Natural variation in chalcone isomerase defines a major locus controlling radial stem growth variation ~~between~~among *Populus nigra* populations

Duruflé Harold*[1], Déjardin Annabelle[1], Jorge Véronique, Pégard Marie[2], Pilate Gilles[1], Rogier Odile1, Sanchez Leopoldo[1], Segura Vincent*[1,3]

[1] INRAE, ONF, BioForA, UMR 0588, 45075 Orléans, France
[2] INRAE, URP3F, UR 0004, 86600 Lusignan, France
[2] UMR AGAP Institut, Univ Montpellier, CIRAD, INRAE, Institut Agro, F-34398 Montpellier, France

*Corresponding authors
Correspondence: harold.durufle@inrae.fr; vincent.segura@inrae.fr

ABSTRACT

Poplar is a promising resource for wood production and the development of lignocellulosic biomass, but currently available varieties have not been optimized for these purposes. Therefore, it is critical to investigate the genetic variability and mechanisms underlying traits that affect biomass yield. Previous studies have shown that target traits in different poplar species are complex, with a small number of genetic factors having relatively low effects compared to medium to high heritability. In this study, a systems biology approach was implemented, combining genomic, transcriptomic, and phenotypic information from a large collection of individuals from natural populations of black poplar from Western Europe. Such an approach identified a QTL and a gene, coding for chalcone isomerase (CHI), as a candidate for controlling radial growth. Additionally, analysis of the structure and diversity of traits as well as CHI gene expression revealed a high allelic fixation index, linked to the geographical origin of the natural populations under study. These findings provide insights into how adaptive traits arise, are selected, and maintained in the populations. Overall, this study contributes to enhancing the use of poplar as a valuable resource for sustainable biomass production.

Keywords: Poplar, GWAS, transcriptomics, systems biology, adaptation

Introduction

Trees play a crucial role in mitigating climate change by sequestering carbon from the atmosphere through photosynthesis, and forest ecosystems are considered the largest terrestrial carbon sinks on Earth (Pan et al., 2011; Harris et al., 2021). The future evolution of carbon sequestration in forests relies heavily on how the growth rate and lifespan of trees respond to the changing climate (Brienen et al., 2020; Zhou, 2022). Trees keep accumulating carbon in their trunks, branches, and roots as they grow, which enables them to capture and store atmospheric carbon for several decades or possibly centuries (Green & Keenan, 2022). The major part of the tree trunk is created by the cambium, and the developing xylem constitutes a complex and dynamic system that generates wood in accordance with the seasonal cycle (Rathgeber et al., 2016). However, we still lack an integrative theory to understand growth patterns because wood formation requires the coordination of many metabolic pathways (Bryant et al., 2023).

Knowing and understanding the links between phenotypes and genetic mutations is a major challenge. Such studies have emerged for poplar, a model for tree biology, genomics, evolutionary and ecological genetics (Jansson & Douglas, 2007; Douglas, 2017). Furthermore, cultivated poplars have commercial value for peeling and veneer, lumber, paper pulp and are also used as bioenergy feedstock due to their high biomass production and favourable cell wall chemistry (Porth & El-Kassaby, 2015; Taylor et al., 2016; An et al., 2021; Abreu et al., 2022). *Populus nigra* is a deciduous tree species native to Europe, Asia and North Africa that occupies riparian ecosystems with diverse climate ranges (De Rigo et al., 2016). The genetic structure of this species in its natural distribution area is not extensively known. Yet, some studies have shown high genetic diversity within populations and low but significant genetic differentiation between river basins, suggesting high levels of gene flow in Western parts of the distribution (Smulders et al., 2008; Dewoody et al., 2015; Wójkiewicz et al., 2021). Seven ancestral genetic clusters were found in the first genome-wide genotyping study of 838 native individuals from 12 Western European river basins (Faivre-Rampant et al., 2016). However, another study of seven species showed that black poplar is highly structured with low diversity within populations (Milesi et al., 2024). These results may be due to the fact that the ecology of the species is strongly influenced by a very dynamic environment, the alluvial banks where it breeds, resulting in a complex structure (Gurnell & Petts, 2006; Alimpić et al., 2022).

Black poplar also shows a wide phenotypic diversity which can be observed on latitudinal clines such as that observed for leaf functional traits in response to drought (Viger et al., 2016) or on leaf morphology and structure (Guet et al., 2015b). Among the observable phenotypes, growth traits and wood production are considered fundamental for the adaptation and productivity of planted forests (Grattapaglia et al., 2009). For example, the biosynthetic pathway of lignin, an essential component of wood, is known to affect abiotic tolerance and growth in *Populus* (Xie et al., 2018). However, few genetic studies have been carried out on traits related to growth, and even fewer at the genomic level using the natural intraspecific diversity of trees. Genetic differentiation between natural populations of *P. trichocarpa* was found for growth and phenology, which was higher than the rather weak differentiation observed at the genome level (Evans et al., 2014; Oubida et al., 2015). This suggests that local adaptation explains patterns of variation in these traits better than genetic drift alone. The adaptive traits of poplar populations show variations depending on the local climate at their geographic origin. Using genome-wide association studies (GWAS) on provenances of *P. trichocarpa*, candidate loci underlying bud phenology and biomass have already been identified (Evans et al., 2014; Zhang et al., 2019). Based on 113 natural *P. tremula* genotypes from Sweden, a study showed significant natural variation in growth and wood-related traits and allowed the identification of genetic markers associated with these traits (Escamez et al., 2023). In this context, the OGDH enzyme (2-oxoglutarate dehydrogenase) was found to be associated with variation in tree volume and constitutes an interesting potential candidate for improving stem volume. Within the same collection of Aspen trees, a major and unique locus was also discovered. It determines the timing of bud formation and facilitates adaptation to different growing seasons and colder climates (Wang et al., 2018). A systems genetics approach

in a subset of the same collection linked natural variation in lignin content and composition to responses to mechanical stimuli and nutrient availability (Luomaranta et al., 2024). Furthermore, QTLs were identified for stem and biomass traits in several mapping populations involving as parental species those typically used to generate cultivated hybrids (*P. deltoides*, *P. nigra* and *P. trichocarpa*). Of note, these studies reported several QTL hotspots for biomass accumulation in different environments (Rae et al., 2008, 2009; Dillen et al., 2009; Monclus et al., 2012).

Although QTL mapping studies in segregating progenies have reported QTL hotspots that explain a large part of genetic variation for growth, the QTL resolution was too limited to identify the underlying candidate genes (Rae et al., 2009). On the other hand, GWAS can make use of the rapid decay of linkage disequilibrium in forest trees (Neale & Kremer, 2011), but most studies carried out so far for growth traits have reported a limited number of loci that individually do not explain a large proportion of the genetic variance of this heritable trait (Mckown et al., 2014; Allwright et al., 2016). Many studies suggest that complex traits are controlled by multiple loci, each with rather small effects (Bradshaw & Stettler, 1995; Grattapaglia et al., 1996; Rae et al., 2007; Wade et al., 2022). To go further in understanding phenotypes and adaptation, the genomics toolbox and statistical methods as systems biology approach made available for research are constantly evolving (Pazhamala et al., 2021). The revolution comes in particular from the applications that “omics” technologies have made possible for plants such as forest trees (Plomion et al., 2016; Borthakur et al., 2022). Thus, with the progression of methodologies and the reduction in the costs of these approaches, a certain number of studies have examined at large scale of endophenotypes like transcriptomic (Chateigner et al., 2020), proteomic (Plomion et al., 2006; Castillejo et al., 2023; Teyssier et al., 2023) or even metabolomic (Rodrigues et al., 2021) in trees. Another study has advocated the use of RNAseq to jointly identify polymorphisms and quantify the transcriptomic variability across natural populations (De Wit et al., 2015). Such an approach could contribute to filling the gap between the genome and phenotypic variation for complex traits and further contribute to the explanation of their missing heritability (Maher, 2008; Chandler et al., 2014).

Here, we report an integrative approach encompassing population genetics and genomics together with transcriptomics to decipher the genetic architecture of secondary growth in ~~GWAS for growth using phenotypic data from~~ natural populations of *P. nigra* ~~evaluated in two common garden experiments together with SNP data from RNAseq (Chateigner et al., 2020; Rogier et al., 2023)~~. By~~We further~~ making use of multi-omic information,~~the transcriptomic data to~~ we dissected a major QTL for stem radial growth identified by GWAS and pinpointed a candidate gene from the flavonoid pathway. Finally, we studied the genomic and transcriptomic diversity of the candidate gene and the phenotypic diversity across the natural populations and could show that the ~~polymorphism~~QTL is involved in growth differentiation, suggesting an implication in local adaptation.

Material and methods

Plant material and field experiments.

The complete plant material and field management was previously described (Guet et al., 2015a; Gebreselassie et al., 2017). Briefly, an initial experimental design based on a total of 1,160 genotypes of *P. nigra*, representative of the species range in Western Europe, was established in two contrasting common gardens located at Orléans (France, ORL, 47°50′N 01°54′E) and Savigliano (Italy, SAV, 44°36′N 07°37′E) in 2008. In both sites, the genotypes were replicated 6 times in a randomized complete block design. A previous study, using a 12k Infinium array (Faivre-Rampant et al., 2016) was used to characterize the genetic diversity within this collection. A subset of 241 genotypes representative of the natural diversity and originating from 10 river basins was selected. Briefly, a population structure analysis on the entire collection with 5,600 SNPs and the model-based ancestry estimation in the ADMIXTURE software (Alexander et al., 2009) highlighted some introgression from the cultivated compartment (Lombardy poplar, ‘Italica’). The 241 genotypes of the present study were selected to minimize such introgression by

setting a threshold of maximum 15% of the ancestral genetic group corresponding to the cultivated genotype.

