New Phytologist

Research

Genome-wide association implicates numerous genes underlying ecological trait variation in natural populations of *Populus trichocarpa*

Athena D. McKown[1]*, Jaroslav Klápště[1,2]*, Robert D. Guy[1], Armando Geraldes[3], Ilga Porth[1,4], Jan Hannemann[5], Michael Friedmann[3], Wellington Muchero[6], Gerald A. Tuskan[6], Jürgen Ehlting[5], Quentin C. B. Cronk[3], Yousry A. El-Kassaby[1], Shawn D. Mansfield[4] and Carl J. Douglas[3]

[1]Department of Forest and Conservation Sciences, Faculty of Forestry, University of British Columbia, Forest Sciences Centre, 2424 Main Mall, Vancouver, BC, V6T 1Z4, Canada; [2]Department of Dendrology and Forest Tree Breeding, Faculty of Forestry and Wood Sciences, Czech University of Life Sciences, Prague 165 21, Czech Republic; [3]Department of Botany, University of British Columbia, Vancouver, BC, V6T 1Z4, Canada; [4]Department of Wood Science, Faculty of Forestry, University of British Columbia, Forest Sciences Centre, 2424 Main Mall, Vancouver, BC, V6T 1Z4, Canada; [5]Department of Biology and Centre for Forest Biology, University of Victoria, Victoria, BC, V8W 3N5, Canada; [6]BioSciences Division, Oak Ridge National Laboratory, Oak Ridge, TN 37831, USA

Author for correspondence:
Athena D. McKown
Tel: +1 604 822 6023
Email: admckown@gmail.com

Received: *7 November 2013*
Accepted: *14 March 2014*

New Phytologist (2014) **203:** 535–553
doi: 10.1111/nph.12815

Key words: biomass, ecophysiology, genome-wide association study (GWAS), phenology, pleiotropy, poplar, single nucleotide polymorphisms (SNP) array, Unified Mixed Model.

Summary

- In order to uncover the genetic basis of phenotypic trait variation, we used 448 unrelated wild accessions of black cottonwood (*Populus trichocarpa*) from much of its range in western North America. Extensive data from large-scale trait phenotyping (with spatial and temporal replications within a common garden) and genotyping (with a 34 K *Populus* single nucleotide polymorphism (SNP) array) of all accessions were used for gene discovery in a genome-wide association study (GWAS).
- We performed GWAS with 40 biomass, ecophysiology and phenology traits and 29 355 filtered SNPs representing 3518 genes. The association analyses were carried out using a Unified Mixed Model accounting for population structure effects among accessions.
- We uncovered 410 significant SNPs using a Bonferroni-corrected threshold ($P < 1.7 \times 10^{-6}$). Markers were found across 19 chromosomes, explained 1–13% of trait variation, and implicated 275 unique genes in trait associations. Phenology had the largest number of associated genes (240 genes), followed by biomass (53 genes) and ecophysiology traits (25 genes).
- The GWAS results propose numerous loci for further investigation. Many traits had significant associations with multiple genes, underscoring their genetic complexity. Genes were also identified with multiple trait associations *within* and/or *across* trait categories. In some cases, traits were genetically correlated while in others they were not.

Introduction

The genetic basis of phenotypic variability is the fundamental underpinning of evolutionary biology and key in understanding factors that define speciation, biogeographical distributions and fitness under natural conditions (Stapley *et al.*, 2010; Savolainen *et al.*, 2013). Achieving such understanding is becoming more attainable as the ability to cast a wider net for gene discovery in traits of interest emerges. In plant biology, the integration of extensive genetic and phenotypic data is finding application in development and improvement of crop species, but is also extending our understanding of the genetics underlying traits of evolutionary and ecological importance (Ingvarsson *et al.*, 2008; Eckert *et al.*, 2009, 2010, 2012; Fournier-Level *et al.*, 2011; Parchman *et al.*, 2012; Olson *et al.*, 2013). Genome-wide association studies (GWAS) can be powerful for identifying putative causal genes, or suites of genes, underlying phenotypic variation, particularly in traits with complex genetic architecture (Vandenkoornhuyse *et al.*, 2010; Ingvarsson & Street, 2011; Savolainen *et al.*, 2013; Sork *et al.*, 2013). Where traits are complex (i.e. involving a number of genes or gene networks), GWAS using high genome coverage of single nucleotide polymorphisms (SNP) markers has been very effective for identifying the genetic architecture underlying variability in these traits (Eckert *et al.*, 2012; Parchman *et al.*, 2012; Riedelsheimer *et al.*, 2012; Morris *et al.*, 2013; Porth *et al.*, 2013a). GWAS can also uncover loci with potential pleiotropic effects that may be important to natural variation within species and their capacity

*These authors contributed equally to this work.

for adaptation (Mackay *et al.*, 2009; Stapley *et al.*, 2010; Porth *et al.*, 2014).

Defining the roles of genotypic and phenotypic variability in adaptation across a landscape are key to understanding the evolution and adaptability of species (Sork *et al.*, 2013). Within tree species, phenotypic variability is influenced by wide geographic distributions and numerous traits are considered to be under polygenic control (Savolainen *et al.*, 2007; Ingvarsson & Street, 2011; Cooke *et al.*, 2012; Sork *et al.*, 2013). High genetic complexity is reported for many adaptive traits in trees, such as cold hardiness, bud break, bud set, cone serotiny, disease resistance and growth (Ruttink *et al.*, 2007; Holliday *et al.*, 2008, 2010; Ingvarsson *et al.*, 2008; Eckert *et al.*, 2009; Ibáñez *et al.*, 2010; Ma *et al.*, 2010; Rohde *et al.*, 2010; Keller *et al.*, 2012; Parchman *et al.*, 2012; La Mantia *et al.*, 2013; Olson *et al.*, 2013). By comparison, the underlying genetic variability for numerous physiological traits considered important in range-wide adaptation of tree species, such as nutrient uptake, leaf anatomy, photosynthetic rate and water-use efficiency (cf. Soolanayakanahally *et al.*, 2009; Chamaillard *et al.*, 2011; Keller *et al.*, 2011; McKown *et al.*, 2014), is only beginning to be explored (González-Martínez *et al.*, 2008; Cumbie *et al.*, 2011).

In this study, we focused on the genetics underlying phenotypic trait variation in black cottonwood (*Populus trichocarpa*), a species of high ecological, scientific and economic value (Cronk, 2005; Tuskan *et al.*, 2006). Like many poplars, *P. trichocarpa* trees are outbreeding, fast growing and often function as pioneers and/or constitute major canopy-forming components of riparian forest ecosystems (Farrar, 1995; Braatne *et al.*, 1996). The species is common throughout the Pacific Northwest of North America and has high natural phenotypic variation relating to its geographical distribution spanning environmental and climatic gradients (Gornall & Guy, 2007; McKown *et al.*, 2014). Trait variation within *P. trichocarpa* relates primarily to its latitudinal distribution and gradients in photoperiodic regime (daylength) and/or temperature across its natural range (McKown *et al.*, 2014). Furthermore, heritability is generally highest in traits that co-vary strongly with these ecological and geographical gradients.

Extensive genomic tools available for *P. trichocarpa* (Tuskan *et al.*, 2006; Geraldes *et al.*, 2013) and high intraspecific variability in traits (McKown *et al.*, 2014) support using the GWAS approach to provide significant insights into the genetic architecture of ecologically important phenotypic variation (Eckert *et al.*, 2010, 2012; Parchman *et al.*, 2012; Morris *et al.*, 2013; Porth *et al.*, 2013a; La Mantia *et al.*, 2013; Olson *et al.*, 2013). Nevertheless, GWAS is challenging to implement using natural populations across a landscape (Ingvarsson & Street, 2011; Neale & Kremer, 2011; Sork *et al.*, 2013). As genetic structure reflects the effects of family relatedness, demography and adaptive history, model-fitting in GWAS as a corrective measure is necessary to balance the risk of false-positives with that of false-negatives (Balding, 2006; Ingvarsson & Street, 2011; Sork *et al.*, 2013). However, attempts to minimize the loss of some associations where relationships exist between loci, demography and geography should be made by assessing corrective measures on a trait-by-trait basis (La Mantia *et al.*, 2013; Porth *et al.*, 2013a,b).

Using accessions originating from wild populations of *P. trichocarpa*, we investigated the genetic basis of intraspecific variation in 40 biomass, ecophysiology and phenology traits in an association genetics framework. We employed GWAS, integrating extensive biological information on quantitative variation in these traits assayed within a common garden over multiple years (McKown *et al.*, 2013, 2014) and SNP genotype data from the same trees obtained using an Illumina iSelect Infinium 34K *Populus* SNP genotyping array developed for *P. trichocarpa* (Geraldes *et al.*, 2013). We predicted that certain traits considered genetically complex, such as growth or bud set, might retrieve multiple associations underscoring the genetic complexity of the trait. Additionally, we expected that genes underlying trait variation would associate repeatedly with the same trait when phenotyped over multiple years. Finally, we expected that the same loci would associate with multiple traits where traits are genetically correlated. Based on the results from our GWAS, we propose numerous key loci for further testing in trait variation, highlighting these as important in the evolution and ecology of *P. trichocarpa*.

Materials and Methods

We performed a GWAS with 448 unrelated individuals using clonal means for 40 traits and 29 355 filtered SNPs (detailed later). Data in the association analysis are publicly available at the University of Victoria PhenoDB website (URL: http://valdes.biol.uvic.ca/phenom) and within the Supporting Information included within this publication (Table S1A,B).

Phenotypic trait measurements

Tree materials were obtained from wild genotypes of *Populus trichocarpa* Torr. & A. Gray originally collected by British Columbia Ministry of Forests, Lands and Natural Resource Operations (FLNRO) spanning the northern two thirds of the species' range (44–60°N, 121–138°W) (Xie *et al.*, 2009; McKown *et al.*, 2013). Phenotyping of individual accessions in the Totem Field common garden, University of British Columbia, was replicated in space (4–20 clonal ramets of similar age and condition) and in time (repeated measurements across years) to confirm the patterns observed in phenotypic traits. This extensive phenotyping effort of all accessions for phenology events, growth and biomass accumulation, photosynthetic gas exchange, leaf traits and stable isotopes has been previously described (McKown *et al.*, 2013, 2014; Tables 1, S1A). Before GWAS, all trait data were checked for normality using a regression model approach. We note that bud set data were analyzed either including all data or removing premature bud set dates (bud set[1]) occurring before the solstice (21 June, day 186) due either to photoperiodic mismatch or other stressors (cf. Soolanayakanahally *et al.*, 2013).

Table 1 Phenotypic traits within three categories (biomass, ecophysiology, phenology) measured in *Populus trichocarpa* accessions indicating number of years measured and total number of significant single nucleotide polymorphisms (SNPs)/genes uncovered using genome-wide association study (GWAS) ($P < 1.7 \times 10^{-6}$)

Category/Trait		Years	SNPs/genes
Biomass traits	Active growth rate (cm d^{-1})	2009–2010	1/1
	Bole fresh mass density (kg m^{-3})	2012	0
	Bole fresh mass (kg)	2012	4/3
	Branches (total number)	2009	27/17
	Height:diameter (H:D; cm:cm)	2009–2011	2/2
	Height (cm)	2008–2011	14/12
	Height gain (cm)	2008–2011	11/10
	Log height growth rate (log cm d^{-1})	2009	5/4
	Log volume growth rate (log cm^3 d^{-1})	2009	3/3
	Volume (cm^3)	2009–2011	13/9
	Volume gain (cm^3)	2009–2011	14/12
	Whole-tree mass (kg)	2012	4/3
Ecophysiology traits	Carbon to nitrogen ratio (C:N; g g^{-1})	2009	1/1
	Chlorophyll content – spring (Chl_{spring}; CCI)	2009	0
	Chlorophyll content – summer (Chl_{summer}; CCI)	2009, 2011	10/7
	Instantaneous water-use efficiency (WUE; μmol CO_2 $mmol^{-1}$ H_2O)	2009	0
	Leaf carbon isotope discrimination (Δ_{leaf}; ‰)	2009	0
	Leaf mass per unit area – spring (LMA_{spring}; mg mm^{-2})	2010–2011	13/7
	Leaf mass per unit area – summer (LMA_{summer}; mg mm^{-2})	2009–2011	0
	Leaf nitrogen content per unit area (N_{area}; mg mm^{-2})	2009	0
	Leaf nitrogen content per unit dry mass (N_{mass}; g g^{-1})	2009	6/5
	Leaf shape (length:width)	2009	6/5
	Leaves per bud (total number)	2011–2012	1/1
	Photosynthetic rate per unit area (A_{max}; μmol CO_2 m^{-2} s^{-1})	2009	0
	Photosynthetic rate per unit dry mass ($A_{max/mass}$; μmol CO_2 g^{-1} s^{-1})	2009	2/2
	Photosynthetic nitrogen-use efficiency (NUE; μmol CO_2 g^{-1} N s^{-1})	2009	0
	Stable carbon isotope ratios ($\delta^{13}C_{wood}$; ‰)	2012	0
	Stable nitrogen isotope ratios ($\delta^{15}N$; ‰)	2009	0
	Stomatal conductance (g_s; mol H_2O m^{-2} s^{-1})	2009	0
Phenology traits	Bud break (Julian date)	2010–2011	8/2
	Bud set (Julian date)	2008–2010	149/104
	Bud set[1] (Julian date)	2009–2010	203/145
	Canopy duration (d)	2009–2010	11/6
	Growth period (d)	2009–2010	68/51
	Height growth cessation (HGC; Julian date)	2009	47/34
	Leaf drop (Julian date)	2008–2010	180/130
	Leaf flush (Julian date)	2010–2012	9/4
	Leaf lifespan (days)	2010	6/6
	Post-bud set period (PBS; d)	2009–2010	56/40
	25% total canopy leaf yellowing (Julian date)	2010	1/1
	50% total canopy leaf yellowing (Julian date)	2010	0
	75% total canopy leaf yellowing (Julian date)	2010	3/2
	100% total canopy leaf yellowing (Julian date)	2010	33/23

[1]Bud set dates occurring before the summer solstice (day 186) removed.

