

# Induced and natural variation affect traits independently in hybrid *Populus*

Weier Guo <sup>1</sup>, H lo se Bastiaanse <sup>1,3</sup>, Julin N. Maloof <sup>2</sup>, Luca Comai <sup>1</sup>, Isabelle M. Henry <sup>1,\*</sup>

<sup>1</sup>Genome Center and Department of Plant Biology, University of California Davis, Davis, CA 95616, USA

<sup>2</sup>Department of Plant Biology, University of California Davis, Davis, CA 95616, USA

<sup>3</sup>Present address: VIB Center for Plant Systems Biology, 9052 Ghent, Belgium, and Department of Plant Biotechnology and Bioinformatics, Ghent University, B-9052 Ghent, Belgium

\*Corresponding author: Department of Plant Biology and Genome Center, University of California Davis, 451 Health Sciences Dr., Davis, CA 95618, USA. Email: imhenry@ucdavis.edu

The genetic control of many plant traits can be highly complex. Both allelic variation (sequence change) and dosage variation (copy number change) contribute to a plant's phenotype. While numerous studies have investigated the effect of allelic or dosage variation, very few have documented both within the same system, leaving their relative contribution to phenotypic effects unclear. The *Populus* genome is highly polymorphic, and poplars are fairly tolerant of gene dosage variation. Here, using a previously established *Populus* hybrid F1 population, we assessed and compared the effect of natural allelic variation and induced dosage variation on biomass, phenology, and leaf morphology traits. We identified QTLs for many of these traits, but our results indicate limited overlap between the QTLs associated with natural allelic variation and induced dosage variation. Additionally, the integration of data from both allelic and dosage variation identifies a larger set of QTLs that together explain a larger percentage of the phenotypic variance. Finally, our results suggest that the effect of the large indels might mask that of allelic QTLs. Our study helps clarify the relationship between allelic and dosage variation and their effects on quantitative traits.

**Keywords:** dosage variation; QTL; poplar; natural variation; trait

## Introduction

Natural allelic variation plays an important role in phenotypic diversity in plants (Alonso-Blanco *et al.* 1999, 2009; Todesco *et al.* 2010; Huang *et al.* 2011, 2012; Huang and Han 2014; Jin *et al.* 2016; Satbhai *et al.* 2017; Zhang *et al.* 2021; Duan *et al.* 2022). The statistical framework raised by R. A. Fisher provides an approach to systematically identify the quantitative trait loci (QTL) responsible for heritable variation (Fisher 1919). In the last decade, the development of new DNA high-throughput sequencing and genotyping technologies has dramatically improved our ability to identify polymorphic genetic markers between individuals or species (Gupta *et al.* 2008; Davey *et al.* 2011; Elshire *et al.* 2011). This, in turn, enables more accurate QTL identification in both plants and animals (McMullen 2003; Wellcome Trust Case Control Consortium 2007; Rafalski 2010; Jamann *et al.* 2015). Despite these technological advances, a wide percentage of the observed phenotypic variance still remains unexplained by the detected QTLs. This is particularly problematic for complex traits with expected polygenic contributions. For example, the QTLs detected through the analysis of biomass-related traits in *Populus* explain, on average, 26% of the observed phenotypic variation (Rae *et al.* 2009). To increase biomass yield through tree breeding, we need to consider other types of heritable variations, aiming for a deeper understanding of the underlying regulatory mechanisms.

Besides allelic variation (sequence variation that does not involve copy number changes), dosage variation can also affect the phenotypic outcomes of many important plant traits. Copy

number variation (CNVs), especially the ones affecting protein-coding regions, have been associated with phenotypic outcomes in multiple plant species (Cook *et al.* 2012; Diaz *et al.* 2012; Li *et al.* 2012; Carbonell-Bejerano *et al.* 2017; Prunier *et al.* 2019). Pan-genomic analyses have identified structural variants across different accessions of multiple plant species, many of which affect important agronomic traits such as flower size, fruit weight, and heat tolerance (Golicz *et al.* 2016; Pinosio *et al.* 2016; Alonge *et al.* 2020; Zmienko *et al.* 2020; Yan *et al.* 2023). Gene deletion and duplication can directly affect expression level (cis-effect), which in turn affects phenotypes. Gene dosage may also affect phenotype through mechanisms explained by the gene balance hypothesis (Birchler and Veitia 2012). Dosage variation can also modulate the expression of genes located outside of indel regions (trans-effect), since many traits are regulated by a complex network comprising multiple genetic components (Birchler and Veitia 2010; Veitia *et al.* 2013).

To increase our understanding of the relative contributions of these two sources of phenotypic variation, we investigated the phenotypic effects of induced dosage variation and natural allelic variation within the same population. We also aimed to document instances of interplay between these two sources of variation. For example, when a locus encodes a protein whose function is dosage sensitive, the CNV-induced expression changes affect the phenotype. However, if allelic variation is also present, such as if one allele is hypomorphic or null, two scenarios are possible: (1) the CNV affecting the deficient allele results in no or little phenotypic variation or (2) the CNV affecting the normal allele results in magnified

phenotypic variation. Either way, focusing on either the allelic variation or the dosage variation alone only addresses part of the mechanisms at play. A more comprehensive approach, which integrates both types of variations may be better suited to fully understand the genetic regulatory factors of complex traits.

*Populus* is an attractive system to study the interplay between allelic and dosage variation. It is dioecious and therefore an obligate outcrosser and its genome are highly polymorphic, both in terms of sequence polymorphisms and CNVs (Tuskan et al. 2006; Pinosio et al. 2016). Pollen irradiation is a widely used approach for inducing indel mutations in plants (Brewbaker and Emery 1961; Yang et al. 2004), starting as early as the 1950s (Nuffer 1957; Mottinger 1970). In tree species, pollen irradiation followed by pollination has been well-established (Osborne 1957; Rudolph 1978). Gamma-induced indels, especially larger ones, are not typically retained in future generations because they are often associated with lethality in the gamete, where the copy number goes down to zero. In clonally propagated crops such as *Populus*, on the other hand, they can be retained indefinitely. In a previous report, we described the establishment of a *Populus* F1 hybrid population (592 lines) from an interspecific cross between a wild-type *P. deltoides* mother and gamma-irradiated pollen from *P. nigra* (Henry et al. 2015b). Whole-genome sequencing analysis revealed that 58% of the F1 lines carry large-scale insertions or deletions (indels). The size of induced indels varies from 250 kb to whole chromosomes. The number of indels per line varies between 0 and 10, with 2.5 indels per individual on average. Indels from different lines can overlap such that each genomic region is covered by 1–31 indels and only 1.6% of the genome (6.2 Mb) is not covered by any indel at all.

Using this resource, we investigated the association between dosage variation across the genome and a variety of phenotypes. This resulted in the identification of “dosage QTLs” associated with biomass, phenology, leaf morphology, and vessel development traits (Bastiaanse et al. 2019, 2020a; Rodriguez-Zaccaro et al. 2021). Since both parental genomes are highly polymorphic, natural allelic variation is expected to play an important role in the observed phenotypic variation, but it was not taken into account in these earlier studies.

Here, we aim to investigate whether allelic variation, and in this case, the differences between the two haplotypes within each parent, also influence these traits (allelic QTLs). Next, we aimed to document the possible interaction between natural allelic variation and induced dosage variation in this population (Fig. 1). This *Populus* clonal system is superior to our study goal since it allows us to obtain replicated phenotypic information easily. In a subset of 343 F1 lines, all offspring of the same two parental clones from this *Populus* population, and detected both dosage and allelic QTLs. Our results suggest a limited overlap between QTLs associated with allelic and dosage variation. A custom method was developed to assess the effect of both allelic and dosage variation in a joint model. The results indicated that allelic and dosage variation affect traits independently. Detection of allelic QTLs in a subset of the population that does not carry large indels resulted in a different set of QTLs, suggesting that large-scale indels might mask the effect of allelic QTLs in the full population. Finally, direct integration of both types of QTLs makes the association between trait values and genetic information stronger.

## Materials and methods

### Data acquisition and preprocessing

Genomic sequencing data, RNA-seq data, and phenotypic information were obtained from previous studies (Henry et al. 2015b;

Zinkgraf et al. 2016; Bastiaanse et al. 2019, 2020a). Briefly, an interspecific cross between wild-type *P. deltoides* and pollen-irradiated *P. nigra* produced 592 F1 hybrid lines. High-coverage Illumina short-read sequences were obtained from the two parental lines with read depth around 45× and 65× for *P. deltoides* and *P. nigra*, respectively. Additionally, low-coverage Illumina genome sequences were obtained from each of the F1 hybrid clones (read depth around 0.5× per line). Leaf RNA sequencing was performed on 166 F1 lines, each in triplicates. The raw RNA-seq reads were pooled per clone and used to assist in haplotype phasing. The collection and statistical analysis of phenotypic information were described in previous studies (Bastiaanse et al. 2019, 2020a). Three categories of phenotypes—leaf morphology, phenology, and biomass—were used in our study (Supplementary File 1).

The preprocessing of sequencing data followed a custom pipeline developed previously. It starts with a demultiplexing step performed using a custom pipeline (<https://github.com/Comai-Lab/allprep>) for separating raw reads into individual libraries. Reads were aligned to the *Populus* reference *P. trichocarpa* v3.0 (Tuskan et al. 2006), using a custom Python script based on Burrows-Wheeler Aligner (Li and Durbin 2009) (<https://comailab.org/data-and-method/bwa-doall-a-package-for-batch-library-processing-and-alignment/>). Bam files were generated in this step, which were used to obtain a mpileup file using a custom Python package (<https://github.com/Comai-Lab/mpileup-tools>) based on Samtools (Li et al. 2009), followed by a simplification step to convert the mpileup file into a parsed-mpileup file.

