Multitrait genome-wide association analysis of *Populus trichocarpa* identifies key polymorphisms controlling morphological and physiological traits

Hari B. Chhetri[1] (iD), David Macaya-Sanz[1], David Kainer[2], Ajaya K. Biswal[3,4], Luke M. Evans[1], Jin-Gui Chen[2], Cassandra Collins[5], Kimberly Hunt[5], Sushree S. Mohanty[3], Todd Rosenstiel[6], David Ryno[4], Kim Winkeler[5], Xiaohan Yang[2], Daniel Jacobson[2], Debra Mohnen[3,4], Wellington Muchero[2], Steven H. Strauss[7], Timothy J. Tschaplinski[2], Gerald A. Tuskan[2] and Stephen P. DiFazio[1] (iD)

[1]Department of Biology, West Virginia University, Morgantown, WV 26506, USA; [2]Biosciences Division, Oak Ridge National Laboratory, Oak Ridge, TN 37831, USA; [3]Department of Biochemistry and Molecular Biology, University of Georgia, Athens, GA 30602, USA; [4]Complex Carbohydrate Research Center, University of Georgia, Athens, GA 30602, USA; [5]ArborGen, Inc., 2011 Broadbank Ct., Ridgeville, SC 29472, USA; [6]Department of Biology, Portland State University, Portland, OR 97207, USA; [7]Department of Forest Ecosystems & Society, Oregon State University, Corvallis, OR 97331, USA

Summary

• Genome-wide association studies (GWAS) have great promise for identifying the loci that contribute to adaptive variation, but the complex genetic architecture of many quantitative traits presents a substantial challenge.

• We measured 14 morphological and physiological traits and identified single nucleotide polymorphism (SNP)-phenotype associations in a *Populus trichocarpa* population distributed from California, USA to British Columbia, Canada. We used whole-genome resequencing data of 882 trees with more than 6.78 million SNPs, coupled with multitrait association to detect polymorphisms with potentially pleiotropic effects. Candidate genes were validated with functional data.

• Broad-sense heritability (H^2) ranged from 0.30 to 0.56 for morphological traits and 0.08 to 0.36 for physiological traits. In total, 4 and 20 gene models were detected using the single-trait and multitrait association methods, respectively. Several of these associations were corroborated by additional lines of evidence, including co-expression networks, metabolite analyses, and direct confirmation of gene function through RNAi.

• Multitrait association identified many more significant associations than single-trait association, potentially revealing pleiotropic effects of individual genes. This approach can be particularly useful for challenging physiological traits such as water-use efficiency or complex traits such as leaf morphology, for which we were able to identify credible candidate genes by combining multitrait association with gene co-expression and co-methylation data.

Introduction

A long-standing question in evolutionary biology is the role of selection in shaping the spatial and temporal patterns of phenotypic variation (Weigel & Nordborg, 2015). In the era of genomics, it is now possible to identify the molecular mechanisms underlying phenotypic variation on the landscape. Due to their wide geographical distribution and climatic gradients, forest trees are an excellent model system for testing how genetic drift, and selection affect genetic variation within a species (Neale & Savolainen, 2004; Ingvarsson & Street, 2011; Neale & Kremer, 2011; Ingvarsson *et al.*, 2016). Forest trees generally have large effective population sizes, extensive gene flow, high genetic variation, and local adaptation (Neale & Savolainen, 2004; González-

Martínez *et al.*, 2006; Ingvarsson & Street, 2011). Of the total land area on Earth, *c.* 31% is occupied by forests, which are of great ecological and economic importance (MacDicken *et al.*, 2016). Therefore, understanding the factors affecting variation in traits that are important for environmental adaptation is of utmost importance, particularly in the context of rapidly changing climates (Aitken *et al.*, 2008).

Association genetics has emerged as a major tool for identifying the genomic regions underlying traits of interest (Ingvarsson & Street, 2011). Using natural populations that have undergone many generations of recombination between ancestral haplotypes allows at fine scale the identification of the genomic region affecting a trait. Nevertheless, one of the major downsides of association mapping is the requirement for large numbers of loci and

individuals (Visscher *et al.*, 2017). With recent advances in sequencing technologies, acquiring genomic data at a whole-genome scale has become much more feasible. Nevertheless, despite the high heritability of many morphological traits, only a small proportion of heritability is explained by SNPs in most GWAS analyses, suggesting insufficient statistical power (Solovieff *et al.*, 2013). Power analyses indicated that most association studies in forest trees are orders of magnitude too small to detect the effects of alleles of small effect and low frequency (Visscher *et al.*, 2017) that collectively account for a large fraction of the heritability of complex traits (Boyle *et al.*, 2017).

While increasing sample size of GWAS populations is clearly desirable, it is costly and, in some cases, may not be feasible. Alternatively, approaches that can improve the power of GWAS from the same inputs can be used. One approach is to use gene- and pathway-based analysis, in which GWAS is performed on a set of SNPs or genes (Kim *et al.*, 2016). Another option is to take a multitrait approach, in which GWAS is performed with multiple related traits combined in a multivariate framework. Recently, the latter approach has gained some popularity because it offers substantial increase in power compared with the standard univariate approach (Porter & O'Reilly, 2017). One of the big advantages of multitrait GWAS is that missing information in one of the phenotypes in the multitrait set can be complemented by the other phenotypes (Ritchie *et al.*, 2015). The increased power of multitrait GWAS depends in part on correlation among traits (Porter & O'Reilly, 2017) and the combination of weak genetic effects across the traits (Casale *et al.*, 2015). Multitrait GWAS also takes advantage of pleiotropic effects of polymorphisms, thereby increasing statistical power even when the traits have low correlation (Broadaway *et al.*, 2016; Hackinger & Zeggini, 2017). Finally, unlike analyses based on principal components, multitrait GWAS effectively captures indirect genetic effects for which a SNP affects one phenotype through its effects on a functionally related phenotype (Stephens, 2013; Porter & O'Reilly, 2017).

Here we used the model species *P. trichocarpa* to explore the utility of multitrait GWAS to detect genetic variants controlling adaptive traits. The genus *Populus* has a wide distribution in the northern hemisphere and is dioecious, wind pollinated, and highly heterozygous. *Populus* are also fast growing, easy to propagate, and demonstrate interspecific hybrid vigour, all of which makes the genus a model system with high economic potential for the production of forest products and biofuels (Jansson & Douglas, 2007; Rubin, 2008). *Populus trichocarpa* is found in the section Tacamahaca (Eckenwalder, 1996) and is distributed from central California to northern British Columbia (BC). It was the first tree genome to be sequenced (Tuskan *et al.*, 2006) and considerable genetic resources are available, including abundant transcriptomes (Sjödin *et al.*, 2009; Geraldes *et al.*, 2011; Zhang J. *et al.*, 2018) and resequencing data (Slavov *et al.*, 2012; Evans *et al.*, 2014), as well as multiple association populations in replicated plantations (Evans *et al.*, 2014; McKown *et al.*, 2014c; Holliday *et al.*, 2016). From these studies, and others, it has been shown that climate plays a major role within this species in shaping genetic variation and driving selection. We show here that multitrait GWAS is a substantially more powerful approach than single-trait GWAS in identifying molecular determinants of quantitative traits, although much remaining heritable variation remains to be identified.

Materials and Methods

Plantation establishment and phenotyping

The *P. trichocarpa* association population consisted of 1084 trees collected from natural populations in western Washington, Oregon (OR) and California States, and in BC (Fig. 1). The trees were clonally propagated from stem cuttings and planted in a common garden in Corvallis, OR in July 2009 (Evans *et al.*, 2014). The plantation consists of three blocks in a completely randomised design and the trees were planted at 2 × 3 m spacing. The plantation was coppiced in December 2010 and again in December 2013. Coppiced plants were allowed to resprout and grow for one season, after which they were pruned to a single leader in January of the following year (2012 and 2015, respectively).

In December 2013, 759 trees were sampled for carbon isotope analysis. Wood cores (12 mm) were taken from breast height of the tree and the 2012 growth ring was selected for analysis. Cross-sections of the wood tissue representing the entire growth ring (*c.* 1.2–1.8 mg) representing early, intermediate and late wood were sampled. The wood samples were oven-dried at 65°C for at least 72 h, weighed and wrapped in a tin capsule before sending to the Appalachian Ecology Laboratory in Frostburg, Maryland, USA for analysis. Carbon isotope composition ($\delta^{13}C$) was estimated as follows:

$$\delta^{13}C = \left(\frac{R_{\text{sample}}}{R_{\text{standard}}} - 1 \right) \times 1000,$$

where, R_{sample} and R_{standard} are the $^{13}C/^{12}C$ ratios in a sample and standard, respectively.

In July 2014, leaf characteristics were measured for 1056 trees (one complete block plus a subset of replicates; Table 1). The first and second fully expanded leaves (counting from the apex) were collected from a branch receiving full sunlight. One of the leaves was used for measuring petiole length and diameter with a digital caliper and then scanned using a hand-held scanner. Images were analysed to estimate leaf area, leaf length, leaf width and leaf perimeter using IMAGEJ software (Schindelin *et al.*, 2015). Dry weights were determined for the same leaves for estimates of specific leaf area (SLA). Leaf chlorophyll (SPAD) was assessed using a SPAD 502 Plus meter (Spectrum Technologies, Aurora, IL, USA) with an average of three replicate measures on leaf section. Abaxial stomatal density was measured by applying clear nail polish to the broadest part of the leaf close to the midrib. A clear piece of tape was then used to capture an imprint of the epidermal leaf surface. These samples were mounted on slides and the number of stomata in 1 mm² area in four random fields was counted at ×400 magnification. Pre-dawn leaf water potential was measured for 964 trees using the cut petiole method (Scholander *et al.*, 2016). Measurements were made on a fully
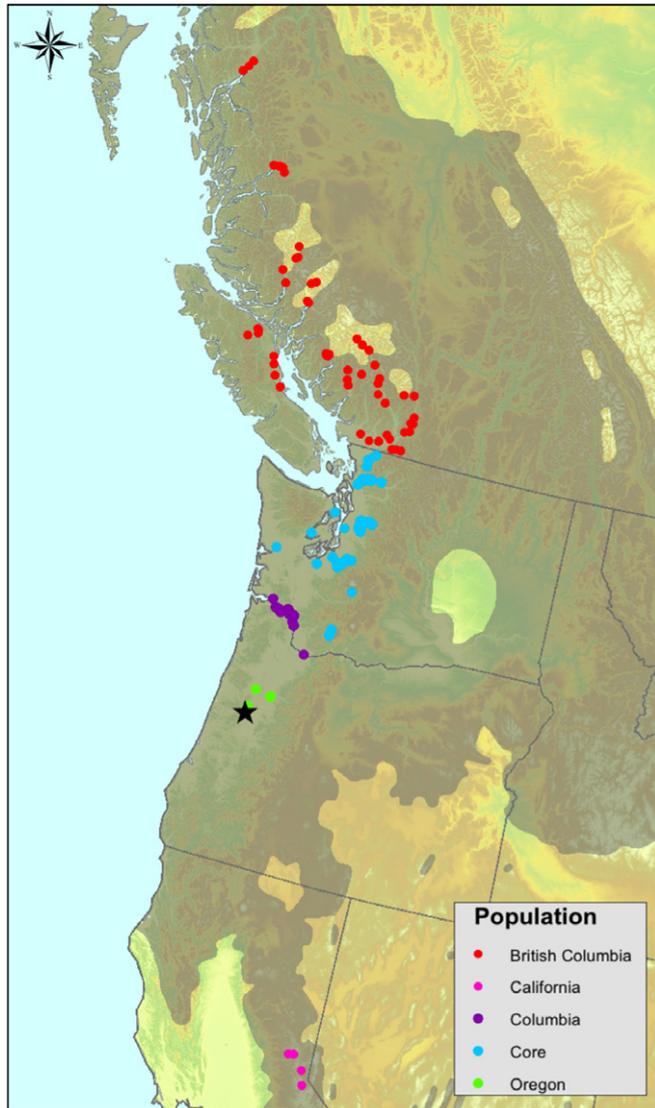
Fig. 1 Source locations of 882 *Populus trichocarpa* genotypes sampled in this study (coloured dots). The trees were grown in a common garden in Corvallis, Oregon, USA (black star).

