

**1 Adaptive Introgression Shapes the Pan-genome of
2 *Populus* Hybrid Zones**

3 Baxter Worthing¹

4 ¹The University of Vermont,

5 **Abstract**

6 Poplar trees...

7 **0.1 Introduction**

8 Hybridization can lead to the exchange of genetic variation across species boundaries,
 9 a process known as introgression. Introgression is a powerful source of genetic novelty,
 10 playing a significant role in the evolutionary history of natural systems (Hedrick,
 11 2013; Suarez-Gonzalez, Lexer, et al., 2018), humans (Racimo et al., 2015) and key
 12 crop varietals (Burgarella et al., 2019; Cheng et al., 2019; Gao et al., 2019; Kou et al.,
 13 2020; Qiao et al., 2021; Sun et al., 2020; Zanini et al., 2022). Admixed individuals, the
 14 recipients of introgressed sequence, harbor novel combinations of alleles originating
 15 from both parental species, and may present phenotypic variation which transcends
 16 that observed in the parental species. In this sense, admixed individuals represent
 17 unique opportunities, not only to better understand the genetic basis of phenotype,
 18 but also to better understand the nature of molecular interactions between divergent
 19 haplotypes. Furthermore, when parental species inhabit differing environments,
 20 patterns of introgression may also reflect the interaction between genetic variation
 21 and environment. Ultimately, these GxG and GxE interactions define the contexts in
 22 which introgression is adaptive. Past research on natural admixed populations sug-
 23 gests that introgression can be an adaptive process, through which genetic variation
 24 related to locally-adaptive phenotypes is passed from one species to another (Gaczorek
 25 et al., 2024; Hamilton & Miller, 2016; Hedrick, 2013; Kremer & Hipp, 2020; Leroy et
 26 al., 2020; Rendón-Anaya et al., 2021; Savolainen et al., 2007; Suarez-Gonzalez, Lexer,
 27 et al., 2018). Closer and more extensive analysis of natural manifestations of adaptive
 28 introgression can reveal the types of genetic variation that are able to cross species
 29 boundaries without a fitness consequence, the regions of the genome that are receptive
 30 to the introduction of foreign sequence, and the degree to which environmental con-
 31 text influences the adaptive nature of introgression. Research in this area is valuable
 32 not only because it sheds light on important questions in evolutionary biology, such
 33 as the evolution of species boundaries and molecular basis of adaptive traits, but
 34 also because it gauges the overall potential for the introduction of divergent genetic
 35 variation into novel genetic and environmental background. Therefore, research in this
 36 area has great potential for broader impact for two main reasons. First, because the
 37 contribution of biotechnology to advances in the fields of medicine and agriculture
 38 increasingly depend on the introduction of genetic sequence into a novel background.
 39 Second, because introgression provides a potentially faster and more reliable route for
 40 the adaptive evolution of phenotype than de novo mutation, meaning that natural sys-
 41 tems and breeding programs alike may depend on introgression in order to facilitate
 42 rapid evolution of adaptive traits in response to changing climates.

43 The size and molecular origin of the genetic variation that is exchanged between
 44 species through adaptive introgression is important to define, as a growing body of
 45 research suggests that genomic structural variation (SV), here defined as genetic
 46 variation larger than 50bp resulting for insertion, deletion, translocation or inversion
 47 of genomic sequence, is often the causal variation underlying ecologically and econom-
 48 ically important traits in many taxa (*CITE*). For instance, SV may play a role in
 49 local adaptation to climate, and past research shows divergence between populations
 50 for SV genotypes resulting from local adaptation (Hämälä et al., 2021; Y. Li et al.,
 51 2024; Z. Li et al., 2023; Songsomboon et al., 2021). At a broader scale, research also
 52 shows that SV maintain genetic differentiation between species (Ferguson et al., 2024;
 53 L. Zhang et al., 2021). However, little is known about the adaptive exchange of SV
 54 genotypes between divergent populations and species. Introgression between crop
 55 varietals and wild relatives is a key factor in the breeding history of many crops,
 56 and recent work shows that SV are often the causal variants involved in this process
 57 (Cheng et al., 2019; Gao et al., 2019; Kou et al., 2020; Qiao et al., 2021; Sun et al.,

58 2020.; Zanini et al., 2022). Introgression is also common in natural systems and recent
 59 work highlights introgression as an important source of adaptive genetic novelty
 60 in many species, particularly forest trees (Hamilton & Miller, 2016; Leroy et al.,
 61 2020; Rendón-Anaya et al., 2021; Suarez-Gonzalez, Lexer, et al., 2018). However,
 62 past work on the genetics of such adaptive introgression in natural systems focused
 63 primarily on SNP genotypes. While hybrid genomes may be porous to variation
 64 at the single nucleotide scale, and introgressed SNP alleles may even present an
 65 adaptive advantage in admixed individuals, it remains unclear if the same is true
 66 for SV genotypes, which have greater potential for deleterious phenotypic effects.
 67 Recently, research has highlighted a potential role for SV in adaptive introgression.
 68 This work suggests that SV may play an important role in both adaptive introgression
 69 (Almarri et al., 2020; Xia et al., 2023; X. Zhang et al., 2021) and in the maintenance
 70 of species boundaries (L. Zhang et al., 2021). If SV is indeed involved in adaptive
 71 introgression, it is worth investigating the molecular nature (eg size, frequency, variant
 72 class, genomic location) of the SV alleles involved, as well as the consistency and
 73 directionality of adaptive introgression involving SV. Admixed individuals may present
 74 mosaics of SV alleles derived from both parental species, or introgression may be
 75 directionally biased toward one species. Similarly, some loci may act as barriers to
 76 introgression, while others are more receptive. Natural hybrid zones may be hotspots
 77 for the de novo generation of SV alleles through processes such as unequal crossing
 78 over. Here, we leverage extensive sampling of natural forest tree hybrid zones, cutting
 79 edge techniques for genotyping SV in admixed individuals and established landscape
 80 genomics analyses to investigate these areas of uncertainty.

81 The study of SV in natural hybrid zones is challenging, as it requires the ability to
 82 accurately genotype SV in a large number of admixed individuals. The majority of
 83 past research on SV in natural hybrid zones has relied on the use of reference genomes
 84 derived from a single individual from only one of the parental species, which can lead
 85 to reference bias (Bock et al., 2023; Feng et al., 2024; Song et al., 2023). Reference
 86 bias occurs when the genomes of resequenced individuals contain regions that are
 87 highly diverged from the reference genome sequence, causing reads originating from
 88 these regions to align incorrectly, or not align at all. This can lead to misinterpre-
 89 tation or under-representation of variation resulting from admixture (Secomandi et
 90 al., 2025). Reference bias is particularly problematic when studying SV, as large
 91 insertions and deletions may be longer than sequencing reads, making variation in
 92 their presence and absence (PAV) challenging to detect. Furthermore, SV is often
 93 species-specific and, because admixed individuals represent a mixture of genetic
 94 variation from two or more species, the presence of reference bias can obscure the
 95 effect that the presence (or absence) of genetic variation has on traits, and can hinder
 96 the identification of SV that is involved in introgression.