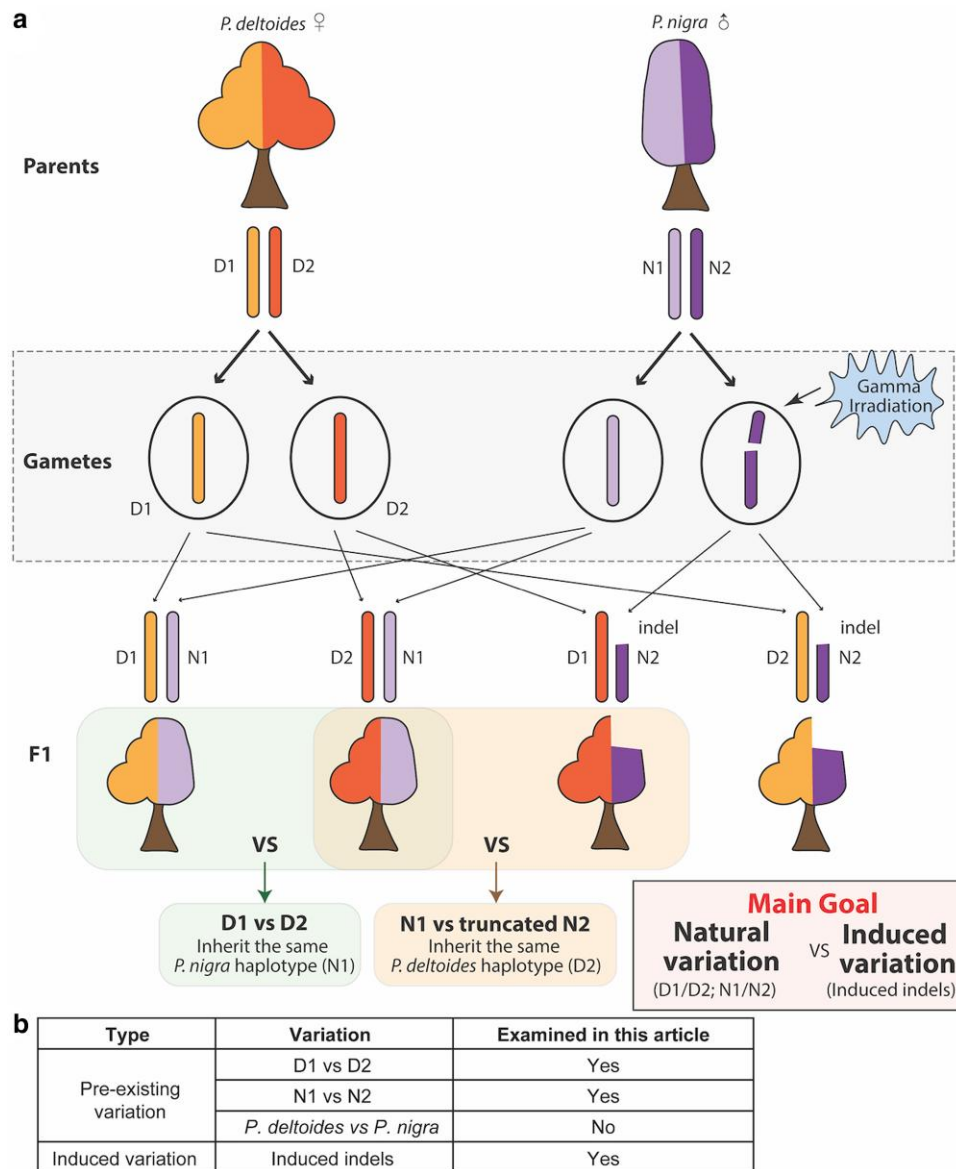
### Haplotype phasing

To describe the parental haplotypes, we identified heterozygous positions in each parent and determined the phasing between these positions, using a custom computational pipeline ([https://github.com/guoweier/QTL\\_manuscript](https://github.com/guoweier/QTL_manuscript)). Specifically, we started by identifying single nucleotide polymorphisms (SNPs) that can distinguish between two haplotypes within a parent (Supplementary Fig. 1a). In short, we selected two lists of SNPs, one for *P. deltoides* and the other for *P. nigra*. The example of *P. deltoides* SNPs selection is shown in Supplementary Fig. 1a. For *P. deltoides*, we selected positions that exhibited heterozygosity in *P. deltoides* and homozygosity in *P. nigra*; or positions that showed heterozygosity in *P. deltoides* with different heterozygous allele combinations in *P. nigra*.

Next, we used RNA-seq data obtained from a subset of 122 F1 individuals to derive phased parental haplotypes (Supplementary Fig. 1b). Briefly, we first used the RNA-seq raw data from the diploid F1 lines for haplotype phasing, after retaining the positions that are at least 20× read depth in the RNA-Seq data. Second, we treated RNA-seq raw data as genomic sequencing data, with the preprocessing approaches that have been described above. Parsed-mpileup file with 122 RNA-seq lines was obtained after running the pipeline. Then, the RNA-seq parsed-mpileup file was used to identify inherited alleles from *P. deltoides* and *P. nigra*, respectively. Finally, we collected the adjacent SNPs combination orders and recorded the order as parental haplotypes when data from more than 90% (109 out of 122) of RNA-seq lines were consistent with it.

### Genotyping

The adjusted phased haplotypes were applied to low-coverage sequencing data for genotyping. Specifically, for each SNP marker, genotype in F1 hybrids was only recorded when it inherited the alternative allele. Recorded genotypic information was then binned (50 SNPs per bin) to increase the robustness of genotype calls. As a control, the same genotyping process was applied to the RNA-seq

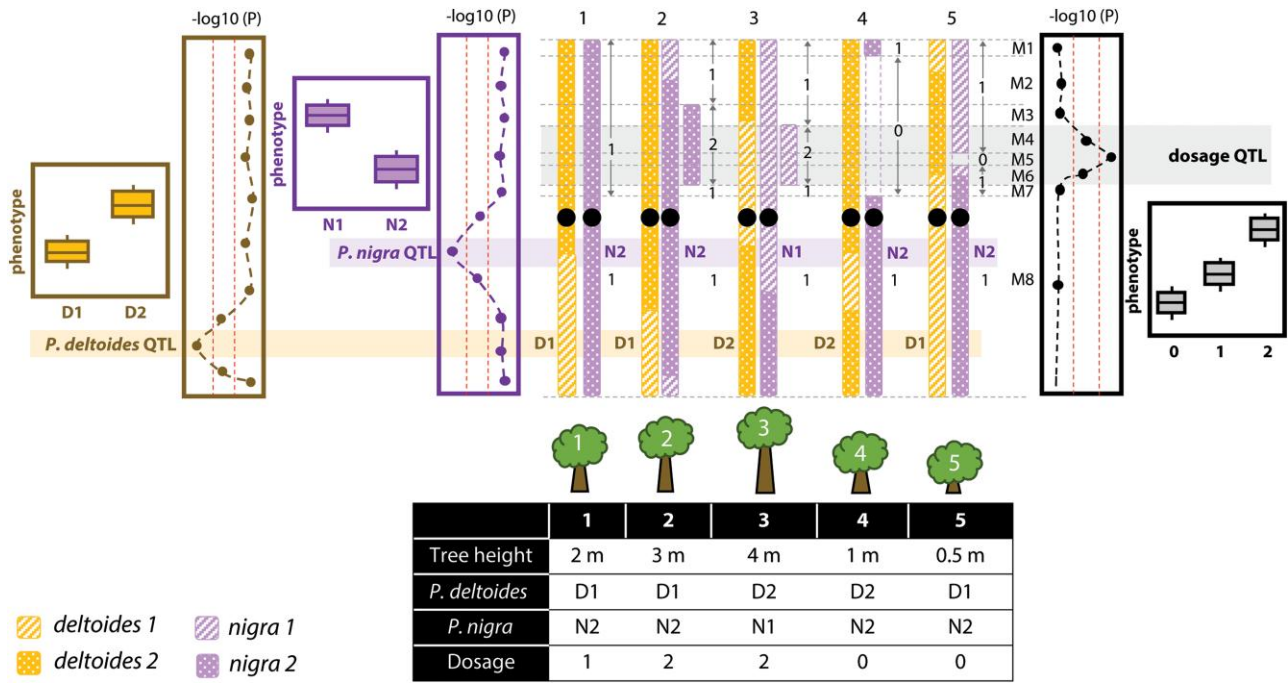


**Fig. 1.** Major goal of this study. a) Illustration of the *Populus* population used in this study and the main goal of this study. The F1 population came from an interspecific cross between *P. deltoides* (female) and *P. nigra* (male). The phenotypic differences can result from i) natural variation (D1/D2, N1/N2); ii) Radiation-induced variation (indels). Our main goal is to investigate the interplay between natural variation and induced variation. D1/D2: *P. deltoides* haplotype1 and haplotype2. N1/N2: *P. nigra* haplotype1 and haplotype2. b) Type of variation examined in this article. For natural variation, we are examining the SNPs between 2 haplotypes within *P. deltoides*, as well as the SNPs between 2 haplotypes within *P. nigra*. We are not testing the species-specific SNPs (*P. deltoides* vs *P. nigra*) in this study. For induced variation, we examined the radiation-induced indels.

data. The transcriptomic genotypes and genomic genotypes were compared manually (all resulting figures can be viewed at [https://github.com/guoweier/QTL\\_manuscript](https://github.com/guoweier/QTL_manuscript)). Next, for the individuals for which both genomic and RNA-seq data were available, we sorted the F1 lines based on the read-depth of the low-coverage genome sequencing data. We then selected a read-depth threshold based on the following: a) Genotypes based on the low-pass genomic data clearly show an expected pattern of recombination along the whole genome and, b) Genotypes obtained from the genomic and RNA-Seq data are consistent. Lines for which only genomic data was available were retained if genomic coverage was above this threshold. As a result, 343 lines were selected to proceed for QTL analysis. Transcriptomic and genomic genotypes comparison of chromosome 1 on the selected F1 line with the lowest read-depth is shown in [Supplementary Fig. 2](#).

## Dosage variation quantification

Methods for quantifying dosage variation have been described in previous studies (Bastiaanse et al. 2019). Shortly, we defined bins based on indels breakpoints and tiled bins along the chromosomes. For each bin, the dosage genotype was determined by comparing the mean read coverage for each individual to the mean of the population. Dosage indicates the total copy number in any given bin. These F1 lines are diploids, so the background dosage number is 2. Since all dosage variation originates from *P. nigra* (Henry et al. 2015b), which is the paternal parent, we decided to only focus on the dosage changes in *P. nigra*. So the normal dosage state is 1, representing the F1 line carrying 1 copy from *P. nigra*. If an F1 clone carries a deletion which occupies 4 bins on chromosome 10, the dosage genotype for these 4 bins was set to 0, while the rest of bins on chromosome 10 were set to 1. Dosage genotypes



**Fig. 2.** Representative illustration of QTL analysis using both allelic and dosage variation information. QTLs detected from sequence variation between two haplotypes (D1/D2) of *P. deltooides*, sequence variation between two haplotypes (N1/N2) of *P. nigra* and dosage variation can all contribute to the same trait (here tree height). *P. deltooides* and *P. nigra* haplotypes were acquired through analysis of allelic variation within each parent (D1/D2 or N1/N2). Dosage information was obtained through the calculation of relative copy number states in each chromosome bin (see details in Material and Methods).

were acquired for all 343 lines for which SNPs genotypes were also obtained. An illustration diagram can be found in Fig. 2.

## QTL analysis

To conduct a QTL analysis that simultaneously includes both allelic and dosage variation we employed a custom Python pipeline available at [https://github.com/guoweier/QTL\\_manuscript](https://github.com/guoweier/QTL_manuscript). We generated a common marker list encompassing three types of variation: the *P. deltooides* haplotype, the *P. nigra* haplotype, and the dosage variation. First, we identified the physical positions of binned markers in *P. deltooides* and *P. nigra* genotypes, respectively. We then imputed genotypes in the unknown regions using information from their flanking binned markers. For example, on the *P. nigra* genotype, marker 1 is Chr01\_1\_10000 with genotype N1 and marker 2 is Chr01\_20000\_30000 with genotype N1. So the genotype in Chr01\_10001\_19999 is N1. If two flanking markers contained different genotypes, or if there was a missing flanking marker, the genomic region in between was assigned as a missing value “NA”. Second, we built a common marker list for the two parents, using *P. deltooides* markers as the reference and imputed *P. nigra* genotypes based on the markers’ physical positions. Last, we applied the common marker to the dosage genotype and obtained the dosage value for each new marker.

Single models were established for analyzing the correlation between phenotypes and each variation type. The model is specified as follows:

$$Y_i = \beta_0 + \beta_1 g_i + \epsilon_i$$

where  $Y_i$  is the phenotype;  $\beta_0$  is the intercept;  $\beta_1$  is the unknown coefficient;  $g_i$  is one of the examining genotypes (*P. deltooides* haplotype or *P. nigra* haplotype or dosage); and  $\epsilon_i$  is the residual variance. *P. deltooides* haplotypes were recorded as D1 or D2. *P. nigra*

haplotypes were recorded as N1 or N2, while deleted regions were recorded as “NA”. The dosage of the *P. nigra* allele was recorded as 0 (deletion), 1 (regular), or 2 (insertion). To establish a suitable threshold for identifying significant QTLs, we employed a permutation test approach (Doerge and Churchill 1996). In short, for each trait and each genotype (*P. deltooides* haplotype or *P. nigra* haplotype or dosage), the phenotype data from the 343 F1 lines were randomized. Next, a linear regression between trait values and marker values was calculated with all the markers along the genome. The maximum t-value was selected. This randomization process was repeated 1,000 times. Then, we selected the top 5 and 1% of maximum t-values. In the observed dataset, the markers with t-values larger than the 5% threshold were considered significant, and those larger than the 1% threshold were considered as confirmed. Adjacent significant markers were considered as belonging to the same QTL.