expanded leaf from the middle of the canopy. Leaves were collected between the the hours of 02:00 h and 05:00 h and a pressure bomb was used in the field to measure the pressure of N_2 gas required to force sap from the cut petiole. Height was measured following the 2015 growing season.

Statistical analyses

All measurements were checked for recording errors and outliers were removed. Data were checked for normality. The phenotypic values were adjusted for any within garden microsite variation using Thin Plate Spline (*tps*) regression using the *Tps* function of the FIELDS package in R. Using *tps*-adjusted phenotypic values, broad-sense heritabilities were estimated for all traits using the genotypes with replicate clonal measurements. Variance components were estimated by fitting the model with the *lmer* and *ranef*

functions of the LME4 package in R, with genotype as a random effect, and error estimated from the residuals of the model:

$$H^2 = \frac{\sigma_G^2}{\sigma_G^2 + \sigma_E^2}$$

Genetic correlation between traits was estimated using Best Linear Unbiased Predictors (BLUPs) from the same model. Using the clonal *tps*-adjusted values, the Pearson correlation was performed using the *cor* function of the STATS package in R. The *prcomp* function of the GGBIPLOT package in R was used to estimate the relationships of the phenotypes using linear combinations (principal components) of the original phenotypic values. Leaf water potential was not included in the principal component analysis (PCA) due to low heritability.

Genotypic data

Preparation of the genotypic data was as described in Evans *et al.* (2014) and Weighill *et al.* (2018). Briefly, whole-genome resequencing was performed for 1053 trees using Illumina genetic analysers at the DOE Joint Genome Institute. Pairwise relatedness was calculated using GCTA (Yang *et al.*, 2011), taking population structure into account. Trees related more closely than first cousins were removed from the analyses. The remaining 882 individuals were used for all subsequent analyses. A genetic relationship matrix was estimated for the remaining trees using Genome-wide Efficient Mixed Model Association (GEMMA), and used as a covariate in GWAS analyses. PCs of all resequencing data were estimated using *smartpca* from EIGENSOFT v.6.1.4 and the first 60 PCs were selected as potential covariates for the association tests. Stepwise regression using the *step* function with default selection criteria (that is both backward and forward selection) of the MASS package in R was used for selecting PCs that were significantly associated with each phenotype or group of phenotypes. All significant PCs were used as covariates for GWAS (Supporting Information Table S1). Finally, SNPs with minor allele frequency ≤ 0.05 and markers with severe departures from Hardy–Weinberg expectations were removed.

Test for association

Association tests were performed using GEMMA (Zhou and Stephens 2012; Zhou & Stephens, 2014). Phenotypic BLUPs, genetic relationship matrix, significant PC axes of the genotypic data and 6781 211 SNPs (remaining SNPs after filtering with MAF < 0.05) were used for the association test. Single-trait GWAS was run for 14 phenotypes (Table 1). The tested model was:

$$y = W\alpha + x\beta + u + \epsilon,$$

where *y* is an *n*-vector of phenotypic BLUP values, where *n* is the number of individuals tested; *W* is an $n \times c$ matrix of covariates; α is a *c*-vector of corresponding coefficients, where *c* is the

Table 1 Broad-sense heritability estimates and the number of single nucleotide polymorphism (SNP)-trait associations for morphological and physiological traits in *Populus trichocarpa*.

Trait	H² (TPS)[a]	H²[b]	N[c]	Total trees[d]	SNPs < 1 × 10⁻⁰⁷[e]	PCs[f]	Chip_H² (± CI)[g]
Morphology							
Height (HT)	0.363	0.320	876	2378 (851)	0	27	1 (± 0.002)
Leaf area (LA)	0.344	0.336	794	1056 (262)	0	23	0.793 (± 0.265)
Leaf aspect ratio (AR)	0.462	0.477	794	1056 (262)	0	20	0.61 (± 0.312)
Leaf dry weight (LD)	0.371	0.360	844	1094 (250)	0	26	0.751 (± 0.251)
Leaf length (LL)	0.370	0.360	794	1056 (262)	0	22	0.766 (± 0.262)
Leaf perimeter (LP)	0.362	0.351	794	1056 (262)	0	22	0.79 (± 0.262)
Leaf width (LW)	0.344	0.346	794	1056 (262)	0	25	0.76 (± 0.266)
Petiole diameter (PD)	0.297	0.184	839	1124 (285)	0	20	0.62 (± 0.263)
Petiole length (PL)	0.561	0.562	839	1124 (285)	0	23	0.881 (± 0.28)
Specific leaf area (SL)	0.371	0.376	784	1010 (226)	2	19	0.746 (± 0.257)
Stomatal density (SD)	0.500	0.493	813	1064 (251)	1	16	0.834 (± 0.267)
Physiology							
Carbon isotope (CI)	0.363	0.375	681	759 (78)	0	15	0.292 (± 0.337)
Leaf water potential (WP)	0.080	0.000	823	964 (141)	0	15	0.322 (± 0.319)
SPAD₂₀₁₄ (SP)	0.310	0.297	839	1124 (285)	1	17	0.566 (± 0.331)

All broad-sense heritability estimates were significantly different from 0 except for WP.
[a]Broad-Sense Heritability with Thin Plate Spline correction (TPS) correction applied to the phenotypic data.
[b]Broad-Sense Heritability without TPS correction.
[c]Number of genotypes.
[d]Number of ramets sampled, with replicates in parentheses.
[e]SNPs with P-values < 1×10^{-7} (suggestive significant SNPs).
[f]Number of SNP PC covariates used in multitrait GWAS.
[g]Mean Chip Heritability values for phenotypes with confidence interval (CI).

number of principal coordinate axes used; x is an n-vector of marker genotypes, β is the effect size of the marker, u is an n-vector of random effects that includes a relatedness matrix and ϵ is an n-vector of errors.

Trait selection for multitrait GWAS

Multitrait combinations were created based on genetic correlations among phenotypes as well as hypothesised structural and functional relationships of the traits. The latter can be important even in the absence of genetic correlations (Stephens, 2013). Pairwise genetic correlations were performed (Fig. 2; Table S2) and the functional relationships were assessed through the relevant literature for the phenotypes before forming 12 multitrait sets (Table 2). For example, leaf area, leaf dry weight, leaf length and leaf width were combined to form a multitrait set because these traits are highly intercorrelated and represent the leaf as a structural unit. Likewise, tree height, leaf area and petiole length were combined because the traits are intercorrelated and all affect interception of photosynthetically active radiation. We also combined traits that did not have high genetic correlations, but presumably had functional relationships. For example, we combined carbon isotope composition, leaf water potential and stomatal density because these traits together provide a broader picture of water-use efficiency (WUE), abiotic stress and gas exchange in plants. Pre-dawn leaf water potential is a measure of water retaining capacity of the plants. Measurement of δ^{13}C composition in wood provides a measure of integrated WUE. Higher composition of δ^{13}C is related to lower carbon isotope discrimination, which in turn is related to high water use efficiency (WUE). The gas exchange process in plant leaves is regulated by stomata in the leaves, and density and number of stomata are key for this mechanism. Multitrait association was conducted with GEMMA using the same model as for single-trait associations, except y is an $n \times d$ matrix of d phenotypes for n individuals.

Analyses of association results

Determining a significance cutoff is one of the biggest challenges for high-dimension analyses such as GWAS (Sham & Purcell, 2014). Here we have chosen a uniformly conservative approach to facilitate comparisons among GWAS methods. We used a P-value cutoff ($\alpha \leq 0.05$) based on the Bonferroni correction criterion of 7.37×10^{-9} and a more liberal P-value cutoff of 1×10^{-7} to identify suggestive associations. These were later cross-referenced to other sources of evidence to highlight robust associations (see Network Analysis below). For the purpose of summarising the results, significant SNPs within 10 kb of one another were merged and counted as a single significant locus. Gene models that were closest to significant SNPs were identified based on v3 of the *P. trichocarpa* genome. Annotation information was obtained from Phytozome, including expression level in different plant tissues and annotations of putative gene functions (Goodstein *et al.*, 2012). Percentage of variance explained (PVE) by SNPs was estimated using the formula in Shim *et al.* (2015).

Network analysis

To gain further insight into possible biological functions of candidate genes identified by the GWAS analysis, we examined the position of the genes in networks constructed for the same

population that was used in this paper. The networks were based on the following: gene co-expression using the JGI Plant Gene Atlas for *P. trichocarpa* (https://phytozome.jgi.doe.gov); GWAS of metabolite profiles determined by GC-MS (Tschaplinski *et al.* 2012), and methylation data for multiple tissues in *P. trichocarpa* (Vining *et al.*, 2012). Details of the underlying data and network construction can be found in Weighill *et al.* (2018). Briefly, candidate genes identified from single- and multitrait GWAS were used as a seed to identify subnetworks that were potentially related to the gene. A merged network was created by combining metabolites at False Discovery Rate (FDR) of 0.1, significant GWAS SNPs linked to the corresponding phenotypes and the co-expression and co-methylation subnetworks. The networks were visualised in CYTOSCAPE v.3.6.1 (Shannon *et al.*, 2003).

