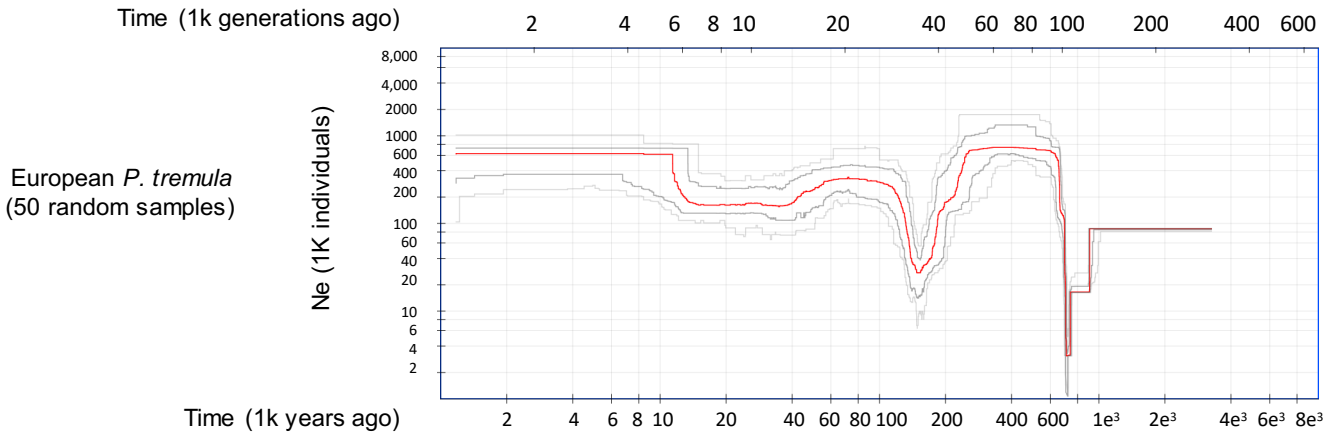
B.

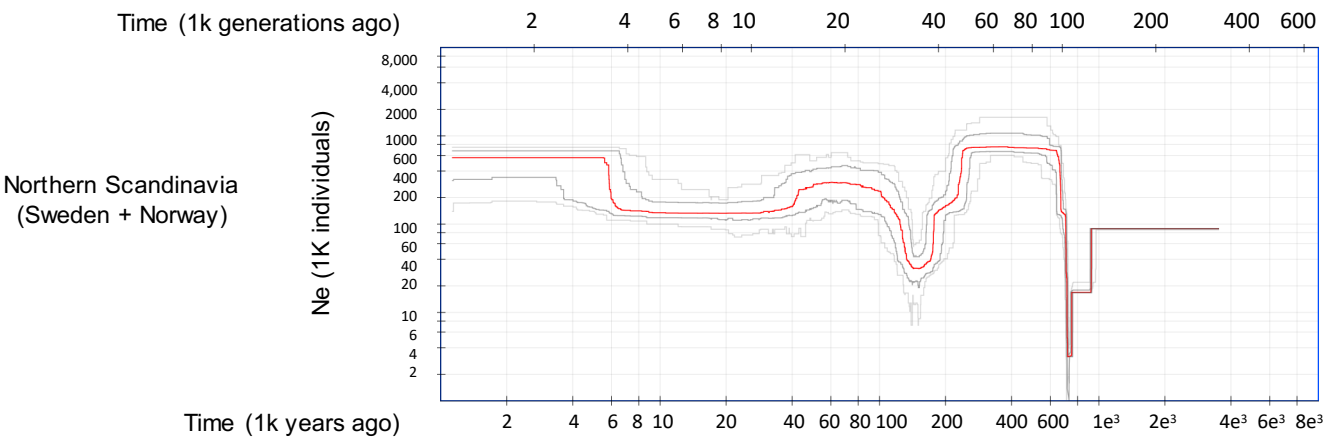

Northern Scandinavia (Sweden & Norway)
iSAFE
0.30
0.25
0.20
0.15
0.10
0.05
0.00
1 2 3 4 5 6 7 8 9 10 11 12 14 15 17 18 19
Chromosome