To investigate how much phenotypic variance can be explained by each single QTL, we performed the QTL mapping using a multivariate model including all markers located underneath that QTL and extracted the adjusted R-square values. For phenotypic variance explained by all QTLs associated with one trait, we took the most significant marker (marker with the largest t-value) underlying each QTL and ran a multivariate model including these selected markers. Integration of QTLs from allelic and dosage variation followed a similar approach. For each trait, we collected the most significant marker from each QTL and fitted these markers into a multivariate model. Adjusted R-square values were recorded.

We designed a custom approach to perform QTL mapping combining all three types of variation. In short, we collected the genotypic information (*P. deltooides* haplotype or *P. nigra* haplotype or dosage) and assigned a State for each combined genotype. There were 10 possible States for the combined variable

(Supplementary File 2). Then, a linear regression was performed using the `lm()` function in R, which is specified as follows:

$$Y_i = \beta_0 + \beta_1 \text{State}_i + \epsilon_i$$

where  $Y_i$  is the phenotype of the  $i$ th individual;  $\beta_0$  is the intercept;  $\beta_1$  is the unknown coefficient;  $\text{State}_i$  is the variable after combining the three genotypes (*P. deltoides* haplotype or *P. nigra* haplotype or dosage) information of the  $i$ th individual; and  $\epsilon_i$  is the residual variance. Next, we performed pairwise comparisons of all present States using the function `pairwisePermutationTest()` in the R package “`rcompanion`” (Mangiafico 2020). Each comparison pair was treated independently, which generated 45 comparisons (Supplementary File 2). For each comparison, the  $P$ -values were collected and adjusted using the Benjamini and Hochberg (BH) method (Benjamini and Hochberg 1995). Adjacent markers were considered to belong to the same QTL. Last we identified the pairs of States that were significantly different to infer the possible genetic factors underlying the observed phenotypic variation. Specifically, QTLs were classified into 6 groups: deletion, deletion + insertion, insertion, *P. deltoides*, *P. deltoides* + *P. nigra*, and *P. nigra* (Supplementary File 2). The proportion of phenotypic variance explained by this custom QTL approach was determined using a method similar to that described above for QTLs from single models.

## Differentially expressed gene analysis and GO enrichment analysis

Differentially expressed genes were identified and the ones located within allelic QTLs were recorded. For each QTL, extreme phenotypic mutants (10 and 90% quantile) were selected, excluding the indel mutants having an indel under the QTL bins. Differential expression analysis was performed using the `limma-voom` method ([https://ucdavis-bioinformatics-training.github.io/2022-April-GGI-DE-in-R/data\\_analysis/DE\\_Analysis\\_with\\_quizzes\\_fixed](https://ucdavis-bioinformatics-training.github.io/2022-April-GGI-DE-in-R/data_analysis/DE_Analysis_with_quizzes_fixed)). Specifically, the estimated read counts were filtered such that only genes having more than 10 reads per million in at least 80% of the libraries were retained.  $P$ -values were adjusted using the Benjamini–Hochberg method (Benjamini and Hochberg 1995). Genes located under the QTL bins and with adjusted  $P$ -value  $< 0.05$  were retained. The annotation information from Phytozome ([https://phytozome-next.jgi.doe.gov/info/Ptrichocarpa\\_v3\\_1](https://phytozome-next.jgi.doe.gov/info/Ptrichocarpa_v3_1)) was added for each gene.

GO terms for *Populus* genes were obtained from Phytozome ([https://phytozome-next.jgi.doe.gov/info/Ptrichocarpa\\_v3\\_1](https://phytozome-next.jgi.doe.gov/info/Ptrichocarpa_v3_1)). Enrichment analysis was performed by comparing GO terms of genes present in QTL bins against the genes expressed in leaf tissue. GO terms were considered suggestively enriched if the adjusted  $P$ -value (BH method)  $< 0.1$ .

## Results

### Deriving combined genotype and dosage information from low-coverage genome data

The *Populus* F1 lines (592) were originally sequenced at a low read depth ( $\sim 0.5\times$  per line), which was sufficient to identify large-scale indels but was not sufficient to reliably haplotype and genotype each individual (Howie et al. 2009; Williams et al. 2012; Martin et al. 2016; Hager et al. 2020). Fortunately, RNA-seq data from 122 of these F1 lines was also available, as well as Illumina short-read sequencing data from two parental lines (*P. deltoides* 45 $\times$ , *P. nigra* 65 $\times$ ) (Henry et al. 2015b; Bastiaanse et al. 2020a). Using these resources, we designed a custom computational process

to derive parental haplotypes and genotype the F1 lines for both parental contributions (Fig. 2 and Supplementary File 3; see Materials and Methods).

The process is divided into 3 steps: parental SNP detection, parental haplotype phasing, and genotyping. Because our population is an F1 population, polymorphisms between the two parental genomes are not informative. Instead, we characterized the 2 pairs of parental haplotypes separately. We first selected 37,556 and 33,035 positions that were heterozygous in the parental clones of *P. deltoides* and *P. nigra*, respectively. Next, we used the RNA-seq reads from 122 diploid F1 lines to derive phased haplotypes for a subset of these SNPs for the two parents separately. Finally, the phased haplotypes were applied to the low-coverage genomic data ( $\sim 0.5\times$  per line) to genotype the remaining F1 individuals. In total, we were able to obtain reliable genotype information for 343 F1 lines (Supplementary Fig. 1c). Last, we generated binned markers (50 SNPs per bin) to increase genotype robustness, and a final common marker set of 507 binned markers was generated for multi-genotype QTL analysis that applied to both the *P. deltoides* and the *P. nigra* genomes (Supplementary Fig. 3).

In terms of dosage variation, among the 343 remaining F1 lines, 54.2% (186 out of 343) were previously characterized to carry at least one indel. Deletions were more prevalent (66.5%) than insertions (33.5%) among these indels, as observed in the original population (Henry et al. 2015b). As described previously, we characterized dosage variation in 546 dosage binned markers, with an average of 6 indels in each dosage marker (Bastiaanse et al. 2019, 2020a; Rodriguez-Zaccaro et al. 2021). Finally, these dosage markers were combined with the natural allelic information to obtain a unified marker list of 507 binned markers, for which we had gathered information about the *P. deltoides* haplotypes, the *P. nigra* haplotypes, and the dosage information for each of the 343 F1 individuals.

### Contributions of natural allelic variation and induced dosage variation on phenotypes can be assigned to QTLs

This population was previously characterized phenotypically (Bastiaanse et al. 2019, 2020a; Rodriguez-Zaccaro et al. 2021) for 3 phenotype categories (38 traits): leaf morphology (22 traits), phenology (7 traits), and biomass (9 traits; Supplementary File 1). In our subset of 343 F1s, using a single model (Trait  $\sim$  Genotype), QTLs were observed for 27 traits. Specifically, 9, 6, and 86 QTLs were identified from *P. deltoides*, *P. nigra*, and dosage genotypes, respectively (Table 1 and Supplementary File 4). Of the dosage QTLs detected here, 77.9% (67 out of 86) were detected in the previous analysis as well (Supplementary Fig. 4; Bastiaanse et al. 2019, 2020a).

Overall comparison of the number of QTLs detected using the three single models reveals that dosage variation has the most pronounced impact on phenotypic variation (Fig. 3, Supplementary Figs. 5 and 6). Interestingly, QTLs observed from the 3 single models did not overlap with each other (Fig. 4), indicating that natural variation in the two parental species, *P. deltoides* and *P. nigra*, and dosage variation may influence these traits independently.

To investigate to what extent indels can affect the identification of allelic QTL results, we selected the 157 lines from this F1 population that did not carry any indels and tested the identification of allelic QTL on this subset. In total, 1 and 8 allelic QTLs were identified from the *P. deltoides* and *P. nigra* parents, respectively (Supplementary Table 1 and Supplementary File 4). Interestingly, there were no common allelic QTLs between the subset population (157 lines) and the full population (343 lines). A subset of both sets

Table 1. QTLs obtained using the three single models.<sup>1</sup>

Category <sup>a</sup>	Traits <sup>b</sup>	Trait ~ <i>P. deltoides</i> <sup>c</sup>			Trait ~ <i>P. nigra</i> <sup>d</sup>			Trait ~ Dosage <sup>e</sup>		
		# of QTL <sup>f</sup>	% explained by single QTL ( $\mu \pm \sigma$ ) <sup>g</sup>	% explained by all QTLs <sup>h</sup>	# of QTL <sup>f</sup>	% explained by single QTL ( $\mu \pm \sigma$ ) <sup>g</sup>	% explained by all QTLs <sup>h</sup>	# of QTL <sup>f</sup>	% explained by single QTL ( $\mu \pm \sigma$ ) <sup>g</sup>	% explained by all QTLs <sup>h</sup>
Biomass	Coppicing_y1	1	4.6 ± 0	4.6	0	NA	NA	2	6.1 ± 1.6	7.0
	Diameter_base <sup>i</sup>	0	NA	NA	0	NA	NA	1	3.4 ± 0	3.4
	Time_serie_diameter_breast_height <sup>k</sup>	0	NA	NA	0	NA	NA	4	3.7 ± 0.1	7.6
	Volume <sup>l</sup>	0	NA	NA	0	NA	NA	1	4.5 ± 0	4.5
	Area_y1_y2	0	NA	NA	0	NA	NA	4	3.1 ± 0.2	9.8
	Circularity_y1_y2	0	NA	NA	0	NA	NA	3	4.4 ± 0.4	11.7
	Horizontal_symmetry_y1_y2	0	NA	NA	0	NA	NA	1	10.8 ± 0	10.8
	Width_y1_y2	1	4.0 ± 0	4.0	1	4.2 ± 0	4.2	4	3.2 ± 0.3	10.0
	Indent_depth_y1_y2	0	NA	NA	0	NA	NA	1	3.2 ± 0	3.2
	Indent_width_y1_y2	0	NA	NA	0	NA	NA	1	4.8 ± 0	4.8
	Num_Indents_y1_y2	0	NA	NA	0	NA	NA	6	3.9 ± 0.7	19.8
	PC1:PC2_y1_y2	0	NA	NA	1	4.1 ± 0	4.1	5	4.2 ± 0.1	10.7
	PC1:PC3_y1_y2	0	NA	NA	0	NA	NA	7	6.2 ± 0.9	23.7
	PC1:PC4_y1_y2	0	NA	NA	1	4.1 ± 0	4.1	4	4.6 ± 1.2	12.8
	PC1_y1_y2	0	NA	NA	0	NA	NA	8	5.6 ± 0.8	21.1
	PC3:PC4_y1_y2	1	4.3 ± 0	4.3	0	NA	NA	2	3.2 ± 0.1	6.6
	PC4_y1_y2	0	NA	NA	0	NA	NA	3	4.4 ± 0.9	13.4
Phenology	Perimeter_y1_y2	0	NA	NA	0	NA	NA	2	3.1 ± 0.1	6.4
	Perimeter2:Area2_y1_y2	1	4.0	4.0	0	NA	NA	0	NA	NA
	Length:width_y1_y2	1	4.7 ± 0	4.7	0	NA	NA	7	6.9 ± 0.6	19.3
	Length_y1_y2	0	NA	NA	0	NA	NA	0	NA	NA
	Bud_burst_y1_y2	0	NA	NA	1	5.6 ± 0	5.6	2	6.1 ± 0.6	6.6
	Color_y1_y2_y3	1	4.1	4.1	0	NA	NA	5	4.6 ± 0.4	14.2
	Drop_y1_y2_y3	1	4.4 ± 0	4.4	0	NA	NA	2	5.9 ± 1.2	7.9
	Green_canopy_duration_y1_y2	0	NA	NA	0	NA	NA	2	4.9 ± 0.1	6.8
	Time_serie_bud_burst_y1_y2	0	NA	NA	2	4.5 ± 0.8	7.2	2	8.6 ± 1.2	17.7
	Time_serie_color_y1_y2_y3	1	3.8 ± 0	3.8	0	NA	NA	3	4.5 ± 0.5	9.7
	Time_serie_drop_y1_y2_y3	1	4.8	4.8	0	NA	NA	3	5.2 ± 1.5	12.7

<sup>a</sup> Three major phenotypic categories (Biomass, Leaf morphology, Phenology) used for QTL analysis in this study.