Functional analysis of the GAUT9 candidate gene

Detailed functional characterisation was conducted for one of the genes highlighted by the multitrait GWAS and network analyses presented here in order to provide experimental validation of the functional roles inferred by our analyses. A 123-bp fragment comprising portions of the coding region and 3′-untranslated region of *Potri.004G111000* (*PtGAUT9.1*) was amplified via PCR from a *P. trichocarpa* cDNA library using the following gene specific primers: PtGAUT9.1-F (CACCCCCGGGTTTG GCCTTTAGACGAATTCC) and PtGAUT9.1-R (TCTAGAG TGACAACTAATGATCGGATCCA). The fragment was cloned into an RNAi cassette and transferred to a binary vector

for *Agrobacterium*-mediated transformation of the *P. deltoides* clone WV94, as previously described (Biswal *et al.*, 2015, 2018a). Measurement of leaf traits was carried out on 10 plants each of 3-month-old WT, empty vector control, and *PdGAUT9.1*-KD lines. RNA isolation and quantitative RT-PCR were performed as previously described (Biswal *et al.*, 2015, 2018a). Briefly, total RNA was isolated using an RNeasy Plant Mini Kit (Qiagen, Valencia, CA, USA). The primers *PdGAUT9.1*-qRT-F (GTGCTTGGCCTCGGATATAA) and *PdGAUT9.1*-qRT-R (GAAACATGAAACCTTGGCTTGA) were used to amplify the target gene, *PdGAUT9.1*. The closely related *PdGAUT9.2* gene was also evaluated to demonstrate the specificity of downregulation using the primers qRT-F (GCGGCATCAATGGTGGATTA) and *PdGAUT9.2*-qRT-R (TTCTATTCCTCGCCACTCTCTC).

We also evaluated the impacts of the associated polymorphism on gene expression in the association population using RNA-seq data from developing xylem, as described by Zhang J. *et al.* (2018). Briefly, normalised gene expression was determined using TopHat2, Cufflinks, and Featurecounts, and normalised via DESeq2, followed by Pearson's correlation with the genotype.

Results

Physiological and morphological trait variation

Broad-sense heritabilities ranged from 0.297 to 0.561 for morphological traits, and from 0.080 to 0.363 for physiological traits
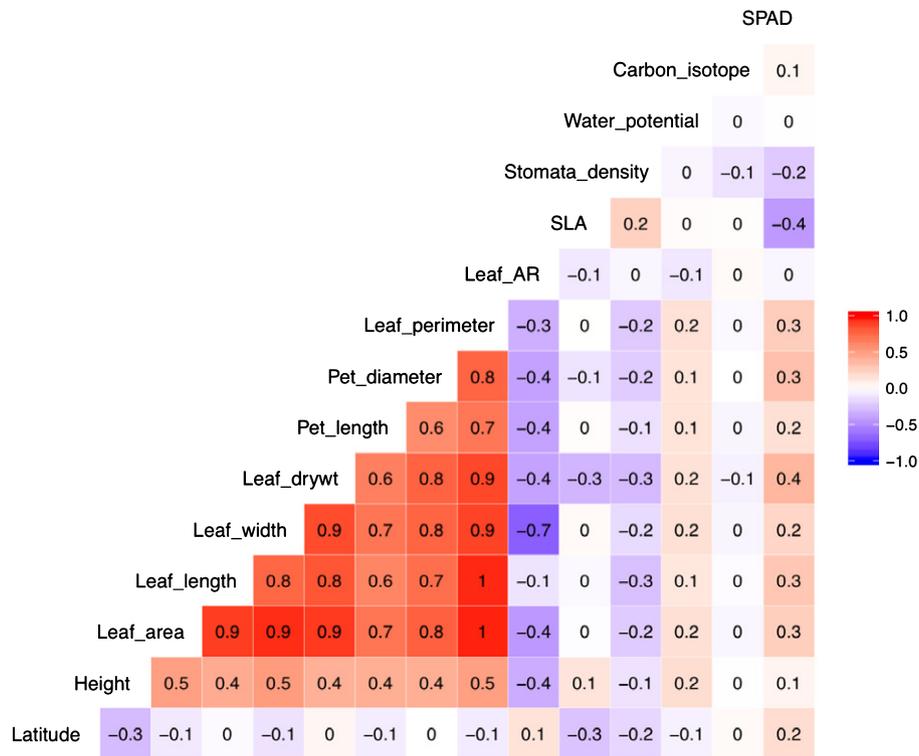
Fig. 2 Pairwise Pearson genetic correlation of selected morphological and physiological traits (traits with at least 681 genotypes) measured in the *Populus trichocarpa* common garden in Corvallis, Oregon, USA. The colour spectrum, bright red to bright blue represents highly positive to highly negative correlations and the number represents the correlation values. Best Linear Unbiased Predictor (BLUP) adjusted values were used. *P*-values are provided in Supporting Information Table S1. SPAD represents leaf greenness. AR, aspect ratio; SLA, specific leaf area.

Table 2 List of traits used for multitrait associations in *Populus trichocarpa*.

Trait combination	Abbreviation	popN[a]	PCs[b]
Carbon isotope, height, leaf area, petiole length	CI_HT_LA_PL	632	14
Carbon isotope, leaf area, stomatal density	CI_LA_SD	603	12
Carbon isotope, leaf area, SPAD, stomatal density	CI_LA_SD_SP	600	12
Carbon isotope, leaf water potential	CI_WP	673	6
Carbon isotope, leaf water potential, stomatal density	CI_SD_WP	638	8
Height, leaf area, petiole length	HT_LA_PL	791	8
Height, petiole diameter, petiole length	HT_PD_PL	839	13
Leaf area, leaf dry weight, leaf length, leaf width	LA_LD_LL_LW	788	14
Leaf area, SPAD, stomatal density	LA_SD_SP	755	14
Leaf aspect ratio, specific leaf area	AR_SL	780	9
Leaf dry weight, petiole diameter, SPAD	LD_PD_SP	831	7
Petiole diameter, petiole length, specific leaf area	PD_PL_SL	781	17

[a]Number of unique genotypes.
[b]Number of single nucleotide polymorphism (SNP) PC covariates used in multitrait GWAS.

(Table 1). Pre-dawn leaf water potential had low broad-sense heritability that was not significantly different from 0 (Table 1). Although TPS regression was used to correct for microsite variation, physiological traits such as pre-dawn leaf water potential appear to be very sensitive to environmental conditions and the microclimatic conditions at the time of sampling.

Most morphological traits were highly intercorrelated, whereas physiological traits were generally not intercorrelated, which is consistent with expectations due to the high measurement error for the latter (Fig. 2; Table S2). We performed PCA to further explore the relationships among traits within the population. The first principal component (PC1) explained >47% of the total variation. PC1 and PC2 together explained 61% of the total variation. Morphological traits were positively weighted towards the PC1 axis (Fig. 3), which also shows a slight negative correlation with the latitude of the provenance ($r = -0.17$, $P \leq 0.001$). Specific leaf area and stomatal density were negatively weighted for PC2, while SPAD was positively weighted along this axis (Fig. 3; Table S3). PC2 generally separated the Columbia population from BC and the core subpopulations (Fig. 3). Most morphological traits were correlated with the latitude of origin (Table S4).

SNP-trait associations

We conducted single-trait GWAS with 6.78 million SNPs for the 14 morphological and physiological traits. We did not identify any SNP that passed Bonferroni correction ($P < 7.37 \times 10^{-9}$). However, we identified a total of four SNPs (Table 1; Fig. 4a) that passed a suggestive association P-value cutoff of 1×10^{-7}. These associated SNP were within or close to four *P. trichocarpa* gene models (Table 3). PVE of significant SNPs ranged from 3.45% to 4.35% (Table S5), although this is likely to be inflated as it is estimated in the discovery population.

Multitrait GWAS for 12 sets of traits identified five SNPs that passed the Bonferroni correction P-value cutoff and 32 SNPs that passed the suggestive association P-value cutoff of 1×10^{-7} (Fig. 4b; Table 4). These SNPs were within or close to 22 *P. trichocarpa* gene models (Table 4). PVE of these SNPs ranged

from 0.0003% to 4.35% for the individual traits comprising the multitrait set (Table S5).

To facilitate the presentation, we divided the multitrait association results into the following three categories based on the correspondence of the multitrait and the single-trait results. First, multitrait GWAS with increased power for the same (or nearby) SNP positions as in the single-trait GWAS (Figs 5a,b, S1). Second, multitrait GWAS with increased power, but with different genomic positions than the single-trait GWAS (Figs 5c,d, S2). Third, multitrait GWAS with reduced power for some loci, but with the same (or nearby) SNP positions as the single-trait GWAS (Figs 5e,f, S3). Each category contained four multitrait combinations. QQ plots showed a clear improvement for multitrait association compared with the corresponding single-trait association (Figs S4–S6).
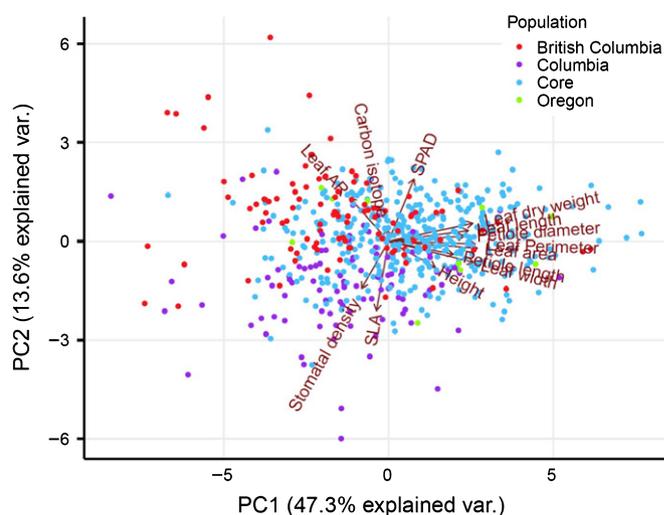
Fig. 3 Principal component analysis (PCA) biplot showing the first and second principal components with individual *Populus trichocarpa* genotypes (the points) coloured by provenance as in Fig. 1, and relative weightings of the explanatory variables indicated by vectors. SPAD represents leaf greenness. AR, aspect ratio; SLA, specific leaf area.
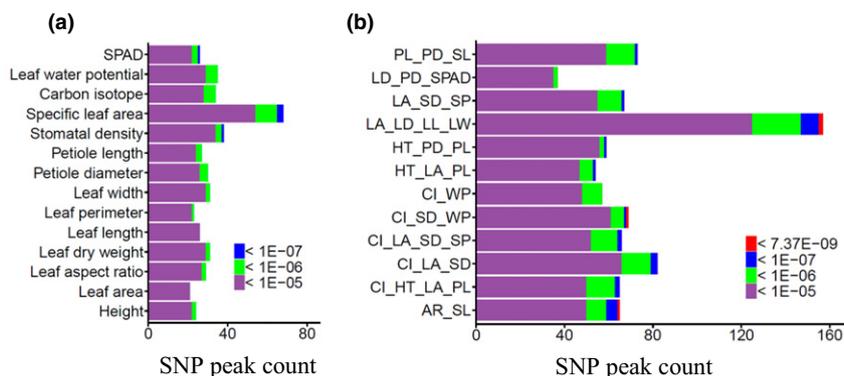
Fig. 4 *Populus trichocarpa* single nucleotide polymorphism (SNP)-trait association peak counts at 10-kb intervals. (a) Single-trait GWAS, (b) multitrait GWAS.