Climate data.

Climatic variations across the locations of origin of the populations were analysed by applying principal component analysis on 19 annual bioclimatic variables obtained from the WorldClim dataset (~~Hijmans et al., 2005~~Fick & Hijmans, 2017). The values used are the 30-year average (19~~67~~0 to ~~199~~2000) with a resolution of 1 km² per grid cell obtained from the GPS location of the original natural populations. The first two principal components (PC1, 42% of explained variance, and PC2, 22% of explained variance) corresponded to the weighted precipitation and temperature variables, respectively.

Phenotyping.

We have described in detail the phenotyping of 21 traits in previous works (Chateigner et al., 2020; Wade et al., 2022). Only the circumference and the basic density of the wood ~~(Infraden)~~ were used in this study. Briefly, trees were pruned at the base after one (SAV) or two years of growth (ORL), to remove a potential cutting effect. Circumference refers to the perimeter of the stem measured at 1-m above the ground with a measuring tape. Measurements were ~~measured~~ carried out on 2-year-old trees in winter 2010-2011 at SAV and in winter 2011-2012 at ORL. Basic density was determined as previously ~~reported as~~ described in (Chateigner et al., 2020). Briefly, it was measured on a piece of wood from the stem section harvested for RNA sequencing (see hereafter) following the Technical Association of Pulp and Paper Industry (TAPPI) standard test method T 258 "Basic density and moisture content of pulpwood". For each site, the phenotypic data were analyzed with a linear mixed model to compute genotypic means adjusted for micro-environmental effects as described in (Gebreselassie et al., 2017). Before the adjustment of the model, a square root transformation was made to ensure the normality and homoscedasticity of the residuals. This transformation was only needed for circumference (not for wood basic density).

Transcriptomic data.

RNA sequencing was carried out on young differentiating xylem and cambium tissues collected in 2015 from two replicates of the 241 genotypes located in two blocks of the Orleans common garden, as described in (Chateigner et al., 2020). Sequencing reads were obtained to provide both transcriptomic and genomic data. Briefly, frozen milled tissue was used to isolate total RNA with RNeasy Plant kit (Qiagen, France), according to manufacturer's recommendations and a treatment with DNase I (Qiagen, France) was made. Samples of young differentiating xylem and cambium tissues of the same tree were pooled in an equimolar extract before sending it for the sequencing at the POPS platform with Illumina Hiseq2000. Reads were mapped to the *P. trichocarpa* v3.0 primary transcripts (available in Phytozome 13, Goodstein et al., 2012.) using bowtie2 v2.4.1 (Langmead & Salzberg, 2012) and only transcripts with at least 1 count in 10% of the samples were kept, yielding 34,229 features. The raw count data were normalized by Trimmed Mean of M-values using the R package edgeR v3.26.4, calculated in counts per millions (CPM) and computed in $log_2(n+1)$. At the end, the CPM were fitted with a linear mixed model including batch and genetic effects to extract their genotypic Best Linear Unbiased Predictors (BLUPs). These genotypic BLUPs of transcripts were used for the rest of our analysis.

Genotypic data.

The full details of genotypic analysis have been described in (Rogier et al., 2023), including software used, data filtering criteria and final SNP selection. Briefly, genotyping data were obtained, using BWA-MEM v0.7.12 to map the reads into the *P. trichocarpa* v3.0 reference genome (available in Phytozome 13, Goodstein et al., 2012.) and the SNPs were called using 3 callers to generate a high-confidence SNP set. The 3 callers were GATK 3.1 (Van der Auwera et al.,

2013), FreeBayes 0.9.20 (Garrison, 2012), and the mpileup command from SAMtools 1.3 (Li et al., 2009). Only the SNPs identified by at least 2 of the 3 callers and with less than 50% of missing values were selected. Missing values were imputed using the Fimpute v.2.2 program (Sargolzaei et al., 2014) and complementary genotyping data previously obtained with a 12 k Illumina Infinium Bead-Chip array (Faivre-Rampant et al., 2016). At the end, we obtained 878,957 SNPs and from these, 440,292 SNPs were retained for this study after filtering for a minimum allele frequency of 0.05.

Genetic analyses

Unless otherwise stated, all analyses have been carried out with R v4.4.1 (R Core Team, 2021) under the RStudio environment (RStudio Team, 2020).

Partition of variance

The following bivariate mixed model was fitted to partition the variance in circumference across the two sites into between- and within-population genetic variation and their interaction with site using the R package breedR v 0.12-5 (Muñoz & Sanchez, 2024):

$$y = \begin{pmatrix} y_1 \\ y_2 \end{pmatrix} = X\beta + Z_b b + Z_w w + \epsilon \quad (1)$$

Where y is a vector of genotypic adjusted means for circumference in ORL and SAV, X, Z_b and Z_w are design matrices relating observations to fixed and random effects, β is the fixed effect of site and b and w are between and within random genetic effects. b and w follow a multivariate normal distribution with mean 0 and variances: $\begin{bmatrix} \sigma^2_{b_1} & \sigma_{b_{12}} \\ \sigma_{b_{21}} & \sigma^2_{b_2} \end{bmatrix} \otimes K_b$ and $\begin{bmatrix} \sigma^2_{w_1} & \sigma_{w_{12}} \\ \sigma_{w_{21}} & \sigma^2_{w_2} \end{bmatrix} \otimes K_w$. K_b and K_w are genomic relationship matrices between and within populations. They were estimated from the full genomic relationship matrix computed with ldak software v5 (Speed et al., 2012), by averaging the kinships per population for K_b and setting the kinships at zero across populations for K_w. The estimated variance-covariance parameters were then used to compute the following variance components: between and within population genetics, and between and within population genetics times environment (Itoh & Yamada, 1990).

Population genetics

F_{ST} was estimated using Weir and Cockerham method (Weir & Cockerham, 1984) and implemented in plink (v1.90b6.3). Q_{ST} was estimated using variance parameters from the previously described mixed-model as: $Q_{ST} = \frac{\sigma^2_b}{\left(\sigma^2_b + 2\sigma^2_w\right)}$.

GWAS

GWAS was performed for circumference in each site with genotypic adjusted means and SNPs, using a linear mixed model as originally proposed by (Yu et al., 2005) and implemented in the R package MM4LMM (Laporte et al., 2022). This model included a random polygenic effect with a covariance structure defined by a genomic relationship matrix computed with the software ldak to account for linkage disequilibrium between SNPs (Speed et al., 2012). We also performed multi-locus GWAS using the multi-locus mixed-model (MLMM) approach implemented in the R package MLMM v0.1.1 (Segura et al., 2012), as well as multi-environment GWAS carried out with the MTMM approach implemented in R (Korte et al., 2012). SNPs were declared as significant according to a Bonferroni corrected threshold of 5%. Linkage disequilibrium between significantly associated SNPs was estimated in R as the squared allelic coefficient.

GWAS were also carried out using transcriptomic data (eQTL analysis) but focusing only on 2 genes of particular interest in this work, because they included significant SNPs in the GWAS.

The analyses were done using both single- and multi-locus approaches, as presented for circumference.

We also looked at associations between our candidate SNP, latitude of origin and climatic data at the population level using a Pearson correlation test.

To validate the findings of present study, ~~F~~further tests were carried with data previously published by (Pégard et al., 2020) on a multi-parental population of *P. nigra* (factorial mating design). This dataset consisted of 629 individuals with genotypic and circumference data. We retrieved 46 SNPs within the interval [chr10:20105000, chr10:20125000] corresponding to the region of interest in the present study, and carried out association tests between these SNPs and the phenotype using a simple linear model: $y = X\beta + \epsilon$.¶

Results

This study investigates variations in stem radial growth among natural populations of black poplar using an integrative approach with multi-omics data. Phenotypic evaluations were conducted in two common garden experiments in France and Italy (Guet et al., 2015a ; Gebreselassie et al., 2017), while genotypic characterization was achieved using SNP data from RNAseq (Rogier et al., 2023). This association between phenotypic and genotypic data identified a major locus, which included two gene models annotated as a protein of unknown function (PUF) and a chalcone isomerase (CHI), respectively. To get more insights into this association, transcriptomic data from the two gene models were integrated, together with secondary traits, such as wood basic density. Finally, data from another population were also analyzed to validate the findings.¶

A QTL controlling radial growth is highlighted by a genome-wide association study.