97 Recently, an increasing number of studies have relied on pan-genome assembly to
 98 overcome reference bias and to accurately genotype SV (Gao et al., 2019; Kang et al.,
 99 2023; Y. Li et al., 2024; Z. Li et al., 2023; Liang et al., 2025; Secomandi et al., 2025;
 100 Songsomboon et al., 2021). A pan-genome assembly is a non-redundant collection
 101 of sequences originating from multiple individuals. This genetic information can be
 102 represented as “nodes” in a pan-genome graph, while the linear sequence of each input
 103 genome is stored as a “path” connecting a series of nodes. In a pan-genome graph,
 104 each node describes the alignment between at least two sequences, given an expected
 105 level of sequence divergence (Kang et al., 2023). Pan-genome graphs can capture
 106 complex variation that is not present in a single reference genome, and can provide a
 107 more accurate representation of the genetic variation present in admixed individuals.
 108 Pan-genome assembly has been used to identify and genotype SV related to ecolog-
 109 ically and economically important traits in a variety of taxa, and has been shown
 110 to be an effective tool for studying adaptive evolution in natural populations (Fang
 111 & Edwards, 2024; Kang et al., 2023; Secomandi et al., 2025). Pan-genome assembly

could help overcome the challenges of studying SV in admixed tree populations, and could provide new insights into the role of SV in adaptive introgression in natural populations of forest trees (Secomandi et al., 2025). Despite this potential, we know of only one study that used pan-genome assembly to examine introgression in the context of hybridization in the wild. Liang et al. (2025) use a pan-genome based approach to genotype SV in a range-wide sampling of the interfertile oak species *Quercus variabilis* and *Q. acutissima*, showing that SV harbor signals of adaptive introgression that differ from those detected from SNP data and identifying adaptive introgression of SV genotypes in a gene rich region of oak chromosome 9 (Liang et al., 2025). In light of this work, it is worth investigating if signatures adaptive introgression on SV exist in other interfertile tree species and asking not only how this adaptive signal is distributed across the whole genome but also how it varies with environmental context across the range of hybridizing species. Here, we add to this nascent line of research by leveraging pan-genome assembly and fine-scale sampling of natural hybrid zones to explore genome-wide relationship between structural variation and introgression between *Populus balsamifera* and *Populus trichocarpa*, two forest tree species that diverged much more recently than the oak species studied by Liang et al. (2025).

Populus balsamifera and *Populus trichocarpa* readily interbreed in nature, and early and advanced generation hybrids have been described in hybrid zones located where their ranges overlap (Chhatre et al., 2018; Geraldes et al., 2014; Suarez-Gonzalez et al., 2016; Suarez-Gonzalez, Hefer, et al., 2018). It is thought that hybridization between these two species facilitates introgression of genetic variation related to locally-adaptive phenotypes from one species to the other (Suarez-Gonzalez et al., 2016; Suarez-Gonzalez, Hefer, et al., 2018). However, studies on controlled interspecific crosses in *Populus* have made it clear that introgression is often biased toward particular regions of the genome, particular *Populus* species, or specific environments (Lexer et al., 2005; Meirmans et al., 2017; Thompson et al., 2010). Therefore, the porosity of the genome to adaptive introgression between these species may be constrained by incompatibilities between the genomic variants involved as well as environmental context, two factors which warrant further investigation. The environmental context of adaptive introgression may be of particular importance in the case of *P. balsamifera* and *P. trichocarpa*. *P. balsamifera* is more commonly found in colder, continental climates while *P. trichocarpa* is more often associated with milder, coastal climates with longer growing seasons (Geraldes et al., 2014; Suarez-Gonzalez, Hefer, et al., 2018). Considering hybrid zones between these species often fall along the boundaries of their ranges (Chhatre et al. (2018); Fetter & Keller (2023)), it seems possible that adaptive introgression helps hybrid individuals persist in environments that are outside of the optimum for either parental species. Suarez-Gonzalez, Hefer, et al. (2018) showed support for this hypothesis, finding that introgressed regions in the genomes of hybrids harbored signs of selection and variation associated with adaptive traits, including phenology. However, the specific adaptive variants within these genomic regions, and the nature of their effects on fitness remain, for the most part, undiscovered. Identification of SV involved in adaptive introgression between these two species would not only contribute to a broader understanding of the molecular basis of adaptive introgression, but would also provide helpful insight into the conservation of these ecologically important species. In trees, long generations times are thought to limit the potential contribution of de novo mutation to adaptive evolution in response to rapid environmental change (Feng et al., 2024; Savolainen et al., 2007). Therefore, adaptive introgression is viewed as one of the few routes for species like *P. balsamifera* and *P. trichocarpa* to adaptively evolve in response to climate change (Feng et al., 2024). If adaptive introgression does help admixed individuals persist in environments outside of their optimum, then the variation involved could guide genetic conservation efforts, such as assisted gene flow. Furthermore, industrial breeding of *Populus* varietals generally involves interspecific crosses, and breeding

167 programs would benefit from a clearer understanding of the genetic variation that is
 168 involved in adaptive introgression between these two species.

169 Here, we take advantage of recent advances in pan-genome assembly methods to
 170 produce a pan-genome reference, comprising 24 diverse haplotypes from *P. balsamifera*,
P. trichocarpa and their hybrids, facilitating unbiased analysis of the sequence
 171 variation that is segregating within natural *Populus* hybrid zones. Using this new
 172 pan-genome assembly, we genotype structural variation across 566 individuals sampled
 173 from within and outside of 6 distinct *P. balsamifera* and *P. trichocarpa* hybrid zones.
 174 We assess the extent of genomic diversity present in these species and their hybrids,
 175 identifying genomic variation that is not present in the *P. trichocarpa* reference
 176 genome. Furthermore, we describe structural variation involved both in introgression
 177 and in genomic divergence between the two species. We shed light on the role that
 178 introgression may play in shaping the pan-genome of these species, and the degree
 179 to which tree genomes are porous to the inter-specific exchange of structural variant
 180 alleles.
 181

182 0.2 Methods

183 0.2.1 Sample collection and whole genome sequencing

184 In January 2020, we collected dormant branch cuttings from 546 poplar trees along
 185 7 distinct latitudinal transects spanning natural contact zones between *Populus* tri-
chocarpa and *Populus balsamifera* Figure 1. Short read whole genome sequencing and
 186 subsequent bioinformatic analyses were performed as described in Bolte et al. (2024).
 187 Briefly, Genomic DNA libraries for all sampled individuals were constructed at the
 188 Duke University Center for Genomic and Computational Biology and sequenced using
 189 an Illumina NovaSeq 6000 instrument with an S4 flow cell with 64 samples per lane
 190 (Illumina Inc., San Diego, USA). De-indexing, QC, trimming adapter sequences, and
 191 sequence pre-processing were completed by the sequencing facility. In addition to
 192 short read sequencing, we also sequenced a subset of 40 sampled individuals with
 193 PacBio HiFi long reads. We harvested tissue from newly-expanded leaves grown under
 194 low light conditions to use for high molecular weight DNA extraction. After confirming
 195 extraction quality through gel electrophoresis and bioanalysis, we submitted HMW
 196 DNA to the University of Maryland Center for Integrative Biosciences Genomics Core
 197 Facility for library preparation and sequencing on the the PacBio Sequel system with
 198 two samples per SMRT flow cell. We sequenced one *P. balsamifera* sample (939) for
 199 an additional run on a full SMRT cell. We also sequenced this individual with Oxford
 200 Nanopore Technology (ONT) platform. We submitted high molecular weight DNA to
 201 the Vermont Integrative Genomics Resource Sequencing Center for library preparation
 202 and sequencing on XXX flowcell (two runs) and XXX flowcell (one run)
 203