Table 3 Genes identified from *Populus trichocarpa* single-trait GWAS.

Gene model[a]	Trait	*P*-value[b]	Functional annotation[a]
Potri.001G371800	Specific leaf area	3.95E-08	NA
Potri.004G111000	Specific leaf area	9.76E-08	Galacturonosyltransferase 9
Potri.008G111800	Stomatal density	8.93E-08	18S pre-ribosomal assembly protein gar2-related
Potri.010G098400	SPAD	4.84E-08	Tetratricopeptide repeat (TPR)-like superfamily protein

NA, not available.
[a]Gene models are annotated using v3.1 of the *P. trichocarpa* genome.
[b]Single nucleotide polymorphism (SNP) *P*-values $< 1 \times 10^{-7}$.

To provide further evidence for the involvement of the associated SNPs in trait variation, we integrated our GWAS results with other independent datasets, including leaf metabolite levels, gene expression and tissue-specific methylation. Three genes that were significantly associated with carbon isotope, leaf area and stomatal density (CI-LA-SD) showed interesting linkages to co-expressed genes and/or metabolites, including a potential regulatory network mediated by YABBY transcription factors, and a possible regulatory network that is associated with phenolic composition (Fig. 6; Tables S6, S7). Similarly, the multitrait associations for leaf shape (leaf area−leaf diameter−leaf length−leaf width, and specific leaf area−leaf aspect ratio) were linked by association with common candidate genes and revealed two large co-expression networks as well as a group of co-expressed enzymes that affected cell wall characteristics (Fig. 7).

Direct evidence of the role of GAUT9 in determining leaf area in *Populus*

One of the genes associated with leaf morphology (LA-LD-LL-LW) was Potri.004G111000, annotated as *galacturonosyltransferase 9* (*GAUT9*). The polymorphism was 1.9 kb downstream of the end of the predicted stop codon (Table S6). The next closest gene was nearly 33 kb away from the SNP, so Potri.004G111000 is the most likely gene to be affected by this polymorphism. Consistent with this result, the associated polymorphism correlated significantly with Potri.004G111000 expression in developing xylem samples from the association population (Fig. S7a; $r = 0.169$, $P < 0.001$), but not in fully expanded leaves (Fig. S7b;

$r = 0.035$, ns). Lack of correlation in leaves could be due to sampling of the wrong developmental stage, but this caveat requires further investigation.

In the process of studying the role of Potri.004G111000 in the recalcitrance of *P. deltoides* xylem, multiple *P. deltoides* RNAi lines were generated and leaf characteristics of wild type, vector control and three RNAi *PdGAUT9.1*-KD lines were compared (Fig. 8). Reducing the GAUT9.1 transcript level by 51–60% (Fig. 8c) resulted in a 43–66% increased leaf length and leaf width at all developmental time points analysed in 3-month-old glasshouse-grown plants (Figs 8d–f).

Discussion

Identification of the genetic underpinnings of adaptive trait variation has been an elusive goal of forest tree research for more than a century (Wheeler *et al.*, 2015). Such efforts have been greatly enhanced in the age of genomics, which potentially enables identification of sequence variants controlling heritable variation. The genus *Populus* has been a focus of much of this effort due to the tremendous investment in genetic and genomic resources in recent years (Jansson & Douglas, 2007). Previous studies have demonstrated that *P. trichocarpa* contains substantial heritable variation that has been shaped by the combined effects of demographic history and selection (Slavov *et al.*, 2012; Evans *et al.*, 2014; Geraldes *et al.*, 2014; Holliday *et al.*, 2016). However, studies focused on GWAS of individual complex traits have mostly failed to uncover variants that control the majority of genetic variation in *P. trichocarpa* (Evans *et al.*, 2014; Geraldes

Table 4 Genes identified based on *Populus trichocarpa* multitrait GWAS.

Gene model[a]	Trait	*P*-value[b]	Functional annotation[a]
Potri.001G173900	Leaf area, leaf dry weight, leaf length, leaf width	1.61E-08	Plant tudor-like RNA-binding protein
Potri.001G174300	Leaf area, leaf dry weight, leaf length, leaf width	9.15E-08	NA
Potri.001G189300	Leaf area, leaf dry weight, leaf length, leaf width	3.30E-08	NA
Potri.001G371800	Leaf aspect ratio, specific leaf area	7.96E-08	NA
Potri.001G411800	Carbon isotope, leaf area, stomatal density	2.41E-08	EF-hand calcium-binding domain containing protein
Potri.002G055400	Leaf area, leaf dry weight, leaf length, leaf width	3.28E-08	Phytochrome interacting factor 4
Potri.002G145100	Carbon isotope, leaf area, stomatal density	2.62E-08	Plant-specific transcription factor YABBY family protein
Potri.003G165400	Leaf aspect ratio, specific leaf area	6.62E-08	Gem-like protein 5
Potri.004G111000	Leaf area, leaf dry weight, leaf length, leaf width	4.72E-08	Galacturonosyltransferase 9
Potri.004G153400	Leaf aspect ratio, specific leaf area	6.59E-08	Similar to RAS-related GTP-binding protein
Potri.005G097900	Leaf area, SPAD, stomatal density	4.29E-08	Similar to oxidoreductase; 2OG-Fe(2) oxygenase family protein
Potri.006G132500	Leaf area, leaf dry weight, leaf length, leaf width	1.57E-12	Ribosomal protein L4/L1 family
Potri.006G134200	Carbon isotope, leaf area, SPAD, stomatal density	5.29E-08	Lysine-ketoglutarate reductase/saccharopine dehydrogenase bifunctional enzyme
Potri.008G121700	Carbon isotope, leaf water potential, stomatal density	5.43E-09	NA
Potri.008G144100	Leaf area, leaf dry weight, leaf length, leaf width	5.37E-08	Regulatory particle triple-A ATPase 6A
Potri.009G015500	Carbon isotope, leaf area, stomatal density	8.76E-09	Mitochondrial transcription termination factor family protein
Potri.012G065600	Leaf area, petiole length, height	3.24E-08	Leo1-like family protein
Potri.014G136400	Petiole diameter, petiole length, specific leaf area	2.66E-08	LRR receptor-like serine/threonine-protein kinase RKF3-related
Potri.016G071700	Carbon isotope, leaf area, SPAD, stomatal density	1.36E-08	NA
Potri.019G021600	Leaf area, leaf dry weight, leaf length, leaf width	8.87E-08	FtsH extracellular protease family
	Carbon isotope, height, leaf area, petiole length	1.56E-08	

NA, not available.

[a]Gene models are annotated using v3.1 of the *P. trichocarpa* genome.

[b]Single nucleotide polymorphism (SNP) *P*-values $< 1 \times 10^{-7}$.

et al., 2014; McKown *et al.*, 2014b, 2018), most likely to be due to a lack of power to detect variants of small effect and/or low allele frequency (Visscher *et al.*, 2017). Here, we attempted to compensate for these problems by using a larger GWAS population and by performing multitrait GWAS coupled with multiple lines of evidence to support the roles of marginally associated loci in the target phenotypes.

Morphological and physiological trait correlations and influence of geography

Largely consistent with previous studies (Evans *et al.*, 2014; McKown *et al.*, 2014a; Holliday *et al.*, 2016), correlations of morphological and physiological traits with latitude in our study suggest that the variation in adaptive traits in *P. trichocarpa* is partly driven by geography. There was a negative correlation of tree height with latitude, indicating that northern provenances grew poorly in our test site. Many leaf traits were also correlated with height as well as latitude, so the functional relationships among these traits cannot be readily discerned. Based on the correlation coefficients, leaves became smaller and thicker with low abaxial stomatal density and high chlorophyll content for trees from higher latitudes. Several other *Populus* studies in common gardens have reported higher nitrogen content, stomatal conductance and photosynthetic assimilation in northern trees (Gornall & Guy, 2007; McKown *et al.*, 2014a; Soolanayakanahally *et al.*, 2015; Elmore *et al.*, 2017; Momayyezi & Guy, 2017). Furthermore, Gornall & Guy (2007) and McKown *et al.* (2014d) found

a negative correlation between abaxial stomata density and latitude, but they further indicated that the northern *P. trichocarpa* trees were amphistomatous, with adaxial stomata density increasing with the latitude. Most trees used in our study lacked adaxial stomata that are likely to reflect the more southerly distribution of our collection (data not shown).