We performed a GWAS for circumference using 428,836 SNPs and detected a significant signal for this trait phenotyped in Savigliano (**Fig. 1a**, **Fig. S1**), with a total of 18 significant SNPs, including 11 on chromosome 10 in strong linkage disequilibrium. Closer examination of this region showed that the signal is distributed over two gene models: Potri.010G212900, annotated as ~~a Beta-Hexosaminidase 1 (Hexo1)~~PUF and Potri.010G213000, annotated as ~~a chalcone isomerase family protein (~~CHI~~)~~ (**Fig. 1b**). In the MLMM approach, the whole signal vanishes out after conditioning on the top SNP, suggesting that a single allele is associated with the trait in the region (**Fig. S~~1~~2**). This top SNP explained~~s more than 50~~ 16% of the phenotypic variation ~~(without accounting for population structure, **Fig. 1c**)~~. While non-significant at the genome-wide level when considering circumference at Orleans (p-value = 0.0025, **Fig. S~~2~~3**), this top SNP still explain~~sed more than 20~~4% of the phenotypic variation in this common garden and its effect ~~i~~was in the same direction as found in Savigliano (**Fig. 1c**). Consequently, a multi-trait GWAS combining phenotypes from the two common gardens confirmed this signal but detected only a total of 7 significant SNPs (**Fig. S~~3~~4**), mainly for the global effect (i.e., common to the two sites). These 7 SNPs identified in the multi-trait GWAS, are included in the 11 detected in single-trait GWAS at Savigliano and constitute our core set of candidate SNPs (**Tab. S1**). Among them, 6 are exonic (4 non-synonymous and 2 synonymous) and 1 is in 5'UTR, unsurprisingly as they come from RNAseq reads. In addition, they are all located on the CHI gene except one. It is worth mentioning that the top SNP is located in an exon of CHI gene and is predicted to be non-synonymous.

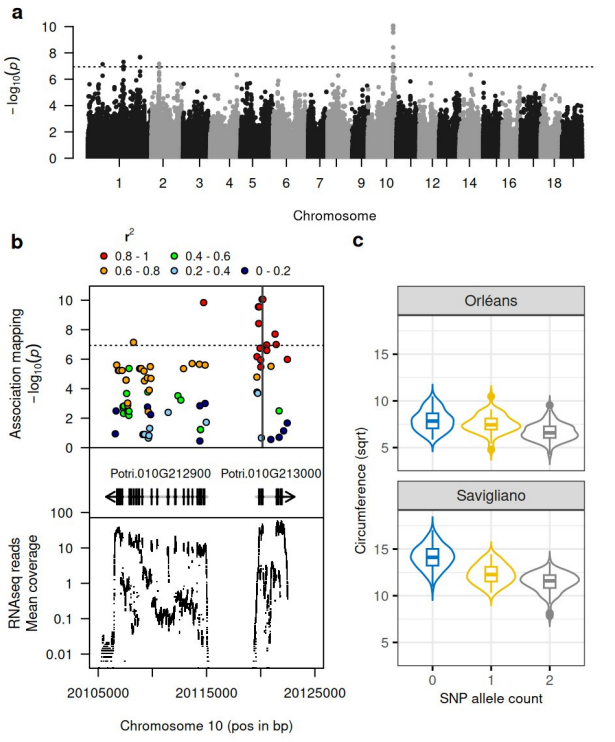

Figure 1 - GWAS of the circumference phenotype. a) Genome-wide Manhattan plot highlighting QTL on chromosome 10 performed using a single locus mixed model and 428,836 SNPs markers from natural *P. nigra* diversity phenotype at Savigliano; b) Manhattan plot~~s~~ focused on SNPs with lowest p-values (colored according to their LD with the top SNP, as estimated with the squared allelic correlation coefficient r2) ~~obtained~~ and concerning 2 gene ~~regions~~models, with corresponding mean coverage of RNAseq reads across individuals; c) Box plot of the circumference in both experimental sites (transformed with a square root, sqrt), depending on the ~~allele~~ count of the alternate allele of the candidate SNP with the lowest p-value (Chr10:20120195 with reference and alternate alleles A and T, respectively).

Gene expression sustains CHI as a candidate gene.

We used the RNAseq data, generated from the xylem and cambium tissues of poplars grown at Orléans as an endophenotype to test whether the expression of our candidate genes correlated with the phenotypes and could be linked to the effect of one of them. Negative correlations were found between gene expressions and phenotypes, and their magnitude was higher for CHI than for ~~Hexo1~~PUF (**Fig. 2a**, **Fig. 2b**), with rR^2 of -0.~~5~~73 (p-value < 2.2e-16) and -0.6~~3~~5 (p-value < 2.2e-16) for circumference evaluated in Savigliano and Orleans, respectively. When using the expression of both genes to jointly explain phenotypes, the correlation between CHI gene expression and circumference was maintained (rR^2 = -0.~~49~~70, p-value < 2.2e-16 at Savigliano and rR^2 = -0.~~2~~50, p-value = 3.41e-16 at Orleans) while it drastically dropped for ~~Hexo1~~PUF (rR^2 = 0.0~~0~~46, p-value = 0.372 at Savigliano and rR^2 = -0.~~0~~523, p-value = 4.53e-04 at Orleans). We also made use of transcriptomic data for CHI and ~~Hexo1~~PUF to perform an eQTL analysis, which highlighted a strong cis control for the 2 genes (**Fig. 2c**). The fact that these two genes are close from each other and in opposite directions on the genome, together with the existence of strong LD in the region (**Fig. 1b**), generates a positive correlation between their expressions (rR^2 = 0.~~19~~43, p-value = 7.1e-12). But, when focusing on the region of interest, we observed different patterns of eQTL signal between the 2 genes (**Fig. 2d**). Interestingly, the pattern of eQTL for CHI gene was similar to the one observed for circumference (**Fig. 2d**, **Fig. 1b**). Altogether, these results supported CHI as a candidate gene for the control of circumference variability.

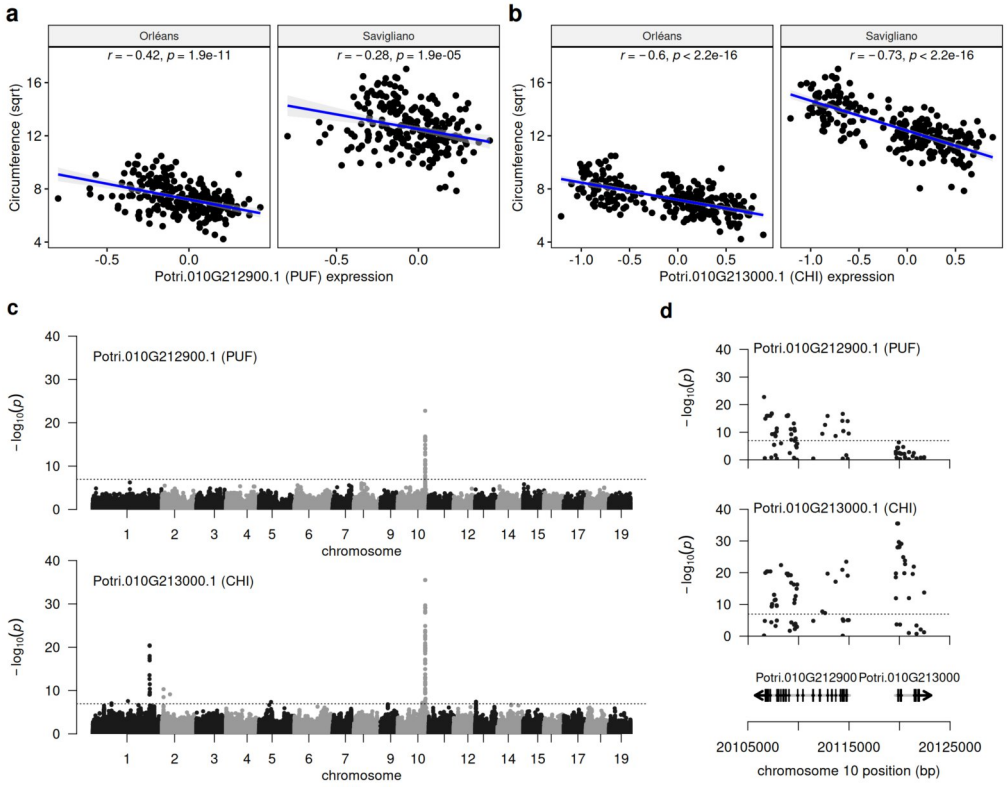

Figure 2 - eQTL analysis sustains CHI (Potri.01G213000) as a candidate gene for the control of circumference variation. a) correlation between the circumference and the expression level of ~~Hexo1~~PUF primary transcript (Potri.010G212900.1) ; b) correlation between the circumference and the expression level of CHI primary transcript (Potri.010G213000). c) Manhattan plot and d) focus on the candidate region of the eQTL analysis using the variations in the expression level of the 2 primary transcripts previously highlighted as phenotypes. Circumference was transformed with a square root (sqrt). The expression level of transcript~~sed~~ have been standardized with a ~~genetic analysis~~mixed linear model (see Material and methods).

Structure of the diversity of the CHI gene highlighted by population-scale analyses.

To further characterize the effect of the top SNP on the phenotypic variability, we partitioned the variance of circumference across locations into between-population and within-population genetic effects, their interaction with location, and a residual term (**Fig. 3**). This analysis showed that a large part of the phenotypic variation (35%) was due to genetic differences between populations, followed by interaction variance between genetics within populations and location (25%), genetic variance across populations (20%), and interaction variance between genetics across populations and location (17%). Interestingly, when the top SNP was included as a fixed effect in this variance partitioning model, it explained up to 24% of the total phenotypic variance, and this part of variability was mainly from the between population genetic component (**Fig. 3**, model 2). This analysis suggests that the QTL, previously identified by GWAS, is driven by differences in radial growth at the population level.