<sup>b</sup> Shortcuts of trait names applied in QTL analysis. The full explanation of traits can be found in Supplementary File 2.

<sup>c-e</sup> Three models used for QTL analysis. Traits represent phenotypic data. *P. deltoides*, *P. nigra* and Dosage represent *P. deltoides* genotypes, *P. nigra* genotypes and Dosage states, respectively.

<sup>f</sup> Number of QTLs observed in each trait.

<sup>g</sup> Phenotypic variance explained by every observed QTL on average.

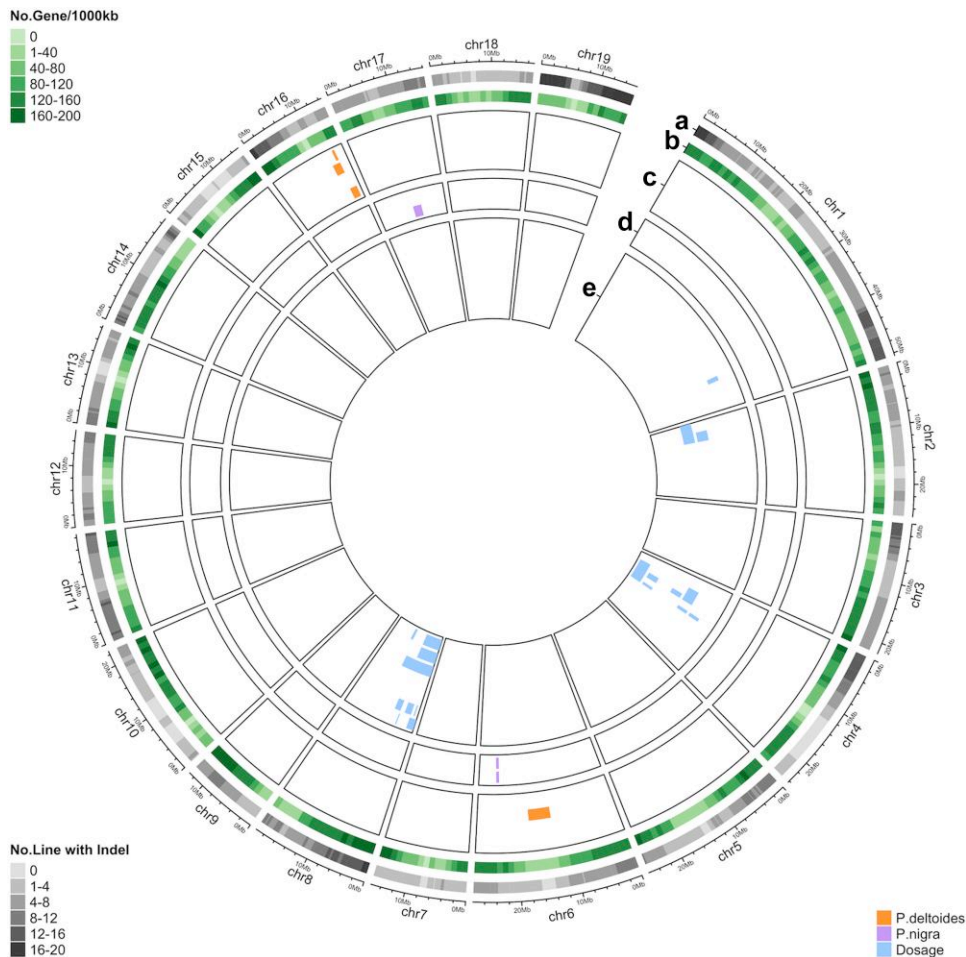
<sup>h</sup> Total phenotypic variance explained by all of the observed QTLs in one trait.

<sup>i</sup> Year 1 refers to the year 2014; Year 2 refers to the year 2015; Year 3 refers to the year 2016.

<sup>j</sup> The 2 biomass traits without year information (Diameter and Volume) are measured at a single time point corresponding to the day of harvest (December 2016).

<sup>k</sup> Biomass trait Time\_serie\_diameter\_breast\_height was the diameter at breast height, measuring as a continuous time series.

<sup>l</sup> Only the 27 traits with observed QTLs were shown here. The complete traits list is in [Supplementary File 2](#).



**Fig. 3.** Observed QTLs for phenology traits using single models. a) Number of lines carrying indels under each bin. b) Gene density across the genome. (c to e) QTLs detected from *P. deltoides* (c), *P. nigra* (d), and dosage (e) genotypes. The traits from outermost to innermost in each track are c) Color\_y1\_y2\_y3, Drop\_y1\_y2\_y3, Time\_serie\_color\_y1\_y2\_y3, Time\_serie\_drop\_y1\_y2\_y3; d) Bud\_burst\_y1\_y2, Time\_serie\_bud\_burst\_y1\_y2; e) Bud\_burst\_y1\_y2, Color\_y1\_y2\_y3, Drop\_y1\_y2\_y3, Green\_canopy\_duration\_y1\_y2, Time\_serie\_bud\_burst\_y1\_y2, Time\_serie\_color\_y1\_y2\_y3, and Time\_serie\_drop\_y1\_y2\_y3. Phenotypic data were obtained from previous reports, and detailed trait information is summarized in Supplementary File 1.

of allelic QTLs overlapped with previously published QTLs. For example, for the allelic QTLs in the full population, *P. nigra* QTLs on chromosomes 6 and 17 for phenology-related traits (bud burst) were consistent with previously reported allelic QTLs (Frewen et al. 2000; Rohde et al. 2011; Fabbrini et al. 2012). For the allelic QTLs in the subset population, *P. nigra* QTLs on chromosome 3 for phenology-related traits (bud burst) and leaf shape were consistent with reported QTLs in *Populus* (Rohde et al. 2011; Xia et al. 2018). These results suggest that the identification of allelic QTL in the full population is significantly affected by the presence of the large-scale indels, which could completely mask the effect of some or all of the allelic QTLs when present.

Coming back to the full population, allelic variation and dosage variation explained 4.94 and 11.27% phenotypic variance, respectively (Fig. 5a). To investigate whether combining the effects of natural allelic variation and induced dosage variation can explain a larger percentage of the observed phenotypic variance, we used a multivariate model to detect allelic and dosage QTLs simultaneously. We first selected 12 traits for which both allelic and dosage variation were associated with detected QTLs (Supplementary File 5). Integration of QTLs from the three single models explained 15.51% of the observed phenotypic variance in these 12 traits. This percentage was significantly higher than the percentage of

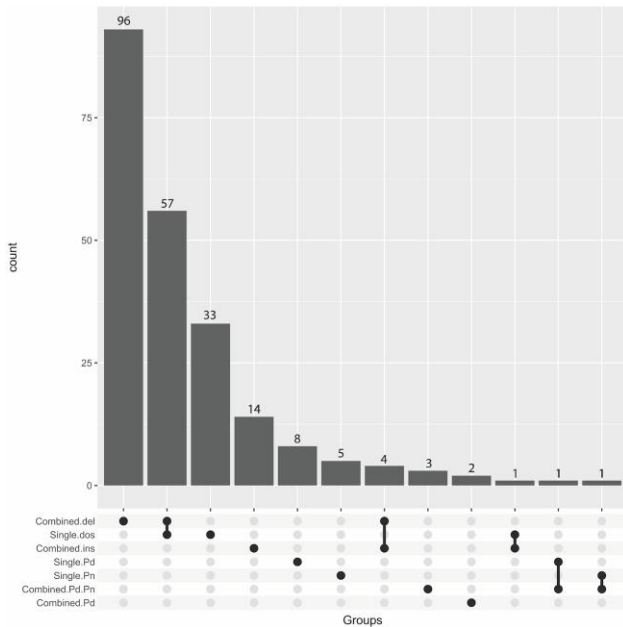
variance explained by either allelic variation alone (Tukey's test,  $P < 0.001$ ) or dosage variation alone (Tukey's test,  $P = 0.019$ ; Fig. 5a and Supplementary File 6).

To investigate the molecular mechanism underlying the detected QTLs, we identified the genes located within the observed QTL regions and examined their differential expression levels based on the leaf transcriptomic data from our previous study (Bastiaanse et al. 2020a) (Supplementary File 7). GO enrichment analysis indicated that differentially expressed genes (DEGs) associated with allelic QTLs were suggestively enriched with translation ( $0.05 < P\text{-value} < 0.1$ ; Supplementary Fig. 7), while DEGs associated with dosage QTLs were significantly enriched with stress response processes (Bastiaanse et al. 2020a).

### A combined univariate model helps refine our understanding of trait regulation

Allelic and dosage variation effects may also interact with each other. For example, dosage effects are expected to be different if the causal gene also carries a loss-of-function allele (Fig. 6). To better understand the interaction between the effects of natural allelic variation and induced dosage variation, we combined the information from the three variation types and assigned each combined genotype to a unique state. For example, D1.N1.1 on

marker 1 represents the individuals with *P. deltooides* haplotype 1, *P. nigra* haplotype 1, and 1 *P. nigra* copy for marker 1. In this model, all individuals fit into one of 10 possible states, and we can