For the most part, we found no clear relationships between wood $\delta^{13}C$ and leaf traits or latitude of origin. This was unlikely to be due to excessive experimental error because these traits all showed significant broad-sense heritability. This finding is consistent with other published reports for *Populus*. For example, a previous field study of *P. trichocarpa* revealed no correlation between $\delta^{13}C$ of wood and location of origin for a wide variety of morphological and physiological traits (McKown *et al.*, 2014a). However, in a glasshouse study of *P. trichocarpa*, intrinsic WUE was correlated with photosynthetic assimilation and leaf mass area (Momayyezi & Guy, 2017). Similarly, *P. balsamifera* showed a positive correlation of wood and leaf $\delta^{13}C$ with latitude in a glasshouse study (Soolanayakanahally *et al.*, 2009), and no correlation with latitude in a field study (Soolanayakanahally *et al.*, 2015). Monclus *et al.* (2009) found a correlation between $\delta^{13}C$ and productivity traits (fresh biomass, height and circumference) for *P. deltoides* × *P. trichocarpa* hybrids but no correlation of leaf $\delta^{13}C$ and productivity for *P. deltoides* × *P. nigra* hybrids (Monclus *et al.*, 2005). The variability in these results may be due to the effect of environments in the common gardens or variation in the genotypic responses to drought (Soolanayakanahally *et al.*, 2015).
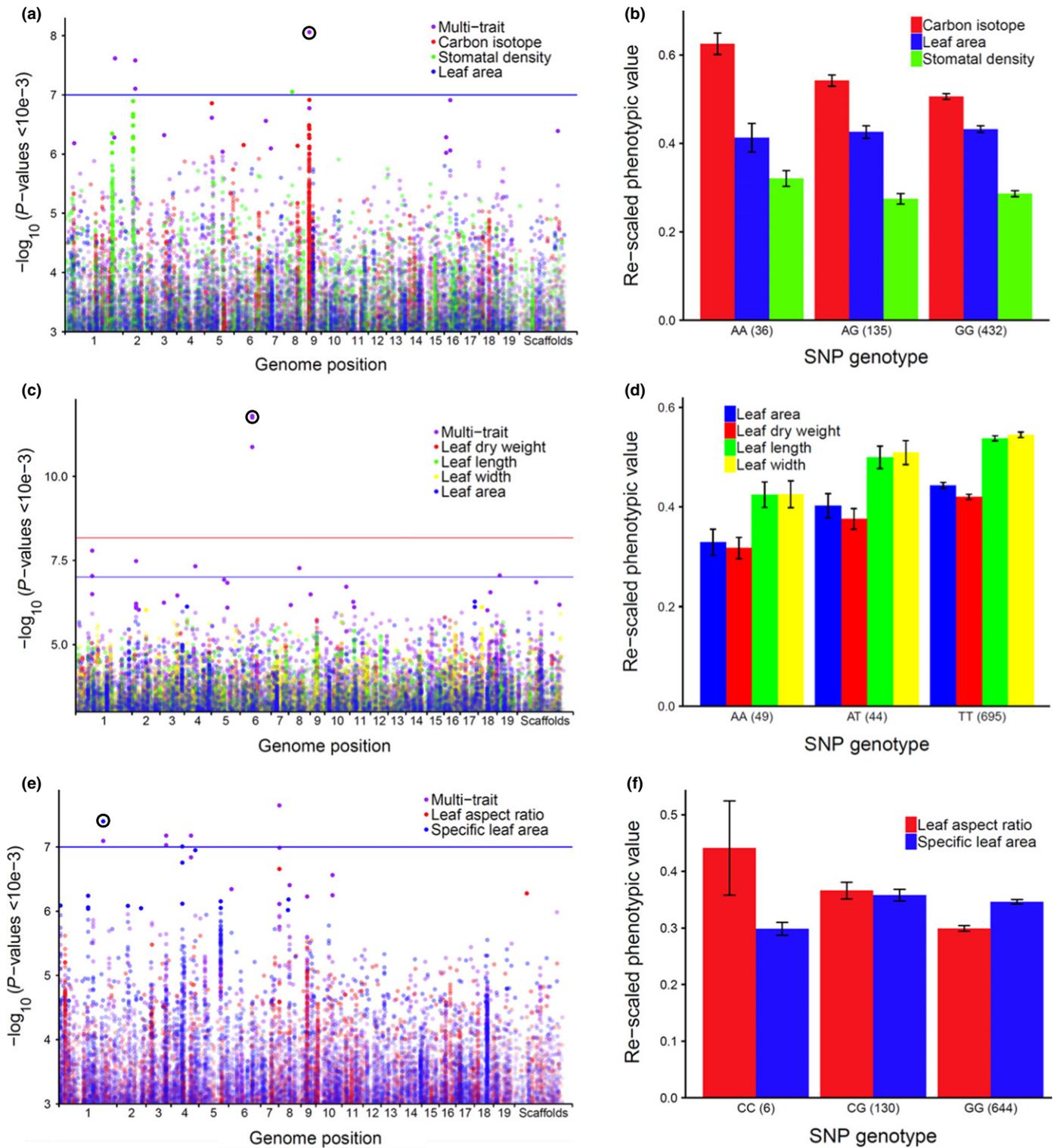
Fig. 5 Manhattan plots comparing GEMMA univariate and multivariate GWAS in *Populus trichocarpa*. The colours of the dots correspond to single-trait or multitrait associations. *P*-values are converted to $-\log_{10}$ (*P*-value). Single nucleotide polymorphisms (SNPs) above red lines passed Bonferroni correction test ($P \leq 7.37 \times 10^{-9}$), SNPs above blue lines are considered suggestive associations ($P \leq 1 \times 10^{-7}$). Only SNPs with $P \leq 1 \times 10^{-3}$) are plotted. (a) Carbon isotope, stomatal density, and leaf area. (b) Allelic effects of SNP near Potri.009G015500 (c) Leaf area, leaf dry weight, leaf length, and leaf width. (d) Allelic effects of SNP near gene *Potri.006G132500*. (e) Leaf aspect ratio and specific leaf area. (f) Allelic effects of SNP near gene *Potri.001G371800*. (b, d, f) SNPs depicted are circled in the corresponding Manhattan plots and the error bars represent ± SEs for rescaled phenotypic values.
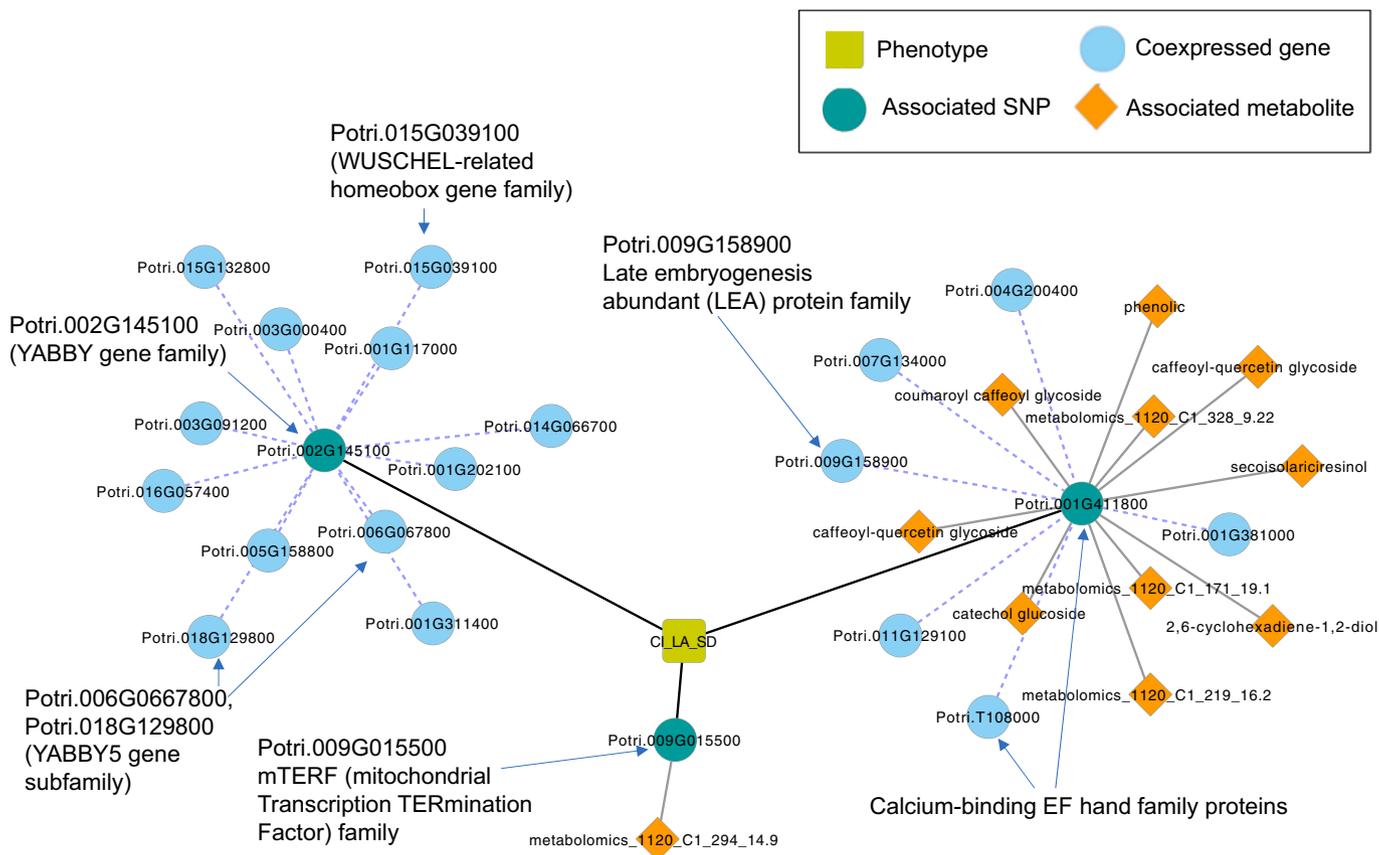
Fig. 6 Merged network for carbon isotope, leaf area and stomatal density (CI_LA_SD) in *Populus trichocarpa*. Networks of co-expressed genes were based on RNA-seq data for 14 tissue types from the Phytozome *Populus* gene atlas. Networks of associated metabolites were based on GWAS for the same population that was used here (Weighill *et al.*, 2018).

Enhanced power with multitrait GWAS

Despite a relatively large sample size and the use of whole-genome SNP data, single-trait GWAS for 14 traits revealed only four loci with suggestive associations and, collectively, these explain only a very small percentage of variance in the phenotypes. By contrast, multitrait GWAS for 12 combinations of a subset of these traits identified 32 SNPs in or near 20 genes. It appears that the overall power of the analysis was moderately improved for most trait combinations compared with the corresponding single-trait GWAS. The power of multitrait GWAS depends on multiple factors, including SNP effect size, direction of effect (positive vs negative), PVE by the SNP, and trait correlations (Zhou & Stephens, 2014; Porter & O'Reilly, 2017). Additionally, because many of our traits were measured without replication, the multitrait analyses may have provided more accurate estimates of the underlying phenotypes due to covariance of some of the traits. Although we used the presumed functional relationship and the correlations of the traits to form the multitrait sets for this study, we cannot determine the relative effects of these factors in these specific analyses, although this has been explored elsewhere through simulation studies (Zhou & Stephens, 2014; Porter & O'Reilly, 2017). Nevertheless, we can gain some insights by examining the PVE of the significant SNPs from multitrait GWAS in the corresponding single-trait analyses

to indirectly infer the contribution of each of the component traits. In the following section we explore this approach using case studies from each of the three categories defined above.