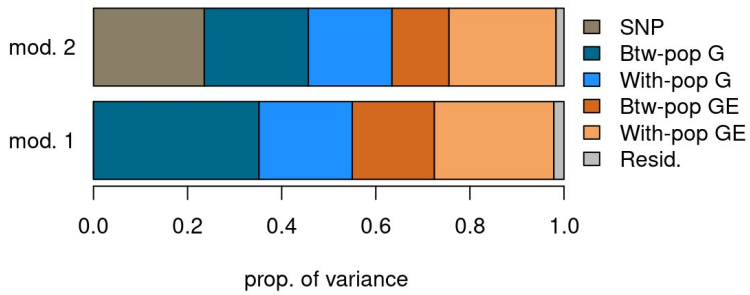

Figure 3 - Partition of phenotypic variance for circumference across two locations using two models: Model 1 (mod.1) refers to the model of variance partition without the top SNP (Chr10:20120195), while model 2 (mod. 2) is the model that includes the top SNP as a cofactor. Btw-pop and With-pop refer to between and within population variances, while G and GE refer to genetic and genetic by environment variance, respectively.

To confirm this observation, we computed the fixation index (F_{ST}) of the 428,836 SNPs and looked at the value of the top SNP detected by the GWAS. This SNP displayed a high F_{ST} value (0.69) well above the 99th percentile (0.28) of the genome-wide F_{ST} distribution (**Fig. 4a**). Such a high fixation index is due to a fixation of the reference allele in several populations mainly from the north-east of the studied area (NL, Kuhkopf, Rhin, Ticino), a fixation of the alternative allele in some population from central (Loire, Val d'Allier) and southern (Ramières) France and southern Italy (Basento), and a balanced situation in intermediate populations between these extremes (Dranse and Paglia) as well as in the population of south-western France (Adour) (**Fig. 4b**). Interestingly, such a genetic differentiation is also observed at the phenotypic level as well as at the transcriptomic level for CHI gene, as highlighted by high Q_{ST} values (**Fig. 4a**) and population differences (**Fig. 4c**). Consequently, associations between SNP and traits (**Fig. 5a**) or gene expression (**Fig. 5b**), as well as correlations between traits and gene expression (**Fig. 5c**), were high and significant when estimated at the population level, except for the trait evaluated at Orleans, which is consistent with the results obtained at the individual level (**Fig. S45**).

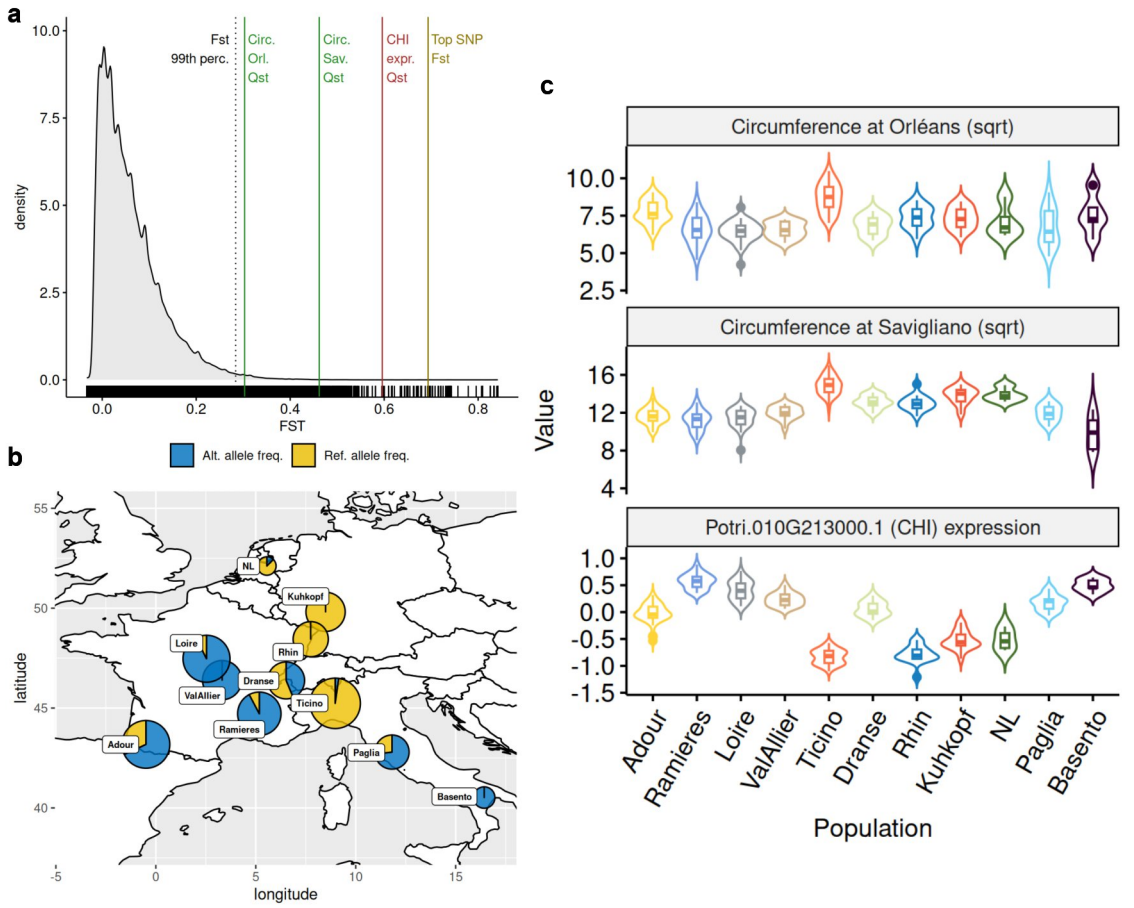

Figure 4 - Structure of the diversity. a) Distribution of genome-wide Fst together with specific values indicated by vertical lines: 99th percentile of the distribution, top SNP (Chr10:20120195) Fst, CHI (Potri.010G213000.1) expression Qst, circumference at Orléans and Savigliano Qsts. b) Geographical origin of populations together with the distribution of alleles within each population for the top SNP (Chr10:20120195), the size of the pie is proportional to the size of the population. c) Distribution of circumferences at Orléans and Savigliano, as well as CHI (Potri.010G213000.1) expression across populations.

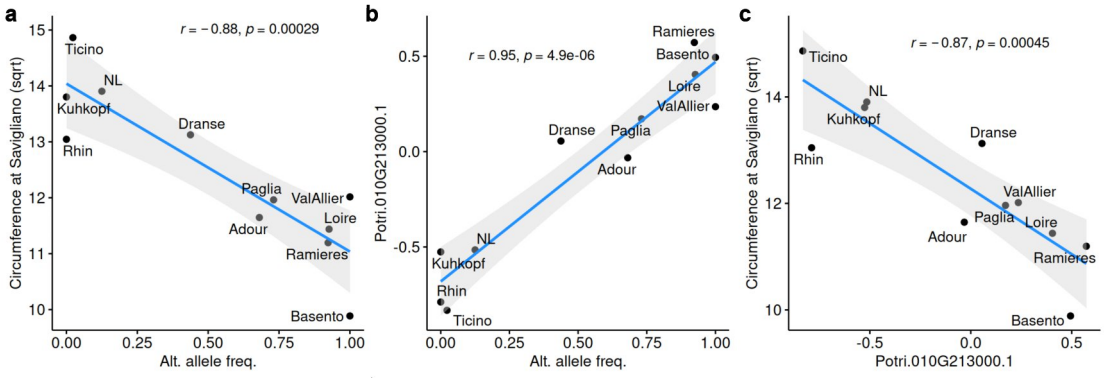

Figure 5 - Associations at the population scale. a) Correlation between circumference at Savigliano and alternate allele frequencies for the top SNP (Chr10:20120195); b) correlation between CHI (Potri.010G213000.1) expression and alternate allele frequencies for the top SNP (Chr10:20120195); c) correlation between circumference at Savigliano and CHI (Potri.010G213000.1) expression.

Validations and interpretations.

To support our findings, we complemented our study by several analyses. First, we looked at co-localizations between the QTL detected in the present study and QTLs previously reported in the literature. Of particular interest, we found in the same genomic region a QTL previously reported by Rae et al. (2009) for several traits related to biomass production in an interspecific poplar progeny, and named poplar biomass locus 3 (PBL3). PBL3 included several QTLs for height and diameter found across multiple years, and it was delimited by two SSRs (ORPM149 and PMGC2786). We retrieved the coordinates of these markers on the *P. trichocarpa* reference genome by blasting their priming sequences. The resulting interval in bp was [17566502, 21189318] (**Fig. S65**). It thus fully includes the QTL reported here which spans the interval [20105000, 20125000] (**Fig. 1b**). Second, we retrieved data from intraspecific crosses of *P. nigra* carried out within the French breeding program and previously used and reported by (Pégard et al., 2020) for genomic prediction. From the SNP set in this previous study, we identified 46 SNPs that fell within the interval and tested associations between each of these SNPs and the phenotype circumference in a panel of 629 individuals resulting from those crosses. The most significant association (p = 2.43e-05) was found for a SNP located at 20 119 788 bp (407 bp from the top SNP) (**Fig. S67**), which was also found significant in the present study with an effect in the same direction (alternative allele associated with an increase in circumference). Finally, to provide some biological interpretation to our findings, we retrieved data on wood basic density measured on samples collected at Orléans. The top SNP displayed a significant association with wood basic density, with a positive effect of the alternate allele, which was thus opposite to the effect found for circumference (**Fig. S8a**). Similarly, a significant positive correlation was found between wood basic density and CHI gene expression while such correlation was negative for circumference (**Fig.5a, Fig.5c, Fig. S8b**).