204 0.2.2 Genome assembly

205 Of the 40 individuals sequenced with HiFi long reads, we selected 16 for whole
 206 genome assembly Figure 1. These samples ranged from 20 to 35x long read coverage.
 207 We performed de-novo genome assembly of the HiFi reads for these samples with
 208 HiCanu (Nurk et al., 2020). We set HiCanu parameters as follows: gSize=“400m”,
 209 lc=“5”, lcer=“0.055”, ovrlp=“350”, mincov=“9”. To detect potential contaminants
 210 in the raw assemblies produced by HiCanu, we used the program Kraken2 to com-
 211 pare the k-mer content of assembled contigs to the “PlusPFP” database of known
 212 taxon-specific k-mers representing archaea, bacteria, viral, human, protozoa, fungi
 213 and plant taxa (Lu et al., 2022). We then used a custom bioawk script to remove
 214 any assembled contig that Kraken2 assigned to a taxonomic unit other than *Populus*,
 215 leaving only unassigned or *Populus*- assigned contigs in each assembly. We assessed
 216 the accuracy and completeness of the decontaminated assemblies using QUAST and
 217 BUSCO (Gurevich et al., 2013; Simão et al., 2015). To further assess the contiguity
 218 and quality of these assemblies, we used minimap2 and BWA to map the original
 219 HiFi reads and additional Illumina short reads back to each assembled haplotype
 (H. Li, 2018; H. Li & Durbin, 2010). We passed these alignments to the program
 220

221 CRAQ (K. Li et al., 2023), which leverages read depth along assembled contigs for
 222 quality assessment. We repeated these quality assessment checks at each subsequent
 223 stage of genome assembly. Phasing, or the separation and concatenation of contigs
 224 belonging to the same parental haplotype is a key step in genome assembly for hybrid
 225 individuals, as divergent haplotypes likely contain important genetic information.
 226 We used mininmap2 to map reads back to assembled contigs and to map assembled
 227 contigs to themselves. We then used the purge haplotigs pipeline (Roach et al.,
 228 2018) to split diploid assemblies into two assembled haplotypes for each individual.
 229 We used RagTag (Alonge et al. 2022) to connect decontaminated, phased contigs
 230 into pseudo-chromosomal scaffolds, guided by alignments of assembled contigs to
 231 the Nisqually1 *P. trichocarpa* reference genome. We visually inspected minimap2
 232 alignments of scaffolded assemblies to the reference genome to identify potential
 233 scaffolding errors. We used RepeatMasker to annotate and mask repetitive regions
 234 and annotated possible coding domains and predicted protein sequences for each
 235 assembly with AUGUSTUS (Smit et al., 2015; Hoff et al., 2019).

236 **0.2.3 Pan-genome assembly**

237 In a pan-genome graph, a non-redundant collection of all input sequences is rep-
 238 resented as “nodes”, while the linear sequence of each input genome is stored as a
 239 “path” connecting a series of nodes. In this approach to graph construction, each
 240 node of the graph w describes the alignment between at least two sequences, given
 241 an expected level of sequence divergence. We constructed a pan-genome graph from
 242 an all-by all alignment of the 24 assembled haplotypes. Any contigs shorter than
 243 100kb were dropped before alignment. We used wfmash (Guarracino et al., 2021) to
 244 perform all-by-all alignments between assembled chromosomes, and seqwish (Garrison
 245 and Guarracino, 2022) to project alignments into a graph pan-genome assembly.
 246 Pan-genome graphs constructed this way can have highly complex topography, which
 247 may hinder downstream applications such a sequence alignment. To combat this
 248 issue, we used the program smoothXG to smooth complex variation in these graphs.
 249 We also assembled a separate graph specifically for subsequent alignment-based
 250 analyses using the minigraph-cactus pipeline (Hickey et al., 2023). We used the
 251 program panacus to asses pan-genome growth and coverage statistics (Parmigiani et
 252 al., 2024). We used the program ODGI (Guarracino et al., 2022) to perform basic
 253 graph quality control and to partition the core (sequence present in all individuals),
 254 shell (sequence variably present or absent across individuals) and singleton sequence
 255 (present in only one individuals) content for each graph and individual represented in
 256 the graphs. We used the program vg (Garrison et al., 2018) to deconstruct each graph
 257 assembly into a VCF containing the variation encapsulated in each graph. We passed
 258 this deconstructed VCF file to vcfbub (github.com/pangenome/vcfbub) to collapse
 259 nested SV sites into the top-level variant for all SV less than 10kb in length. Using
 260 bcftools (Danecek et al., 2021) we split biallelic sites into multiallelic records for this
 261 deconstructed vcf.

262 **0.2.4 Genotyping Pan-genomic Variation**

263 We aligned sequencing reads from additional samples to pan-genome graph assembly
 264 to genotype SNPs and SVs. We aligned PacBio HiFi reads for 40 total individuals
 265 (including the 16 used to construct the pan-genome graph) to the minigraph-cactus
 266 graph using the program GraphAligner (Rautiainen & Marschall, 2020). To character-
 267 ize sequence variation across a broader set of individuals, we also mapped Illumina
 268 short reads for 575 individuals to the pan-genome graph using the program giraffe
 269 (Sirén et al., 2021). We then called SV and SNPs from these alignments using the
 270 vg call algorithm (Garrison et al., 2018). We used bcftools to split multiallelic calls
 271 into separate records, normalize alleles against the *P. trichocarpa* reference and merge
 272 calls for all samples into one VCF representing SV and SNP genotypes in the *P.*
trichocarpa reference coordinate space.

274 0.2.5 Analysis of Structural Variation and Introgression

275 We used ADMIXTURE results from Bolte et al (2024) to assign each individual an
 276 admixture proportion at K=2 and K=4. The genetic data used for this ADMIX-
 277 TURE analysis came from SNP calls from mappings of the same illumina reads used
 278 in this study, however they were mapped to the *P. trichocarpa* reference genome.
 279 We used this same genetic data set to perform local ancestry inference (LAI) using
 280 LOTR (Dias-Alves et al., 2018). These ADMIXTURE and LAI data sets were used to
 281 visualize and analyze patterns of local and global genomic ancestry throughout this
 282 study.

283 We used ODGI to flatten the Nisqually1 *P. trichocarpa* reference genome path in
 284 the cactus graph. We then used a bed file representation of this path and reference
 285 gene annotations to genotype PAV of reference sequence across all paths of the cactus
 286 graph. We used R version 4.1.0 to identify core and shell gene sequences from this
 287 data and to perform PCA on the PAV genotypes (R Core Team, 2024). We also ran a
 288 similar PAV PCA with the merged vcf representing all 575 sequenced samples in this
 289 study, using plink2 (Chang et al., 2015). We used the --allele-weights flag to extract
 290 the loading of each allele for each SV on the first 3 principal components.

291 To find SV potentially involved in introgression, we employed custom scripts to
 292 identify ancestry blocks along the chromosomes of all 335 admixed individuals in this
 293 study. An individual was considered admixed if its k=2 ADMIXTURE score indicated
 294 at least 5 percent admixture genome-wide. We used Bedtools (Quinlan & Hall, 2010)
 295 to find SV that overlapped ancestry blocks in each admixed individual. We then
 296 scored each SV allele based on its prevalence in ancestry blocks that contrasted the
 297 average global ancestry of an individual.

298 For all SV that were outliers in principal component or LAI analyses, we used Bed-
 299 tools to check for overlap with annotated genes. We then used plantgenie to analyze
 300 PFAM enrichment of gene sets (Sundell et al., 2015).

301 0.3 Results

302 0.3.1 Whole genome sequencing and assembly.