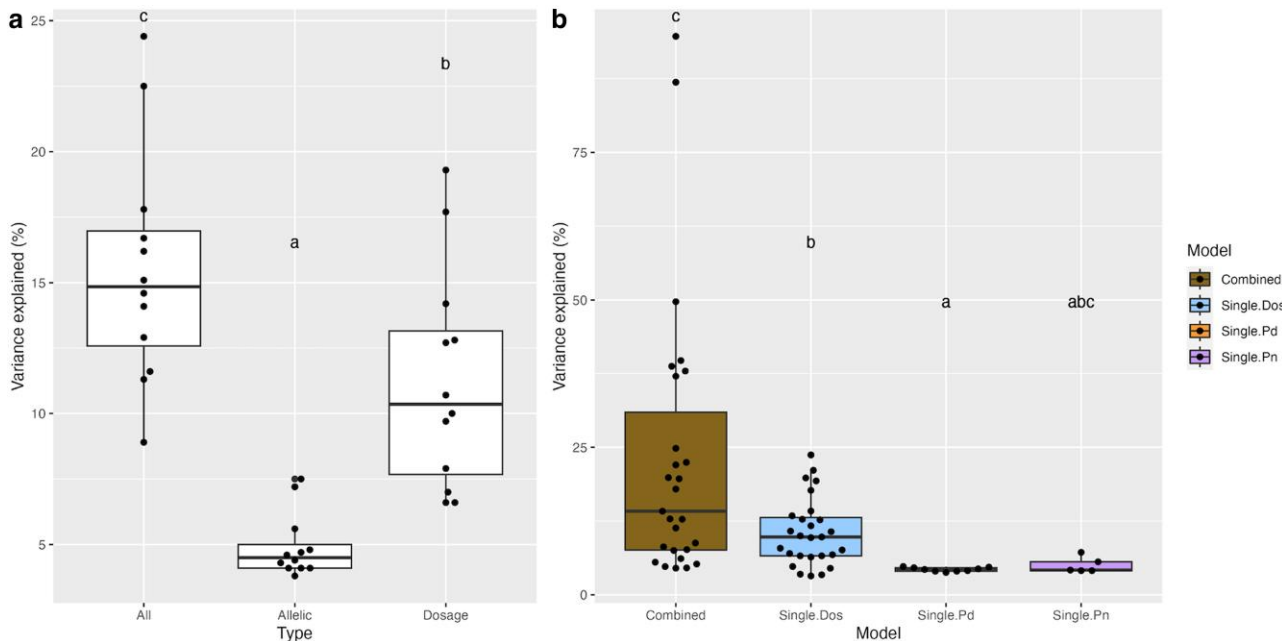


**Fig. 4.** Number of QTLs detected from single and combined models. Single.dos: QTLs from the single model Trait ~ Dosage. Single.Pd: QTLs from the single model Trait ~ *P. deltooides* haplotypes. Single.Pn: QTLs from the single model Trait ~ *P. nigra* haplotypes. Combined.del: QTLs associated with a deletion. Combined.ins: QTLs associated with an insertion. Combined.Pd: QTLs associated with the *P. deltooides* haplotypes. Combined.Pn: QTLs associated with an insertion and the *P. deltooides* and *P. nigra* haplotypes.

incorporate these integrated genotypic states into a univariate model, such as Trait ~ States (Supplementary File 2). Next, pairwise comparisons can be performed between groups in the different genotype states using linear regression. Loci exhibit significant phenotypic differences through pairwise comparison and were assigned as QTLs. We categorized these QTLs into 6 groups (deletion, insertion, deletion + insertion, *P. deltooides*, *P. nigra*, and *P. deltooides* + *P. nigra*), according to the phenotypic differences between compared genotypic states (see Materials and Methods).

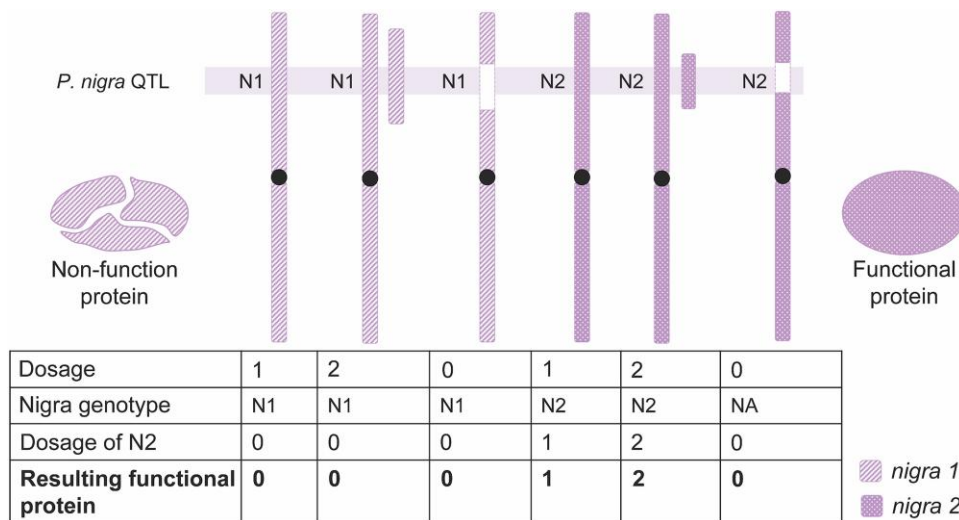
In total, we observed 163 QTLs from the combined model that belonged to 4 different groups [deletion, insertion, *P. deltooides*, and *P. deltooides* + *P. nigra* (Table 2 and Supplementary File 4)]. Among these 4 groups, most QTLs were associated with deletions (Fig. 4). This result is consistent with expectation from single models, since dosage variation was associated with QTLs much more often than allelic differences (Fig. 4). These findings are also illustrated in the Circos plots, where deletions (Fig. 7c, Supplementary Figs. 8c and 9c) are associated with most QTLs, followed by insertions (Fig. 7d, Supplementary Figs. 8d and 9d), and allelic variation (Fig. 7e, Supplementary Figs. 8e and 9e). These observations confirmed that dosage variation drives phenotypic variation for most traits in our population, while variation in parental haplotype did not strongly modulate the effects of dosage variation.

The combined model detected only a few instances where the QTLs observed by different genotypes overlapped (Fig. 4). These QTLs were associated with leaf shape and localized on chromosome 17 (Supplementary Fig. 9), where they were associated with both deletions and insertions. This result is consistent with the outcome from the single model analysis, indicating that dosage and allelic variation may independently affect the examined traits.



**Fig. 5.** Phenotypic variance explained using the single and combined models. a) Phenotypic variance explained by allelic and dosage variation using single models. 12 traits were selected for the observation of both allelic and dosage QTLs. On the x-axis, All represents the variance explained by all QTLs identified using the three single models. Allelic represents the variance explained by the collection of QTLs from *P. deltooides* and *P. nigra* haplotypes. Dosage represents the variance explained by dosage variation. b) Comparison of the percentage of phenotypic variance explained by the single and combined models. On the x-axis, Combined represents the variance explained by QTLs observed from the combined model. Single.Dos, Single.Pd and Single.Pn represent QTLs identified from three single models associated with dosage variation, *P. deltooides* haplotypes, and *P. nigra* haplotypes, respectively. Statistical significance was calculated through pairwise permutation tests ( $P$ -value < 0.05).





**Fig. 6.** Representative diagram of possible interplay between *P. nigra* haplotypes and dosage variation. N1 (*P. nigra 1*) encodes a non-functional protein, while N2 (*P. nigra 2*) encodes a functional protein. Copy number changes on N1 have no effect on phenotypes, while copy number changes on N2 result in dramatic differences on phenotypic outcomes. Chromosomes inherited from *P. deltoides* (not shown) are always present in one copy.

**Table 2.** QTLs obtained using the combined model.

Phenotype (# of traits)	Groups <sup>a</sup>	Total # of QTL	# of traits with QTL	Variance explained by single QTL ( $\mu \pm \sigma$ ) (%)	Variance explained by all QTLs of a trait ( $\mu \pm \sigma$ ) (%)
Biomass (9)	deletion	14	6	$7.6 \pm 6.7$	$12.3 \pm 5.6$
	deletion	84	12	$6.4 \pm 4.2$	$27.1 \pm 30.2$
	insertion	12	5	$13.0 \pm 12.0$	
Phenology (7)	<i>P. deltoides</i>	2	2	$6.9 \pm 1.1$	
	deletion	39	6	$5.8 \pm 4.4$	$24.5 \pm 14.1$
	insertion	7	3	$10.3 \pm 7.0$	
	<i>P. deltoides</i> + <i>P. nigra</i>	5	3	$5.0 \pm 1.1$	

<sup>a</sup> The observed QTLs were categorized into groups based on their origin: deletion, insertion, deletion + insertion, *P. deltoides*, *P. nigra*, *P. deltoides* + *P. nigra*. This table only shows groups for which QTLs were identified.

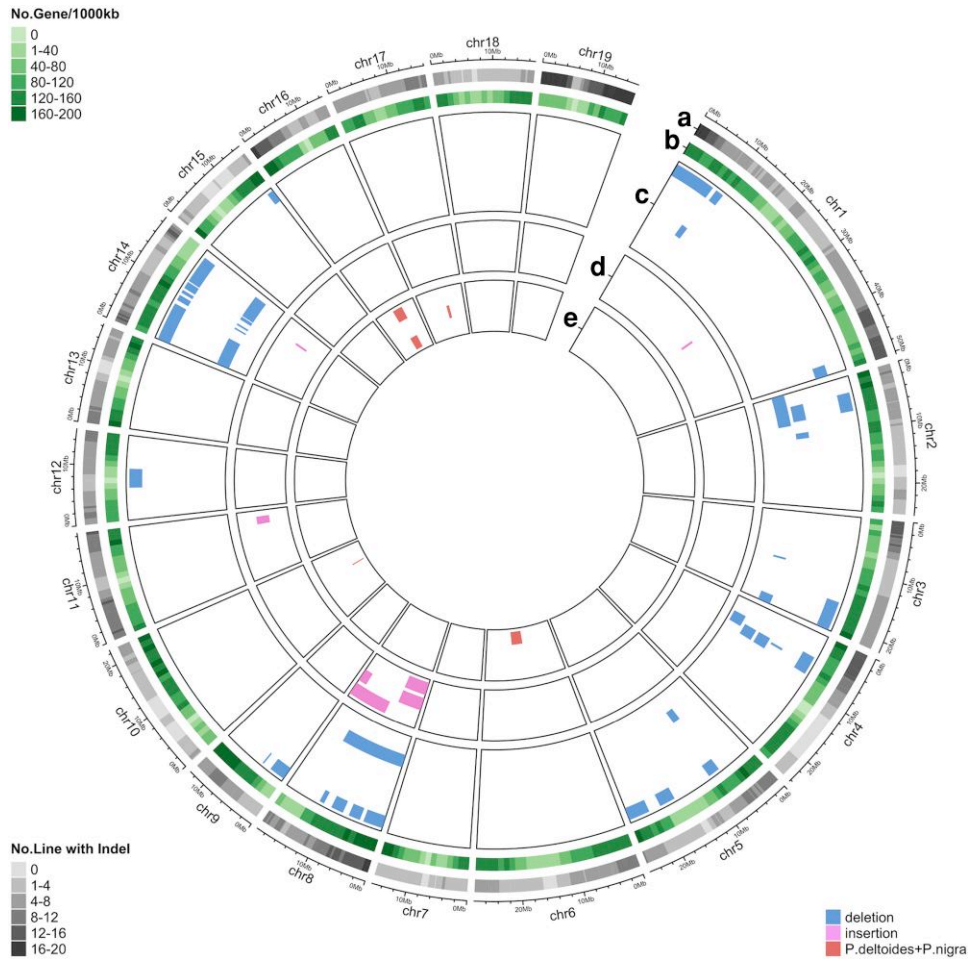
Finally, we investigated the percentage of phenotypic variance explained by the QTLs identified using the combined model. To calculate phenotypic variance for each trait, QTLs belonging to the same trait were merged. Merged QTLs explained on average 23.2% of the phenotypic variance, which is significantly higher than the variance explained from dosage variation only (on average 10.6%) or *P. deltoides* haplotype variation (on average of 4.3%) (permutation test,  $P$ -value < 0.05), and is suggestively higher than only *P. nigra* haplotypes (on average of 5.1%; permutation test,  $P$ -value < 0.1) (Fig. 5b and Supplementary File 6). Meanwhile, we observed that the integration of QTLs from all three single models explained a smaller percentage of the phenotypic variance than the QTLs from the combined model (12.2% vs 23.2%; permutation test,  $P$ -value < 0.05). Presumably, the increase originates from the QTLs identified using the combined model but not identified using the single models. Some of these QTLs were shown to be suggestive ( $0.05 < P$ -value < 0.1) when using the single models (Fig. 8a, chromosomes 3, 4), while others were not identified at all using the single model (Fig. 7, chromosome 14). These findings confirm the advantage of using a combined model approach.