SNP genotyping

A total of 448 unrelated, phenotyped *P. trichocarpa* accessions (with > 0.03% genetic distance) were successfully genotyped with a 34K *Populus* Illumina Infinium® SNP genotyping array designed for *P. trichocarpa* (Table S1B). Full details of SNP discovery/selection, array development, performance and data filtering criteria are given in Geraldes *et al.* (2011, 2013). Candidate gene selection for the chip resulted in the inclusion of 34 131 SNP markers within 3543 genes and intergenic regions (± 2 kb up- or downstream from the longest transcript) across the genome. Genotyping was carried out as described by Geraldes *et al.* (2013) and array hybridizations performed at Oak Ridge National Laboratory (ORNL, TN). Genotype calls were filtered with GenomeStudio v2010.3 (http://support.illumina.com/array/array_software/genomestudio.ilmn).

Only SNPs with GenTrain score ≥ 0.5 and genotypes with GenCall score ≥ 0.15 were exported, criteria maximizing genotype call accuracy while minimizing missing data (Geraldes *et al.*, 2013). We further excluded SNPs with minor allele frequency < 0.05 and call rate < 0.9. Following this filtering process, we used 29 355 SNPs representing 3518 genes for associations. Each significant trait-associated SNP identified by GWAS was visually inspected for quality using the corresponding clustering plot

(GenomeStudio v2010.3). The 'Nisqually-1' genome sequence *P. trichocarpa* v2.2 SNP positions and gene models described in Geraldes *et al.* (2013) were translated into v3.0 positions by aligning sequences flanking the SNP with the latest *Populus* reference genome assembly on Phytozome 9.1 (http://www.phytozome.net/).

Population structure analysis

We evaluated the effects of genetic structure within our population using the Unified Mixed Model framework (Balding, 2006; Yu *et al.*, 2006) and compared log likelihood values between models with the Bayesian Information Criterion (BIC) (Yu *et al.*, 2006). We assessed a number of options for population structure fit on a trait-by-trait basis. We constructed family relatedness using a kinship (K) model and population structure using a principal component analysis (P) model or a clustering matrix (Q) model. We also calculated combinations of structures (P + K, Q + K), and a 'simple' model (i.e. simple linear regression without any additional correction).

SNPs used for population and kinship estimates were further filtered for Hardy-Weinberg Equilibrium using the 'Chisq' function in the R package 'HardyWeinberg' (Graffelman & Morales, 2008) and for linkage disequilibrium (LD) at $r^2 < 0.2$ (Wang *et al.*, 2009). Following these filtering criteria, 8749 SNPs (distributed throughout the genome) were used to fit all model analyses. The K model was calculated following Loiselle *et al.* (1995) and the relationship matrix was estimated by first multiplying the kinship matrix by two, then setting diagonal elements as one and negative off-diagonal elements as zero (Yu *et al.*, 2006). The 'nearPD' function implemented in the R package 'Matrix' (Higham, 2002) was used to obtain the positive definite relationship matrix required in the mixed model framework. The P model was done using the 'prcomp' function implemented in the base R package (R Core Development Team, 2011) and significant principal components (PC) were selected according to the broken-stick rule (Jackson, 1993) implemented in the R package 'vegan'. Within our population, only PC1 was significant. The parametric clustering model-based inference (Q matrix) was performed using the R package 'popgen' (Marchini, 2013) which implements both the uncorrelated allele frequency model of STRUCTURE (Pritchard *et al.*, 2000) when using the function 'ps', and the correlated allele frequency model (Falush *et al.*, 2003) by using the function 'ps' and 'popdiv' in conjunction. The number of populations tested ranged from $K = 1$ to $K = 10$ populations. Both the burn-in period and the number of sampling iterations after the burn-in period were set to 60 000, and thinning was set at the default (1). For each scenario (K), 20 runs were performed to obtain both mean and standard deviation for the log likelihood value to construct a delta coefficient for the most probable number of populations (Evanno *et al.*, 2005). While the uncorrelated allele frequency model did not detect any population structure (i.e. no peak appeared indicating the best fit was reached in the scenario considering $K = 1$, results confirmed with GENELAND; Guillot *et al.*, 2005), the correlated allele frequency model detected $K = 5$. We used the $K = 5$ cluster results from the correlated allele frequency for the Q matrix in our GWAS.

We evaluated the model fit on a trait-by-trait basis using the Bayesian Information Criterion (BIC) where the lowest BIC value indicates the best model fit. Among all studied traits, BIC selected the simple, P or Q models depending on the trait (Table S2). In no case was the kinship (K) component within the K, P + K or Q + K models considered the best fit for the data structure. This lack of importance of the K component confirmed the absence of familial relatedness within the study population (see also La Mantia *et al.*, 2013; Porth *et al.*, 2013a). By comparison, QQ plots (i.e. the ranking of observed P-values from smallest to highest against the expected values) showed that inclusion of the K component generated a uniform distribution of P-values (simply reflecting the tested null hypothesis that no marker is a causal variant) and indicated a substantial decrease in the power to detect true positives (Figs S1–S3). We consider this result likely to be related to the presence of linkage between the actual true positives and other SNPs due to dense SNP coverage, rather than to the confounding effect of population structure in our sample set (Pearson & Manolio, 2008). In such a case, the QQ plot may fail to identify the real source of deviation from the null hypothesis and thus risks exaggerating confounding factors resulting in an excess of false-negatives.

Association genetics

We used the GLM procedure implemented in TASSEL (Bradbury *et al.*, 2007) to perform the association analysis as follows:

$$\boldsymbol{y} = \mu + \mathbf{S}\alpha + \mathbf{X}\beta + \boldsymbol{e} \qquad \text{Eqn 1}$$

($\boldsymbol{y}$, vector of measurements; μ, overall population mean; $\mathbf{S}$ and $\mathbf{X}$, index matrices assigning fixed effects for both SNP genotype and population to the measurements, respectively; $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$, vectors of fixed effects for both SNP genotype and population, respectively; e, residual effect). Following the GWAS, we used Bonferroni multiple testing correction ($\alpha/29\,355$) rather than the false discovery rate (FDR) correction owing to nonindependence of the tests where test statistics were correlated due to LD between SNPs used in the array (cf. Schwartzman & Lin, 2011). We considered SNP–trait associations significant at $\alpha = 0.05/29\,355$ where $P < 1.7 \times 10^{-6}$ and report these. As subsidiary signal, we also included trait associations at $\alpha = 0.1/29\,355$ where $P < 3.4 \times 10^{-6}$ in the Supporting Information if the SNP in question was already considered significant by association to another trait at the lower cut-off of $P < 1.7 \times 10^{-6}$. Composite pairwise LD between all significant trait-associated SNPs was calculated based on genotype correlations (Weir *et al.*, 2004).

Cumulative R^2 of significant SNPs

In order to address the total phenotypic variance accounted for by all trait-associated SNPs on a trait-by-trait basis, we calculated a 'cumulative R^2' metric. These values were obtained by the difference in R^2 between full and reduced models (Ingvarsson *et al.*,

2008). The full model comprises all significant SNPs detected by GWAS for the trait in question and population structure (as selected by BIC, see 'Population structure analysis' above) while the reduced model contains only population structure. Analysis was performed using the 'glm' function and R^2 values were extracted using the 'RsquareAdj' function implemented in the R package 'vegan' (Peres-Neto *et al.*, 2006). We then repeated this test using $P < 3.4 \times 10^{-6}$ to include our subsidiary SNP association information (see above).

Genetic correlations between phenotypic traits

In order to confirm that trait correlation was not solely responsible for detection of potential functional pleiotropy, we assessed the pairwise genetic correlations of all traits to identify a common genetic basis for independent variation (Porth *et al.*, 2013b). These genetic correlations are 'broad-sense' (i.e. using phenotype trait data from all clonal replicates) and based on clonal best linear unbiased predictions (BLUPs) using PC1 from the PCA for structure correction (McKown *et al.*, 2014). The broad-sense genetic correlation matrix was performed using the 'cor' function in the 'stats' R package and Pearson product-moment correlations were estimated following:

$$r_{G_{x,y}} = \frac{Cov_{gxgy}}{\sqrt{Var_{gx} \times Var_{gy}}} \quad \text{Eqn 2}$$

(Cov_{gxgy}, covariance between clonal BLUPs of traits x and y; Var_{gx}, variance in clonal BLUPs for trait x; Var_{gy},variance in clonal BLUPs for trait y). The clonal breeding values were obtained from linear mixed model results presented in McKown *et al.* (2014).

Tests for Gene Ontology enrichment

All genes uncovered by GWAS were tested for Gene Ontology (GO) enrichment using 'function' and 'process' categorizations against the available genes from the SNP array (i.e. genes included in GWAS following SNP filtering). We tested all genes, and subgroupings of genes based on individual trait categories or groupings of categories. Significant GO terms were determined with GOTermFinder software (http://go.princeton.edu/cgi-bin/GOTermFinder) using FDR correction for multiple comparisons (Boyle *et al.*, 2004).

Results

SNP discovery through GWAS

The GWAS using 29K SNPs uncovered a total of 1118 significant SNP–trait associations (involving 410 unique SNPs) across the three studied trait categories (i.e. biomass, ecophysiology and phenology). Most traits required population structure correction (either P model or Q matrix; 65 out of 71 tests), decided on a trait-by-trait basis using BIC model selection (Table S2). Significant trait-associated SNP markers were found across all 19 chromosomes with the highest numbers of significant SNPs ($n \geq 35$) on chromosomes 2, 6 and 9 (Fig. 1, Table S3). The number of trait-associated SNPs/chromosome was significantly different from the number of SNPs/chromosome on the array (after filtering) (χ^2 test, $P = 0.0075$) and trait-associated SNP distribution across chromosomes did not correlate strongly with the density of SNPs/chromosome on the array (using 500 kb windows along each chromosome; $r^2 = 0.13$). Most trait-associated SNP markers were located in noncoding regions (78%) while a smaller number of SNP markers were within coding regions (nonsynonymous = 10%, synonymous = 12%) (Tables S3, S4). This largely reflected the relative distribution of the SNPs used and no enrichment based on position within gene region was found (χ^2 test, not significant).

In total, 275 genes were identified with at least one significant trait-associated SNP (Tables S4, S5). Where multiple trait-associated SNPs within a gene were retrieved, a range in LD values between such SNPs was observed ($r^2 = 0$–1.0; Table S6). This variability in LD within genes is likely due to the high variability in recombination rate throughout the genome (Slavov *et al.*, 2012). Nevertheless, on average, LD within genes was high ($r^2 = 0.73$; Table S6). Among the 18 genes with low or no LD between trait-associated SNPs ($r^2 = 0$–0.3), 12 had multiple associations within the same trait category while six had associations across trait categories (Tables S5, S6). Some trait-associated SNPs

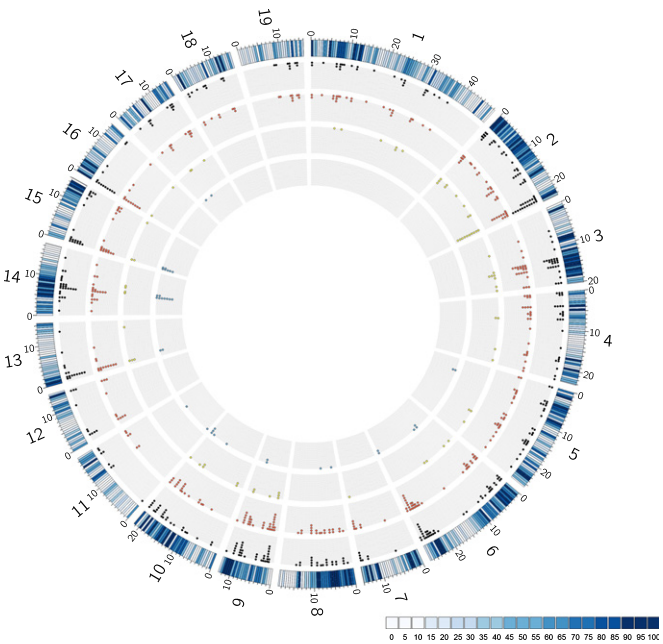

Fig. 1 Genomic distribution of single nucleotide polymorphisms (SNPs) on the 34K *Populus* genotyping array and significant trait-associated SNPs uncovered using genome-wide association study (GWAS) across 19 chromosomes in *P. trichocarpa*. SNP density on the array per 500 kB windows on each chromosome is illustrated by a heat map (outermost ring). All SNPs retrieved by GWAS are indicated in black (second ring). These are further distinguished by trait category where SNPs related to phenology traits are marked in red (third ring), biomass traits in yellow (fourth ring) and ecophysiology traits in blue (fifth, inner ring). Image courtesy of N. Farzaneh.

located in different genes but within the same genomic regions also showed moderate to complete linkage ($r^2 = 0.35–1.0$; Table S6). Among these, two genomic regions had multiple associations within the same trait category while four had associations across trait categories.

The 410 significant SNPs within 275 genes were associated with 30 of the 40 assayed biomass, ecophysiology and phenology traits (Table 1). Total numbers of identified SNP–trait associations varied, depending on the trait, and SNP markers explained between 1.2 and 13.2% of the phenotypic variation, depending on the association (average $r^2 = 0.037$; Table S5). The phenology category retrieved the largest number of SNP–trait associations whereas both the biomass and ecophysiology categories had far fewer associations (Tables 1, S5). SNP–trait associations at $P < 1.7 \times 10^{-6}$ identified 53 genes associated with biomass (20 genes were solely associated with biomass traits), 25 genes associated with ecophysiology (15 genes solely with ecophysiology traits), and 240 genes associated with phenology (200 genes solely with phenology traits) (Fig. 2, Table S5). Correspondingly, the cumulative proportion of phenotypic variance explained by significant SNPs (cumulative R^2) was highest within the phenology category and lower in both biomass and ecophysiology categories (Figs 3, S4).

Among phenology traits, bud set, growth period, height growth cessation, post-bud set period, 100% leaf yellowing and leaf drop had the greatest number of associations. Within the biomass category, branch numbers, height/height gain and volume/volume gain yielded the most SNP associations. The highest numbers of associations among ecophysiology traits included leaf mass per area of preformed leaves (LMA_{spring}), nitrogen per unit mass (N_{mass}), summer chlorophyll content (Chl_{summer}) and leaf shape. Many genes were repeatedly associated with the same trait where year-to-year data existed and/or with multiple traits *within* trait categories, particularly among phenology traits (see 'Genes with effects on phenology' below). GWAS further identified genes with significant associations *across* two trait categories (42 out of 275) and three genes with associations across all trait categories.