303 Illumina sequencing yielded on average of 8358 MB of sequence, corresponding to
 304 an average sequencing depth of 21x. PacBio HiFi sequencing yielded on average
 305 6368 MB of sequence, corresponding to an average sequencing depth of 16x. Sample
 306 939 was sequenced to a depth of roughly 60X HiFi and roughly 20X ONT reads.
 307 Table 1 shows summary statistics for the 16 whole genome assemblies included in the
 308 pan-genome. Assembly quality and contiguity were generally high, however alternate
 309 haplotypes were not always fully assembled (Table 1).

310 0.3.2 Pan-genome assembly

311 The minigraph-cactus pangenome graph contained a total of 102,907,131 nodes
 312 comprising 1.2 Gbp of sequence. Of this, 243 Mbp was present across all individuals
 313 (core genome), while 423Mbp was present in some but not all individuals (shell
 314 genome). The remaining 537 Mbp represented singleton sequence. Figure 2 shows how
 315 the size of the sequence classes changes as individuals are added to the pan-genome.
 316 Figure 3 shows the proportions of core, shell and singleton sequence within each
 317 sample. ?@fig-4 shows each genome in the pan-genome graph plotted on the first two
 318 principal components of a PCA on PAV of reference sequence.

319 0.3.3 Pan-genomic Variation

320 3708 annotated genes were variably present or absent across the 17 genomes used
 321 to construct the pan- =genome. These genes were significantly enriched for various
 322 protein families including NB-ARC domain, D-mannose binding lectin, sulfotrans-
 323 ferases, cupins, and Cytochrome P450. Of these shell genes, 167 were present in all
 324 of the pure *P. trichocarpa* individuals and none of the pure *P. balsamifera* samples
 325 making up the pan-genome graph. These *P. trichocarpa* - specific genes were enriched

326 for protein families, including sulfotransferases, serine carboxypeptidases, Mlo family
 327 proteins, D-mannose binding lectin, Glycosyl hydrolases and absistic acid (ABA)
 328 induced proteins.

329 The cactus pan-genome graph contained 1,773,427 SV greater than 20bp in length.
 330 Figure 4 shows the distribution of these variants across the genome. The distribution
 331 of frequencies of the non-reference allele for these SV is shown in Figure 5.

332 The first two principal components of a PCA on PAV genotyped from short read
 333 alignments to the pan-genome graph are shown in Figure 6. The loading of each allele
 334 for each SV on the first principal component of a PCA on PAV genotyped from short
 335 read alignments to the pan-genome graph is shown in ?@fig-8. The SV that were
 336 outliers on PC1 overlapped 1362 annotated genes. These genes were significantly
 337 enriched for protein families involved in development and growth (K-Box proteins,
 338 KNOX proteins), cold tolerance (DEAD/DEAH box helicases), and phenology (SRF
 339 transcription factors).

340 **0.3.4 SV and Introgression**

341 We identified 126,973 SV that overlapped ancestry blocks in admixed individuals, and
 342 found that 30,721 of these were associated with blocks of either *P. trichocarpa* or *P.*
 343 *balsamifera* ancestry in admixed individuals.

344 **0.3.5 Tables**

Assembly Statistics												
haplotype	Assembly size				Assembly size completeness				Assembly quality			
	length	Number of Scaffolds	N50	N50_n	gaps	N_count	Coverage Rate	busco_complete	busco_duplicated	busco_fragmented	Low Confidence Rate	Mean Assembly Quality Index
247	406044802	19	19100130	8	559	55900	0.9986	96.7	21.7	0.7	0.0000231000	90.450
247_alt	376222526	22	21729037	7	1470	147000	0.9970	92.7	16.8	0.9	0.0001421499	81.550
365	400343671	19	20009943	8	193	19300	0.9993	96.7	19.6	0.6	0.0000448000	93.800
365_alt	382988961	25	19752833	8	743	74300	0.9974	95.0	19.0	0.8	0.0001006478	86.900
368	408927241	19	20528059	8	479	47900	0.9984	96.6	20.6	0.7	0.0001896474	91.600
368_alt	376380057	26	19967418	8	1331	133100	0.9962	92.6	18.3	0.9	0.0001708698	82.850
406	406883189	20	21305652	8	934	93400	0.9986	97.6	19.4	0.6	0.0000080500	86.050
406_alt	343567265	23	17785992	8	5810	581000	0.9903	74.0	11.2	3.2	0.0001700919	58.550
439	399397382	19	19486512	8	566	56600	0.9985	96.7	19.4	0.7	0.0000117000	90.800
439_alt	356224784	21	19609853	8	2205	220500	0.9948	88.6	16.1	1.1	0.0000465000	76.050
515	395807708	19	21661901	7	1296	129600	0.9975	97.3	19.4	0.6	0.0000448000	81.700
515_alt	286914403	20	14090572	8	6632	663200	0.9881	60.1	8.0	3.5	0.0002017710	52.850
562	402113775	21	19407464	8	1535	153500	0.9977	97.2	19.0	0.6	0.0000486000	80.500
562_alt	294685596	22	15156518	8	7591	759100	0.9843	60.0	7.0	3.9	0.0001118243	53.050
566	404344743	20	20403993	8	1550	155000	0.9971	96.9	19.7	0.8	0.0000449000	80.200
566_alt	295750810	20	14501527	8	6773	677300	0.9873	62.3	8.3	4.1	0.0000791000	54.100
712	406603955	21	20302875	8	550	55000	0.9986	96.5	20.8	0.5	0.0000614000	91.350
712_alt	374451786	20	20382624	8	1553	155300	0.9961	92.4	18.2	0.9	0.0000495000	81.900
762	400616755	20	21889695	7	853	85300	0.9978	96.4	21.4	0.6	0.0003550476	60.400
762_alt	365455960	23	18068783	8	2445	244500	0.9946	90.5	17.8	1.2	0.0003571648	54.650
776	412523377	30	20099238	8	1192	119200	0.9974	97.4	20.3	0.5	0.0000362000	81.550
776_alt	407700287	24	20201693	8	4121	412100	0.9904	92.9	17.3	1.5	0.0000911000	65.400
801	410936669	21	19562430	8	323	32300	0.9988	96.2	20.3	0.6	0.0000448000	92.850
801_alt	381097682	22	19323978	8	1138	113800	0.9963	93.0	18.4	0.9	0.0001059650	84.350
809	408711020	22	20677021	7	796	79600	0.9981	96.9	20.9	0.9	0.0001127936	85.600
809_alt	365592262	24	18971862	8	2034	203400	0.9963	89.9	16.9	1.1	0.0000788000	73.950
822	405664496	21	19776368	8	740	74000	0.9981	96.7	20.5	0.7	0.0000317000	88.900
822_alt	347527765	21	18042639	8	2151	215100	0.9958	86.8	15.6	1.0	0.0001331520	74.800
860	416270270	20	23271044	7	941	94100	0.9970	97.1	21.0	0.8	0.0002180482	87.150
860_alt	377658316	22	19870789	8	2457	245700	0.9951	90.1	15.8	1.3	0.0001090960	73.150
969	399016342	20	20036743	8	1808	180800	0.9957	96.4	18.5	0.7	0.0000709000	56.750
969_alt	277863190	21	15795860	7	7380	738000	0.9841	58.9	7.5	3.9	0.0001591467	45.135
939	422799018	19	24674954	7	117	11700	0.9974	98.1	19.4	0.6	0.0000517000	95.950
939_alt	411479088	19	22188346	8	288	28800	0.9964	97.5	19.5	0.7	0.0002849355	94.200
nisqually_1	389204664	19	21678634	7	59	590000	NA	98.2	19.6	0.5	NA	NA