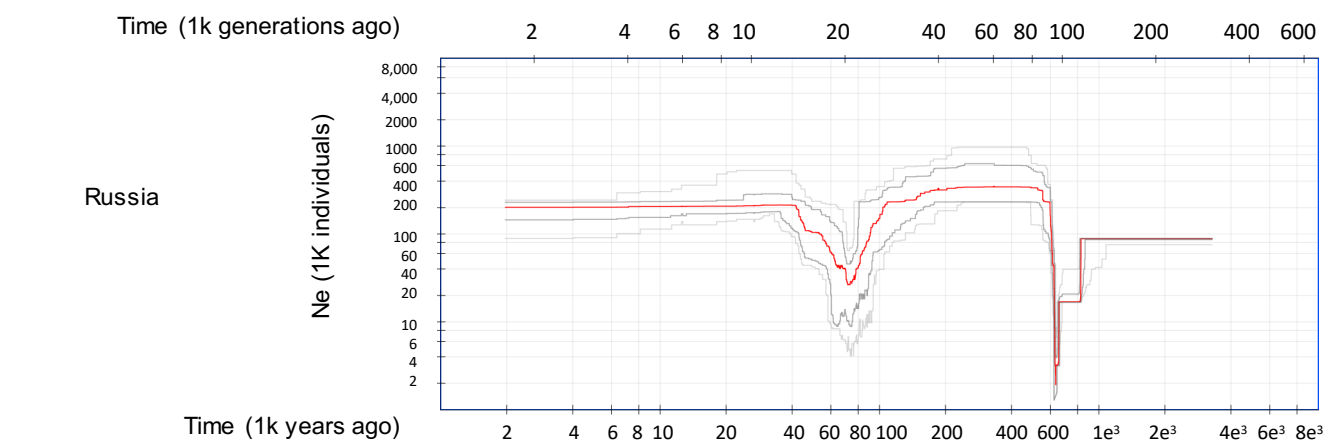
C.


Nordic Population (Sweden, Norway, Russia)
iSAFE
0.30
0.25
0.20
0.15
0.10
0.05
0.00
1 2 3 4 5 6 7 8 9 10 11 12 14 15 17 18 19
Chromosome

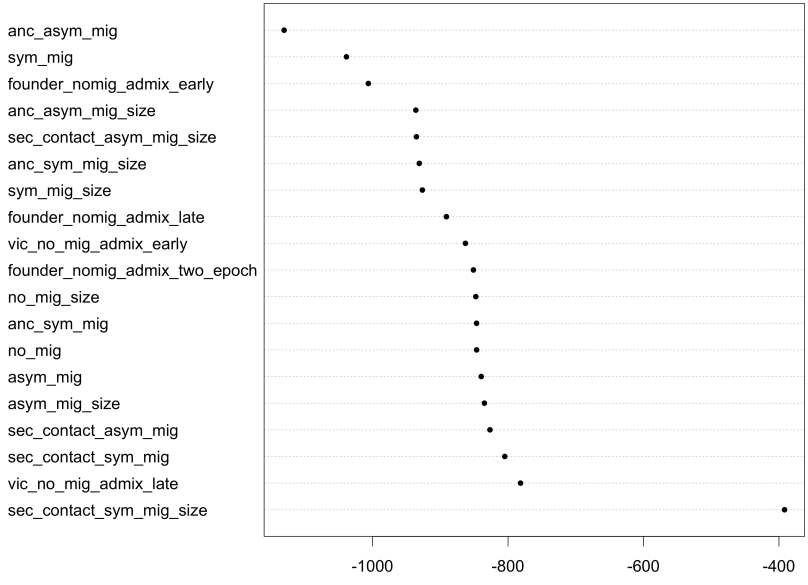

Nordic Populations
Central Europe
Norway+Iceland
Scottish Population
-log10(FDR)
0
10
20
30
40
1
2
3
4
5
6
7
9
11
13
15
17
19

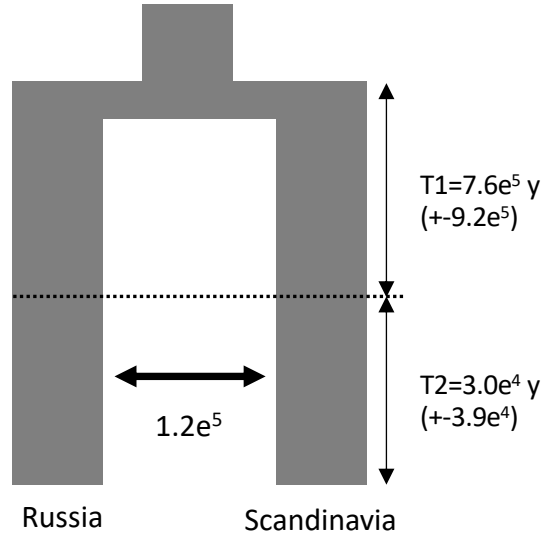

European *P. tremula*
(50 random samples)
Time (1k generations ago)
2 4 6 8 10 20 40 60 80 100 200 400 600
Ne (1K individuals)
8,000
4,000
2000
1000
600
400
200
100
60
40
20
10
6
4
2
Time (1k years ago)
2 4 6 8 10 20 40 60 80 100 200 400 600 1e³ 2e³ 4e³ 6e³ 8e³