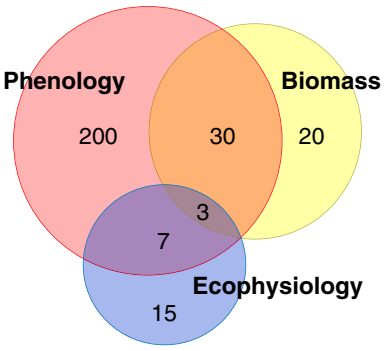

Fig. 2 Diagram depicting 275 unique genes identified through genome-wide association study (GWAS). Numbers of associations are arranged in circles by trait category (biomass, ecophysiology, phenology), with circle size representing relative proportion of significant genes and circle overlaps representing numbers of genes associated with more than one trait category. See Tables 3–6 and Supporting Information Table S5 for detailed information on significant SNP-trait associations and gene identities.

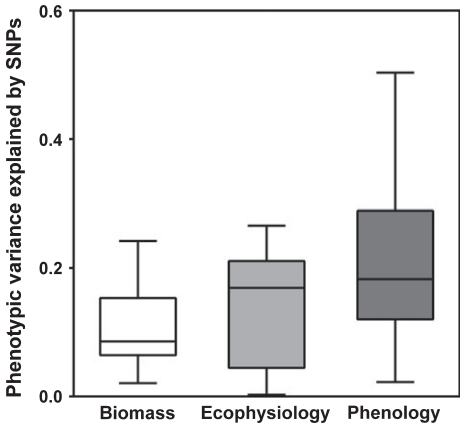

Fig. 3 Quantile distribution of proportions of phenotypic variance explained by significant single nucleotide polymorphism (SNP)–trait associations for each trait (cumulative R^2) within each trait category (biomass, ecophysiology, phenology). Cumulative R^2 values for individual traits are shown in Fig. S4.

The genes uncovered by GWAS were largely transcription factors/regulators, transferases, kinases, transporters, hydrolases and other/unknown gene functions (based on the *Arabidopsis* homologs) (Tables 2, S5). These genes were tested for enrichment of Gene Ontology (GO) terms using all results, phenology-related, biomass-related, ecophysiology-related and multiple category-related (Table S7). Significant enrichment was only found considering genes associated in the biomass-related group (auxin binding (GO:0010011), hormone binding (GO:0042562)) and genes with associations *across* trait categories (substrate-specific channel activity (GO:0022838), nitrate transmembrane transporter activity (GO:0015112), channel activity (GO:0015267), passive transmembrane transporter activity (GO:0022803)). Other high-ranking GO terms included response to red/far red light, binding (e.g. DNA, hormone, kinase, protein) and circadian rhythm but were not significantly enriched after multiple testing correction.

Genes underlying phenotypic variation

From the large number of SNP–trait associations, we highlight examples of specific loci providing the *Arabidopsis* homologue annotation, location information (i.e. chromosome/SNP/feature), allelic variation among accessions and the underlying phenotypic variability (Table S8). We focused on genes associated: to biomass or ecophysiology traits; to phenology traits; and with multiple traits *within* and/or *across* trait categories. Full SNP results, marker r^2 and LD values are available in Tables S4–S6.

Genes with effects on biomass or ecophysiology A small number of genes (35 out of 275) exhibited significant associations only with variation in biomass or ecophysiology traits (Fig. 2, Tables 3, 4, S5). These encompassed a range of functions, such as transcription factors, kinases, phytochrome, transporters and binding elements. In many cases, genes were retrieved either by year-to-year data from the same trait and/or from multiple traits

Table 2 General functional classifications of genes identified by genome-wide association study (GWAS) with significant single nucleotide polymorphism (SNP) markers associated to growth, ecophysiology and phenology traits

Putative function[1]	Number[2]
Actin-related[3]	3
Apoptosis	1
Aquaporin[3]	3
Binding - other[3]	8
Calmodulin[3]	4
Cell division	1
Cell wall metabolism	7
Cytochrome[3]	4
Cytoskeleton[3]	4
Dehydratase/dehydrogenase	8
DNA repair	1
Hydrolase	12
Ion binding	8
Ion transporter[3]	8
Kinase[3]	15
Laccase[3]	3
Ligase[3]	6
Membrane[3]	4
Other[3,4]	27
Oxygenase/oxidase	5
Peroxidase[3]	1
Phosphatase	3
Phytochrome	1
Phytohormone	4
Protease[3]	5
Protein binding[4]	8
Ribosome	1
RNA binding	2
Senescence	2
Transcription factor/regulator[3,4]	62
Transferase[3]	20
Transporter[3]	9
Unknown[3]	20
Zinc finger[3]	5

[1]Functional gene prediction based on Geraldes *et al.* (2013).
[2]Number of genes with SNPs associated with trait variation. Full details are given in Table S5.
[3]Includes genes with associations across two trait categories.
[4]Includes genes with associations across three trait categories.

in the same category. Within the biomass-related category, Potri.010G250500 (protein binding *EXO70G1*; *EXOCYST SUBUNIT EXO70 FAMILY PROTEIN G1*) was associated with active growth rate, height (2009–2011), and bole and whole-tree mass (Table 3). Effects of the SNP (10_22286918; intergenic) linked the common allele with substantially greater biomass overall. Without any apparent geographic pattern, accessions homozygous for the common allele were 39% taller (each year) and had greater bole and whole tree mass (76% and 89%, respectively) compared to the minor homozygotes, with heterozygous accessions being intermediate compared to both homozygotes (Table S8). Among the genes uncovered within the ecophysiology category, leaf N content (N_{mass}) and the correlated C : N ratio were associated with Potri.010G221600 (*EMB1144*; *EMBRYO DEFECTIVE 1144* chorismate synthase) (Table 4). Accessions homozygous for the minor allele (SNP 10_20651512; 3′UTR) had 14% lower N_{mass} and 1.2× greater C : N ratio compared to the other accessions, but no difference was observed comparing heterozygotes with homozygotes of the major allele (Table S8). Another gene, Potri.011G024000 (*SPK1*; *SPIKE1*), was associated with maximum photosynthetic rate per unit mass ($A_{max/mass}$) (Table 4). Allelic effects of the SNP (11_2007822; intron) linked the minor homozygotes with 22% higher photosynthesis than the major homozygotes, and heterozygotes with having intermediate trait values (Table S8).

Genes with effects on phenology The majority of genes uncovered by GWAS had SNPs associated with phenology traits (Fig. 2, Tables 5, S5). Genes ranged in function, including cytochromes, hydrolases, ion binding/transport, transcription factors/regulators and transferases. Many encoded proteins putatively related to light perception, photoperiod and/or circadian rhythm, or were phytohormone-related/response proteins (involving auxin, cytokinin, gibberellin, abscisic acid, and ethylene). Numerous genes were repeatedly associated with phenology traits across different years (Tables 5, 6, S5). Among 89 genes associated with phenology traits measured over multiple years, 80 genes were found to be associated with the *same* trait in at least 2 yr. In addition, we found 13 genes with associations to the same trait in multiple years just below our stringent cutoff criteria ($P < 3.4 \times 10^{-6}$). Similarly, GWAS uncovered 185 genes with associations to 2–7 *different* phenology traits, and an additional 15 genes had multiple phenology trait associations detected just below our stringent cutoff criteria ($P < 3.4 \times 10^{-6}$). Analyses using all bud set dates available for the population vs removing premature bud set dates occurring before the solstice (i.e. bud set[1]) largely resulted in the same SNP–trait associations; however, a handful of genes were found only using bud set[1].

Relating to light perception, Potri.010G215200 (*PRR7*; *PSEUDO-RESPONSE REGULATOR 7* transcription regulator), was associated with fall phenology events of bud set (2008–2010), growth period, height growth cessation and leaf drop (Table 5). Allelic effects of the SNP (10_202495; coding sequence, nonsynonymous) linked the minor homozygote accessions with earlier height growth cessation, bud set and leaf drop (32, 36 and 23 d, respectively) and correspondingly shorter growth period (51 d) compared to the major homozygotes, with the heterozygous state intermediate to both homozygotes (Table S8). Among the phytohormone-related genes, Potri.018G033600 (*GA3OX1*; *GIBBERELLIN 3-OXIDASE 1*) was linked with multiple phenology traits. A single SNP (18_2683640; intergenic) was associated with bud set (2008–2010), growth period (2009–2010), height growth cessation, post-bud set period and leaf drop (Table 5). The minor homozygotes showed later height growth cessation, bud set and leaf drop (25, 31, and 31 d, respectively) resulting in a longer growth period (44 d) and shorter post-bud set period (28 d) compared to the major homozygotes with the heterozygous state intermediate to both homozygotes (Table S8). Transcription factor Potri.009G017400 (*BLH1*; *BEL1-LIKE HOMEODOMAIN 1*) was associated with bud set (2009–2010),

Table 3 Genes identified by genome-wide association study (GWAS) with single nucleotide polymorphism (SNP) markers associated to biomass traits

Gene model[1]	Trait[2]	AT homolog	Annotated description[1]
Potri.001G256100	Volume gain	AT3G21070	*NADK1* (*NAD KINASE 1*)
Potri.001G323100	Height gain	AT3G26810	*AFB2* (*AUXIN SIGNALING F-BOX 2*)
Potri.001G345500	Branches	AT5G40440	*MKK3* (*MITOGEN-ACTIVATED PROTEIN KINASE KINASE 3*)
Potri.002G005800	Volume gain	AT1G76420	*CUC3* (*CUP SHAPED COTYLEDON3*)
Potri.002G052100	Height gain	AT4G02780	*GA1* (*GIBBERELLIC ACID REQUIRING 1*)
Potri.002G111900	Log volume growth rate	AT1G50010	*TUA2* (*TUBULIN ALPHA-2 CHAIN*)
Potri.003G059400	Active growth rate	AT1G15490	Hydrolase, alpha/beta fold family protein
Potri.003G139300	Volume, Volume gain	AT1G64380	AP2 domain-containing transcription factor
Potri.003G195300	Height gain	AT3G54390	Transcription factor GT-2
Potri.005G142300	Log height growth rate	AT2G23300	Leucine-rich repeat transmembrane protein kinase
Potri.006G150400	Branches	AT2G19580	*TET2* (*TETRASPANIN2*)
Potri.010G019000	Log height growth rate	AT3G06350	*MEE32* (*MATERNAL EFFECT EMBRYO ARREST 32*)
Potri.010G250500	Bole mass, Height, Whole-tree mass	AT4G31540	*EXO70G1* (*EXOCYST SUBUNIT EXO70 FAMILY PROTEIN G1*)
Potri.013G123800	H : D	AT1G75840	*ARAC5* (*RAC-LIKE GTP BINDING PROTEIN 5*)
Potri.014G134800	Height	AT3G62980	*TIR1* (*TRANSPORT INHIBITOR RESPONSE 1*)
Potri.014G141400	Log height growth rate	AT4G18880	*HSFA4A* (*HEAT SHOCK TRANSCRIPTION FACTOR A4A*)
Potri.015G127200	Volume gain	AT4G25240	*SKS1* (*SKU5 SIMILAR 1*)
Potri.016G000300	H : D	AT2G44190	*EDE1* (*ENDOSPERM DEFECTIVE 1*)
Potri.016G128300	Log volume growth rate	AT2G38470	*WRKY33* (*WRKY DNA-BINDING PROTEIN 33*)
Potri.018G076400	Log height growth rate	AT3G24450	Copper-binding family protein

[1]Poplar gene models are annotated to v3 of the genome. See Table S5 for full gene details, associated SNPs, and complete annotation description.
[2]See Table 1 for trait explanations and units.
H : D, height : diameter.

Table 4 Genes identified by genome-wide association study (GWAS) with single nucleotide polymorphism (SNP) markers associated to ecophysiology traits

Gene model[1]	Trait[2]	AT homolog	Annotated description[1]
Potri.005G072700	LMA_{spring}	AT4G31700	*RPS6* (*RIBOSOMAL PROTEIN S6*)
Potri.005G073000	LMA_{spring}	AT5G65270	*RABA4A* (*RAB GTPASE HOMOLOG A4A*)
Potri.006G097300	Chl_{summer}	AT2G38090	MYB family transcription factor
Potri.006G116900	Chl_{summer}	AT5G03760	*CSLA9* (*CELLULOSE SYNTHASE LIKE A9*)
Potri.008G105200	Leaves per bud	AT2G18790	*PHYB* (*PHYTOCHROME B*)
Potri.009G110500	LMA_{spring}	AT2G16050	Thioredoxin-related/zinc ion binding
Potri.010G121500	Leaf shape	AT1G25380	Mitochondrial FAD carrier protein
Potri.010G221600	C : N, N_{mass}	AT1G48850	*EMB1144* (*EMBRYO DEFECTIVE 1144*)
Potri.011G024000	$A_{max/mass}$	AT4G16340	*SPK1* (*SPIKE1*)
Potri.011G107900	Leaf shape	AT2G34250	Protein transport protein SEC61 subunit alpha
Potri.013G032500	Leaf shape	AT3G47590	Esterase/lipase/thioesterase family protein
Potri.014G103600	LMA_{spring}	AT2G46710	RAC GTPase activating protein, putative
Potri.014G116800	LMA_{spring}	AT2G47180	*GOLS1* (*GALACTINOL SYNTHASE 1*)
Potri.015G009100	Chl_{summer}	AT4G27740	Yippee putative zinc-binding protein
Potri.018G019900	Leaf shape	AT5G10930	*CIPK5* (*CBL-INTERACTING PROTEIN KINASE 5*)

[1]Poplar gene models are annotated to v3 of the genome. See Table S5 for full gene details, associated SNPs, and complete annotation description.
[2]See Table 1 for trait explanations and units.
$A_{max/mass}$, assimilation rate per unit mass; C:N, carbon:nitrogen ratio; Chl, chlorophyll; LMA, leaf mass per area; N_{mass}, nitrogen per unit mass.

growth period, post-bud set period, leaf yellowing and leaf drop (2008,10) (Table 5). Both significant SNP markers (09_2874013; coding sequence, synonymous/09_2874898; intron) are in high pairwise LD ($r^2 = 0.95$). The double minor homozygote accessions had earlier bud set, canopy yellowing and leaf drop (22, 17, and 12 d, respectively), and subsequently shorter growth period (26 d) and longer post-bud set period (18 d) compared to the double major homozygote accessions (Table S8). The common heterozygote had equivalent trait values to the double major homozygote, while the less common heterozygote (6 trees total) showed trait values similar to the double minor homozygote.