345 format: svg} {fig-

346 format: svg}

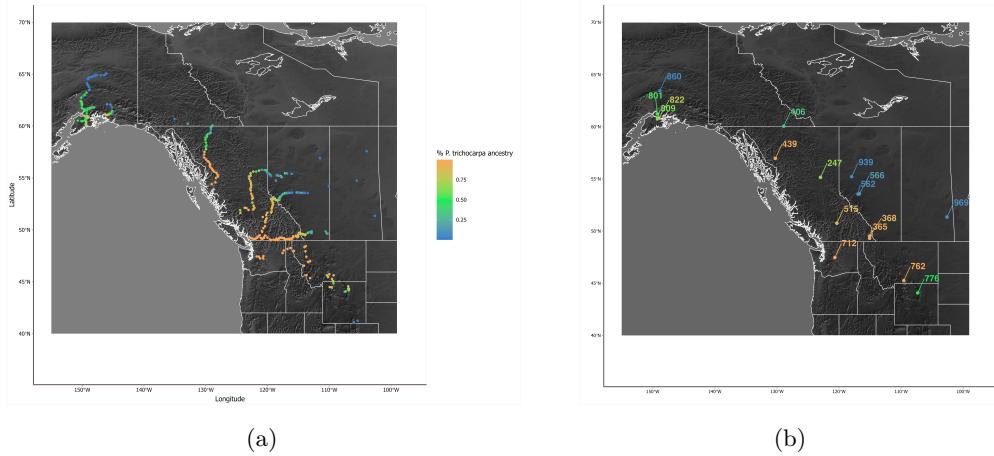


Figure 1: (a): sampling locations for 575 individuals used in this study, color indicates ancestry based in ADMIXTURE analysis (b): A subset of 16 individuals used for whole genome assembly

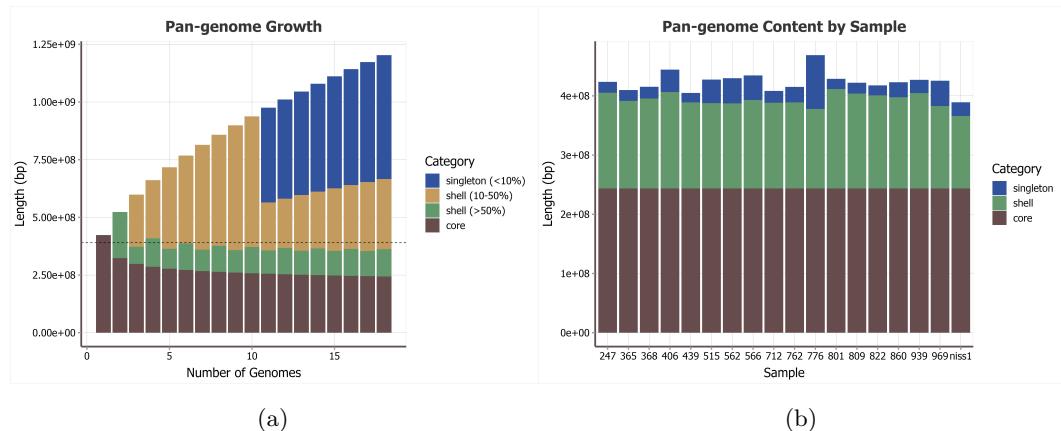


Figure 2: (a): Pan-genome growth curve visualizes the change in core and shell genome size as samples are added. (b): The relative length of the core, shell and singleton portions of the pan-genome for each sample represented

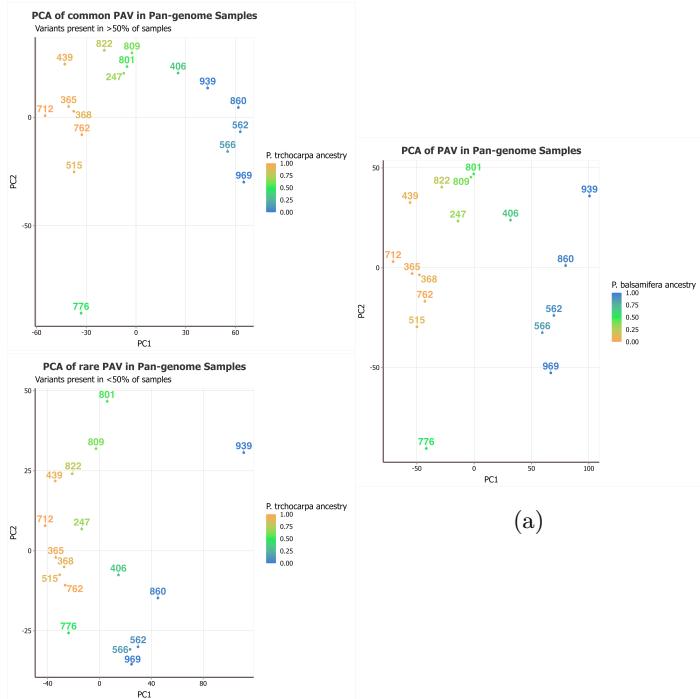


Figure 3: The first two principal components of a PCA on presence/absence variation in the pan-genome. Results of PCA on common (a), rare (b) and all (c) PAV are shown. Color indicates ancestry based in ADMIXTURE analysis.

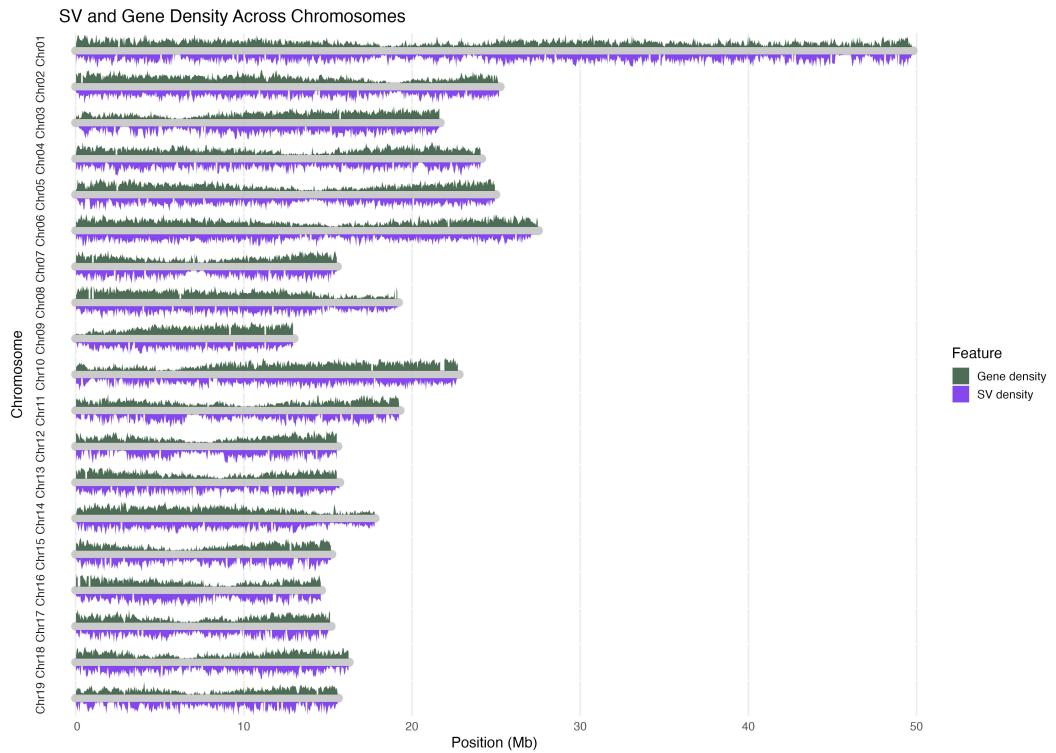


Figure 4: The density of SV larger than 20bp in length across the genome (purple) compared to the density of annotated genes (green)