Discussion

We made use of ~~growth~~circumference data collected in two common garden experiments together with transcriptome-wide SNP data to search for genetic associations between genotype and phenotype in *P. nigra*. Such analysis pinpointed a small genomic region located at the distal end of chromosome 10 which encompassed 2 gene models, of which one was annotated as a chalcone isomerase (CHI). Transcriptomic data within one of the two common gardens further supported an implication of CHI in the phenotypic variation. Because the black poplar collection was structured into subpopulations corresponding to the geographic origins of the accessions, we further focused on differences between subpopulations and found that CHI diversity is a main driver of growth differences at the subpopulation scale. Such findings suggest an implication of this gene in local adaptation. Finally, we seek to validate our results with data from previous works and found that our significant loci match a previously reported QTL hotspot for biomass accumulation in an interspecific poplar family (Rae et al., 2009). We further validated the effect of the QTL in an independent panel with a *P. nigra* pedigree from the French breeding program (Pégard et al., 2020).

The strongest effect in the GWAS was found for the phenotypic data collected in the common garden (SAV) where the genetic variability for growth was the largest. This site enabled a better expression of the phenotypic variability for growth. Unfortunately, transcriptomic evaluation was carried out in the other common garden (ORL). Consequently, it is hard to conclude on the interplay between SNP variation and gene expression to explain the variation in growth. Indeed, if we run a mediation analysis, as proposed by Sasaki et al. (2018), using phenotypic data from SAV, we cannot conclude that the expression of CHI mediates the genetic association (data not shown). While if we repeat such analysis with phenotypic data from ORL we find that the association is mediated by CHI expression, although the association with growth at ORL is not significant genome-wide. Yet, the fact that gene expression data were collected from a different site than the one in which the GWAS is significant and on trees of different ages, underlines the robustness of the results.

Another complication with the loci detected originates from the confounding effect of population structure. Indeed, the phenotype, gene expression as well as polymorphisms display a significant variability across populations which drives the correlations and associations between them. Such a situation is not ideal for association mapping, and the confounding effect attributed to population structure has to be specifically handled by the statistical model applied, usually a linear mixed model with a random polygenic effect having as covariance the genomic relationship between individuals (Yu et al., 2005). In addition, we used the program ldak to account for LD between polymorphisms in the estimation of the relationships (Speed et al., 2012), required to be independent, because our SNPs come from RNAseq and are thus clustered by genes with potentially some strong LD between neighbouring SNPs. The correction applied within the linear mixed model with such a matrix appeared to be milder than the one achieved with a regular ~~GRM~~genomic relationship matrix such as the one estimated following (VanRaden, 2008), resulting in a significant signal. Please also note that we did not consider including a fixed effect of the population structure in the model, which would inevitably clean the signal, since the phenotype is heavily structured. Such a complicated situation underlines the need to validate the association, which was achieved through two main approaches. First, we found that the detected locus falls within a QTL hotspot for biomass previously reported in several mapping populations (Rae et al., 2008, 2009; Dillen et al., 2009; Monclus et al., 2012). Second, we have shown that one of the significant SNP affects also the growth in a large collection of *P. nigra* from a breeding pedigree previously used for testing genomic prediction in black poplar (Pégard et al., 2020). While statistically significant, the effect of the SNP is lower than in the natural populations. This could potentially be explained by GxE interaction since we already found that the SNP effect is different between the two common garden experiments and the *P. nigra* pedigree from Pégard et al. (2020) was evaluated in a different location within a quite different climatic area (oceanic climate).

Another way of validating the locus would be to gain insights into the biological mechanism relating the polymorphisms to the trait through the expression of CHI. Considering the polymorphisms, we could identify four non-synonymous SNPs significantly associated with the phenotype, one in the first exon and three in the second exon of the gene, including the top SNP (**Tab. S1**). Interestingly, some of these SNPs (**Tab. S1**) are part of nucleotide triplets that map to stop codons (Chr10-20120172) or involve nucleotides very close to them (Chr10-20120195). In addition, there are two alternative transcripts for CHI gene: Potri.010G213000.3 and Potri.010G213000.2. The latter is shorter (the last exon could be missing), most likely due to the presence of SNPs linked to stop codons, implying that a variant is associated with a truncated and probably nonfunctional protein. We could thus hypothesize that one or several of these SNPs affect the enzymatic activity of CHI, for which a cellular response could be an overexpression of the gene as compensation. This would be consistent with the observed positive relationship between the most significant polymorphism and the expression of CHI (**Fig. 5b**). Also, the decrease in the enzymatic activity of CHI for individuals carrying the alternate allele of the top SNP could be consistent with the decrease in growth observed in these individuals (**Fig. 1c**). Another hypothesis to explain the negative correlation between growth and CHI expression could be a trade-off between growth and wood quality (Novaes et al., 2010). ~~To test this hypothesis, we retrieved data on wood density measured on samples collected at Orléans. The top SNP displayed a significant association with wood density, with a positive effect of the alternate allele, which was thus opposite to the effect found for circumference (**Fig. S7a**). Similarly, a significant positive correlation was found between wood density and CHI gene expression while such correlation was positive for circumference (**Fig. S7b**).~~ ~~These results~~Wood basic density data provided some evidence for the effect of CHI on the trade-off between wood growth and density. CHI is a key enzyme in the flavonoid biosynthetic pathway, where it catalyzes the cyclization of a central intermediate for the production of major flavonoids such as flavanones, flavonols, and anthocyanins. Flavonoids play essential roles in defence, pigmentation, and environmental adaptation, and CHI could thus be involved in these processes. However, this metabolic pathway competes with lignin biosynthesis, as they share the common precursor p-coumaroyl-CoA (Mahon et al., 2022). Interestingly, a pre~~ceding~~vious study revealed that silencing hydroxycinnamoyl-CoA shi-

kimate/quinate hydroxycinnamoyl transferase (HCT) in Arabidopsis thaliana involves an accumulation of flavonoids and a reduction of plant growth (Besseau et al., 2007). However, this relationship between lignin and growth was later found to be unrelated to flavonoids (Li et al., 2010). Another interesting hypothesis to relate fllavonoids and growth could be the inhibitory effect of flavonoids on auxin transport, as reported in Arabidopsis thaliana (Brown et al., 2001). To test this hypothesis it would be interesting to collect phenotypic data on the roots of trees of the populations under study. ~~This is consistent with our results since CHI is one of the first key enzymes of the flavonoid pathway (Dare et al., 2020). HCT is a key enzyme of the lignin pathway, which is well documented and consists of a metabolic grid that modifies phenylalanine in multiple steps to ultimately produce the monolignols p-coumaryl, coniferyl, and sinapyl alcohols. In a previous study, screening a wide diversity of populations and focusing on the lignin biosynthesis pathway made it possible to identify common and rare functional variants in several genes (Marroni et al., 2011), including a natural defective allele for HCT (Vanholme et al., 2013). However, no effect on growth could be detected in this work.~~ ¶

When looking at the loci diversity at the population level we found a strong differentiation, far above the genome-wide level (**Fig. 4**). Such a differentiation is thus more likely to result from differential selection than genetic drift. Of particular interest, the differentiation across natural populations was also found for CHI expression and circumference (**Fig. 4**), and we could show that the top SNP contributed mainly to the between-population component of genetic variation for growth (**Fig. 3**). As a result, highly significant correlations were found between allele frequencies, gene expression, and phenotype at the population level (**Fig. 5**). When looking at the repartition of alleles on a map representing the geographic origin of the populations, a clear North-East versus South-West differentiation appears. Such a tendency was confirmed by the significant correlation found between latitude of origin and allele frequencies (**Fig. S8a**). One could thus hypothesize that the differentiation could be related to climatic differences across Western Europe, which was confirmed by the significant correlation detected between allelic frequencies and a temperature proxy of the climate of origin (**Fig. S8b**). If we go back to the phenotypic data across populations, it's worth noting that the southern populations with the alternate allele fixed display a lower growth and higher wood density. These data support the idea that southern populations are growing slowly as an adaptation to high summer temperatures, which ultimately underlines the adaptive relevance of the locus reported here.

This work strengthens the interest in combining transcriptomics with genomics data across large natural populations to unravel locus and genes involved in key adaptive processes such as the trade-off between growth and wood formation. Such results provide some guidance to breed future varieties of trees with improved efficiency to store carbon.

Acknowledgements

We thank GBFOR (INRAE, Forest Genetics and Biomass Facility), https://doi.org/10.15454/1.5572308287502317E12 for management of the common garden experiment.

Funding

This work was done within the SYBIOPOP project (ANR-13-JSV6-0001) funded by the French National Research Agency (ANR). The platform POPS benefits from the support of the LabEx Saclay Plant Sciences-SPS (ANR-10-LABX-0040-SPS).