Individual genes with effects across trait categories GWAS identified 40 genes with SNPs associated with variation in two trait categories, and three genes with SNPs associated across all trait categories (Fig. 2, Tables 6, S5). All genes with multiple trait category effects had associations to phenology events, particularly bud set and leaf drop. One example, Potri.001G057400 (*HK3*; *HISTIDINE KINASE 3* cytokinin receptor) was associated with

Table 5 Selected genes identified by genome-wide association study (GWAS) with significant single nucleotide polymorphism (SNP) markers associated to phenology traits across multiple years and/or multiple phenology traits[2]

Gene model[2]	Trait[3]	AT homolog	Annotated description[2]
Potri.001G000600	Bud set	AT1G55570	*SKS12* (*SKU5 SIMILAR 12*)
Potri.001G110800	Bud set[1], Leaf drop	AT4G25480	*DREB1A* (*DEHYDRATION RESPONSE ELEMENT B1A*)
Potri.001G190800	Leaf drop	AT2G19770	*PRF3* (*PROFILIN3*)
Potri.001G252600	Bud set, Leaf drop	AT5G58620	zinc finger (CCCH-type) family protein
Potri.001G327100	Bud set[1], Canopy duration, Growth period, Leaf lifespan	AT3G27010	*TCP20* (*TEOSINTE BRANCHED 1, CYCLOIDEA, PCF (TCP)-DOMAIN FAMILY PROTEIN 20*)
Potri.001G375500	Bud set, Growth period, Leaf drop, PBS	AT1G53210	sodium/calcium exchanger family protein
Potri.002G013400	Bud set, Growth period, PBS, Leaf drop	AT5G42250	alcohol dehydrogenase, putative
Potri.002G055400	Bud set[1], Leaf drop	AT3G59060	*PIL6* (*PHYTOCHROME INTERACTING FACTOR 3-LIKE 6*)
Potri.002G074400	Canopy duration, Growth period	AT1G43890	*RAB18* (*RAB GTPASE HOMOLOG B18*)
Potri.002G099800	Leaf drop, PBS	AT1G78300	*GRF2* (*GENERAL REGULATORY FACTOR 2*)
Potri.002G184300	Bud set, Growth period	AT1G02305	cathepsin B-like cysteine protease
Potri.002G242500	Bud set, Growth period, Leaf drop, PBS	AT2G32720	*CB5-B* (*CYTOCHROME B5 ISOFORM B*)
Potri.002G242700	Bud set, PBS	AT5G48740	leucine-rich repeat family protein
Potri.003G050100	Leaf drop	AT1G52150	*ATHB-15*
Potri.003G126900	Bud set, HGC, Leaf drop	AT4G23100	*GSH1* (*GLUTAMATE-CYSTEINE LIGASE*)
Potri.003G128100	Bud set, Leaf drop	AT4G23340	2OG-Fe(II) oxygenase family protein
Potri.003G131700	Bud set[1], Leaf drop	AT4G23500	glycoside hydrolase family 28 protein
Potri.003G173000	Bud set, Leaf drop	NA	unknown function
Potri.004G002700	Bud set[1], Leaf drop	AT2G32950	*COP1* (*CONSTITUTIVE PHOTOMORPHOGENIC 1*)
Potri.004G013400	Bud set, Growth period, PBS, Leaf drop	AT1G11790	*ADT1* (*AROGENATE DEHYDRATASE 1*)
Potri.004G116100	Bud set, Leaf drop	AT3G02150	*PTF1* (*PLASTID TRANSCRIPTION FACTOR 1*)
Potri.004G168600	100% Leaf yellowing, Leaf drop	AT4G38770	*PRP4* (*PROLINE-RICH PROTEIN 4*)
Potri.004G174400	Bud set[1], Canopy duration, Growth period, PBS	AT4G38620	*MYB4* (*MYB DOMAIN PROTEIN 4*)
Potri.005G086400	Bud set[1], Leaf drop	AT4G39410	*WRKY13* (*WRKY DNA-BINDING PROTEIN 13*)
Potri.005G111600	Leaf drop	AT2G17840	*ERD7* (*EARLY-RESPONSIVE TO DEHYDRATION 7*)
Potri.005G138400	Bud set, Leaf flush, Leaf drop	AT5G67030	*ABA1* (*ABA DEFICIENT 1*)
Potri.005G140200	Bud set, HGC, Leaf drop	AT2G23380	*CLF* (*CURLY LEAF*)
Potri.005G156500	Bud set, HGC, PBS	NA	Unknown function
Potri.005G166100	Bud set[1], Leaf drop	AT5G65170	VQ motif-containing protein
Potri.005G170500	Bud set, Leaf drop	AT1G77920	*TGA7*
Potri.006G008300	Bud set	NA	Protease inhibitor
Potri.006G039000	Bud set, Canopy duration, Growth period, 100% Leaf yellowing, Leaf drop	AT5G06950	*AHBP-1B*; *CAMP-RESPONSE ELEMENT BINDING PROTEIN-RELATED*
Potri.006G054500	Bud set, Growth period, Leaf drop	AT3G57600	*DREB2F* (*DEHYDRATION RESPONSIVE ELEMENT BINDING PROTEIN 2F*)
Potri.006G057700	Bud set, Growth period, PBS	AT3G12160	*RABA4D* (*RAB GTPASE HOMOLOG A4D*)
Potri.006G209200	Bud set[1], Leaf drop	AT5G22380	*ANAC090* (*NAC DOMAIN CONTAINING PROTEIN 90*)
Potri.006G241600	Bud set, Leaf drop, PBS	AT5G11520	*ASP3* (*ASPARTATE AMINOTRANSFERASE 3*)
Potri.006G249900	Bud set, HGC, Leaf drop	AT2G25600	*SPIK* (*SHAKER POLLEN INWARD K+ CHANNEL*)
Potri.006G263000	Bud set	AT2G37585	glycosyltransferase family 14 protein
Potri.006G264500	100% Leaf yellowing, Leaf drop	AT5G10840	endomembrane protein 70
Potri.006G264600	Bud set, 100% Leaf yellowing, Leaf drop	AT2G25060	plastocyanin-like domain-containing protein
Potri.007G076500	Bud set[1], Leaf drop	AT4G39350	*CESA2* (*CELLULOSE SYNTHASE A2*)
Potri.008G086800	Bud set[1], Leaf drop	AT1G26820	*RNS3* (*RIBONUCLEASE 3*)
Potri.008G138400	Bud set[1], Leaf drop	AT1G14720	*XTR2* (*XYLOGLUCAN ENDOTRANSGLYCOSYLASE RELATED 2*)
Potri.008G140700	Bud set, Growth period, PBS	AT2G01980	*NHX7* (*NA+/H+ ANTIPORTER 7*)
Potri.008G161900	Bud set, Leaf drop	AT5G43650	basic helix-loop-helix (bHLH) family protein
Potri.008G162800	Bud set, Growth period, HGC, 100% Leaf yellowing, Leaf drop, PBS	AT3G23090	*TPX2* (*TARGETING PROTEIN FOR XKLP2*)
Potri.008G195500	Bud set[1], Leaf drop	AT3G07630	*ADT2* (*AROGENATE DEHYDRATASE 2*)
Potri.009G006500	100% Leaf yellowing, Leaf drop	AT2G28110	*FRA8* (*FRAGILE FIBER 8*)
Potri.009G011000	Bud set, 100% Leaf yellowing, Leaf drop	AT2G28315	DUF707, protein of unknown function
Potri.009G014500	Leaf drop	AT5G60690	*REV* (*REVOLUTA*)
Potri.009G017400	Bud set, Growth period, 100% Leaf yellowing, Leaf drop, PBS	AT2G35940	*BLH1* (*BEL1-LIKE HOMEODOMAIN 1*)
Potri.009G021800	Bud set[1], Leaf drop	AT2G26930	*CDPMEK* (4-(*CYTIDINE 5\'-PHOSPHO*)-2-C-METHYL-D-ERITHRITOL KINASE)

Table 5 (Continued)

Gene model[2]	Trait[3]	AT homolog	Annotated description[2]
Potri.009G035000	Bud set, Leaf drop	AT3G46640	*PCL1* (*PHYTOCLOCK 1*)
Potri.009G099800	Bud set, HGC, Leaf drop	AT4G34050	*CAFFEOYL COENZYME A O-METHYLTRANSFERASE 1*
Potri.009G106000	Bud set, HGC, Leaf drop	AT2G15780	plastocyanin-like domain-containing protein
Potri.010G077000	Bud set, Growth period, Leaf drop	AT5G43650	basic helix-loop-helix (bHLH) family protein
Potri.010G093900	Bud set, Growth period, Leaf drop	AT1G14310	haloacid dehalogenase-like hydrolase family protein
Potri.010G179300	Bud set, Growth period, HGC, Leaf drop, PBS	AT5G16250	unknown protein
Potri.010G187600	Growth period, Leaf drop	AT3G55990	*TBL28* (*TRICHOME BIREFRINGENCE-LIKE 28*)
Potri.010G212900	Bud set, Leaf drop	AT3G55260	*HEXO1* (*BETA-HEXOSAMINIDASE 1*)
Potri.010G215200	Bud set, Growth period, HGC, Leaf drop	AT5G02810	*PRR7* (*PSEUDO-RESPONSE REGULATOR 7*)
Potri.011G094400	Bud set, Growth period	AT5G55180	glycosyl hydrolase family 17 protein
Potri.011G140300	Growth period, HGC	AT1G17200	integral membrane family protein
Potri.011G153300	Bud set, Growth period	AT2G46770	*ANAC043* (*NAC DOMAIN CONTAINING PROTEIN 43*)
Potri.012G014500	Bud set, Growth period, Leaf drop	AT3G49220	pectinesterase
Potri.012G088200	Leaf drop	AT5G03340	*CDC48* (*CELL DIVISION PROTEIN 48*)
Potri.012G132400	75%,100% Leaf yellowing	AT5G51810	*GA20OX2* (*GIBBERELLIN 20 OXIDASE 2*)
Potri.013G013100	Bud set, Growth period, HGC, Leaf drop, PBS	AT5G27920	F-box family protein
Potri.013G062400	Leaf drop	NA	Dehydrin
Potri.014G047000	Bud set[1], Leaf drop	AT2G44840	*ERF7* (*ETHYLENE-RESPONSIVE ELEMENT BINDING FACTOR 7*)
Potri.014G087600	Bud set, PBS	AT5G41390	*PLAC8* family
Potri.014G129400	Leaf drop	AT3G62820	pectin methylesterase inhibitor family protein
Potri.014G160000	Bud set, PBS	AT1G04980	*PDIL2-2*
Potri.015G008300	Bud set, Growth period, PBS	AT1G55580	*LAS* (*LATERAL SUPPRESSOR*)
Potri.015G013700	Bud set[1], Leaf drop	AT3G49220	pectinesterase family protein
Potri.015G078600	Bud set[1], PBS	AT5G63000	uncharacterized conserved protein
Potri.015G105000	Bud set, Growth period	AT5G23720	*PHS1* (*PROPYZAMIDE-HYPERSENSITIVE 1*)
Potri.015G125500	Bud set, Growth period, Leaf drop	AT5G23260	*TT16* (*TRANSPARENT TESTA16*)
Potri.015G129100	Bud set, PBS	AT4G22680	*MYB85* (*MYB DOMAIN PROTEIN 85*)
Potri.015G136400	Bud set, 100% Leaf yellowing, Leaf drop	AT5G51990	*DREB1D* (*DEHYDRATION-RESPONSIVE ELEMENT-BINDING PROTEIN 1D*)
Potri.016G000100	Bud set, 100% Leaf yellowing, Leaf drop	AT1G80260	*EMB1427* (*EMBRYO DEFECTIVE 1427*)
Potri.016G000200	Bud set, Leaf drop	AT1G79610	*NHX6* (*NA+/H+ ANTIPORTER 6*)
Potri.016G134600	Leaf drop	AT3G51630	*WNK5* (*WITH NO LYSINE (K) KINASE 5*)
Potri.017G042200	Bud set, Leaf drop	AT3G21175	*ZML1* (*ZIM-LIKE 1*)
Potri.017G086200	Bud set, Growth period, 75%,100% Leaf yellowing, Leaf drop, Leaf lifespan, PBS	AT5G61430	*ANAC100* (*NAC DOMAIN CONTAINING PROTEIN 100*)
Potri.017G090800	Bud set, Leaf drop	AT5G15470	*GAUT14* (*GALACTURONOSYLTRANSFERASE 14*)
Potri.018G033600	Bud set, Growth period, HGC, Leaf drop, PBS	AT1G15550	*GA3OX1* (*GIBBERELLIN 3-OXIDASE 1*)
Potri.018G090100	Bud set, Growth period, Leaf drop	AT2G36460	fructose-bisphosphate aldolase, putative
Potri.019G076800	Bud set	AT1G71692	*AGL12* (*AGAMOUS-LIKE 12*)

[1]Indicates association only retrieved with bud set dates following the summer solstice (occurrences before day 186 removed).
[2]Poplar gene models are annotated to v3 of the genome. See Table S5 for full association results with all phenology traits, gene details, associated SNPs, and complete annotation description.
[3]See Table 1 for trait explanations and units.
HGC, Height growth cessation; PBS, post-bud set period.

leaf flush, bud set (2009–2010), growth period, height growth cessation, post-bud set period and height gain (Table 6). Allelic effects of the SNP (01_4368872; intron) linked the minor homozygotes accessions with earlier leaf flushing (6 d), later height growth cessation and bud set (18, 24 d, respectively), longer growth period (36 d), shorter post-bud set period (22 d), and correspondingly greater height gain (30%) compared to the major homozygote accessions (Table S8). The heterozygous state also showed earlier leaf flushing (3 d) but other traits were equivalent to the major homozygotes.