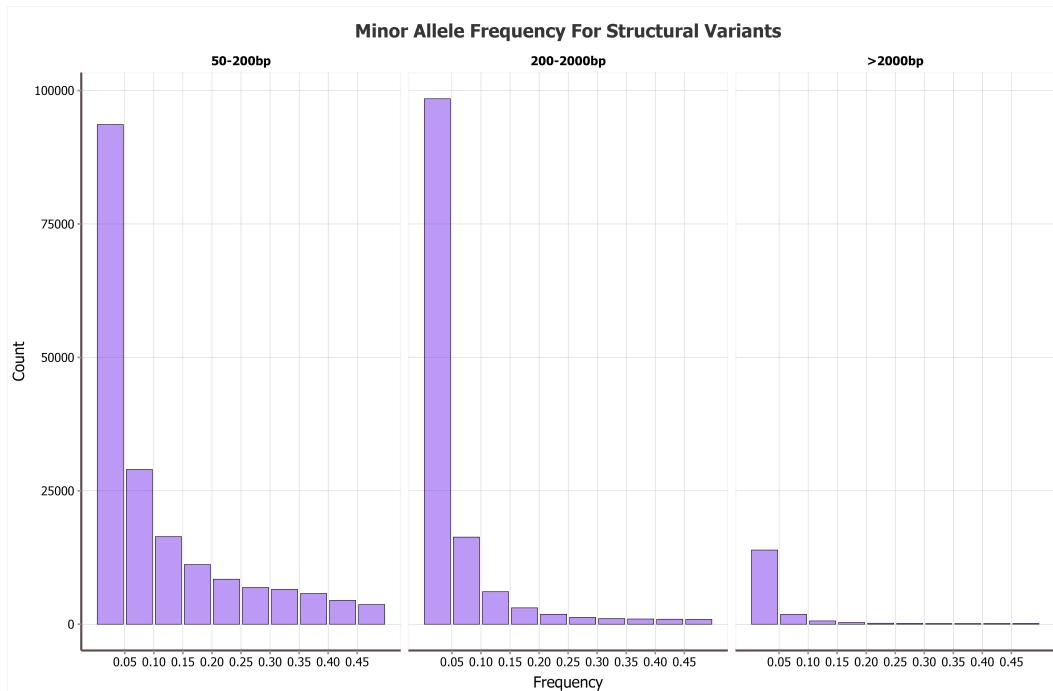


Figure 5: The frequency distribution of the minor allele for SV of different size classes

1 Figures References

- 347
- 348
- Almarri, M. A., Bergström, A., Prado-Martinez, J., Yang, F., Fu, B., Dunham, A.
349 S., et al. (2020). Population structure, stratification, and introgression of human
350 structural variation. *Cell*, 182(1), 189–199.
- Bock, D. G., Cai, Z., Elphinstone, C., Gonzalez-Segovia, E., Hirabayashi, K., Huang,
352 K., et al. (2023). Genomics of plant speciation. *Plant Communications*, 4(5).
- Burgarella, C., Barnaud, A., Kane, N. A., Jankowski, F., Scarcelli, N., Billot, C., et
354 al. (2019). Adaptive introgression: An untapped evolutionary mechanism for crop
355 adaptation. *Frontiers in Plant Science*, 10, 4.
- Chang, C. C., Chow, C. C., Tellier, L. C., Vattikuti, S., Purcell, S. M., & Lee, J. J.
357 (2015). Second-generation PLINK: Rising to the challenge of larger and richer
358 datasets. *Gigascience*, 4(1), s13742–015.
- Cheng, H., Liu, J., Wen, J., Nie, X., Xu, L., Chen, N., et al. (2019). Frequent intra-
360 and inter-species introgression shapes the landscape of genetic variation in bread
361 wheat. *Genome Biology*, 20, 1–16.
- Chhatre, V. E., Evans, L. M., DiFazio, S. P., & Keller, S. R. (2018). Adaptive intro-
363 gression and maintenance of a trispecies hybrid complex in range-edge populations
364 of populus. *Molecular Ecology*, 27(23), 4820–4838.
- Danecek, P., Bonfield, J. K., Liddle, J., Marshall, J., Ohan, V., Pollard, M. O.,
366 et al. (2021). Twelve years of SAMtools and BCFtools. *GigaScience*, 10(2).
367 <https://doi.org/10.1093/gigascience/giab008>
- Dias-Alves, T., Mairal, J., & Blum, M. G. (2018). Loter: A software package to infer
369 local ancestry for a wide range of species. *Molecular Biology and Evolution*, 35(9),
370 2318–2326.
- 371

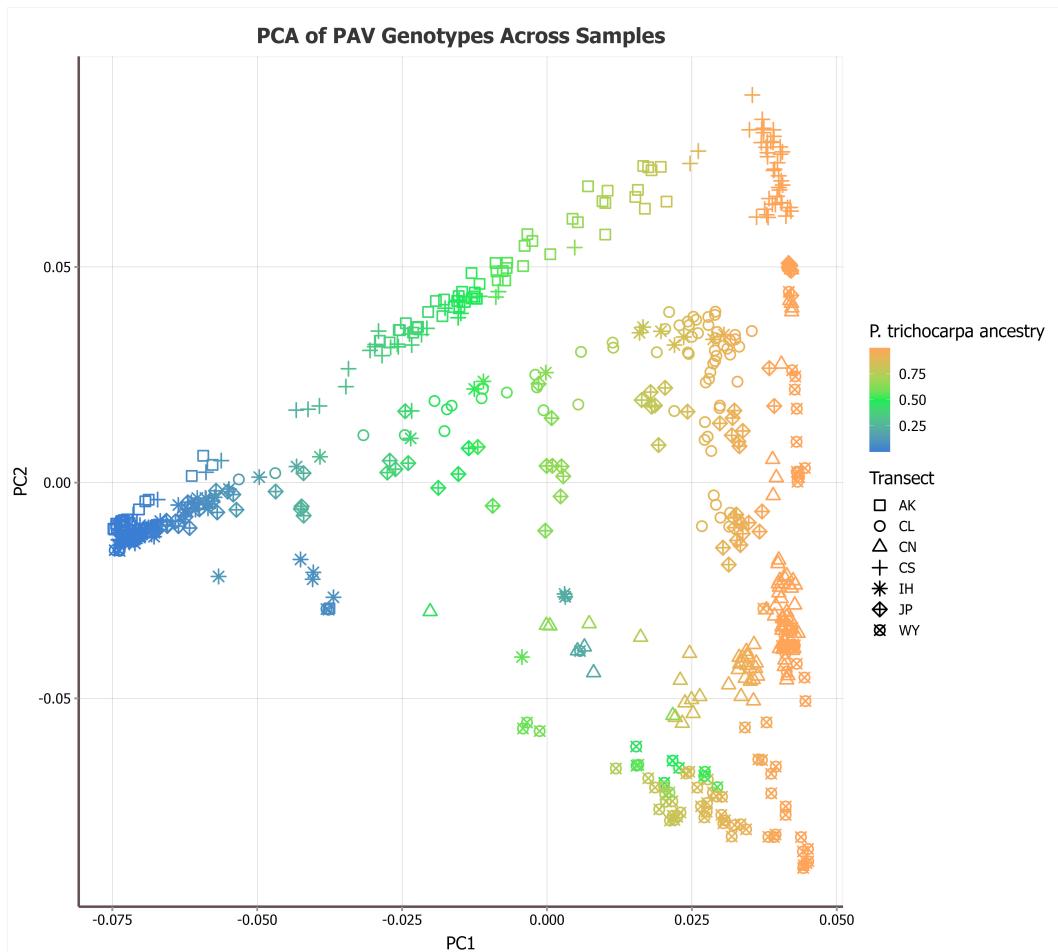


Figure 6: The first two principal components of a PCA of PAV genotyped from short read alignments to the pan-genome graph. Color indicates ancestry based on ADMIXTURE analysis. Shape indicates which of the latitudinal transects the individual was sampled from.