Category 1

This category includes multitrait GWAS with increased power for the same (or nearby) SNP positions as in the single-trait GWAS. The multitrait GWAS that includes carbon isotope, leaf area and stomatal density (CI-LA-SD) well represents this category (Fig. 5a). The multitrait GWAS detected four suggestive SNPs, each of which was in approximately the same position as a non-significant peak from at least one of the corresponding single-trait GWASs. Examination of the effect plot for the SNP with the highest significance (Fig. 5b) suggests that the multitrait GWAS may be capturing a pleiotropic effect in this case, as two of the traits, $\delta^{13}C$ and stomatal density, both had higher means for homozygotes for the major allele compared with the other genotype classes. A similar pattern is evident for the peaks on Chr01 and Chr02, although in these cases the allelic effects are in opposite configurations for leaf area and stomatal density (Fig. S8), possibly reflecting a weak negative correlation between these traits.

Examination of provisional annotations and direct functional linkages for the genes closest to the associated SNPs provides
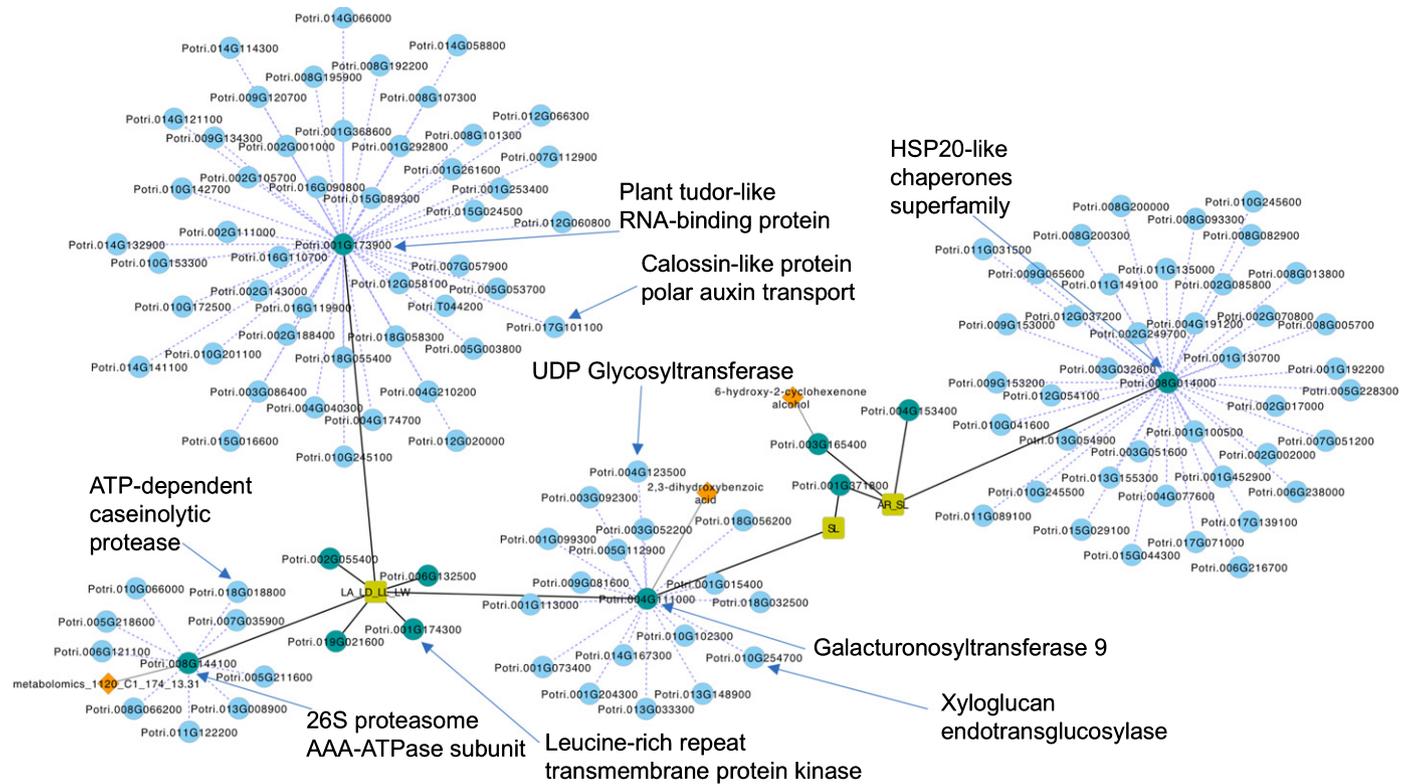
Fig. 7 Merged network for leaf area, dry weight, leaf length, and leaf width (LA_LD_LL_LW) and leaf aspect ratio-specific leaf area (AR_SL) in *Populus trichocarpa*. Symbols are as defined in Fig. 6.

further insight into the possible mechanisms by which these loci affect these three traits. For example, Potri.002G145100, a putative YABBY-1-related axial regulator, was co-expressed with 12 other genes, including another YABBY-5 transcription factor and a WUSCHEL-related homeobox gene family member, possibly representing a large regulatory network (Fig. 6).

Another possible regulatory network contained Potri.001G411800, an EF-hand Calcium-Binding Domain protein. In *P. trichocarpa*, Potri.001G411800 has moderate expression in early and late dormant bud early male and female development, root tip, young, immature and first fully expanded leaves (Goodstein *et al.*, 2012). This gene was co-expressed with six other genes, including another EF-hand family protein, Potri.011G129100. Calcium is an important second messenger in eukaryotes and has important roles in cell signalling and response to biotic and abiotic stresses and developmental cues (Sanders *et al.*, 2002; Chen *et al.*, 2015; Ranty *et al.*, 2016; Zhu, 2016; Edel *et al.*, 2017). The EF-hand motif is the most common and highly conserved calcium-binding motif (Lewit-Bentley & Réty, 2000; Zeng *et al.*, 2017).

This co-expression network provides further evidence that Potri.001G411800 is involved in responses to abiotic stress. It is co-expressed with a late embryogenesis abundant (LEA) hydroxyproline-rich glycoprotein (Potri.009G158900), a group that has a major role in responses to drought, salinity and, osmotic and temperature-related stresses (Gao & Lan, 2016; Magwanga *et al.*, 2018). Potri.001G411800 is also associated with 10 different metabolites in the same population, including several that are

related to plant development and stress responses (Fig. 6). For example, at least five of the metabolites are identified as flavonoids or flavonoid glycosides, including caffeoyl-quercetin glycoside, coumaroyl caffeoyl glycoside, and catechol glycoside. Flavonoids are known to have antioxidant properties that are induced under abiotic and biotic environmental stresses (Hernández *et al.*, 2009). Quercetin glycosides also play an important role in plant growth and development (Parvez *et al.*, 2004). More importantly, they are known to have a role in osmotic adjustment in which the deleterious effect of water deficit is minimised by the active accumulation of solutes such as glycosides and phenolics as a response to drought (Tschaplinski *et al.*, 2019).

Category 2

This category includes cases in which the multitrait GWAS had increased power, but the associated loci did not overlap with peaks in the single-trait GWAS. The multitrait set that includes leaf area, leaf dry weight, leaf length and leaf width (LA-LD-LL-LW) well represents this category (Fig. 5c). We detected 10 significant SNPs in the multitrait GWAS compared with none in the corresponding single-trait GWAS, and the peaks were largely non-overlapping. The locus with the highest association in the multitrait analysis showed similar patterns of genotypic means for all four traits, consistent with an additive effect (Fig. 5d). Most of the other cases had low minor allele frequency and high variation among phenotypes for homozygotes for the minor allele (Fig. S9; Table S5).
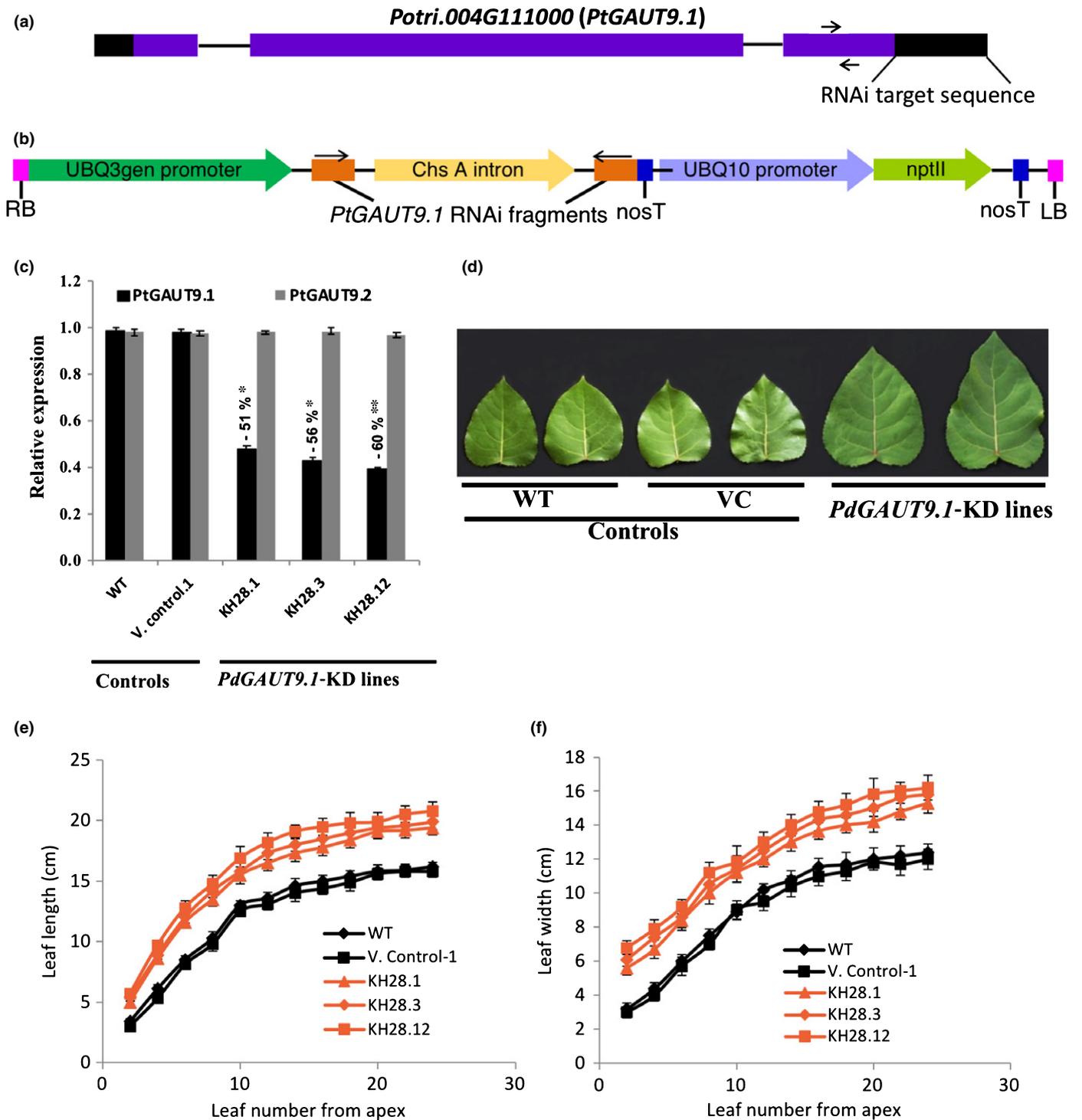
Fig. 8 Effects of PdGAUT9.1 downregulation on leaf size in *Populus deltoides*. (a) Gene model for *PtGAUT9.1* (*Potri.004G111000*) from *Populus trichocarpa* v3.0 genome. Black boxes indicate the 5′ and 3′ untranslated regions (UTRs); purple boxes indicate exons and lines indicate introns. The indicated RNAi targeted sequence was 123 bp. The sequences used for quantitative RT-PCR are indicated by arrows. (b) Schematic representation of *PtGAUT9.1* RNAi silencing construct used to generate *P. deltoides* PdGAUT9.1-KD transgenic lines. (c) Relative transcript abundance of *PdGAUT9.1* (*Potri.004G111000*) and *PdGAUT9.2* (*Potri.017G106800*) as determined by quantitative RT-PCR analysis of leaf RNA from glasshouse-grown 3-month-old poplar WT, vector control (V. control.1) and *PdGAUT9.1-KD* lines (KH28.1, KH28.3 and KH28.12). Expression of *PdGAUT9.1* in poplar WT was set to 1 and *18S rRNA* was used as a reference gene. Error bars are ± SE, $n = 6$. Differences were tested by one-way ANOVA (*, $P < 0.05$; **, $P < 0.001$). (d) Leaf phenotype (the sixth leaf from the apex) of *P. deltoides* control (WT and VC) and *PdGAUT9.1*-KD line (KH28.12) from 3-month-old plants. (e) Length and (f) width of leaves from different developmental stages of three different 3-month-old *GAUT9*-KD transgenic lines (KH28.1, KH28.3 and KH28.12). Every other leaf of 10 plants was measured starting with the second leaf from the apex. The error bars represent ± SEs for leaf length and leaf width.