Northern Scandinavia
(Sweden + Norway)
Time (1k generations ago)
2 4 6 8 10 20 40 60 80 100 200 400 600
Ne (1K individuals)
8,000
4,000
2000
1000
600
400
200
100
60
40
20
10
6
4
2
Time (1k years ago)
2 4 6 8 10 20 40 60 80 100 200 400 600 1e³ 2e³ 4e³ 6e³ 8e³

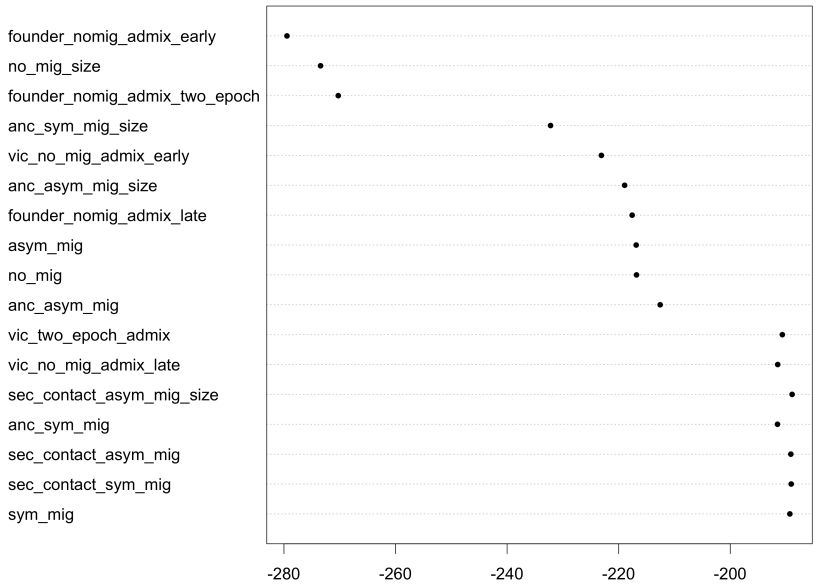

Russia
Time (1k generations ago)
2 4 6 8 10 20 40 60 80 100 200 400 600
Ne (1K individuals)
8,000
4,000
2000
1000
600
400
200
100
60
40
20
10
6
4
2
Time (1k years ago)
2 4 6 8 10 20 40 60 80 100 200 400 600 1e³ 2e³ 4e³ 6e³ 8e³

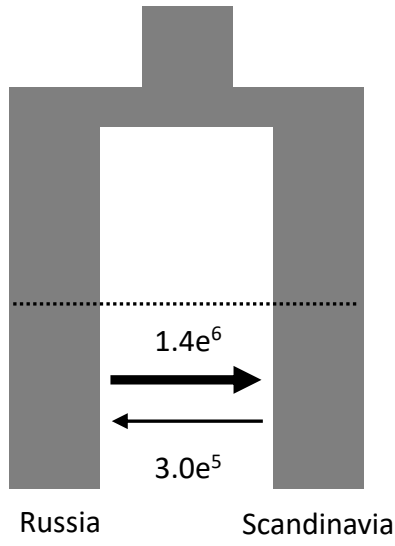
A.


chromosome 10
anc_asym_mig
sym_mig
founder_nomig_admix_early
anc_asym_mig_size
sec_contact_asym_mig_size
anc_sym_mig_size
sym_mig_size
founder_nomig_admix_late
vic_no_mig_admix_early
founder_nomig_admix_two_epoch
no_mig_size
anc_sym_mig
no_mig
asym_mig
asym_mig_size
sec_contact_asym_mig
sec_contact_sym_mig
vic_no_mig_admix_late
sec_contact_sym_mig_size
-1000
-800
-600
-400


secondary contact, symetric migration
T1=7.6e5 y
(+-9.2e5)
T2=3.0e4 y
(+-3.9e4)
1.2e5
Russia
Scandinavia

B.


selective sweep
founder_nomig_admix_early
no_mig_size
founder_nomig_admix_two_epoch
anc_sym_mig_size
vic_no_mig_admix_early
anc_asym_mig_size
founder_nomig_admix_late
asym_mig
no_mig
anc_asym_mig
vic_two_epoch_admix
vic_no_mig_admix_late
sec_contact_asym_mig_size
anc_sym_mig
sec_contact_asym_mig
sec_contact_sym_mig
sym_mig
-280
-260
-240
-220
-200


secondary contact, asymetric migration
1.4e6
3.0e5
Russia
Scandinavia