- 372 Fang, B., & Edwards, S. V. (2024). Fitness consequences of structural variation
 373 inferred from a house finch pangenome. *Proceedings of the National Academy of Sciences*, 121(47), e2409943121.
- 375 Feng, J., Dan, X., Cui, Y., Gong, Y., Peng, M., Sang, Y., et al. (2024). Integrating
 376 evolutionary genomics of forest trees to inform future tree breeding amid rapid
 377 climate change. *Plant Communications*, 5(10).
- 378 Ferguson, S., Jones, A., Murray, K., Andrew, R. L., Schwessinger, B., Bothwell, H., &
 379 Borevitz, J. (2024). Exploring the role of polymorphic interspecies structural variants
 380 in reproductive isolation and adaptive divergence in eucalyptus. *GigaScience*,
 381 13, giae029.
- 382 Fetter, K. C., & Keller, S. R. (2023). Admixture mapping and selection scans identify
 383 genomic regions associated with stomatal patterning and disease resistance in
 384 hybrid poplars. *Ecology and Evolution*, 13(10), e10579.
- 385 Gaczorek, T., Dudek, K., Fritz, U., Bahri-Sfar, L., Baird, S., Bonhomme, F., et al.
 386 (2024). Widespread adaptive introgression of major histocompatibility complex
 387 genes across vertebrate hybrid zones. *Molecular Biology and Evolution*, 41(10),
 388 msae201.
- 389 Gao, L., Gonda, I., Sun, H., Ma, Q., Bao, K., Tieman, D. M., et al. (2019). The
 390 tomato pan-genome uncovers new genes and a rare allele regulating fruit flavor.
 391 *Nature Genetics*, 51(6), 1044–1051.
- 392 Garrison, E., Sirén, J., Novak, A. M., Hickey, G., Eizenga, J. M., Dawson, E. T.,
 393 et al. (2018). Variation graph toolkit improves read mapping by representing
 394 genetic variation in the reference. *Nature Biotechnology*, 36(9), 875–879.
 395 <https://doi.org/10.1038/nbt.4227>
- 396 Geraldes, A., Farzaneh, N., Grassa, C. J., McKown, A. D., Guy, R. D., Mansfield,
 397 S. D., et al. (2014). Landscape genomics of populus trichocarpa: The role of
 398 hybridization, limited gene flow, and natural selection in shaping patterns of
 399 population structure. *Evolution*, 68(11), 3260–3280.
- 400 Gurevich, A., Saveliev, V., Vyahhi, N., & Tesler, G. (2013). QUAST: Quality assess-
 401 ment tool for genome assemblies. *Bioinformatics*, 29(8), 1072–1075.
- 402 Hämälä, T., Wafula, E. K., Guiltinan, M. J., Ralph, P. E., Depamphilis, C. W., &
 403 Tiffin, P. (2021). Genomic structural variants constrain and facilitate adaptation
 404 in natural populations of theobroma cacao, the chocolate tree. *Proceedings of the National Academy of Sciences*, 118(35), e2102914118.
- 405 Hamilton, J. A., & Miller, J. M. (2016). Adaptive introgression as a resource for
 406 management and genetic conservation in a changing climate. *Conservation Biology*,
 407 30(1), 33–41.
- 408 Hedrick, P. W. (2013). Adaptive introgression in animals: Examples and compar-
 409 ison to new mutation and standing variation as sources of adaptive variation.
 410 *Molecular Ecology*, 22(18), 4606–4618.
- 411 Hickey, G., Monlong, J., Ebler, J., Novak, A. M., Eizenga, J. M., Gao, Y., et al.
 412 (2023). Pangenome graph construction from genome alignments with Minigraph-
 413 Cactus. *Nature Biotechnology*, 42(4), 663–673. <https://doi.org/10.1038/s41587-023-01793-w>
- 414 Kang, M., Wu, H., Liu, H., Liu, W., Zhu, M., Han, Y., et al. (2023). The pan-genome
 415 and local adaptation of arabidopsis thaliana. *Nature Communications*, 14(1), 6259.
- 416 Kou, Y., Liao, Y., Toivainen, T., Lv, Y., Tian, X., Emerson, J., et al. (2020). Evolu-
 417 tionary genomics of structural variation in asian rice (*oryza sativa*) domestication.
 418 *Molecular Biology and Evolution*, 37(12), 3507–3524.
- 419 Kremer, A., & Hipp, A. L. (2020). Oaks: An evolutionary success story. *New Phytolo-
 420 gist*, 226(4), 987–1011.
- 421 Leroy, T., Louvet, J.-M., Lalanne, C., Le Provost, G., Labadie, K., Aury, J.-M., et
 422 al. (2020). Adaptive introgression as a driver of local adaptation to climate in
 423 european white oaks. *New Phytologist*, 226(4), 1171–1182.