## Discussion

Identifying candidate genes underlying a target trait is a crucial step toward understanding the mechanisms affecting the trait, and for applying this knowledge to plant breeding. Quantitative

trait loci (QTL) analysis, which typically correlates SNP to traits or phenotype-associated features such as gene expression and RNA alternative splicing (Brem et al. 2002; Li et al. 2016), is an efficient approach for this endeavor. Besides SNPs, other genetic features such as dosage variation (Bastiaanse et al. 2019, 2020a; Rodriguez-Zaccaro et al. 2021) can affect traits of interest. A unique *Populus* population, which carries natural allelic variation and induced dosage variation was previously established (Henry et al. 2015b). Previous analysis demonstrated few point mutations and small indels in this population (Henry et al. 2015b), indicating that preexisting SNPs and induced large-scale indels are the major sources of genetic variation in this population and presumably drive the observed phenotypic variation. In our study, we aimed to investigate the effects of natural allelic variation and induced dosage variation on quantitative traits. In general, our results indicate no overlap between QTLs from natural and dosage variation in our system.

A single model approach was used to describe the correlation between each source of variation and target traits. *P. deltoides* and *P. nigra* genotypic information allowed for the identification of QTLs between different haplotypes within each parental species. Compared with previous QTL analysis in other *Populus* cross populations (Rae et al. 2009; Rohde et al. 2011; Fabbrini et al. 2012), our study found fewer allelic QTLs. As demonstrated by our research identifying QTLs in the subset of trees that do not carry large indels, this may be because the presence of many large



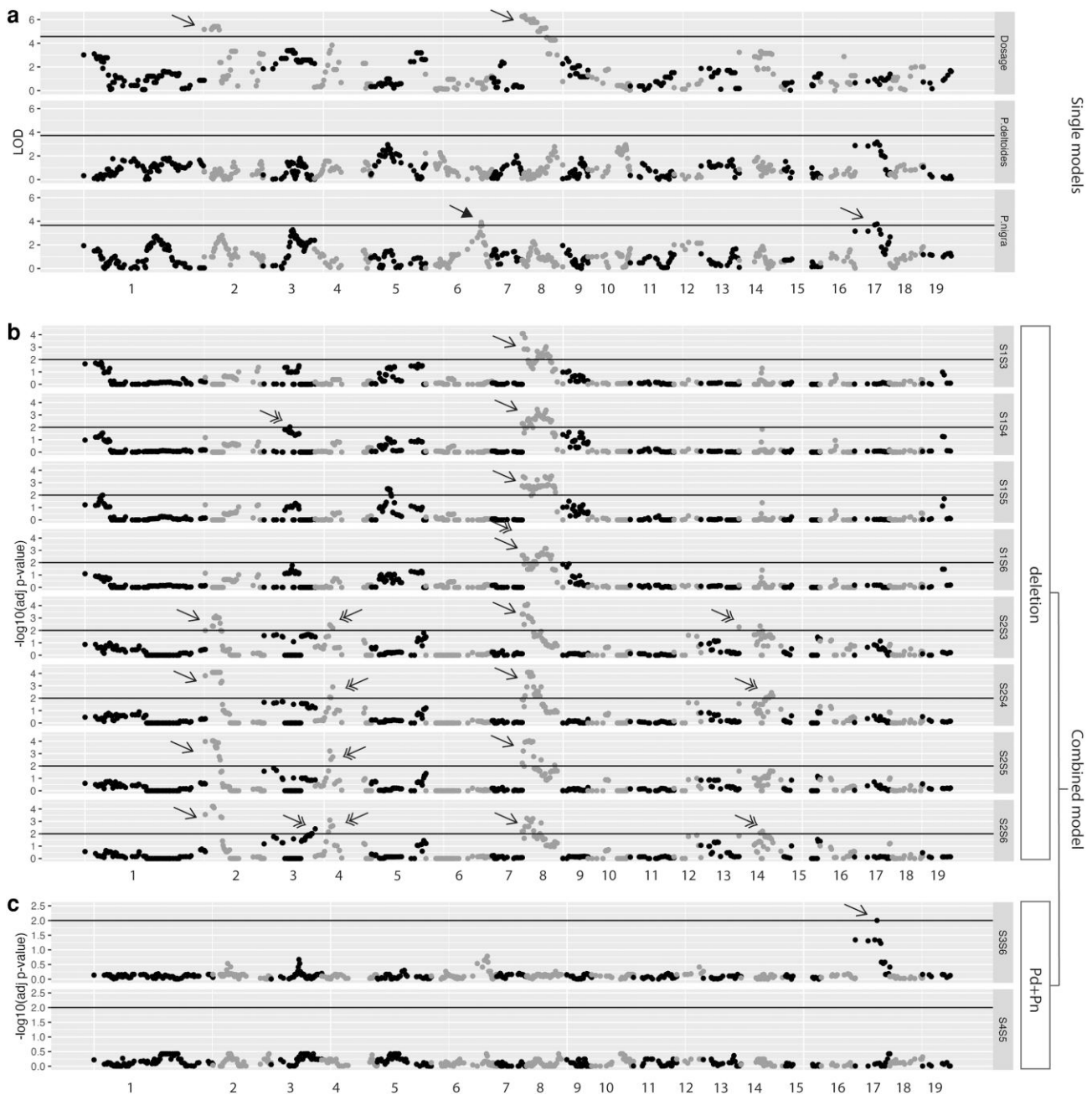
**Fig. 7.** Observed QTLs for the phenology traits using the combined model. a) Number of lines carrying indels under each bin. b) Gene density across the genome. (c to e) QTLs detected based on variation in deletion (c), insertion (d), and *P. deltooides* + *P. nigra* haplotypes (e). The traits from outermost to innermost in each track are c) Bud\_burst\_y1\_y2, Color\_y1\_y2\_y3, Drop\_y1\_y2\_y3, Green\_canopy\_duration\_y1\_y2, Time\_serie\_bud\_burst\_y1\_y2, Time\_serie\_drop\_y1\_y2\_y3; d) Color\_y1\_y2\_y3, Drop\_y1\_y2\_y3, Time\_serie\_drop\_y1\_y2\_y3; e) Color\_y1\_y2\_y3, and Time\_serie\_bud\_burst\_y1\_y2, Time\_serie\_color\_y1\_y2\_y3.

indels may mask the observation of QTLs associated with natural allelic variation. For example, dosage-sensitive genes can play the trans-regulatory factors and affect large numbers of genes across the genome (Bastiaanse et al. 2020a). Interestingly, we found no overlap between the *P. deltooides* QTLs and the *P. nigra* QTLs. A previous study (Rohde et al. 2011) also reported no overlap between *P. deltooides* and *P. nigra* QTLs when the two species were used as the two parents of the same population (*P. deltooides* × *P. nigra*), which is consistent with our results. However, in the same study, shared QTLs were observed if *P. deltooides* and *P. nigra* were used in different crosses (Rohde et al. 2011). This might be because, if both *P. deltooides* and *P. nigra* carry genetic variation at the same location and both parental genotypes affect the trait, the source of phenotypic variation is more difficult to identify. Instead, when they are crossed with other *Populus* species, which do not carry variations that affect the trait, QTLs can be detected. With the current data, it is difficult to determine if the pathways that control these three phenotypic categories—biomass, leaf morphology, and phenology—are similar or not.

Dosage variation was induced by  $\gamma$  irradiation of *P. nigra* pollen and all resulting indels are located on the *P. nigra* chromosomes (Henry et al. 2015b). Therefore, we expected to observe some overlap between *P. nigra* allelic QTLs and dosage QTLs. For example, if the *P. nigra* QTL is associated with alleles affecting gene expression

levels, then dosage and allelic variation would have similar effects, with decreased protein level to 0 in the case of deletion or increased levels to two-folds in the case of an insertion. According to this model, both *P. nigra* QTL and dosage QTL act through dosage-dependent regulation of the target trait. The dosage-dependent behavior is consistent with additivity and has been described as the basis for quantitative variation (Lukens and Doebley 1999; Frary et al. 2000).

Surprisingly, dosage QTLs and allelic QTLs do not overlap (Fig. 4). There can be multiple reasons for this outcome, depending on the mechanisms underlying the QTL at hand. For loci that display only allelic QTL, the impact of 1× to 2× constitutive dosage variation might be insufficient to affect protein function, whereas allelic variation could potentially affect gene function through more drastic modifications, such as significantly altering the expression pattern, or directly affecting the protein function if there are changes in the amino acid sequence. It is also possible that dosage variation at those loci was absent or too infrequent in the indel population for the detection of a dosage QTL effect. Indeed, over 50% of the *P. nigra* loci are connected to fewer than 5 indels (Henry et al. 2015b), limiting the statistical power of our dosage QTL analysis. Finally, gene dosage compensation is another possible explanation, in which the structural gene dosage effect is canceled by an inverse regulatory effect, exerted either within the same locus or from an unlinked region (Birchler et al.



**Fig. 8.** Examples of QTLs identified using the single and combined models on the phenology-related trait *Time\_serie\_bud\_burst\_y1\_y2*. Each dot represents a genetic marker. The X-axis indicates genomic positions. The y-axis indicates the LOD scores (a) or adjusted P-value with negative log<sub>10</sub> fold (b, c). Dots above the horizontal lines were selected as QTLs. These QTLs were categorized as observed only in the single models (triangle arrow), only in the combined model (double arrows) and in both single and combined models (single arrows). a) LOD scores of genetic markers from three single models. Top: Trait ~ Dosage; Middle: Trait ~ *P. deltoides* haplotype; Bottom: Trait ~ *P. nigra* haplotype. (b, c) Combined model categorized QTLs into 6 groups based on their origin. Deletion (b) and *P. deltoides* + *P. nigra* (Pd + Pn) (c) groups are shown here because QTLs were identified in these comparisons. b) Each plot represents one pairwise comparison between two genotype states for which genotypes are equal but dosage varies. c) Each plot represents one pairwise comparison between two genotype states for which dosage is equal but the *P. deltoides* and *P. nigra* haplotypes vary.

1990; Birchler and Veitia 2012). The combination of these two opposite effects would result in no significant change of gene expression. Conversely, for loci for which only dosage QTLs were detected, it is possible that natural allelic variation is not present at these loci, or that it has too subtle an impact to affect the associated phenotype. The gene balance hypothesis can explain the success in detecting dosage QTLs and the failure of detecting allelic QTLs in the case of genes encoding proteins that are part of multisubunit complexes. According to this hypothesis, traits

regulated by multisubunit complexes are particularly sensitive to dosage. Copy number variations involving the genes encoding these subunits can perturb their stoichiometry, leading to a dramatic alteration in the protein complex function and, ultimately, impacting the connected traits (Birchler and Veitia 2012). On the other hand, sequence variation with subtle effects would be difficult to identify (Birchler and Veitia 2021).

Integration of QTLs from dosage and allelic variation, compared to either allelic QTLs or dosage QTLs alone, significantly improved

the percentage of variance explained (Fig. 5a). These results suggest that a large proportion of the phenotypic variation was caused by the induced large-scale indels, but not all of it. Some of the phenotypic variation is caused by natural allelic variation, and taking both the allelic and dosage variation into account improves phenotypic prediction. However, the integration of all identified QTLs from the single models explained, on average, only 12.2% of the observed phenotypic variance, indicating that the majority of the variance remains unexplained. This could be due to the interaction between allelic and dosage variation. For example, dosage effects are expected to be allele-sensitive if the responsible gene is heterozygous for a null allele (Fig. 6). As a result, single models focusing solely on natural allelic variation or induced dosage variation are not able to identify these interactive effects.