Some genes had extensive complexity in both genetic variation and the resulting phenotype. Potri.014G102700 (*CYP78A9*; *CYTOCHROME P450 78A9*) had numerous SNPs associated across spring traits, including phenology events bud break (2010–2011), leaf flush (2010–2012) and the ecophysiology trait LMA_{spring} (2010–2011) (Table 6). The six significant SNPs (14_8045578/14_8045889/14_8046287; intergenic, upstream; 14_8047714; coding sequence, synonymous, 14_8048878/14_8049068; intergenic, downstream) are in moderate to high pairwise LD (average $r^2 = 0.47$, range = 0.19–0.99) (Table S6).

Table 6 Genes with significant single nucleotide polymorphism (SNP) markers associated with traits from 2 to 3 categories (phenology, biomass, ecophysiology)

Gene model[2]	Phenology[3]	Biomass[3]	Ecophys.[3]	AT homolog	Annotated description[2]
Potri.001G057400	Bud set, Growth period, HGC, Leaf flush, PBS	Height gain		AT1G27320	*HK3* (*HISTIDINE KINASE 3*)
Potri.001G093800	Bud set[1]	Branches		AT4G11090	Unknown protein
Potri.001G320800	Bud set, Leaf drop, PBS	Branches		AT5G60490	*FLA12* (*FASCICLIN-LIKE ARABINOGALACTAN-PROTEIN 12*)
Potri.002G002000	Bud set, 100% Leaf yellowing, Leaf drop	Bole mass, Whole-tree mass		AT1G21050	DUF617, protein of unknown function
Potri.002G165900	Bud set, Growth period, HGC, Leaf drop, PBS	Branches		AT2G46225	*ABIL1* (*ABI-1-LIKE 1*)
Potri.002G206400	Bud set, Growth period, 100% Leaf yellowing, Leaf drop, PBS	Height, Volume gain		AT2G47750	*GH3.9* (*PUTATIVE INDOLE-3-ACETIC ACID-AMIDO SYNTHETASE GH3.9*)
Potri.002G257900	Bud set, Leaf drop	Branches, Volume		AT5G44030	*CESA4* (*CELLULOSE SYNTHASE A4*)
Potri.003G128600	Bud set, Growth period	Volume, Volume gain		AT1G01620	*PIP1C* (*PLASMA MEMBRANE INTRINSIC PROTEIN 1C*)
Potri.003G143600	Bud set, Growth period, HGC, Leaf drop, PBS	Height, Height gain		AT5G28540	*BIP1/HSP70 PROTEIN*
Potri.003G152700	Bud set[1], Leaf drop	Branches		NA	Unknown function
Potri.003G214200	Bud set, Growth period, HGC, 100% Leaf yellowing	Branches		AT5G13000	*GSL12* (*GLUCAN SYNTHASE-LIKE 12*)
Potri.004G089800	Bud set, Leaf drop, PBS	Branches, Height gain, Volume, Volume gain		AT2G01570	*RGA1* (*REPRESSOR OF GA1-3 1*)
Potri.004G174500	Bud set, PBS	Volume		AT4G35000	*APX3* (*ASCORBATE PEROXIDASE 3*)
Potri.004G230500	Bud set, Growth period, Leaf drop	Branches, Volume, Volume gain		AT1G10320	DUF3594; PHD Zn-finger protein
Potri.005G141200	Bud set	Bole mass, Height, Height gain, Whole-tree mass		AT5G67200	leucine-rich repeat transmembrane protein kinase
Potri.005G199600	Bud set[1], Leaf drop	Branches		AT1G71790	F-actin capping protein beta subunit family protein
Potri.006G038600	Bud set, Growth period, PBS	Height gain, Volume gain		AT2G41200	unknown protein
Potri.006G068400	Bud set, Growth period, Leaf drop	Branches, Height gain		AT5G35410	*SOS2* (*SALT OVERLY SENSITIVE 2*); *CBL-INTERACTING PROTEIN KINASE 24*
Potri.006G158400	Bud set, HGC, Leaf drop	Branches		AT1G03390	transferase activity
Potri.006G275500	Bud break		LMA_{spring}	AT5G10630	*EF-1-alpha* (*ELONGATION FACTOR 1-alpha*)
Potri.007G010700	Bud set, Leaf drop	Volume, Volume gain		AT5G10470	Kinesin (KAR3 subfamily)
Potri.008G038900	Bud set[1]		Leaf shape	AT3G54810	zinc finger (GATA type) family protein
Potri.009G008500	Bud set, Growth period, HGC, 100% Leaf yellowing, Leaf drop, PBS	Height		AT5G60770	*NRT2.4* (*NITRATE TRANSPORTER 2:4*)
Potri.009G008600	Bud set, Growth period, HGC, 100% Leaf yellowing, Leaf drop, PBS	Height		AT1G08090	*NRT2:1* (*NITRATE TRANSPORTER 2:1*)
Potri.009G034500	Bud set, PBS	Height gain		AT2G29130	*LAC2* (*LACCASE 2*)
Potri.009G136600	Bud set, Growth period, HGC, Leaf drop, PBS	Volume		AT4G35100	*PIP3* (*PLASMA MEMBRANE INTRINSIC PROTEIN 3*)
Potri.010G165700	Bud set	Branches		AT3G01140	*MYB106* (*MYB DOMAIN PROTEIN 106*)
Potri.010G184000	Bud set, Growth period, HGC, 100% Leaf yellowing, Leaf drop, Leaf lifespan, PBS	Branches, Volume gain		AT2G40320	*TBL33* (*TRICHOME BIREFRINGENCE-LIKE 33*)
Potri.010G250600	Bud set, HGC, 100% Leaf yellowing, Leaf drop		$A_{max/mass}$, N_{mass}	AT1G51630	*MSR2* (*MANNAN SYNTHESIS RELATED 2*)

Table 6 (Continued)

Gene model[2]	Phenology[3]	Biomass[3]	Ecophys.[3]	AT homolog	Annotated description[2]
Potri.010G254400	Bud set, HGC, Leaf drop		N_{mass}	AT3G54540	*GCN4* (*GENERAL CONTROL NON-REPRESSIBLE 4*)
Potri.013G021700	Bud set, HGC, Leaf drop, PBS	Branches, Volume, Volume gain		AT4G14950	*VMP1* (*VACUOLE MEMBRANE PROTEIN 1*)
Potri.014G102700	Bud break, Canopy duration, Leaf flush		LMA_{spring}	AT3G61880	*CYP78A9* (*CYTOCHROME P450 78A9*)
Potri.014G109800	Bud set, Growth period	Log volume growth rate		AT1G02305	cathepsin B-like cysteine protease
Potri.014G113700	Bud set, Growth period, PBS	Height		AT4G01840	*KCO5* (*CA2+ ACTIVATED OUTWARD RECTIFYING K+ CHANNEL 5*)
Potri.015G002300	Bud set, Leaf drop	Height	Chl_{summer}	AT5G24470	*PRR5* (*PSEUDO-RESPONSE REGULATOR 5*)
Potri.015G002600	Bud set, HGC, Leaf drop, PBS	Height	Chl_{summer}	AT5G24520	*TTG1* (*TRANSPARENT TESTA GLABRA 1*)
Potri.015G004100	Bud set, HGC, 100% Leaf yellowing, Leaf drop, PBS		Chl_{summer}, N_{mass}	AT3G49530	*ANAC062* (*ARABIDOPSIS NAC DOMAIN CONTAINING PROTEIN 62*)
Potri.015G009300	Bud set, Growth period, 100% Leaf yellowing, Leaf drop, PBS		Chl_{summer}	AT4G24060	Dof-type zinc finger domain-containing protein
Potri.017G040800	Bud set, Growth period, 100% Leaf yellowing, Leaf drop	Height	N_{mass}	AT4G15210	*BAM5* (*BETA-AMYLASE 5*)
Potri.017G079600	Bud set[1], Leaf drop	Branches		AT1G74690	*IQD13* (*IQ-DOMAIN 13*)

[1]Association only retrieved with bud set dates following the summer solstice (occurrences before day 186 removed).
[2]Poplar gene models are annotated to v3 of the genome. See Table S5 for full gene details, associated SNPs, and complete annotation description.
[3]See Table 1 for trait explanations and units.
Chl, chlorophyll; Ecophys, ecophysiology; HGC, height growth cessation; LMA, leaf mass per area; N_{mass}, nitrogen per unit mass; PBS, post-bud set period.

Genetic variation was highly complex and different combinations of the six SNPs resulted in 36 genetic variants (haplotypes), all with varying phenotypes (not shown). Individual SNPs had differing effects on phenotypic traits among accessions homozygous for the major or minor allele (depending on the SNP) and did not appear to show phenotypic change in the same direction. Each SNP resulted in variable bud break (6–10 d), leaf flush (2–8 d), canopy duration (11–18 d) and LMA_{spring} (9–13%) with heterozygote accessions intermediate to both homozygotes (Table S8).

Multiple, linked genes with effects across trait categories In some cases, GWAS uncovered significant SNPs with multiple trait associations in high pairwise LD to SNPs in other genes (Fig. 4, Table S9). These formed 'gene blocks' comprising adjacent genes or genes within up to 17 kb on the individual chromosome. Blocks involved 1–6 trait-associated SNPs per gene primarily located within intronic, flanking regions and coding sequence. Pairwise LD between SNPs from different genes ranged from moderate to complete linkage ($r^2 = 0.45-1.0$) (Table S9). Not unexpectedly, genes in linkage often had similar phenotypic effects but dissimilar annotated functions. One cluster *c.* 8 kb in length included Potri.010G250600 (*MSR2*; *MANNAN SYNTHESIS RELATED 2*) and Potri.010G254400 (*GCN4*; *GENERAL CONTROL NON-REPRESSIBLE 4* transporter) with associations to four phenology traits (bud set (2009–2010), height growth cessation, 100% leaf yellowing and leaf drop (2008,10)) and three ecophysiology traits ($A_{max/mass}$, N_{mass} and Chl_{summer}) (Fig. 4a, Tables 6, S9). Allelic effects of the significant SNPs from Potri.010G250600 (10_22291570; 5′UTR/10_22295252; 3′UTR) and Potri.010G254400 (10_22492661; 5′UTR) showed similar phenotypic change (i.e. in the same direction) when the SNPs were analyzed independently (Table S8). The minor homozygote accessions had earlier phenology events and greater leaf N content/chlorophyll/photosynthetic rates compared to the major homozygote accessions while the heterozygous accessions ranged in phenotypic effect from equivalency to either homozygote to divergent phenotypes. Combined, the two genes had 10 genetic variants (haplotypes) with different allelic combinations of the three SNPs and varying phenotypes observed (not shown).

In other clusters, phenotypic effects varied depending on the SNP, but in total, associations spanned all three categories across the linked genes. One region *c.* 17 kb in length included a putative light-response gene Potri.015G002300 (*PRR5*; *PSEUDO-RESPONSE REGULATOR 5*), Potri.015G002600 (*TTG1*; *TRANSPARENT TESTA GLABRA 1*, protein binding) and Potri.015G004100 (*ANAC062*; *NAC-DOMAIN PROTEIN 62* transcription factor) (Fig. 4b, Tables 6, S9). SNP alleles from Potri.015G002300 (15_141448; coding sequence, nonsynonymous/15_141921; coding sequence, synonymous/15_142205 coding sequence, nonsynonymous), Potri.015G002600 (15_162241; intron/15_163004; coding sequence, synonymous) and Potri.015G004100 (15_277979; coding sequence,

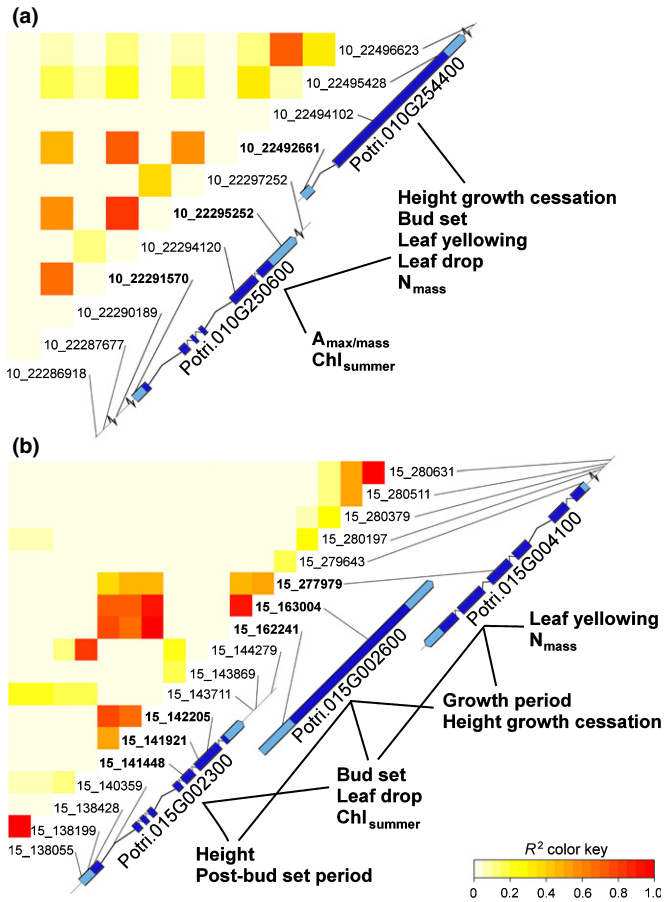

Fig. 4 Linkage of single nucleotide polymorphism (SNP) markers in gene blocks with associations to different traits. (a) Potri.010G250600 (*MSR2*; *MANNAN SYNTHESIS RELATED 2*) and Potri.010G254400 (*GCN4*; *GENERAL CONTROL NON-REPRESSIBLE 4*) have multiple associations with seven traits across two trait categories. Genes are not immediately adjacent; total physical length is 8 kb. (b) Potri.015G002300 (*PRR5*; *PSEUDO-RESPONSE REGULATOR 5*), Potri.015G002600 (*TTG1*; *TRANSPARENT TESTA GLABRA 1*) and Potri.015G004100 (*ANAC062*; *NAC-DOMAIN PROTEIN 62*) have multiple associations with nine traits across three trait categories. Genes are not immediately adjacent; total physical length is 17 kb. Gene scaling and SNP locations are accurate within genes but distances between gene markers are not indicated. Gene regions are identified by coding (dark blue), intron (solid line), 3′UTR/5′UTR (light blue), and noncoding (lines extending beyond UTR regions). Hatch marks on noncoding regions indicate extensive segments of intergenic regions that could not be illustrated within the figure.

nonsynonymous) affected varying combinations of phenology traits (bud set (2008–2010), growth period, height growth cessation, 100% leaf yellowing and leaf drop (2008,10), post-bud set period], two ecophysiology traits (Chl_{summer} (2009,11), N_{mass}) and one biomass trait (tree height). The underlying genetic variation was relatively complex and different combinations of the six SNPs resulted in 13 genetic variants (haplotypes) with diverse effects on the phenotypes (not shown).