This is likely to be when joint estimation of the variances provides more power to detect differences among the genotypic classes, resulting in significant multitrait associations for loci that showed no association with the individual component traits.

This case study also provides an example of the use of multiple lines of evidence to provide further support for relatively weak associations (Weighill *et al.*, 2018). For example, the SNP Chr04_9996091 had a *P*-value of 4.72×10^{-8}, which does not pass a strict FDR with α = 0.05, so we classified this as a 'suggestive' association. However, we have multiple lines of evidence that the closest gene model, Potri.004G111000, is involved with leaf development. This gene encodes a putative galacturonosyl-transferase that is moderately expressed in the first fully expanded leaf, young and immature leaves, and pre-dormant and fully open vegetative buds (Goodstein *et al.*, 2012). The closest *Arabidopsis thaliana* homolog of this gene, AT3G02350, encodes a gene annotated as galacturonosyltransferase 9 (*GAUT9*), for which enzyme activity has not yet been established. *GAUT9* belongs to the *GAUT* gene family of proven and putative pectin homogalacturonan (HG) galacturnosyltransferases (Sterling *et al.*, 2006; Atmodjo *et al.*, 2013; Biswal *et al.*, 2018b; Voiniciuc *et al.*, 2018). Downregulation of the *PdGAUT9.1* gene caused increased leaf length and width in both developing and mature leaves of glasshouse-grown *P. deltoides*, confirming the role of the gene in leaf development in *Populus*.

Other *GAUT* genes have also been shown to affect cell wall properties and leaf size in *Populus*. Downregulation of a *GAUT12* homolog in *P. deltoides* showed decreased xylan and pectin content in the cell wall and increased biomass yield (Biswal *et al.*, 2015), while overexpression showed a reduction in overall plant productivity and resulted in smaller leaves, reduced xylem cell numbers and size, and an increase in the amount of xylose and galacturonic acid in the cell wall (Biswal *et al.*, 2018b). Downregulation of *GAUT4* in *P. deltoides* resulted in decreased pectic homogalacturonan and rhamnogalacturonan II and increased plant height, diameter, leaf area, and biomass (Biswal *et al.*, 2018a). Additional evidence supporting increased leaf growth with decreases in pectin is provided by reports of increased expansion of Arabidopsis rosette leaves resulting from overexpression of polygalacturonase, an enzyme that degrades pectic homogalacturonan (Rui *et al.*, 2017).

Co-expression analysis lends further support for the involvement of this gene in cell wall biosynthesis. Potri.004G111000 (GAUT9) was co-expressed with 17 other gene models and one metabolite that are cell-wall related (Fig. 7; Tables S6, S7). For example, Potri.004G123500, is annotated as a member of the uridine diphosphate (UDP) glycosyltransferase (UGT) superfamily. In *P. trichocarpa*, this gene had moderate expression in pre-, early and late dormant buds, young and immature and first fully expanded leaves, and stem nodes and internodes (Goodstein *et al.*, 2012). Another gene, Potri.010G102300, encodes a xyloglucan endotransglucosylase that is a member of the Glycoside Hydrolase Family 16 and which is also expected to affect cell wall properties (Nishikubo *et al.*, 2011; Yang *et al.*, 2014).

Two other genes with suggestive associations to leaf morphology characteristics were co-expressed with a large number of other genes with potential regulatory functions. Potri.008G144100 is similar to the AAA-ATPase subunit of the 26S proteasome complex, and was co-expressed with eight other genes with putative roles in protein degradation or synthesis, plus an ATP-dependent caseinolytic protease (Potri.018G018800) potentially involved in lipid processing (Fig. 7; Table S7). Another gene associated with leaf morphology, Potri.001G173900, encodes a Tudor-like RNA-binding protein with conserved ENT and Agenet domains. There is emerging evidence that the latter domain may be involved in transcriptional regulation in Arabidopsis (Zhang C. *et al.*, 2018). This gene was co-expressed with 49 other genes with putative roles in protein degradation or RNA regulation, as well as nine genes with annotations related to carbohydrate metabolism (Tables S6, S7). These two genes are excellent candidates as master regulators of leaf morphology, possibly mediated by cell wall modification.

Drost *et al.* (2015) identified a major QTL peak on Chr10 for leaf width in an interspecific *P. trichocarpa* × *P. deltoides* pseudobackcross family. This peak is in close proximity to one of our GWAS peaks for LA−LD−LL−LW. An ADP-ribosylation factor GTPase (*PtARF1*) was the prime candidate gene in the hybrid family based on eQTL analysis and functional assays (Drost *et al.*, 2015). However, this gene was over 200 kb from the closest associated SNP in our study (Potri.010G254700, a leucine-rich repeat transmembrane protein kinase). Furthermore, *PtARF1* did not appear in our networks, indicating that it was not even weakly associated with leaf morphology in our population, and was not co-expressed or co-methylated with any weakly associated genes. This may indicate that different mechanisms control leaf morphology within *P. trichocarpa* compared with interspecific hybrids. However, the hypothesised mechanism for *PtARF1* focuses on its role in vesicle-mediated trafficking of the PIN protein to regulate auxin gradients (Drost *et al.*, 2015), which is broadly consistent with the genes in our network that affect cell wall extensibility and carbohydrate metabolism. Intriguingly, the co-expression network of Potri.001G173900 includes a gene (Potri.017G101100) for which the best homolog in Arabidopsis (AT3G02260) is a putative calossin-like protein required for polar auxin transport.

Category 3

This case study includes multitrait GWAS sets that had lower top SNP-trait association signals compared with the corresponding single-trait GWAS for some loci. The multitrait GWAS set that included leaf aspect ratio and specific leaf area (AR-SL) well represents this category (Fig. 5e). We detected four loci with suggestive associations for multitrait GWAS compared with two for single-trait GWAS for specific leaf area (Fig. 5e). For locus Chr01_38557469, the association for SL was stronger than that for the AR−SL combination, and that locus explained 3.78% of the variation in SL (Table S5). By contrast, there was no hint of an association for AR at that locus, possibly due to high variation in the minor allele homozygous individuals (Fig. 5f). The loci for which the multitrait association showed the lowest *P*-value follow a similar pattern to those from categories 1 and 2, with both traits showing differences among the genotypes (Fig. S10). For the

AR-SL multitrait we detected four gene models, out of which one gene model overlapped with the single-trait GWAS for specific leaf area (Fig. 7). One of these, Potri.008G014000, belongs to the HSP-20-like chaperones superfamily. This gene is co-expressed with 49 other genes in *P. trichocarpa*, including 22 with putative roles in protein degradation or RNA processing, suggesting that this is another important regulatory network for leaf morphology (Table S7).

Comparison with previous GWAS studies in *P. trichocarpa*

We compared all genes identified from single-trait as well as the multitrait GWAS with the previous GWAS studies in *P. trichocarpa* using 34K *Populus* SNP array data (McKown *et al.*, 2014b,c), but despite the moderate (0.2–0.4) heritabilities of most of the comparable traits such as leaf traits, height, chlorophyll content and stomatal density in these studies, none of the 22 genes we identified in our study overlapped with the previous studies. This difference might be due to the difference in the genotypes and the common garden used for our study, or to higher phenotyping error in the present study. It also likely to reflect the more targeted genome sampling in the previous studies, which only assayed 3543 genes that were preselected based on annotations and other functional information (Geraldes *et al.*, 2011). Nevertheless, we believe that the whole-genome resequencing of 882 trees used in our study allowed the detection of robust genetic variants underlying some phenotypic traits. Some corroboration for these associations was provided by patterns of expression and co-expression, intersection with genetic control of metabolites, and direct confirmation of mutant phenotypes.

Conclusions

We have presented one of the most comprehensive GWAS studies to date for *P. trichocarpa* in terms of the size of the SNP dataset and the number of genotypes. Taking advantage of the natural variation present in the population and the power of multitrait association, we detected candidate genes that were associated with adaptive morphological and physiological traits. Some of these may represent genes with potentially pleiotropic effects on adaptive traits including leaf morphology, and WUE. These have great potential for further functional characterisation and can be a suitable target for breeding programmes as they capture functional and structural relationships among the traits that are not apparent with single-trait GWAS. Furthermore, the network analysis added an extra layer of information that provided further independent lines of evidence supporting the involvement of these genes in their associated phenotypes and provides clues about possible mechanisms of action. This step is important in functional annotation, which remains a major challenge for recalcitrant model organisms such as forest trees.