We next developed a combined model including all variation types. We categorized the QTLs into 6 groups based on the following types of variation: deletion, insertion, deletion + insertion, *P. deltoides* haplotypes, *P. nigra* haplotypes, and *P. deltoides* + *P. nigra* haplotypes. Most QTLs were associated with dosage-related groups, with deletions being the most common cause, followed by insertions. QTLs associated with allelic variation (*P. deltoides*, *P. nigra*, and *P. deltoides* + *P. nigra* haplotypes) were the least common. Most QTLs were observed within dosage-related groups. Possibly, this is because dosage variants were newly induced and have not experienced selection. There was no overlap between allelic and dosage QTLs, which is consistent with the results obtained using the single models.

Taken together, we investigated the contribution of natural allelic variation and induced dosage variation in F1 *Populus* hybrids on quantitative traits. We found no overlap between allelic and dosage variation QTLs, suggesting that the naturally occurring sequence polymorphisms and the induced structural variation influence the traits under different constraints and through different mechanisms. Integrating the QTLs from allelic and dosage variation significantly increased the proportion of phenotypic explained variance compared to considering only allelic or dosage QTLs. A new method was designed to include all types of variation simultaneously for QTL analysis, and it was applied to investigate the interaction between allelic and dosage variation in detail. This novel approach significantly increased the explained proportion of phenotypic variance and revealed that genomic fragment deletion had the most pronounced effect on traits. The future direction would be to identify responsible genes within the QTL intervals as a next step toward helping the development of *Populus* clones with commercial benefits.

## Data availability

The sequences reported in this paper were previously deposited (Henry et al. 2015b; Bastiaanse et al. 2020a) and can be found in the National Center for Biotechnology Information BioProject Database (BioProject ID: PRJNA241273 and PRJNA646735) (Henry et al. 2015a; Bastiaanse et al. 2020b).

Supplemental material available at G3 online.

## Acknowledgments

We thank Meric Lieberman for assistance on bioinformatics.

## Funding

This work was supported by the U.S. Department of Energy, Office of Science, Office of Biological and Environmental Research (B.E.R.) Grant nr. DESC0007183 to L.C., and by the National Science

Foundation, Plant Genome Research Program award IOS-1956429, and National Science Foundation, Plant Genome Integrative Organismal Systems (IOS) Grant PGRP IOS-2055260.

## Conflicts of interest

The authors declare no competing interest.

## Literature cited

- Alonge M, Wang X, Benoit M, Soyk S, Pereira L, Zhang L, Suresh H, Ramakrishnan S, Maumus F, Ciren D, et al. 2020. Major impacts of widespread structural variation on gene expression and crop improvement in tomato. *Cell*. 182(1):145–161.e23. doi:10.1016/j.cell.2020.05.021.
- Alonso-Blanco C, Aarts MG, Bentsink L, Keurentjes JJ, Reymond M, Vreugdenhil D, Koornneef M. 2009. What has natural variation taught us about plant development, physiology, and adaptation? *Plant Cell*. 21(7):1877–1896. doi:10.1105/tpc.109.068114.
- Alonso-Blanco C, Blankestijn-de Vries H, Hanhart CJ, Koornneef M. 1999. Natural allelic variation at seed size loci in relation to other life history traits of *Arabidopsis thaliana*. *Proc Natl Acad Sci U S A*. 96(8):4710–4717. doi:10.1073/pnas.96.8.4710.
- Bastiaanse HLS, Henry IM, Tsai H, Lieberman M, Canning C, Comai L, Groover A. 2020a. A systems genetics approach to deciphering the effect of dosage variation on leaf morphology in *Populus*. *Plant Cell*. 33(4):940–960. doi:10.1093/plcell/koaa016.
- Bastiaanse HLS, Henry IM, Tsai H, Lieberman MC, Canning C, Comai L, Groover AT. 2020b. A systems genetics approach to deciphering the effect of dosage variation on leaf morphology in *Populus*. GenBank BioProject PRJNA646735. <https://identifiers.org/bioproject:PRJNA241273>.
- Bastiaanse H, Zinkgraf M, Canning C, Tsai H, Lieberman M, Comai L, Henry I, Groover A. 2019. A comprehensive genomic scan reveals gene dosage balance impacts on quantitative traits in *Populus* trees. *Proc Natl Acad Sci U S A*. 116(27):13690–13699. doi:10.1073/pnas.1903229116.
- Benjamini Y, Hochberg Y. 1995. Controlling the false discovery rate: a practical and powerful approach to multiple testing. *J R Stat Soc Series B Stat Methodol*. 57(1):289–300. doi:10.1111/j.2517-6161.1995.tb02031.x.
- Birchler JA, Hiebert JC, Paigen K. 1990. Analysis of autosomal dosage compensation involving the alcohol dehydrogenase locus in *Drosophila melanogaster*. *Genetics*. 124(3):679–686. doi:10.1093/genetics/124.3.677.
- Birchler JA, Veitia RA. 2010. The gene balance hypothesis: implications for gene regulation, quantitative traits and evolution. *New Phytol*. 186(1):54–62. doi:10.1111/j.1469-8137.2009.03087.x.
- Birchler JA, Veitia RA. 2012. Gene balance hypothesis: connecting issues of dosage sensitivity across biological disciplines. *Proc Natl Acad Sci U S A*. 109(37):14746–14753. doi:10.1073/pnas.1207726109.
- Birchler JA, Veitia RA. 2021. One hundred years of gene balance: how stoichiometric issues affect gene expression, genome evolution, and quantitative traits. *Cytogenet Genome Res*. 161(10–11): 529–550. doi:10.1159/000519592.
- Brem RB, Yvert G, Clinton R, Kruglyak L. 2002. Genetic dissection of transcriptional regulation in budding yeast. *Science*. 296(5568): 752–755. doi:10.1126/science.1069516.
- Brewbaker JL, Emery GC. 1961. Pollen radiobotany. *Radiat Bot*. 1: 101–154. doi:10.1016/S0033-7560(61)80015-X.
- Carbonell-Bejerano P, Royo C, Torres-Pérez R, Grimplet J, Fernandez L, Franco-Zorilla JM, Lijavetzky D, Baroja E, Martínez J,