Gene with potential pleiotropic effects on unrelated traits In some cases, SNPs were associated with multiple traits that were genetically uncorrelated. The genes Potri.001G057400 (*HK3*; *HISTIDINE KINASE 3*) and Potri.005G138400 (*ABA1*; *ABA DEFICIENT 1*) each had single SNPs associated with numerous phenology traits including leaf flush, which was not correlated with any of the other associated phenology traits (see 'Genes with effects on biomass or ecophysiology' and 'Genes with effects on phenology' above; Tables 5, 6, S10). In other instances, GWAS uncovered separate SNPs within the same gene associated with different traits or suites of traits. Potri.008G038900 (encoding a homolog of *Arabidopsis* zinc finger (GATA type) family protein) had different SNPs associated with either leaf shape or bud set while Potri.014G109800 (encoding a homolog of *Arabidopsis* cathepsin B-like cysteine protease) had different SNPs associated with either log volume growth rate or multiple phenology traits (Table 6). Both cases lacked trait correlation, as neither leaf shape nor log volume growth rate is genetically correlated to any phenology trait (Table S10). Many single genes or gene clusters had SNPs with associations to $A_{max/mass}$, N_{mass} and Chl_{summer}, which are themselves are correlated, but not to any phenology or biomass trait (Tables 6, S10). For example, Potri.015G009300 (encoding a homolog of the *Arabidopsis* Dof-type zinc finger domain-containing protein) had SNPs associated with phenology traits and the ecophysiology trait Chl_{summer} while Potri.017G040800 (*BAM5*; *BETA-AMYLASE 5*) had SNPs associated with biomass and phenology traits and the ecophysiology trait N_{mass}. In addition, the two genes blocks previously described (see earlier) also included uncorrelated ecophysiology trait associations (Potri.010G250600/Potri.010G254400 and Potri.015G002300/Potri.015G002600/Potri.015G004100).

Discussion

In this study, GWAS combining extensive genomic and phenotypic information from natural populations of *P. trichocarpa* uncovered numerous loci underlying variation in biomass, ecophysiology and phenology traits based on: a large collection of individuals spanning much of the natural species range; detailed, replicated trait phenotyping studies; and the largest genome-wide dataset of genetic polymorphisms in *P. trichocarpa* to date.

Genes underlying biomass and ecophysiology

Certain genes implicated by GWAS in determining rates of growth, whole-plant biomass and ecologically related physiological traits in *P. trichocarpa* may have some relationship to the associated phenotype while other associations implicate differing involvement or functionality for *P. trichocarpa* genes compared to their annotated *Arabidopsis* gene homologues that were used for poplar gene annotations (solely based on sequence homologies). For instance, Potri.010G250500 (*EXO70G1*; *EXOCYST SUBUNIT EXO70 FAMILY PROTEIN G1*) was associated with a major effect on biomass variation in the intercorrelated, complex traits of height and tree mass (Table 3). Notably, Potri.010G250500 is the upstream, neighboring gene to Potri.010G250600 (highlighted as potentially pleiotropic and linked to another high-effect gene Potri.010G254400; Fig. 4a). While Potri.010G250500 is unlinked to this gene block, it had substantial effects on tree biomass and may be related and/or

affected by the potentially pleiotropic action of this genomic region. In other species, the specific function of *EXO70G1* is unknown but EXO70 proteins are thought to be involved in auxin efflux carrier recycling contributing to polar auxin transport (Drdová *et al.*, 2013). *Arabidopsis* mutants in a related exocyst component (*EXO70A1*) show reduced fertility and altered cellular development/organogenesis (Synek *et al.*, 2006). Additional associations implicate potentially novel functionality in *P. trichocarpa* related to phenotypic variation. For instance, Potri.010G019000 (*MEE32*; *MATERNAL EFFECT EMBRYO ARREST 32*) was associated with log height growth in *P. trichocarpa* (Table 3) and has only been previously linked with tension wood growth in *P. tremula* (Andersson-Gunnerås *et al.*, 2006). Another example linked Potri.008G105200 (*PHYB*; *PHYTOCHROME B*) with the number of preformed leaves in terminal buds but not phenology (see later).

Despite high intraspecific variation among accessions of *P. trichocarpa* in ecophysiology and biomass/growth-related traits, we found fewer associations relative to phenology and lower total phenotypic variance accounted for by trait-associated SNPs (explained by cumulative R^2) (Figs 2, 3, S4). This may be due to lack of sufficient genomic coverage (i.e. SNPs not on the genotyping array) and would be ameliorated by using a broader sampling of genetic variation. Another possibility might relate to the effects of rare alleles which are hard to detect using GWAS (Ingvarsson & Street, 2011). A third possibility is loss of associations where relationships between SNP loci and geography exist (Balding, 2006) and as identified by PCA for the present study population (McKown *et al.*, 2014). Finally, the heritability values of many biomass and ecophysiology traits are low to moderate suggesting a high local environment-response component (McKown *et al.*, 2014; Fig. S5), and thus, detecting underlying genetic variation in these traits may be inherently difficult using GWAS.

Genes underlying phenology

The greatest number of genetic associations in *P. trichocarpa* involved phenology and also provided the highest cumulative R^2 values found in any studied trait category (Figs 2, 3, S4) supporting high genetic complexity in such traits. The large number of genes involved in phenology is not necessarily unwarranted. In previous studies, numerous genes have also been found that control the bud activity–dormancy cycle in *Populus* (Ruttink *et al.*, 2007; Jackson, 2009; Ma *et al.*, 2010; Rohde *et al.*, 2010, 2011; Olson *et al.*, 2013) and distantly-related *Salix* (Ghelardini *et al.*, 2014). Within this study, most SNPs provided a small contribution to the overall trait, suggesting that the evolution of variation in phenological traits involves numerous loci with small effects (cf. Rockman, 2012). This complex genetic architecture for many phenology traits in *P. trichocarpa* reflects the activity–dormancy cycle of the meristem. The whole-plant switch from active growth to quiescence is intricate and triggered by a number of signals, including daylength, temperature and environmental stressors (Cooke *et al.*, 2012).

In our study, the genetic complexity of phenology traits was observed in the broad span of putative functions in associated genes, particularly late summer and fall phenology traits of bud set and leaf drop (Tables 5, 6, S5). Loci implicated by GWAS underlying phenology trait variation included multiple genes related to environmental response in *Arabidopsis*, such as light perception, hormone signaling, heat shock stress, cold response, water relations and drought stress. Others were related to different types of signaling in *Arabidopsis*, such as calmodulins/calcium, ion transport, phosphatases and kinases. We note that gene numbers and cumulative R^2 identified for phenology traits did not necessarily relate to trait heritability (Figs S4, S5, Table S1). For instance, both bud break and leaf flush ($H^2 = 0.88$, 0.85, respectively) yielded only five genes (one with associations to both traits), while bud set and leaf drop ($H^2 = 0.74$, 0.60, respectively) yielded 222 genes (80 with associations to both traits). Correspondingly, cumulative R^2 was much higher in bud set and leaf drop compared to bud break and leaf flush (Fig. S4), despite the similarly high heritability values (Fig. S5).

The timing of individual phenology events within our population is generally correlated across years but actual dates shifted annually depending on the timing of seasonal environmental cues (McKown *et al.*, 2013). Strong genetic correlations between different phenology events exist and traits tend to be highly intercorrelated within a 'season' but not across seasons (i.e. spring vs late summer/fall; Table S10). Nevertheless, these correlations are not necessarily predictive and intraspecific phenotypic variation in phenology can be somewhat modified from year to year depending on environmental conditions (e.g. the timing of bud set and leaf drop is not fixed). Thus, retrieving repeated associations between SNPs and phenology traits measured across different years supports the biological relevance of these genes. Genes, such as Potri.009G017400 (*BLH1*; *BEL1-LIKE HOMEODOMAIN 1*), Potri.010G215200 (*PRR7*; *PSEUDO-RESPONSE REGULATOR 7*) and Potri.018G033600 (*GA3OX1*; *GIBBERELLIN 3-OXIDASE 1*), were each associated with multiple late summer/fall phenology traits across numerous years and have some precedent for understanding phenology timing. In *Arabidopsis*, *BLH1* regulates the high irradiance response of *PHYTOCHROME A* (*PHYA*) (Staneloni *et al.*, 2009) and modulates signaling by abscisic acid during development (Kim *et al.*, 2013). *BLH1* is also linked to the initiation of bud formation in *P. tremula* × *P. alba* (Ruttink *et al.*, 2007) and is related to late summer *Melampsora* susceptibility in *P. trichocarpa* (La Mantia *et al.*, 2013). *PRR7* is a core clock gene in circadian rhythm determination within *Arabidopsis* through transcription–translation feedback loops (Haydon *et al.*, 2013) and may participate in a similar role in *Populus*. Likewise, *GA3OX1* is implicated in photoperiodic perception (Song *et al.*, 2013) and seed dormancy (Footitt *et al.*, 2013) in *Arabidopsis*, and gibberellins also have well-established roles in the transition to dormancy in *Populus* (Ruttink *et al.*, 2007).

Other genes identified have a characterized function in *Arabidopsis*, but are novel loci for understanding phenotypic variation in *P. trichocarpa*. Potri.014G102700 (*CYP78A9*; *CYTOCHROME P450 78A9*) was repeatedly associated with spring phenology events while Potri.002G242500 (*CB5-B*; *CYTOCHROME B5 ISOFORM B*) was repeatedly found with

late summer/fall phenology events. Although *CB5-B* is not known to relate to phenology, *CYP78A* genes in *Arabidopsis* are generally related to plant size, fertility, and the timing of bud opening and organ abscission (Sotelo-Silveira *et al.*, 2013). Potri.014G102700 (*CYP78A9*) also showed high genetic complexity with variable effects (among related traits), and may be an example of 'conditional neutrality' or 'antagonistic pleiotropy' where different alleles might be favorable depending on the environment (Savolainen *et al.*, 2013). Some genes associated with phenology highlighted links to nutrient availability. Two genes in high linkage, Potri.009G008500 (*NRT2.4*; *NITRATE TRANSPORTER 2:4*) and Potri.009G008600 (*NRT2:1*; *NITRATE TRANSPORTER 2:1*), were associated with late summer/fall phenology events across all years (Tables 6, S5, S9) suggesting that nitrate transporters or nitrogen availability/allocation might affect the regulation of these events. Nitrate transporters have been implicated in nitrogen sensing and auxin signal transduction, and *NRT2.4* is highly expressed in numerous aboveground tissues in *Populus*, including the meristem (Bai *et al.*, 2013). However, neither gene has been previously invoked in phenology in any species.

Many light-associated genes previously implicated in *Populus* phenology were found in our association study while others were not, despite inclusion on the SNP array. This has also been reported in the sister-species *P. balsamifera*, where significant phenology-related SNPs did not necessarily correspond with SNPs uncovered in other association studies for *Populus* (Olson *et al.*, 2013). In addition to previously discussed genes, our GWAS uncovered *COP1* (*CONSTITUTIVE PHOTOMORPHOGENIC 1*), *FAR1* (*FAR-RED IMPAIRED RESPONSE 1*), *PCL1* (*PHYTOCLOCK 1*), *PIL6* (*PHYTOCHROME INTERACTING FACTOR 3-LIKE 6*) and *PRR5* (*PSEUDO-RESPONSE REGULATOR 5*) (Table S5). Yet, notable genes were not among the associations, including *PHYA*, *PHYB*, *CCA1* (*CIRCADIAN CLOCK-ASSOCIATED1*), *FRI* (*FRIGIDA*), *GI* (*GIGANTEA*), *LATE ELONGATED HYPOCOTYL (LHY),* and *TOC1* (*TIMING OF CHLOROPHYLL a/b BINDING PROTEIN/PRR1*) (Ruttink *et al.*, 2007; Ingvarsson *et al.*, 2008; Ma *et al.*, 2010; Rohde *et al.*, 2011; Cooke *et al.*, 2012; Fabbrini *et al.*, 2012; Keller *et al.*, 2012; Olson *et al.*, 2013). SNPs from *CCA1*, *LHY*, *TOC1/PRR1* were retrieved by phenology traits using the simple model (not shown); thus, it is possible that their signal was diminished by correcting for population structure in the mixed model, as these loci have known relationships with geography (McKown *et al.*, 2014; A. Geraldes, unpublished). Nevertheless, it also suggests that variation within these genes does not underlie intraspecific variation of such traits in *P. trichocarpa*, as observed in the closely related *P. balsamifera* (Olson *et al.*, 2013) and more distantly related *P. nigra* (Rohde *et al.*, 2011).

Genes with associations across trait categories and potential functional pleiotropy

One of the significant findings of this study were the numerous genes with multiple significant associations to different traits, including associations across trait categories (Fig. 2, Tables 3–6), and blocks of linked genes with shared genotype–phenotype associations (Fig. 4). We consider these to be indications of pleiotropy in a broad sense (cf. Mackay *et al.*, 2009). The repeated occurrence of ecophysiology traits associated with pleiotropic loci was notable (Table 6), particularly as these had little or no correlative relationship to biomass and/or phenology traits. Other examples of potentially pleiotropic loci have also been uncovered in *P. trichocarpa* (Porth *et al.*, 2014), including a set of genes affecting phenology, wood fiber properties and disease resistance (I. Porth & J. Klápště, unpublished). In this case, phenology traits and fiber properties are functionally uncorrelated traits and the evolution of pleiotropy suggests that the developmental integration of these different traits might have led to their genetic integration (evidenced as trait co-selection; cf. Cheverud, 1996).