- 426 Lexer, C., Fay, M., Joseph, J., Nica, M.-S., & Heinze, B. (2005). Barrier to gene
 427 flow between two ecologically divergent *populus* species, p. Alba (white poplar)
 428 and p. Tremula (european aspen): The role of ecology and life history in gene
 429 introgression. *Molecular Ecology*, 14(4), 1045–1057.
- 430 Li, H. (2018). Minimap2: Pairwise alignment for nucleotide sequences. *Bioinformatics*,
 431 34(18), 3094–3100.
- 432 Li, H., & Durbin, R. (2010). Fast and accurate long-read alignment with burrows–
 433 wheeler transform. *Bioinformatics*, 26(5), 589–595.
- 434 Li, K., Xu, P., Wang, J., Yi, X., & Jiao, Y. (2023). Identification of errors in draft
 435 genome assemblies at single-nucleotide resolution for quality assessment and
 436 improvement. *Nature Communications*, 14(1), 6556.
- 437 Li, Y., Yao, J., Sang, H., Wang, Q., Su, L., Zhao, X., et al. (2024). Pan-genome analy-
 438 sis highlights the role of structural variation in the evolution and environmental
 439 adaptation of asian honeybees. *Molecular Ecology Resources*, 24(2), e13905.
- 440 Li, Z., Liu, X., Wang, C., Li, Z., Jiang, B., Zhang, R., et al. (2023). The pig
 441 pangenome provides insights into the roles of coding structural variations in
 442 genetic diversity and adaptation. *Genome Research*, 33(10), 1833–1847.
- 443 Liang, Y.-Y., Liu, H., Lin, Q.-Q., Shi, Y., Zhou, B.-F., Wang, J.-S., et al. (2025).
 444 Pan-genome analysis reveals local adaptation to climate driven by intro-
 445 gression in oak species. *Molecular Biology and Evolution*, 42(5), msaf088.
<https://doi.org/10.1093/molbev/msaf088>
- 446 Lu, J., Rincon, N., Wood, D. E., Breitwieser, F. P., Pockrandt, C., Langmead, B.,
 447 et al. (2022). Metagenome analysis using the Kraken software suite. *Nature
 448 Protocols*, 17(12), 2815–2839. <https://doi.org/10.1038/s41596-022-00738-y>
- 450 Meirmans, P., Godbout, J., Lamothe, M., Thompson, S., & Isabel, N. (2017). History
 451 rather than hybridization determines population structure and adaptation in
 452 *populus balsamifera*. *Journal of Evolutionary Biology*, 30(11), 2044–2058.
- 453 Nurk, S., Walenz, B. P., Rhie, A., Vollger, M. R., Logsdon, G. A., Grothe, R., et al.
 454 (2020). HiCanu: Accurate assembly of segmental duplications, satellites, and allelic
 455 variants from high-fidelity long reads. *Genome Research*, 30(9), 1291–1305.
- 456 Parmigiani, L., Garrison, E., Stoye, J., Marschall, T., & Doerr, D. (2024). Pana-
 457 cus: fast and exact pangenome growth and core size estimation. *Bioinformatics*.
<https://doi.org/10.1093/bioinformatics/btae720>
- 459 Qiao, Q., Edger, P. P., Xue, L., Qiong, L., Lu, J., Zhang, Y., et al. (2021). Evolution-
 460 ary history and pan-genome dynamics of strawberry (*fragaria* spp.). *Proceedings of
 461 the National Academy of Sciences*, 118(45), e2105431118.
- 462 Quinlan, A. R., & Hall, I. M. (2010). BEDTools: A flexible suite of utilities for
 463 comparing genomic features. *Bioinformatics*, 26(6), 841–842.
- 464 R Core Team. (2024). *R: A language and environment for statistical computing*.
 465 Vienna, Austria: R Foundation for Statistical Computing. Retrieved from
<https://www.R-project.org/>
- 467 Racimo, F., Sankararaman, S., Nielsen, R., & Huerta-Sánchez, E. (2015). Evidence
 468 for archaic adaptive introgression in humans. *Nature Reviews Genetics*, 16(6),
 469 359–371.
- 470 Rautiainen, M., & Marschall, T. (2020). GraphAligner: rapid and versatile sequence-
 471 to-graph alignment. *Genome Biology*, 21(1). [https://doi.org/10.1186/
 472 s13059-020-02157-2](https://doi.org/10.1186/s13059-020-02157-2)
- 473 Rendón-Anaya, M., Wilson, J., Sveinsson, S., Fedorkov, A., Cottrell, J., Bailey, M.
 474 E., et al. (2021). Adaptive introgression facilitates adaptation to high latitudes
 475 in european aspen (*populus tremula l.*). *Molecular Biology and Evolution*, 38(11),
 476 5034–5050.
- 477 Savolainen, O., Pyhäjärvi, T., & Knürr, T. (2007). Gene flow and local adaptation in
 478 trees. *Annu. Rev. Ecol. Evol. Syst.*, 38(1), 595–619.

- 479 Secomandi, S., Gallo, G. R., Rossi, R., Rodríguez Fernandes, C., Jarvis, E. D.,
 480 Bonisoli-Alquati, A., et al. (2025). Pangenome graphs and their applications in
 481 biodiversity genomics. *Nature Genetics*, 1–14.
- 482 Simão, F. A., Waterhouse, R. M., Ioannidis, P., Kriventseva, E. V., & Zdobnov, E. M.
 483 (2015). BUSCO: Assessing genome assembly and annotation completeness with
 484 single-copy orthologs. *Bioinformatics*, 31(19), 3210–3212.
- 485 Sirén, J., Monlong, J., Chang, X., Novak, A. M., Eizenga, J. M., Markello, C.,
 486 et al. (2021). Pangenomics enables genotyping of known structural variants
 487 in 5202 diverse genomes. *Science*, 374(6574). <https://doi.org/10.1126/science.abg8871>
- 488 Song, B., Ning, W., Wei, D., Jiang, M., Zhu, K., Wang, X., et al. (2023). Plant
 489 genome resequencing and population genomics: Current status and future
 490 prospects. *Molecular Plant*, 16(8), 1252–1268.
- 491 Songsomboon, K., Brenton, Z., Heuser, J., Kresovich, S., Shakoor, N., Mockler, T.,
 492 & Cooper, E. A. (2021). Genomic patterns of structural variation among diverse
 493 genotypes of sorghum bicolor and a potential role for deletions in local adaptation.
 494 *G3*, 11(7), jkab154.
- 495 Suarez-Gonzalez, A., Hefer, C. A., Christe, C., Corea, O., Lexer, C., Cronk, Q. C., &
 496 Douglas, C. J. (2016). Genomic and functional approaches reveal a case of adap-
 497 tive introgression from *populus balsamifera* (balsam poplar) in *p. átrichocarpa*
 498 (black cottonwood). *Molecular Ecology*, 25(11), 2427–2442.
- 499 Suarez-Gonzalez, A., Lexer, C., & Cronk, Q. C. (2018). Adaptive introgression: A
 500 plant perspective. *Biology Letters*, 14(3), 20170688.
- 501 Suarez-Gonzalez, A., Hefer, C. A., Lexer, C., Douglas, C. J., & Cronk, Q. C. (2018).
 502 Introgression from *populus balsamifera* underlies adaptively significant variation
 503 and range boundaries in *p. trichocarpa*. *New Phytologist*, 217(1), 416–427.
- 504 Sun, X., Jiao, C., Schwaninger, H., Chao, C. T., Ma, Y., Duan, N., et al. (2020).
 505 Phased diploid genome assemblies and pan-genomes provide insights into the
 506 genetic history of apple domestication. *Nature Genetics*, 52(12), 1423–1432.
- 507 Sundell, D., Mannapperuma, C., Netotea, S., Delhomme, N., Lin, Y.-C., Sjodin, A., et
 508 al. (2015). The plant genome integrative explorer resource: PlantGenIE. org.
- 509 Thompson, S. L., Lamothe, M., Meirmans, P. G., Perinet, P., & Isabel, N. (2010).
 510 Repeated unidirectional introgression towards *populus balsamifera* in contact zones
 511 of exotic and native poplars. *Molecular Ecology*, 19(1), 132–145.
- 512 Xia, X., Zhang, F., Li, S., Luo, X., Peng, L., Dong, Z., et al. (2023). Structural
 513 variation and introgression from wild populations in east asian cattle genomes
 514 confer adaptation to local environment. *Genome Biology*, 24(1), 211.
- 515 Zanini, S. F., Bayer, P. E., Wells, R., Snowdon, R. J., Batley, J., Varshney, R. K., et
 516 al. (2022). Pangenomics in crop improvement—from coding structural variations
 517 to finding regulatory variants with pangenome graphs. *The Plant Genome*, 15(1),
 518 e20177.
- 519 Zhang, L., Reifová, R., Halenková, Z., & Gompert, Z. (2021). How important are
 520 structural variants for speciation? *Genes*, 12(7), 1084.
- 521 Zhang, X., Liu, T., Wang, J., Wang, P., Qiu, Y., Zhao, W., et al. (2021). Pan-genome
 522 of raphanus highlights genetic variation and introgression among domesticated,
 523 wild, and weedy radishes. *Molecular Plant*, 14(12), 2032–2055.
- 524