- García-Escudero E, et al. 2017. Catastrophic unbalanced genome rearrangements cause somatic loss of berry color in grapevine. *Plant Physiol.* 175(2):786–801. doi:[10.1104/pp.17.00715](https://doi.org/10.1104/pp.17.00715).
- Cook DE, Lee TG, Guo X, Melito S, Wang K, Bayless AM, Wang J, Hughes TJ, Willis DK, Clemente TE, et al. 2012. Copy number variation of multiple genes at Rhg1 mediates nematode resistance in soybean. *Science.* 338(6111):1206–1209. doi:[10.1126/science.1228746](https://doi.org/10.1126/science.1228746).
- Davey JW, Hohenlohe PA, Etter PD, Boone JQ, Catchen JM, Blaxter ML. 2011. Genome-wide genetic marker discovery and genotyping using next-generation sequencing. *Nat Rev Genet.* 12(7):499–510. doi:[10.1038/nrg3012](https://doi.org/10.1038/nrg3012).
- Díaz A, Zikhali M, Turner AS, Isaac P, Laurie DA. 2012. Copy number variation affecting the Photoperiod-B1 and Vernalization-A1 genes is associated with altered flowering time in wheat (*Triticum aestivum*). *PLoS One.* 7(3):e33234. doi:[10.1371/journal.pone.0033234](https://doi.org/10.1371/journal.pone.0033234).
- Doerge RW, Churchill GA. 1996. Permutation tests for multiple loci affecting a quantitative character. *Genetics.* 142(1):285–294. doi:[10.1093/genetics/142.1.285](https://doi.org/10.1093/genetics/142.1.285).
- Duan Z, Zhang M, Zhang Z, Liang S, Fan L, Yang X, Yuan Y, Pan Y, Zhou G, Liu S, et al. 2022. Natural allelic variation of GmSTO5 controlling seed size and quality in soybean. *Plant Biotechnol J.* 20(9):1807–1818. doi:[10.1111/pbi.13865](https://doi.org/10.1111/pbi.13865).
- Elshire RJ, Glaubitz JC, Sun Q, Poland JA, Kawamoto K, Buckler ES, Mitchell SE. 2011. A robust, simple genotyping-by-sequencing (GBS) approach for high diversity species. *PLoS One.* 6(5):e19379. doi:[10.1371/journal.pone.0019379](https://doi.org/10.1371/journal.pone.0019379).
- Fabbrini F, Gaudet M, Bastien C, Zaina G, Harfouche A, Beritognolo I, Marron N, Morgante M, Scarascia-Mugnozza G, Sabatti M. 2012. Phenotypic plasticity, QTL mapping and genomic characterization of bud set in black poplar. *BMC Plant Biol.* 12:47. doi:[10.1186/1471-2229-12-47](https://doi.org/10.1186/1471-2229-12-47).
- Fisher RA. 1919. XV.—the correlation between relatives on the supposition of Mendelian inheritance. *Earth Environ Sci Trans R Soc Edinb.* 52(2):399–433. doi:[10.1017/S0080456800012163](https://doi.org/10.1017/S0080456800012163).
- Frary A, Nesbitt TC, Grandillo S, Knaap E, Cong B, Liu J, Meller J, Elber R, Alpert KB, Tanksley SD. 2000. Fw2.2: a quantitative trait locus key to the evolution of tomato fruit size. *Science.* 289(5476):85–88. doi:[10.1126/science.289.5476.85](https://doi.org/10.1126/science.289.5476.85).
- Frewen BE, Chen TH, Howe GT, Davis J, Rohde A, Boerjan W, Bradshaw HD Jr. 2000. Quantitative trait loci and candidate gene mapping of bud set and bud flush in populus. *Genetics.* 154(2):837–845. doi:[10.1093/genetics/154.2.837](https://doi.org/10.1093/genetics/154.2.837).
- Golicz AA, Bayer PE, Barker GC, Edger PP, Kim H, Martinez PA, Chan CK, Severn-Ellis A, McCombie WR, Parkin IA, et al. 2016. The pan-genome of an agronomically important crop plant Brassica oleracea. *Nat Commun.* 7(1):13390. doi:[10.1038/ncomms13390](https://doi.org/10.1038/ncomms13390).
- Gupta PK, Rustgi S, Mir RR. 2008. Array-based high-throughput DNA markers for crop improvement. *Heredity (Edinb).* 101(1):5–18. doi:[10.1038/hdy.2008.35](https://doi.org/10.1038/hdy.2008.35).
- Hager P, Mewes HW, Rohlf M, Klein C, Jeske T. 2020. SmartPhase: accurate and fast phasing of heterozygous variant pairs for genetic diagnosis of rare diseases. *PLoS Comput Biol.* 16(2):e1007613. doi:[10.1371/journal.pcbi.1007613](https://doi.org/10.1371/journal.pcbi.1007613).
- Henry IM, Zinkgraf MS, Groover AT, Comai L. 2015a. A dosage-based resource for functional genomics in PoplarOrganism: Populus x canadensis. GenBank BioProject PRJNA241273. <https://identifiers.org/bioproject:PRJNA241273>.
- Henry IM, Zinkgraf MS, Groover AT, Comai L. 2015b. A system for dosage-based functional genomics in poplar. *Plant Cell.* 27(9):2370–2383. doi:[10.1105/tpc.15.00349](https://doi.org/10.1105/tpc.15.00349).
- Howie BN, Donnelly P, Marchini J. 2009. A flexible and accurate genotype imputation method for the next generation of genome-wide association studies. *PLoS Genet.* 5(6):e1000529. doi:[10.1371/journal.pgen.1000529](https://doi.org/10.1371/journal.pgen.1000529).
- Huang X, Effgen S, Meyer RC, Theres K, Koornneef M. 2012. Epistatic natural allelic variation reveals a function of AGAMOUS-LIKE6 in axillary bud formation in Arabidopsis. *Plant Cell.* 24(6):2364–2379. doi:[10.1105/tpc.112.099168](https://doi.org/10.1105/tpc.112.099168).
- Huang X, Han B. 2014. Natural variations and genome-wide association studies in crop plants. *Annu Rev Plant Biol.* 65(1):531–551. doi:[10.1146/annurev-arplant-050213-035715](https://doi.org/10.1146/annurev-arplant-050213-035715).
- Huang X, Paulo MJ, Boer M, Effgen S, Keizer P, Koornneef M, van Eeuwijk FA. 2011. Analysis of natural allelic variation in Arabidopsis using a multiparent recombinant inbred line population. *Proc Natl Acad Sci U S A.* 108(11):4488–4493. doi:[10.1073/pnas.1100465108](https://doi.org/10.1073/pnas.1100465108).
- Jamann TM, Balint-Kurti PJ, Holland JB. 2015. QTL mapping using high-throughput sequencing. *Methods Mol Biol.* 1284:257–285. doi:[10.1007/978-1-4939-2444-8\\_13](https://doi.org/10.1007/978-1-4939-2444-8_13).
- Jin J-Q, Yao M-Z, Ma C-L, Ma J-Q, Chen L. 2016. Natural allelic variations of TCS1 play a crucial role in caffeine biosynthesis of tea plant and its related species. *Plant Physiol Biochem.* 100:18–26. doi:[10.1016/j.plaphy.2015.12.020](https://doi.org/10.1016/j.plaphy.2015.12.020).
- Li H, Durbin R. 2009. Fast and accurate short read alignment with Burrows–Wheeler transform. *Bioinformatics.* 25(14):1754–1760. doi:[10.1093/bioinformatics/btp324](https://doi.org/10.1093/bioinformatics/btp324).
- Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, Marth G, Abecasis G, Durbin R; 1000 Genome Project Data Processing Subgroup. 2009. The sequence alignment/map format and SAMtools. *Bioinformatics.* 25(16):2078–2079. doi:[10.1093/bioinformatics/btp352](https://doi.org/10.1093/bioinformatics/btp352).
- Li YI, van de Geijn B, Raj A, Knowles DA, Petti AA, Golan D, Gilad Y, Pritchard JK. 2016. RNA splicing is a primary link between genetic variation and disease. *Science.* 352(6285):600–604. doi:[10.1126/science.aad9417](https://doi.org/10.1126/science.aad9417).
- Li Y, Xiao J, Wu J, Duan J, Liu Y, Ye X, Zhang X, Guo X, Gu Y, Zhang L, et al. 2012. A tandem segmental duplication (TSD) in green revolution gene Rht-D1b region underlies plant height variation. *New Phytol.* 196(1):282–291. doi:[10.1111/j.1469-8137.2012.04243.x](https://doi.org/10.1111/j.1469-8137.2012.04243.x).
- Lukens LN, Doebley J. 1999. Epistatic and environmental interactions for quantitative trait loci involved in maize evolution. *Genet Res.* 74(3):291–302. doi:[10.1017/S0016672399004073](https://doi.org/10.1017/S0016672399004073).
- Mangiafico S. 2020. rcompanion: Functions to support extension education program evaluation. R package version. doi:[10.32614/CRAN.package.rcompanion](https://doi.org/10.32614/CRAN.package.rcompanion).
- Martin M, Patterson M, Garg S, Fischer SO, Pisanti N, Klau GW, Schöenhuth A, Marschall T. 2016. WhatsHap: fast and accurate read-based phasing. *bioRxiv.* 085050. doi:[10.1101/085050](https://doi.org/10.1101/085050).
- McMullen MD. 2003. Quantitative trait locus analysis as a gene discovery tool. *Methods Mol Biol.* 236:141–154. doi:[10.1385/1-59259-413-1:141](https://doi.org/10.1385/1-59259-413-1:141).
- Mottinger JP. 1970. The effects of X rays on the bronze and shrunken loci in maize. *Genetics.* 64(2):259–271. doi:[10.1093/genetics/64.2.259](https://doi.org/10.1093/genetics/64.2.259).
- Nuffer MG. 1957. Additional evidence on the effect of X-ray and ultraviolet radiation on mutation in maize. *Genetics.* 42(3):273–282. doi:[10.1093/genetics/42.3.273](https://doi.org/10.1093/genetics/42.3.273).
- Osborne TS. 1957. Past, present and potential uses of radiation in southern plant breeding. In: *Proceedings of Ninth Oak Ridge Regional Symposium.* p. 5–10.
- Pinosio S, Giacomello S, Faivre-Rampant P, Taylor G, Jorge V, Le Paslier MC, Zaina G, Bastien C, Cattonaro F, Marroni F, et al. 2016. Characterization of the poplar pan-genome by genome-wide identification of structural variation. *Mol Biol Evol.* 33(10):2706–2719. doi:[10.1093/molbev/msw161](https://doi.org/10.1093/molbev/msw161).
- Prunier J, Giguère I, Ryan N, Guy R, Soolanayakanahally R, Isabel N, MacKay J, Porth I. 2019. Gene copy number variations involved in

- balsam poplar (*Populus balsamifera* L.) adaptive variations. *Mol Ecol*. 28(6):1476–1490. doi:[10.1111/mec.14836](https://doi.org/10.1111/mec.14836).
- Rae AM, Street NR, Robinson KM, Harris N, Taylor G. 2009. Five QTL hotspots for yield in short rotation coppice bioenergy poplar: the poplar Biomass Loci. *BMC Plant Biol*. 9(1):23. doi:[10.1186/1471-2229-9-23](https://doi.org/10.1186/1471-2229-9-23).
- Rafalski JA. 2010. Association genetics in crop improvement. *Curr Opin Plant Biol*. 13(2):174–180. doi:[10.1016/j.pbi.2009.12.004](https://doi.org/10.1016/j.pbi.2009.12.004).
- Rodriguez-Zaccaro FD, Henry IM, Groover A. 2021. Genetic regulation of vessel morphology in *Populus*. *Front Plant Sci*. 12:705596. doi:[10.3389/fpls.2021.705596](https://doi.org/10.3389/fpls.2021.705596).
- Rohde A, Storme V, Jorge V, Gaudet M, Vitacolonna N, Fabbri F, Ruttink T, Zaina G, Marron N, Dillen S, et al. 2011. Bud set in poplar—genetic dissection of a complex trait in natural and hybrid populations. *New Phytol*. 189(1):106–121. doi:[10.1111/j.1469-8137.2010.03469.x](https://doi.org/10.1111/j.1469-8137.2010.03469.x).
- Rudolph TD. 1978. Seed yield and quality in *Populus tremuloides* following pollination with gamma-irradiated pollen. *Can J Bot*. 56(23):2967–2972. doi:[10.1139/b78-359](https://doi.org/10.1139/b78-359).
- Satbhai SB, Setzer C, Freynschlag F, Slovak R, Kerdaffrec E, Busch W. 2017. Natural allelic variation of *FRO2* modulates *Arabidopsis* root growth under iron deficiency. *Nat Commun*. 8(1):15603. doi:[10.1038/ncomms15603](https://doi.org/10.1038/ncomms15603).
- Todesco M, Balasubramanian S, Hu TT, Traw MB, Horton M, Epple P, Kuhns C, Sureshkumar S, Schwartz C, Lanz C, et al. 2010. Natural allelic variation underlying a major fitness trade-off in *Arabidopsis thaliana*. *Nature*. 465(7298):632–636. doi:[10.1038/nature09083](https://doi.org/10.1038/nature09083).
- Tuskan GA, Difazio S, Jansson S, Bohlmann J, Grigoriev I, Hellsten U, Putnam N, Ralph S, Rombauts S, Salamov A, et al. 2006. The genome of black cottonwood, *Populus trichocarpa* (Torr. & Gray). *Science*. 313(5793):1596–1604. doi:[10.1126/science.1128691](https://doi.org/10.1126/science.1128691).
- Veitia RA, Bottani S, Birchler JA. 2013. Gene dosage effects: nonlinearities, genetic interactions, and dosage compensation. *Trends Genet*. 29(7):385–393. doi:[10.1016/j.tig.2013.04.004](https://doi.org/10.1016/j.tig.2013.04.004).
- Wellcome Trust Case Control Consortium. 2007. Genome-wide association study of 14,000 cases of seven common diseases and 3,000 shared controls. *Nature*. 447(7145):661–678. doi:[10.1038/nature05911](https://doi.org/10.1038/nature05911).
- Williams AL, Patterson N, Glessner J, Hakonarson H, Reich D. 2012. Phasing of many thousands of genotyped samples. *Am J Hum Genet*. 91(2):238–251. doi:[10.1016/j.ajhg.2012.06.013](https://doi.org/10.1016/j.ajhg.2012.06.013).
- Xia W, Xiao Z, Cao P, Zhang Y, Du K, Wang N. 2018. Construction of a high-density genetic map and its application for leaf shape QTL mapping in poplar. *Planta*. 248(5):1173–1185. doi:[10.1007/s00425-018-2958-y](https://doi.org/10.1007/s00425-018-2958-y).
- Yan H, Sun M, Zhang Z, Jin Y, Zhang A, Lin C, Wu B, He M, Xu B, Wang J, et al. 2023. Pangenomic analysis identifies structural variation associated with heat tolerance in pearl millet. *Nat Genet*. 55(3):507–518. doi:[10.1038/s41588-023-01302-4](https://doi.org/10.1038/s41588-023-01302-4).
- Yang C, Mulligan BJ, Wilson ZA. 2004. Molecular genetic analysis of pollen irradiation mutagenesis in *Arabidopsis*. *New Phytol*. 164(2):279–288. doi:[10.1111/j.1469-8137.2004.01182.x](https://doi.org/10.1111/j.1469-8137.2004.01182.x).
- Zhang S, Zhu L, Shen C, Ji Z, Zhang H, Zhang T, Li Y, Yu J, Yang N, HeY, et al. 2021. Natural allelic variation in a modulator of auxin homeostasis improves grain yield and nitrogen use efficiency in rice. *Plant Cell*. 33(3):566–580. doi:[10.1093/plcell/koaa037](https://doi.org/10.1093/plcell/koaa037).
- Zinkgraf M, Haiby K, Lieberman MC, Comai L, Henry IM, Groover A. 2016. Creation and genomic analysis of irradiation hybrids in populus. *Curr Protoc Plant Biol*. 1(2):431–450. doi:[10.1002/cppb.20025](https://doi.org/10.1002/cppb.20025).
- Zmienko A, Marszałek-Zenczak M, Wojciechowski P, Samelak-Czajka A, Luczak M, Kozłowski P, Karłowski WM, Figlerowicz M. 2020. AthCNV: a map of DNA copy number variations in the *Arabidopsis* genome. *Plant Cell*. 32(6):1797–1819. doi:[10.1105/tpc.19.00640](https://doi.org/10.1105/tpc.19.00640).

Editor: D.-J. de Koning