The pleiotropic loci in this study provide novel candidates underlying phenotypic variation in *P. trichocarpa* and suggest the presence of genomic regions with importance for environmental response in *P. trichocarpa*. The gene block with Potri.010G250600 (*MSR2*; *MANNAN SYNTHESIS RELATED 2*) and Potri.010G254400 (*GCN4*; *GENERAL CONTROL NONREPRESSIBLE 4*) is potentially pleiotropic in *P. trichocarpa* (Fig. 4a) but the individual genes are not known to be functionally related or pleiotropic within other plant species. In *Arabidopsis*, *MSR2* is localized to the Golgi apparatus, and has been implicated in mannan biosynthesis in a number of tissues, including developing vascular tissue, leaves, stems and flowers (Wang *et al.*, 2013). The *Arabidopsis* transporter *GCN4* is a putative ATP-binding transporter family protein but is not fully characterized in any plant species. Within *P. trichocarpa*, Potri.010G254400 (*GCN4*) is also associated with rates of *Melampsora* infection (La Mantia *et al.*, 2013) and may play a role in disease resistance/susceptibility.

Another block (Fig. 4b) with Potri.015G002300 (*PRR5*; *PSEUDO-RESPONSE REGULATOR 5*), Potri.015G002600 (*TTG1*; *TRANSPARENT TESTA GLABRA 1*) and Potri.015G004100 (*ANAC062*; *NAC-DOMAIN PROTEIN 62*) suggests genes related to environmental sensing and stress response may have pleiotropic activity. *PRR5* is highly upregulated with the onset of short days in *P. tremula* × *P. alba* (Ruttink *et al.*, 2007). It has also has been implicated in growth cessation and bud set in association studies of *P. tremula* × *P. alba* (Ruttink *et al.*, 2007) and *P. tremula* (Ma *et al.*, 2010), and is associated with cell wall crystallinity in *P. trichocarpa* (Porth *et al.*, 2013a). In *Arabidopsis*, *PRR5* plays a role in directly regulating circadian clock genes (Nakamichi *et al.*, 2012). Other direct targets of this regulator include transcription factors involved in flowering, hypocotyl extension and cold-stress responses, suggesting that *PRR5* has light-mediated effects on many physiological processes. Both *TTG1* and *ANAC062* are transcription factors associated with stress responses. *TTG1* affects many plant processes, including flavonoid biosynthesis, response to abscisic acid and root growth in relation to water stress in *Arabidopsis* (Nguyen *et al.*, 2013). Likewise, *ANAC062* is a membrane-associated stress response transcription factor in *Arabidopsis* and involved in abscisic acid response, cold stress and salinity tolerance (Seo & Park, 2010).

The potential pleiotropic loci detected by GWAS in this study spanned a number of functions and may have effects by acting

upstream of signaling pathways that affect multiple traits (such as hormone signaling) or by directly targeting multiple genes for regulation. Within gene blocks with pleiotropic effects, such genomic regions may contain individual genes involved in signaling whose direct targets are in linkage, linked genes with similar functionality, or may represent genes with adaptive influence resulting in linkage through selective forces (Yeaman, 2013).

Conclusions

Employing the GWAS approach to scan the *P. trichocarpa* genome for significant allelic variation underlying important biomass, ecophysiology and phenology traits, we identified numerous individual genes and genomic regions where allelic variation was associated with intraspecific trait variation. The large number of SNP–trait associations highlights the polygenic nature of phenology traits in particular (Fig. 2). It is unlikely, however, that all contributing SNPs or genes are acting equally. Some may be large-effect quantitative trait nucleotides (QTNs) (Rockman, 2012; Martin & Orgogozo, 2013). The complexity of genetic trait architecture also encompasses nonadditive genetic effects such as epistasis (Hansen, 2013) and gene × environment interactions, which might modify the resulting gene effect (Hill, 2010). We noted that many allelic frequencies often accompanied phenotypic change in the same direction, suggestive of directional epistasis (Hansen, 2013) or constitute 'hotspots' where particular genes repeatedly are elements of phenotypic variation in similar traits (Martin & Orgogozo, 2013). Yet, we need to be cautious about the discrepancies between *functional* vs *statistical* epistasis (i.e. relative independence from population variation, cf. Hansen, 2013). The employed linear model in GWAS assumes only additive effects and may be partially fitting epistasis, which we cannot clearly dissect, and thus can exaggerate the 'additive' effect of the detected causative variants. Further work is required to differentiate between phenotypic variation related to epistasis and large-effect QTNs that constitute dispersed adaptive modifications, and more numerous, smaller-effect allelic variations. In the case of the latter, these may be employed to 'fine tune' a phenotype (Martin & Orgogozo, 2013) and/or encompass smaller trait changes required for local adaptation (Savolainen *et al.*, 2007).

Our association results suggest a number of markers with potential ecological effects in *P. trichocarpa*. Many genes identified by GWAS are considered to affect growth and development and/or to respond to signaling and environmental stressors. Numerous loci, including potentially pleiotropic loci, have also been retrieved in parallel F_{ST} outlier studies indicating adaptive potential (A. Geraldes, unpublished; I. Porth & J. Klápště, unpublished). The extensive results from SNP–trait associations within this study highlight multiple avenues for further work, such as investigating functional roles of the genes implicated, genetic pleiotropy between genetically correlated and uncorrelated traits, relationships of genes with geography and local adaptation, and operative roles of important SNP variants in noncoding regions. Conclusively, this study presents an essential platform for future detailed exploration aimed at understanding species-wide ecology and evolution, particularly where numerous genetic mechanisms are invoked.

Acknowledgements

We thank L. E. Gunter, M. S. Azam, E. Drewes, N. Farzaneh, L. Liao, E. Moreno, L. Muenter and L. Quamme for data monitoring, collection and image presentation. We also thank anonymous reviewers for their suggestions and revisions in improving the manuscript. This work was supported by the Genome British Columbia Applied Genomics Innovation Program (Project 103BIO) and Genome Canada Large-Scale Applied Research Project (Project 168BIO) funds to R.D.G., J.E., Q.C.B.C., Y.A.E-K., S.D.M. and C.J.D. and by funds within the BioEnergy Science Center, a US Department of Energy Bioenergy Research Facility under contract DE–AC05–00OR22725.

References

Andersson-Gunneråas S, Mellerowicz EJ, Love J, Segerman B, Ohmiya Y, Coutinho PM, Nilsson P, Henrissat B, Moritz T, Sundberg B. 2006. Biosynthesis of cellulose-enriched tension wood in *Populus*: global analysis of transcripts and metabolites identifies biochemical and developmental regulators in secondary wall biosynthesis. *Plant Journal* **45**: 144–165.

Bai H, Euring D, Volmer K, Janz D, Polle A. 2013. The nitrate transporter (NRT) gene family in poplar. *PLoS ONE* **8**: e72126.

Balding DJ. 2006. A tutorial on statistical methods for population association studies. *Nature Reviews Genetics* **7**: 781–791.

Boyle EI, Weng S, Gollub J, Jin H, Botstein D, Cherry JM, Sherlock G. 2004. GO:TermFinder–open source software for accessing Gene Ontology information and finding significantly enriched Gene Ontology terms associated with a list of genes. *Bioinformatics* **20**: 3710–3715.

Braatne JH, Rood SB, Heilman PE. 1996. Life history, ecology and conservation of riparian cottonwoods in North America. In: Stettler RF, Bradshaw HD Jr, Heilman PE, Hinckley TM, eds. *Biology of Populus and its implications for management and conservation*. Ottawa, Canada: NRC Research Press, 57–85.

Bradbury PJ, Zhang Z, Kroon DE, Casstevens TM, Ramdoss Y, Buckler ES. 2007. TASSEL: software for association mapping of complex traits in diverse samples. *Bioinformatics* **23**: 2633–2635.

Chamaillard S, Fichot R, Vincent-Barbaroux C, Bastien C, Depierreux C, Dreyer E, Villar M, Brignolas F. 2011. Variations in bulk leaf carbon isotope discrimination, growth and related leaf traits among three *Populus nigra* L. populations. *Tree Physiology* **31**: 1076–1087.

Cheverud JM. 1996. Developmental integration and the evolution of pleiotropy. *American Zoologist* **36**: 44–50.

Cooke JEK, Eriksson ME, Junttila O. 2012. The dynamic nature of bud dormancy in trees: environmental control and molecular mechanisms. *Plant, Cell & Environment* **35**: 1707–1728.

Cronk QCB. 2005. Plant eco-devo: the potential of poplar as a model organism. *New Phytologist* **166**: 39–48.

Cumbie WP, Eckert A, Wegrzyn J, Whetten R, Neale D, Goldfarb B. 2011. Association genetics of carbon isotope discrimination, height and foliar nitrogen in a natural population of *Pinus taeda* L. *Heredity* **107**: 105–114.

Drdová EJ, Synek L, Pečenková T, Hála M, Kulich I, Fowler JE, Murphy AS, Zárský V. 2013. The exocyst complex contributes to PIN auxin efflux carrier recycling and polar auxin transport in *Arabidopsis*. *Plant Journal* **73**: 709–719.

Eckert AJ, Bower AD, González-Martínez SC, Wegrzyn JL, Coop G, Neale DB. 2010. Back to nature: ecological genomics of loblolly pine (*Pinus taeda*, Pinaceae). *Molecular Ecology* **19**: 3789–3805.

Eckert AJ, Bower AD, Wegrzyn JL, Pande B, Jermstad KD, Krutovsky KV, St. Clair JB, Neale DB. 2009. Association genetics of coastal Douglas fir (*Pseudotsuga menziesii* var. *menziesii*, Pinaceae). I. Cold-hardiness related traits. *Genetics* **182**: 1289–1302.

Eckert AJ, Wegrzyn JL, Cumbie WP, Goldfarb B, Huber DA, Tolstikov V, Fiehn O, Neale DB. 2012. Association genetics of the loblolly pine (*Pinus taeda*, Pinaceae) metabolome. *New Phytologist* **193**: 890–902.

Evanno G, Regnaut S, Goudet J. 2005. Detecting the number of clusters of individuals using the software STRUCTURE: a simulation study. *Molecular Ecology* **14**: 2611–2620.

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 Biology* **12**: 47.

Falush D, Stephens M, Pritchard JK. 2003. Inference of population structure using multilocus genotype data: linked loci and correlated allele frequencies. *Genetics* **164**: 1567–1587.

Farrar JL. 1995. *Trees in Canada*. Ottawa, Canada: Natural Resources Canada and Fitzhenry and Whiteside Limited.

Footitt S, Huang Z, Clay HA, Mead A, Finch-Savage WE. 2013. Temperature, light and nitrate sensing coordinate *Arabidopsis* seed dormancy cycling, resulting in winter and summer annual phenotypes. *Plant Journal* **74**: 1003–1015.

Fournier-Level A, Korte A, Cooper MD, Nordborg M, Schmitt J, Wilczek AM. 2011. A map of local adaptation in *Arabidopsis thaliana*. *Science* **334**: 86–89.

Geraldes A, DiFazio SP, Slavov GT, Ranjan P, Muchero W, Hannemann J, Gunter LE, Wymore AM, Grassa CJ, Farzaneh N *et al.* 2013. A 34K SNP genotyping array for *Populus trichocarpa*: design, application to the study of natural populations and transferability to other *Populus* species. *Molecular Ecology Resources* **13**: 306–323.

Geraldes A, Pang J, Thiessen N, Cezard T, Moore R, Zhao Y, Tam A, Wang S, Friedmann M, Birol I *et al.* 2011. SNP discovery in black cottonwood (*Populus trichocarpa*) by population transcriptome resequencing. *Molecular Ecology Resources* **11**: 81–92.

Ghelardini L, Berlin S, Weih M, Lagercrantz U, Gyllenstrand N, Rönnberg-Wästljung AC. 2014. Genetic architecture of spring and autumn phenology in *Salix*. *BMC Plant Biology* **14**: 31.

González-Martínez SC, Huber D, Ersoz E, Davis JM, Neale DB. 2008. Association genetics in *Pinus taeda* L. II. Carbon isotope discrimination. *Heredity* **101**: 19–26.

Gornall JL, Guy RD. 2007. Geographic variation in ecophysiological traits of black cottonwood (*Populus trichocarpa*). *Canadian Journal of Botany* **85**: 1202–1213.

Graffelman J, Morales J. 2008. Graphical test for Hardy-Weinberg equilibrium based on the temporary plot. *Human Heredity* **65**: 77–84.

Guillot G, Mortier F, Estoup A. 2005. GENELAND: a computer package for landscape genetics. *Molecular Ecology Notes* **5**: 712–715.

Hansen TF. 2013. Why epistasis is important for selection and adaptation. *Evolution* **67**: 3501–3511.

Haydon MJ, Mielczarek O, Robertson FC, Hubbard KE, Webb AA. 2013. Photosynthetic entrainment of the *Arabidopsis thaliana* circadian clock. *Nature* **502**: 689–692.

Higham N. 2002. Computing the nearest correlation matrix – a problem from finance. *IMA Journal of Numerical Analysis* **22**: 329–343.

Hill WG. 2010. Understanding and using quantitative genetic variation. *Philosophical Transactions of the Royal Society B* **365**: 73–85.

Holliday JA, Ralph SG, White R, Bohlmann J, Aitken SN. 2008. Global monitoring of autumn gene expression within and among phenotypically divergent populations of Sitka spruce (*Picea sitchensis*). *New Phytologist* **178**: 103–122.

Holliday JA, Ritland K, Aitken SN. 2010. Widespread, ecologically relevant genetic markers developed from association mapping of climate-related traits in Sitka spruce (*Picea sitchensis*). *New Phytologist* **188**: 501–514.

Ibáñez C, Kozarewa I, Johansson M, Ögren E, Rohde A, Eriksson ME. 2010. Circadian clock components regulate entry and affect exit of seasonal dormancy as well as winter hardiness in *Populus* trees. *Plant Physiology* **153**: 1823–1833.