Conflict of interest disclosure

The authors declare that they comply with the PCI rule of having no financial conflicts of interest in relation to the content of the article.

Data, scripts, code, and supplementary information availability

Data are available online: https://doi.org/10.57745/DSBTGG.

Information about the RNA-seq project, from which the gene expression data come, is available in the Gene Expression Omnibus (GEO) from NCBI (accession number: GSE128482). Raw sequences (FASTQ) are available in the Sequence Read Archive (SRA) from NCBI (accession number: SRP188754).

Information on the studied genotypes is available in the GnpIS Information System (Steinbach et al., 2013) via the FAIDARE data portal (https://urgi.versailles.inra.fr/faidare/), using the keys "Black poplar" and "POPULUS NIGRA RNASEQ PANEL" for the fields "Crops" and "Germplasm list", respectively.

Scripts and code are available online: https://forgemia.inra.fr/vincent.segura/sybiopop_chi.git.

Supplementary information is available online: https://doi.org/10.5281/zenodo.13950469

References

Abreu M, Silva L, Ribeiro B, Ferreira A, Alves L, Paixão SM, Gouveia L, Moura P, Carvalheiro F, Duarte LC, et al. 2022. Low Indirect Land Use Change (ILUC) Energy Crops to Bioenergy and Biofuels—A Review. *Energies*, **15**, 4348. https://doi.org/10.3390/en15124348

Alexander DH, Novembre J, Lange K. 2009. Fast model-based estimation of ancestry in unrelated individuals. *Genome Research*, **19**,1655–1664. https://doi.org/10.1101/gr.094052.109¶

Alimpić F, Milovanović J, Pielech R, Hinkov G, Jansson R, Dufour S, Beza M, Bilir N, del Blanco LS, Božič G, et al. 2022. The status and role of genetic diversity of trees for the conservation and management of riparian ecosystems: A European experts' perspective. *Journal of Applied Ecology*, **59**, 2476–2485. https://doi.org/10.1111/1365-2664.14247

Allwright MR, Payne A, Emiliani G, Milner S, Viger M, Rouse F, Keurentjes JJB, Bérard A, Wildhagen H, Faivre-Rampant P, et al. 2016. Biomass traits and candidate genes for bioenergy revealed through association genetics in coppiced European Populus nigra (L.). *Biotechnology for Biofuels and Bioproducts*, **9**, 1–22. https://doi.org/10.1186/s13068-016-0603-1

An Y, Liu Y, Liu Y, Lu M, Kang X, Mansfield SD, Zeng W, Zhang J. 2021. Opportunities and barriers for biofuel and bioenergy production from poplar. *GCB Bioenergy*, **13**, 905–913. https://doi.org/10.1111/gcbb.12829

Besseau S, Hoffmann L, Geoffroy P, Lapierre C, Pollet B, Legrand M. 2007. Flavonoid Accumulation in Arabidopsis Repressed in Lignin Synthesis Affects Auxin Transport and Plant Growth. *The Plant Cell*, **19**, 148–162. https://doi.org/10.1105/tpc.106.044495

Borthakur D, Busov V, Cao XH, Du Q, Gailing O, Isik F, Ko JH, Li C, Li Q, Niu S, et al. 2022. Current status and trends in forest genomics. *Forestry Research*, **2**. https://doi.org/10.48130/FR-2022-0011

Bradshaw HD, Stettler RF. 1995. Molecular genetics of growth and development in Populus. IV. Mapping QTLs with large effects on growth, form, and phenology traits in a forest tree. *Genetics*, **139**, 963–973. https://doi.org/10.1093/genetics/139.2.963

Brienen RJW, Caldwell L, Duchesne L, Voelker S, Barichivich J, Baliva M, Ceccantini G, Di Filippo A, Helama S, Locosselli GM, et al. 2020. Forest carbon sink neutralized by pervasive growth-lifespan trade-offs. *Nature Communications*, **11**, 1–10. https://doi.org/10.1038/s41467-020-17966-z

Brown DE, Rashotte AM, Murphy AS, Normanly J, Tague BW, Peer WA, Taiz L, Muday GK, 2001. Flavonoids Act as Negative Regulators of Auxin Transport in Vivo in Arabidopsis, *Plant Physiology*, **126**, 524–535, https://doi.org/10.1104/pp.126.2.524¶

Bryant N, Zhang J, Feng K, Shu M, Ployet R, Chen J-G, Muchero W, Yoo CG, Tschaplinski TJ, Pu Y, et al. 2023. Novel candidate genes for lignin structure identified through genome-wide association study of naturally varying Populus trichocarpa. *Frontiers in Plant Science*, **14**, 1373. https://doi.org/10.3389/fpls.2023.1153113

572 Castillejo MA, Pascual J, Jorrín-Novo J V., Balbuena TS. 2023. Proteomics research in forest
573 trees: A 2012-2022 update. *Frontiers in Plant Science*, **14**, 1130665.
574 https://doi.org/10.3389/fpls.2023.1130665
575 Chandler CH, Chari S, Tack D, Dworkin I. 2014. Causes and consequences of genetic back-
576 ground effects illuminated by integrative genomic analysis. *Genetics*, **196**, 1321–1336.
577 https://doi.org/10.1534/genetics.113.159426
578 Chateigner A, Lesage-Descauses M-C, Rogier O, Jorge V, Leplé J-C, Brunaud V, Roux CP-L,
579 Soubigou-Taconnat L, Martin-Magniette M-L, Sanchez L, et al. 2020. Gene expression predic-
580 tions and networks in natural populations supports the omnigenic theory. *BMC Genomics*, **21**.
581 https://doi.org/10.1186/s12864-020-06809-2
582 Dare AP, Tomes S, McGhie TK, van Klink JW, Sandanayaka M, Hallett IC, Atkinson RG. 2020.
583 Overexpression of chalcone isomerase in apple reduces phloridzin accumulation and in-
584 creases susceptibility to herbivory by two-spotted mites. *The Plant Journal*, **103**, 293–307.
585 https://doi.org/10.1111/tpj.14729
586 Dewoody J, Trewin H, Taylor G. 2015. Genetic and morphological differentiation in Populus nigra
587 L.: isolation by colonization or isolation by adaptation? *Molecular Ecology*, **24**, 2641–2655.
588 https://doi.org/10.1111/mec.13192
589 Dillen SY, Storme V, Marron N, Bastien C, Neyrinck S, Steenackers M, Ceulemans R, Boerjan
590 W. 2009. Genomic regions involved in productivity of two interspecific poplar families in Eu-
591 rope. 1. Stem height, circumference and volume. *Tree Genetics and Genomes*, **5**, 147–164.
592 https://doi.org/10.1007/s11295-009-0199-8
593 Douglas CJ. 2017. Populus as a Model Tree. In: Comparative and Evolutionary Genomics of An-
594 giosperm Trees. Springer, Cham, 61–84. https://doi.org/10.1007/7397_2016_12
595 Escamez S, Robinson KM, Luomaranta M, Gandla ML, Mähler N, Yassin Z, Grahn T, Scheepers
596 G, Stener LG, Jansson S, et al. 2023. Genetic markers and tree properties predicting wood
597 biorefining potential in aspen (Populus tremula) bioenergy feedstock. *Biotechnology for Biofu-*
598 *els and Bioproducts*, **16**, 1–16. https://doi.org/10.1186/s13068-023-02315-1
599 Evans LM, Slavov GT, Rodgers-Melnick E, Martin J, Ranjan P, Muchero W, Brunner AM,
600 Schackwitz W, Gunter L, Chen JG, et al. 2014. Population genomics of Populus trichocarpa
601 identifies signatures of selection and adaptive trait associations. *Nature Genetics*, **46**,
602 1089–1096. https://doi.org/10.1038/ng.3075
603 Faivre-Rampant P, Zaina G, Jorge V, Giacomello S, Segura V, Scalabrin S, Guérin V, De Paoli
604 E, Aluome C, Viger M, et al. 2016. New resources for genetic studies in Populus nigra:
605 genome-wide SNP discovery and development of a 12k Infinium array. *Molecular Ecology Re-*
606 *sources*, **16**, 1023–1036. https://doi.org/10.1111/1755-0998.12513
607 Fick SE, Hijmans RJ. 2017. WorldClim 2: new 1-km spatial resolution climate surfaces for global
608 land areas. *International journal of climatology*, **37**, 4302-4315.
609 https://doi.org/10.1002/joc.5086¶
610 Garrison E, Marth G. 2012. Haplotype-based variant detection from short-read sequencing. arXiv
611 preprint. **arXiv**,1207. https://arxiv.org/abs/1207.3907.¶
612 Gebreselassie MN, Ader K, Boizot N, Millier F, Charpentier JP, Alves A, Simões R, Rodrigues
613 JC, Bodineau G, Fabbrini F, et al. 2017. Near-infrared spectroscopy enables the genetic anal-
614 ysis of chemical properties in a large set of wood samples from Populus nigra (L.) natural
615 populations. *Industrial Crops and Products*, **107**, 159–171. https://doi.org/10.1016/j.ind-
616 crop.2017.05.013
617 Goodstein DM, Shu S, Howson R, Neupane R, Hayes RD, Fazo J, Mitros T, Dirks W, Hellsten U,
618 Putnam N, Rokhsar DS. 2012. Phytozome: a comparative platform for green plant genomics,
619 *Nucleic Acids Res.* **40** (D1): D1178-D1186. https://doi.org/10.1093/nar/gkr944¶
620 Grattapaglia D, Bertolucci FLG, Penchel R, Sederoff RR. 1996. Genetic Mapping of Quantitative
621 Trait Loci Controlling Growth and Wood Quality Traits in Eucalyptus grandis Using a Maternal
622 Half-sib Family and RAPD Markers. *Genetics*, **144**, 1205–1214. https://doi.org/10.1093/genet-
623 ics/144.3.1205

Grattapaglia D, Plomion C, Kirst M, Sederoff RR. 2009. Genomics of growth traits in forest trees. *Current Opinion in Plant Biology*, **12**, 148–156. https://doi.org/10.1016/j.pbi.2008.12.008
Green JK, Keenan TF. 2022. The limits of forest carbon sequestration. *Science*, **376**, 692–693. https://doi.org/10.1126/science.abo6547
Guet J, Fabbrini F, Fichot R, Sabatti M, Bastien C, Brignolas F. 2015a. Genetic variation for leaf morphology, leaf structure and leaf carbon isotope discrimination in European populations of black poplar (Populus nigra L.). *Tree Physiology*, **35**, 850–863. https://doi.org/10.1093/treephys/tpv056
Guet J, Fichot R, Lédée C, Laurans F, Cochard H, Delzon S, Bastien C, Brignolas F. 2015b. Stem xylem resistance to cavitation is related to xylem structure but not to growth and water-use efficiency at the within-population level in Populus nigra L. *Journal of Experimental Botany*, **66**, 4643–4652. https://doi.org/10.1093/jxb/erv232
Gurnell A, Petts G. 2006. Trees as riparian engineers: The Tagliamento River, Italy. *Earth Surface Processes and Landforms*, **31**, 1558–1574. https://doi.org/10.1002/esp.1342
Harris NL, Gibbs DA, Baccini A, Birdsey RA, de Bruin S, Farina M, Fatoyinbo L, Hansen MC, Herold M, Houghton RA, et al. 2021. Global maps of twenty-first century forest carbon fluxes. *Nature Climate Change*, **11**, 234–240. https://doi.org/10.1038/s41558-020-00976-6
Hijmans RJ, Cameron SE, Parra JL, Jones PG, Jarvis A. 2005. Very high resolution interpolated climate surfaces for global land areas. *International Journal of Climatology*, **25**, 1965–1978. https://doi.org/10.1002/joc.1276¶
Jansson S, Douglas CJ. 2007. Populus: A Model System for Plant Biology. *Annual Review of Plant Biology*, **58**, 435–458. https://doi.org/10.1146/annurev.arplant.58.032806.103956
Itoh Y, Yamada Y. 1990. Relationships between genotype x environment interaction and genetic correlation of the same trait measured in different environments. *Theoret. Appl. Genetics* **80**, 11–16. https://doi.org/10.1007/BF00224009¶
Korte A, Vilhjálmsson BJ, Segura V, Platt A, Long Q, Nordborg M. 2012. A mixed-model approach for genome-wide association studies of correlated traits in structured populations. *Nature Genetics*, **44**, 1066–1071. https://doi.org/10.1038/ng.2376
Langmead B, Salzberg SL. 2012. Fast gapped-read alignment with Bowtie 2. *Nature Methods*, **9**, 357–359. https://doi.org/10.1038/nmeth.1923
Laporte F, Charcosset A, Mary-Huard T. 2022. Efficient ReML inference in variance component mixed models using a Min-Max algorithm. *PLOS Computational Biology*, **18**, e1009659. https://doi.org/10.1371/journal.pcbi.1009659
Li H. 2011. A statistical framework for SNP calling, mutation discovery, association mapping and population genetical parameter estimation from sequencing data. *Bioinformatics*, **27**, 2987–2993. https://doi.org/10.1093/bioinformatics/btr509¶
Li X, Bonawitz ND, Weng J-K, Chapple C. 2010. The Growth Reduction Associated with Repressed Lignin Biosynthesis in Arabidopsis thaliana Is Independent of Flavonoids. *The Plant Cell*, **22**, 1620–1632. https://doi.org/10.1105/tpc.110.074161¶
Luomaranta M, Grones C, Choudhary S, Milhinhos A, Kalman TA, Nilsson O, Robinson KM, Street NR, Tuominen H. 2024. Systems genetic analysis of lignin biosynthesis in Populus tremula. *New Phytologist*, **243**, 2157–2174. https://doi.org/10.1111/nph.19993
Maher B. 2008. Personal genomes: The case of the missing heritability. *Nature*, **456**, 18–21. https://doi.org/10.1038/456018a
Mahon EL, de Vries L, Jang S-K, Middar S, Kim H, Unda F, Ralph J, Mansfield SD. 2022. Exogenous chalcone synthase expression in developing poplar xylem incorporates naringenin into lignins, *Plant Physiology*, **188**, 984–996, https://doi.org/10.1093/plphys/kiab499¶
Marroni F, Pinosio S, Di Centa E, Jurman I, Boerjan W, Felice N, Cattonaro F, Morgante M. 2011. Large-scale detection of rare variants via pooled multiplexed next-generation sequencing: towards next-generation Ecotilling. *The Plant Journal*, **67**, 736–745. https://doi.org/10.1111/j.1365-313X.2011.04627.x¶
Mckown AD, Guy RD, Klápště J, Geraldes A, Friedmann M, Cronk QCB, El-Kassaby YA, Mansfield SD, Douglas CJ. 2014. Geographical and environmental gradients shape phenotypic trait

677 variation and genetic structure in Populus trichocarpa. *New Phytologist*, **201**, 1263–1276.
678 https://doi.org/10.1111/nph.12601
679 Milesi P, Kastally C, Dauphin B, Cervantes S, Bagnoli F, Budde KB, Cavers S, Fady B, Faivre-
680 Rampant P, González-Martínez SC, et al. 2024 Resilience of genetic diversity in forest trees
681 over the Quaternary. *Nature Communications*, **15**, 8538. https://doi.org/10.1038/s41467-024-
682 52612-y
683 Monclus R, Leplé JC, Bastien C, Bert PF, Villar M, Marron N, Brignolas F, Jorge V. 2012. Inte-
684 grating genome annotation and QTL position to identify candidate genes for productivity, ar-
685 chitecture and water-use efficiency in Populus spp. *BMC plant biology*, **12**, 173.
686 https://doi.org/10.1186/1471-2229-12-173
687 Muñoz F, Sanchez L. 2024. breedR: Statistical Methods for Forest Genetic Resources Analysts.
688 R package version 0.12-5, https://github.com/famuvie/breedR.¶
689 Neale DB, Kremer A. 2011. Forest tree genomics: growing resources and applications. Nature
690 Reviews Genetics 2011 12:2 12: 111–122. https://doi.org/10.1038/nrg2931
691 Novaes E, Kirst M, Chiang V, Winter-Sederoff H, Sederoff R. 2010. Lignin and Biomass: A Neg-
692 ative Correlation for Wood Formation and Lignin Content in Trees. *Plant Physiology*, **154**,
693 555–561. https://doi.org/10.1104/pp.110.161281
694 Oubida RW, Gantulga D, Zhang M, Zhou L, Bawa R, Holliday JA. 2015. Partitioning of multivari-
695 ate phenotypes using regression trees reveals complex patterns of adaptation to climate
696 across the range of black cottonwood (Populus trichocarpa). *Frontiers in Plant Science*, **6**.
697 https://doi.org/10.3389/fpls.2015.00181
698 Pan Y, Birdsey RA, Fang J, Houghton R, Kauppi PE, Kurz WA, Phillips OL, Shvidenko A, Lewis
699 SL, Canadell JG, et al. 2011. A large and persistent carbon sink in the world's forests. *Sci-*
700 *ence*, **333**, 988–993. https://doi.org/10.1126/science.1201609
701 Pazhamala LT, Kudapa H, Weckwerth W, Millar AH, Varshney RK. 2021. Systems biology for
702 crop improvement. *The Plant Genome*, **14**, e20098. https://doi.org/10.1002/tpg2.20098
703 Pégard M, Segura V, Muñoz F, Bastien C, Jorge V, Sanchez L. 2020. Favorable Conditions for
704 Genomic Evaluation to Outperform Classical Pedigree Evaluation Highlighted by a Proof-of-
705 Concept Study in Poplar. *Frontiers in Plant Science*, **11**, 581954.
706 https://doi.org/10.3389/fpls.2020.581954
707 Plomion C, Bastien C, Bogeat-Triboulot MB, Bouffier L, Déjardin A, Duplessis S, Fady B, Heuertz
708 M, Le Gac AL, Le Provost G, et al. 2016. Forest tree genomics: 10 achievements from the
709 past 10 years and future prospects. *Annals of Forest Science*, **73**, 77–103.
710 https://doi.org/10.1007/s13595-015-0488-3
711 Plomion C, Lalanne C, Claverol S, Meddour H, Kohler A, Bogeat-Triboulot MB, Barre A, Le
712 Provost G, Dumazet H, Jacob D, et al. 2006. Mapping the proteome of poplar and application
713 to the discovery of drought-stress responsive proteins. *Proteomics*, **6**, 6509–6527.
714 https://doi.org/10.1002/pmic.200600362
715 Porth I, El-Kassaby YA. 2015. Using Populus as a lignocellulosic feedstock for bioethanol.
716 *Biotechnology Journal*, **10**, 510–524. https://doi.org/10.1002/biot.201400194
717 R Core Team. 2021. R: A Language and Environment for Statistical Computing.
718 Rae AM, Pinel MPC, Bastien C, Sabatti M, Street NR, Tucker J, Dixon C, Marron N, Dillen SY,
719 Taylor G. 2008. QTL for yield in bioenergy Populus: Identifying G×E interactions from growth
720 at three contrasting sites. *Tree Genetics and Genomes*, **4**, 97–112.
721 https://doi.org/10.1007/s11295-007-0091-3
722 Rae AM, Street NR, Robinson KM, Harris N, Taylor G. 2009. Five QTL hotspots for yield in short
723 rotation coppice bioenergy poplar: The Poplar Biomass Loci. *BMC Plant Biology*, **9**, 1–13.
724 https://doi.org/10.1186/1471-2229-9-23
725 Rae AM, Tricker PJ, Bunn SM, Taylor G. 2007. Adaptation of tree growth to elevated CO2: quan-
726 titative trait loci for biomass in Populus. *New Phytologist*, **175**, 59–69.
727 https://doi.org/10.1111/j.1469-8137.2007.02091.x
728 Rathgeber CBK, Cuny HE, Fonti P. 2016. Biological basis of tree-ring formation: A crash course.
729 *Frontiers in Plant Science*, **7**, 734. https://doi.org/10.3389/fpls.2016.00734