Ingvarsson PK, Garcia MV, Luquez V, Hall D, Jansson S. 2008. Nucleotide polymorphism and phenotypic associations within and around the phytochrome B2 locus in European aspen (*Populus tremula*, Salicaceae). *Genetics* **178**: 2217–2226.

Ingvarsson PK, Street NR. 2011. Association genetics of complex traits in plants. *New Phytologist* **189**: 909–922.

Jackson DA. 1993. Stopping rules in Principal Components Analysis: a comparison of heuristical and statistical approaches. *Ecology* **74**: 2204–2214.

Jackson SD. 2009. Plant responses to photoperiod. *New Phytologist* **181**: 517–531.

Keller SR, Levsen N, Olson MS, Tiffin P. 2012. Local adaptation in the flowering-time gene network of balsam poplar, *Populus balsamifera* L. *Molecular Biology and Evolution* **29**: 3143–3152.

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

Kim D, Cho YH, Ryu H, Kim Y, Kim TH, Hwang I. 2013. *BLH1* and *KNAT3* modulate ABA responses during germination and early seedling development in *Arabidopsis*. *Plant Journal* **75**: 755–766.

La Mantia J, Klápště J, El-Kassaby YA, Azam S, Guy RD, Douglas CJ, Mansfield SD, Hamelin R. 2013. Association analysis identifies *Melampsora* × *columbiana* poplar leaf rust resistance SNPs. *PLoS ONE* **8**: e78423.

Loiselle BA, Sork VL, Nason J, Graham C. 1995. Spatial genetic structure of a tropical understory shrub, *Psychotria officinalis* (Rubiaceae). *American Journal of Botany* **82**: 1420–1425.

Ma XF, Hall D, St. Onge KR, Jansson S, Ingvarsson PK. 2010. Genetic differentiation, clinal variation and phenotypic associations with growth cessation across the *Populus tremula* photoperiodic pathway. *Genetics* **186**: 1033–1044.

Mackay TFC, Stone EA, Ayroles JF. 2009. The genetics of quantitative traits: challenges and prospects. *Nature Reviews Genetics* **10**: 565–577.

Marchini JL. 2013. *POPGEN: An R package for statistical and population genetic inference. Version 1.0-3*. [WWW document] URL http://www.stats.ox.ac.uk/~marchini/software.html [accessed 7 April 2013].

Martin A, Orgogozo V. 2013. The loci of repeated evolution: a catalog of genetic hotspots of phenotypic variation. *Evolution* **67**: 1235–1250.

McKown AD, Guy RD, Azam MS, Drewes EC, Quamme L. 2013. Seasonality and phenology alter functional leaf traits. *Oecologia* **172**: 653–665.

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 variation and genetic structure in *Populus trichocarpa*. *New Phytologist* **201**: 1263–1276.

Morris GP, Ramu P, Deshpande SP, Hash CT, Shah T, Upadhyaya HD, Riera-Lizarazu O, Brown PJ, Acharya CB, Mitchell SE *et al.* 2013. Population genomic and genome-wide association studies of agroclimatic traits in sorghum. *Proceedings of the National Academy of Sciences, USA* **110**: 453–458.

Nakamichi N, Kiba T, Kamioka M, Suzuki T, Yamashino T, Higashiyama T, Sakakibara H, Mizuno T. 2012. Transcriptional repressor PRR5 directly regulates clock-output pathways. *Proceedings of the National Academy of Sciences, USA* **109**: 17 123–17 128.

Neale DB, Kremer A. 2011. Forest tree genomics: growing resources and applications. *Nature Reviews Genetics* **12**: 111–122.

Nguyen HN, Kim JH, Hyun WY, Nguyen NT, Hong SW, Lee H. 2013. TTG1-mediated flavonols biosynthesis alleviates root growth inhibition in response to ABA. *Plant Cell Reports* **32**: 503–514.

Olson MS, Levsen N, Soolanayakanahally RY, Guy RD, Schroeder WR, Keller SR, Tiffin P. 2013. The adaptive potential of *Populus balsamifera* L. to phenology requirements in a warmer global climate. *Molecular Ecology* **22**: 1214–1230.

Parchman TL, Gompert Z, Mudge J, Schilkey FD, Benkman CW, Buerkle CA. 2012. Genome-wide association genetics of an adaptive trait in lodgepole pine. *Molecular Ecology* **21**: 2991–3005.

Pearson TA, Manolio TA. 2008. How to interpret a genome-wide association study. *JAMA* **299**: 1335–1344.

Peres-Neto PR, Legendre P, Dray S, Borcard D. 2006. Variation partitioning of species data matrices: estimation and comparison of fractions. *Ecology* **87**: 2614–2625.

Porth I, Klápště J, McKown AD, La Mantia J, Hamelin RC, Skyba O, Unda F, Friedmann MC, Cronk QCB, Ehlting J *et al.* 2014. Extensive functional

pleiotropy of *REVOLUTA* substantiated through forward genetics. *Plant Physiology* **164**: 548–554.

Porth I, Klápště J, Skyba O, Hannemann J, McKown AD, Guy RD, DiFazio SP, Muchero W, Ranjan P, Tuskan GA *et al.* 2013a. Genome-wide association mapping for wood characteristics in *Populus* identifies an array of candidate single nucleotide polymorphisms. *New Phytologist* **200**: 710–726.

Porth I, Klápště J, Skyba O, Lai BS, Geraldes A, Muchero W, Tuskan GA, Douglas CJ, El-Kassaby YA, Mansfield SD. 2013b. *Populus trichocarpa* cell wall chemistry and ultrastructure trait variation, genetic control and genetic correlations. *New Phytologist* **197**: 777–790.

Pritchard JK, Stephens M, Donnelly P. 2000. Inference of population structure using multilocus genotype data. *Genetics* **155**: 945–959.

R Core Development Team. 2011. *R: a language and environment for statistical computing*. Vienna, Austria: R Foundation for Statistical Computing.

Riedelsheimer C, Lisec J, Czedik-Eysenberg A, Sulpice R, Flis A, Grieder C, Altmann T, Stitt M, Willmitzer L, Melchinger AE. 2012. Genome-wide association mapping of leaf metabolic profiles for dissecting complex traits in maize. *Proceedings of the National Academy of Sciences, USA* **109**: 8872–8877.

Rockman MV. 2012. The QTN program and the alleles that matter for evolution: all that's gold does not glitter. *Evolution* **66**: 1–17.

Rohde A, Bastien C, Boerjan W. 2011. Temperature signals contribute to the timing of photoperiodic growth cessation and bud set in poplar. *Tree Physiology* **31**: 472–482.

Rohde A, Storme V, Jorge V, Gaudet M, Vitacolonna N, Fabbrini F, Ruttink T, Zaina G, Marron N, Dillen S *et al.* 2010. Bud set in poplar – genetic dissection of a complex trait in natural and hybrid populations. *New Phytologist* **189**: 106–121.

Ruttink T, Arend M, Morreel K, Storme V, Rombauts S, Fromm J, Bhalerao RP, Boerjan W, Rohde A. 2007. A molecular timetable for apical bud formation and dormancy induction in poplar. *Plant Cell* **19**: 2370–2390.

Savolainen O, Lascoux M, Merilä J. 2013. Ecological genomics of local adaptation. *Nature Reviews Genetics* **14**: 807–820.

Savolainen O, Pyhäjärvi T, Knürr T. 2007. Gene flow and local adaptation in trees. *Annual Review of Ecology Evolution and Systematics* **38**: 595–619.

Schwartzman A, Lin X. 2011. The effect of correlation in false discovery rate estimation. *Biometrika* **98**: 199–214.

Seo PJ, Park CH. 2010. A membrane-bound NAC transcription factor as integrator of biotic and abiotic stress signals. *Plant Signaling & Behavior* **5**: 481–483.

Slavov GT, DiFazio SP, Martin J, Schackwitz W, Muchero W, Rodgers-Melnick E, Lipphardt MF, Pennacchio CP, Hellsten U, Pennacchio LA *et al.* 2012. Genome resequencing reveals multiscale geographic structure and extensive linkage disequilibrium in the forest tree *Populus trichocarpa*. *New Phytologist* **196**: 713–725.

Song YH, Ito S, Imaizumi T. 2013. Flowering time regulation: photoperiod- and temperature-sensing in leaves. *Trends in Plant Science* **18**: 575–583.

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

Soolanayakanahally RY, Guy RD, Silim SN, Song M. 2013. Timing of photoperiodic competency causes phenological mismatch in balsam poplar (*Populus balsamifera* L.). *Plant, Cell & Environment* **36**: 116–127.

Sork VL, Aitken SN, Dyer RJ, Eckert AJ, Legendre P, Neale DB. 2013. Putting the landscape into the genomics of trees: approaches for understanding local adaptation and population responses to changing climate. *Tree Genetics & Genomes* **9**: 901–911.

Sotelo-Silveira M, Cucinotta M, Chauvin AL, Chávez Montes RA, Colombo L, Marsch-Martínez N, de Folter S. 2013. Cytochrome P450 CYP78A9 is involved in *Arabidopsis* reproductive development. *Plant Physiology* **162**: 779–799.

Staneloni RJ, Rodriguez-Batiller MJ, Legisa D, Scarpin MR, Agalou A, Cerdán PD, Meijer AH, Ouwerkerk PB, Casal JJ. 2009. Bell-like homeodomain selectively regulates the high-irradiance response of phytochrome A. *Proceedings of the National Academy of Sciences, USA* **106**: 13 624–13 629.

Stapley J, Reger J, Feulner PG, Smadja C, Galindo J, Ekblom R, Bennison C, Ball AD, Beckerman AP, Slate J. 2010. Adaptation genomics: the next generation. *Trends in Ecology & Evolution* **25**: 705–712.

Synek L, Schlager N, Eliás M, Quentin M, Hauser MT, Zárský V. 2006. AtEXO70A1, a member of a family of putative exocyst subunits specifically expanded in land plants, is important for polar growth and plant development. *Plant Journal* **48**: 54–72.

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**: 1596–1604.

Vandenkoornhuyse P, Dufresne A, Quaiser A, Gouesbet G, Binet F, Francez AJ, Mahé S, Bormans M, Lagadeuc Y, Couée I. 2010. Integration of molecular functions at the ecosystemic level: breakthroughs and future goals of environmental genomics and post-genomics. *Ecology Letters* **13**: 776–791.

Wang D, Sun Y, Stang P, Berlin JA, Wilcox MA, Li Q. 2009. Comparison of methods for correcting population stratification in a genome-wide association study of rheumatoid arthritis: principal component analysis versus multidimensional scaling. *BMC Proceedings* **3**(Suppl 7): S109.

Wang Y, Mortimer JC, Davis J, Dupree P, Keegstra K. 2013. Identification of an additional protein involved in mannan biosynthesis. *Plant Journal* **73**: 105–117.

Weir BS, Hill WG, Cardon LR. 2004. Allelic association patterns for a dense SNP map. *Genetic Epidemiology* **24**: 442–450.

Xie C-Y, Ying CC, Yanchuk AD, Holowachuk DL. 2009. Ecotypic mode of regional differentiation caused by restricted gene migration: a case in black cottonwood (*Populus trichocarpa*) along the Pacific Northwest coast. *Canadian Journal of Forest Research* **39**: 519–526.

Yeaman S. 2013. Genomic rearrangements and the evolution of clusters of locally adaptive loci. *Proceedings of the National Academy of Sciences, USA* **110**: E1743–E1751.

Yu J, Pressoir G, Briggs WH, Vroh Bi I, Yamasaki M, Doebley JF, McMullen MD, Gaut BS, Nielsen DM, Holland JB *et al.* 2006. A unified mixed-model method for association mapping that accounts for multiple levels of relatedness. *Nature Genetics* **38**: 203–208.

Supporting Information

Additional supporting information may be found in the online version of this article.

Fig. S1 Compiled Q–Q plots for all *phenology* traits indicating five structure model choices (simple, P, Q, K, P + K, Q + K).

Fig. S2 Compiled Q–Q plots for all *biomass* traits indicating five structure model choices (simple, P, Q, K, P + K, Q + K).

Fig. S3 Compiled Q–Q plots for all *ecophysiology* traits indicating five structure model choices (simple, P, Q, K, P + K, Q + K).

Fig. S4 Proportion of the phenotypic variance explained by significant SNP–trait associations (cumulative R^2) for each trait within each trait category at $P < 1.7 \times 10^{-6}$.

Fig. S5 Proportion of the phenotypic variance explained by significant SNP–trait associations (cumulative R^2) vs broad-sense heritability values (H^2) of phenotypic traits.

Table S1 (a) Phenotypic trait data in *Populus trichocarpa* used for genome-wide association study; (b) genotypic SNP data *Populus trichocarpa* used for genome-wide association study

Table S2 Bayesian Information Criterion (BIC) indicating log likelihood values for model selection in genome-wide association study

Table S3 Comparison of significant genes and SNPs from genome-wide association study by chromosome

Table S4 Annotations of SNPs uncovered from 'Nisqually-1' genome *Populus trichocarpa* v2.2 to v3.0

Table S5 Full details of SNP–trait associations using genome-wide association studies indicating population structure correction model, associated *P*-values, significance (α), marker R^2, closest *Arabidopsis* homolog and putative gene function/annotation

Table S6 Pairwise linkage disequilibrium (LD) r^2 values between all SNP markers significant at $P < 1.7 \times 10^{-6}$

Table S7 Comparison of Gene Ontology (GO) terms (categorized as function or process) assigned to genes uncovered through SNP discovery ($P < 1.7 \times 10^{-6}$) and genes on the array

Table S8 Candidate genes/SNPs with multiple associations to traits and average phenotypic value associated with each allelic variant

Table S9 Pairwise linkage disequilibrium (LD) r^2 values between SNP markers of different genes with effects across multiple trait categories

Table S10 Genetic trait correlations among biomass, ecophysiology, and phenology traits indicating Pearson's Product-Moment Correlation coefficients (*r*)

Please note: Wiley Blackwell are not responsible for the content or functionality of any supporting information supplied by the authors. Any queries (other than missing material) should be directed to the *New Phytologist* Central Office.