de Rigo D, Enescu CM, Durrant TH, Caudullo G. 2016. Populus nigra in Europe: distribution, habitat, usage and threats. In: SanMiguel-Ayanz J, de Rigo D, Caudullo G, Houston Durrant T, Mauri A (eds.), European Atlas of Forest Tree Species. Publ. Off. EU, Luxembourg, pp. e0182a4+. https://w3id.org/mtv/FISE-Comm/v01/e01aa69

Rodrigues AM, Miguel C, Chaves I, António C. 2021. Mass spectrometry-based forest tree metabolomics. *Mass Spectrometry Reviews*, **40**, 126–157. https://doi.org/10.1002/mas.21603

Rogier O, Chateigner A, Lesage-Descauses MC, Mandin C, Brunaud V, Caius J, Soubigou-Taconnat L, Almeida-Falcon J, Bastien C, Benoit V, et al. 2023. RNAseq based variant dataset in a black poplar association panel. *BMC research notes*, **16**, 248. https://doi.org/10.1186/s13104-023-06521-w

RStudio Team. 2020. RStudio: Integrated Development Environment for R.

Sargolzaei M, Chesnais JP, Schenkel FS. 2014. A new approach for efficient genotype imputation using information from relatives. *BMC Genomics*, **15**, 1–12. https://doi.org/10.1186/1471-2164-15-478

Sasaki E, Frommlet F, Nordborg M. 2018. GWAS with Heterogeneous Data: Estimating the Fraction of Phenotypic Variation Mediated by Gene Expression Data. *G3 Genes|Genomes|Genetics*, **8**, 3059–3068. https://doi.org/10.1534/g3.118.200571

Segura V, Vilhjálmsson BJ, Platt A, Korte A, Seren Ü, Long Q, Nordborg M. 2012. An efficient multi-locus mixed-model approach for genome-wide association studies in structured populations. *Nature Genetics*, **44**, 825–830. https://doi.org/10.1038/ng.2314

Smulders MJM, Cottrell JE, Lefèvre F, van der Schoot J, Arens P, Vosman B, Tabbener HE, Grassi F, Fossati T, Castiglione S, et al. 2008. Structure of the genetic diversity in black poplar (Populus nigra L.) populations across European river systems: Consequences for conservation and restoration. *Forest Ecology and Management*, **255**, 1388–1399. https://doi.org/10.1016/j.foreco.2007.10.063

Speed D, Hemani G, Johnson MR, Balding DJ. 2012. Improved heritability estimation from genome-wide SNPs. *American Journal of Human Genetics*, **91**, 1011–1021. https://doi.org/10.1016/j.ajhg.2012.10.010

Steinbach D, Alaux M, Amselem J, Choisne N, Durand S, Flores R, Keliet A-O, Kimmel E, Lapalu N, Luyten I, et al. 2013. GnpIS: an information system to integrate genetic and genomic data from plants and fungi. *Database*, **2013**, bat058. https://doi.org/10.1093/database/bat058

Taylor G, Allwright MR, Smith HK, Polle A, Wildhagen H, Hertzberg M, Bhalerao R, Joost Keurentjes JJB, Scalabrin S, Scaglione D, et al. 2016. Bioenergy trees: Genetic and genomic strategies to improve yield. In: Barth S, Murphy-Bokern D, Kalinina O, Taylor G, Jones M (eds) Perennial Biomass Crops for a Resource-Constrained World. Springer, Cham. https://doi.org/10.1007/978-3-319-44530-4_15

Teyssier C, Rogier O, Claverol S, Gautier F, Lelu-Walter M-A, Duruflé H. 2023. Comprehensive Organ-Specific Profiling of Douglas Fir (Pseudotsuga menziesii) Proteome. *Biomolecules*, **13**, 1400. https://doi.org/10.3390/biom13091400

Van der Auwera GA, Carneiro MO, Hartl C, Poplin R, del Angel G, Levy-Moonshine A, Jordan T, Shakir K, Roazen D, Thibault J, et al. 2013. From FastQ Data to High-Confidence Variant Calls: The Genome Analysis Toolkit Best Practices Pipeline. *Current Protocols in Bioinformatics*, **43**: 11.10.1-11.10.33. https://doi.org/10.1002/0471250953.bi1110s43¶

~~Vanholme B, Cesarino I, Goeminne G, Kim H, Marroni F, Van Acker R, Vanholme R, Morreel K, Ivens B, Pinosio S, et al. 2013. Breeding with rare defective alleles (BRDA): a natural Populus nigra HCT mutant with modified lignin as a case study. *New Phytologist*, **198**, 765–776. https://doi.org/10.1111/nph.12179¶~~

VanRaden PM. 2008. Efficient Methods to Compute Genomic Predictions. Journal of Dairy Science 91: 4414–4423. https://doi.org/10.3168/jds.2007-0980

Viger M, Smith HK, Cohen D, Dewoody J, Trewin H, Steenackers M, Bastien C, Taylor G. 2016. Adaptive mechanisms and genomic plasticity for drought tolerance identified in European black poplar (Populus nigra L.). *Tree Physiology*, **36**, 909–928. https://doi.org/10.1093/treephys/tpw017

Wade AR, Duruflé H, Sanchez L, Segura V. 2022. eQTLs are key players in the integration of genomic and transcriptomic data for phenotype prediction. *BMC Genomics*, **23**, 2021.09.07.459279. https://doi.org/10.1186/s12864-022-08690-7

Wang J, Ding J, Tan B, Robinson KM, Michelson IH, Johansson A, Nystedt B, Scofield DG, Nilsson O, Jansson S, et al. 2018. A major locus controls local adaptation and adaptive life history variation in a perennial plant. *Genome Biology*, **19**, 1–17. https://doi.org/10.1186/s13059-018-1444-y

Weir BS, Cockerham CC. 1984. Estimating F-Statistics for the Analysis of Population Structure. *Evolution*, **38**, 1358. https://doi.org/10.2307/2408641

de Wit P, Pespeni MH, Palumbi SR. 2015. SNP genotyping and population genomics from expressed sequences – current advances and future possibilities. *Molecular Ecology*, **24**, 2310–2323. https://doi.org/10.1111/mec.13165

Wójkiewicz B, Lewandowski A, Żukowska WB, Litkowiec M, Wachowiak W. 2021. Low effective population size and high spatial genetic structure of black poplar populations from the Oder valley in Poland. *Annals of Forest Science*, **78**, 1–24. https://doi.org/10.1007/s13595-021-01055-2

Xie M, Zhang J, Tschaplinski TJ, Tuskan GA, Chen JG, Muchero W. 2018. Regulation of lignin biosynthesis and its role in growth-defense tradeoffs. *Frontiers in Plant Science*, **9**, 407961. https://doi.org/10.3389/fpls.2018.01427

Yu J, Pressoir G, Briggs WH, Bi IV, Yamasaki M, Doebley JF, McMullen MD, Gaut BS, Nielsen DM, Holland JB, et al. 2005. A unified mixed-model method for association mapping that accounts for multiple levels of relatedness. *Nature Genetics*, **38**, 203–208. https://doi.org/10.1038/ng1702

Zhang M, Suren H, Holliday JA. 2019. Phenotypic and Genomic Local Adaptation across Latitude and Altitude in Populus trichocarpa. *Genome Biology and Evolution*, **11**, 2256–2272. https://doi.org/10.1093/gbe/evz151

Zhou Y. 2022. Understanding urban plant phenology for sustainable cities and planet. *Nature Climate Change,* **12,** 302–304. https://doi.org/10.1038/s41558-022-01331-7