Acknowledgements

Author contributions

HBC, DM-S, AKB, LME, DK, TJT and SPD collected data and performed analyses; AKB, DM, CC, LME, KH, SSM, DR, KW and XY produced and characterised transgenics; J-GC, TR, DJ, WM, SHS, GAT, TJT and SPD conceived and designed the study; HBC and SPD wrote the manuscript; DK, LME, J-GC, DJ, GAT, SHS and DM edited the manuscript.

ORCID

Hari B. Chhetri ⓘ https://orcid.org/0000-0001-6820-8789
Stephen P. DiFazio ⓘ https://orcid.org/0000-0003-4077-1590

References

Aitken SN, Yeaman S, Holliday JA, Wang T, Curtis-McLane S. 2008. Adaptation, migration or extirpation: climate change outcomes for tree populations. *Evolutionary Applications* 1: 95–111.

Atmodjo MA, Hao Z, Mohnen D. 2013. Evolving views of pectin biosynthesis. *Annual Review of Plant Biology* 64: 747–779.

Biswal AK, Atmodjo MA, Li M, Baxter HL, Yoo CG, Pu Y, Lee YC, Mazarei M, Black IM, Zhang JY *et al.* 2018a. Sugar release and growth of biofuel crops are improved by downregulation of pectin biosynthesis. *Nature Biotechnology* 36: 249–257.

Biswal AK, Atmodjo MA, Pattathil S, Amos RA, Yang X, Winkeler K, Collins C, Mohanty SS, Ryno D, Tan L *et al.* 2018b. Working towards recalcitrance mechanisms: increased xylan and homogalacturonan production by

overexpression of *GAlactUronosylTransferase12 (GAUT12)* causes increased recalcitrance and decreased growth in *Populus*. *Biotechnology for Biofuels* 11: 1–26.

Biswal AK, Hao Z, Pattathil S, Yang X, Winkeler K, Collins C, Mohanty SS, Richardson EA, Gelineo-Albersheim I, Hunt K *et al.* 2015. Downregulation of GAUT12 in *Populus deltoides* by RNA silencing results in reduced recalcitrance, increased growth and reduced xylan and pectin in a woody biofuel feedstock. *Biotechnology for Biofuels* 8: 41.

Boyle EA, Li YI, Pritchard JK. 2017. An expanded view of complex traits: from polygenic to omnigenic. *Cell* 169: 1177–1186.

Broadaway KA, Cutler DJ, Duncan R, Moore JL, Ware EB, Jhun MA, Bielak LF, Zhao W, Smith JA, Peyser PA *et al.* 2016. A statistical approach for testing cross-phenotype effects of rare variants. *American Journal of Human Genetics* 98: 525–540.

Casale FP, Rakitsch B, Lippert C, Stegle O. 2015. Efficient set tests for the genetic analysis of correlated traits. *Nature Methods* 12: 755–758.

Chen C, Sun X, Duanmu H, Zhu D, Yu Y, Cao L, Liu A, Jia B, Xiao J, Zhu Y. 2015. GsCML27, a gene encoding a calcium-binding Ef-hand protein from *Glycine soja*, plays differential roles in plant responses to bicarbonate, salt and osmotic stresses. *PLoS ONE* 10: 1–19.

Drost DR, Puranik S, Novaes E, Novaes CRDB, Dervinis C, Gailing O, Kirst M. 2015. Genetical genomics of *Populus* leaf shape variation. *BMC Plant Biology* 15: 166.

Eckenwalder JE. 1996. Systematics and evolution of *Populus*. In: Stettler RF, Bradshaw HD, Heilman PE, Hinckley TM, eds. *Biology of* Populus *and its implications for management and conservation*. Ottawa, Canada: NRC Research Press, 7–32.

Edel KH, Marchadier E, Brownlee C, Kudla J, Hetherington AM. 2017. The evolution of calcium-based signalling in plants. *Current Biology* 27: R667–R679.

Elmore AJ, Craine JM, Nelson DM, Guinn SM. 2017. Continental scale variability of foliar nitrogen and carbon isotopes in *Populus balsamifera* and their relationships with climate. *Scientific Reports* 7: 7759.

Evans LM, Slavov GT, Rodgers-Melnick E, Martin J, Ranjan P, Muchero W, Brunner AM, Schackwitz W, Gunter L, Chen J-G *et al.* 2014. Population genomics of *Populus trichocarpa* identifies signatures of selection and adaptive trait associations. *Nature Genetics* 46: 1089–1096.

Gao J, Lan T. 2016. Functional characterization of the late embryogenesis abundant (LEA) protein gene family from *Pinus tabuliformis* (Pinaceae) in *Escherichia coli*. *Scientific Reports* 6: 1–10.

Geraldes A, Farzaneh N, Grassa CJ, McKown AD, Guy RD, Mansfield SD, Douglas CJ, Cronk QCB. 2014. Landscape genomics of *Populus trichocarpa*: the role of hybridization, limited gene flow, and natural selection in shaping patterns of population structure. *Evolution* 68: 3260–3280.

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.

González-martínez SC, Krutovsky KV, Neale DB. 2006. Forest-tree population genomics and adaptive evolution. *New Phytologist* 170: 227–238.

Goodstein DM, Shu S, Howson R, Neupane R, Hayes RD, Fazo J, Mitros T, Dirks W, Hellsten U, Putnam N *et al.* 2012. Phytozome: a comparative platform for green plant genomics. *Nucleic Acids Research* 40: 1178–1186.

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

Hackinger S, Zeggini E. 2017. Statistical methods to detect pleiotropy in human complex traits. *Open Biology* 7: 170125.

Hernández I, Alegre L, Van Breusegem F, Munné-Bosch S. 2009. How relevant are flavonoids as antioxidants in plants? *Trends in Plant Science* 14: 125–132.

Holliday JA, Zhou L, Bawa R, Zhang M, Oubida RW. 2016. Evidence for extensive parallelism but divergent genomic architecture of adaptation along altitudinal and latitudinal gradients in *Populus trichocarpa*. *New Phytologist* 209: 1240–1251.

Ingvarsson PK, Hvidsten TR, Street NR. 2016. Towards integration of population and comparative genomics in forest trees. *New Phytologist* 212: 338–344.

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

Jansson S, Douglas CJ. 2007. *Populus*: a model system for plant biology. *Annual Review of Plant Biology* 58: 435–458.

Kim J, Zhang Y, Pan W. 2016. Powerful and adaptive testing for multi-trait and multi-SNP associations with GWAS and sequencing data. *Genetics* 203: 715–731.

Lewit-Bentley A, Réty S. 2000. EF-hand calcium-binding proteins. *Current Opinion in Structural Biology* 10: 637–643.

MacDicken K, Jobsson O, Pina L, Marklund S, Maulo VC, Adhikari Y, Garzuglia M, Lindquist E, Reams G, Mundhenk P *et al.* 2016. *Global forest resources assessment 2015*. Rome: Food and Agriculture Organizations of the United Nations.

Magwanga RO, Lu P, Kirungu JN, Lu H, Wang X, Cai X, Zhou Z, Zhang Z, Salih H, Wang K *et al.* 2018. Characterization of the late embryogenesis abundant (LEA) proteins family and their role in drought stress tolerance in upland cotton. *BMC Genetics* 19: 1–31.

McKown AD, Guy RD, Klápště J, Geraldes A, Friedmann M, Cronk QCB, El-Kassaby YA, Mansfield SD, Douglas CJ. 2014a. Geographical and environmental gradients shape phenotypic trait variation and genetic structure in *Populus trichocarpa*. *New Phytologist* 201: 1263–1276.

McKown AD, Guy RD, Quamme L, Klápště J, La Mantia J, Constabel CP, El-Kassaby YA, Hamelin RC, Zifkin M, Azam MS. 2014b. Association genetics, geography and ecophysiology link stomatal patterning in *Populus trichocarpa* with carbon gain and disease resistance trade-offs. *Molecular Ecology* 23: 5771–5790.

McKown AD, Klápště J, Guy RD, El-Kassaby YA, Mansfield SD. 2018. Ecological genomics of variation in bud-break phenology and mechanisms of response to climate warming in *Populus trichocarpa*. *New Phytologist* 220: 300–316.

McKown AD, Klápště J, Guy RD, Geraldes A, Porth I, Hannemann J, Friedmann M, Muchero W, Tuskan GA, Ehlting J *et al.* 2014c. Genome-wide association implicates numerous genes underlying ecological trait variation in natural populations of *Populus trichocarpa*. *New Phytologist* 203: 535–553.

Momayyezi M, Guy RD. 2017. Substantial role for carbonic anhydrase in latitudinal variation in mesophyll conductance of *Populus trichocarpa* Torr. & Gray. *Plant, Cell & Environment* 40: 138–149.

Monclus R, Dreyer E, Delmotte FM, Villar M, Delay D, Boudouresque E, Petit JM, Marron N, Brechet C, Brignolas F. 2005. Productivity, leaf traits and carbon isotope discrimination in 29 *Populus deltoides* x *P. nigra* clones. *New Phytologist* 167: 53–62.

Monclus R, Villar M, Barbaroux C, Bastien C, Fichot R, Delmotte FM, Delay D, Petit JM, Brchet C, Dreyer E *et al.* 2009. Productivity, water-use efficiency and tolerance to moderate water deficit correlate in 33 poplar genotypes from a *Populus deltoides* × *Populus trichocarpa* F$_1$ progeny. *Tree Physiology* 29: 1329–1339.

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

Neale DB, Savolainen O. 2004. Association genetics of complex traits in conifers. *New Phytologist* 9: 325–330.

Nishikubo N, Takahashi J, Roos AA, Derba-Maceluch M, Piens K, Brumer H, Teeri TT, Stalbrand H, Mellerowicz EJ. 2011. Xyloglucan endo-transglycosylase-mediated xyloglucan rearrangements in developing wood of hybrid aspen. *Plant Physiology* 155: 399–413.

Parvez MM, Tomita-Yokotani K, Fujii Y, Konishi T, Iwashina T. 2004. Effects of quercetin and its seven derivatives on the growth of *Arabidopsis thaliana* and *Neurospora crassa*. *Biochemical Systematics and Ecology* 32: 631–635.

Porter HF, O'Reilly PF. 2017. Multivariate simulation framework reveals performance of multi-trait GWAS methods. *Scientific Reports* 7: 1–12.

Ranty B, Aldon D, Cotelle V, Galaud J-P, Thuleau P, Mazars C. 2016. Calcium sensors as key hubs in plant responses to biotic and abiotic stresses. *Frontiers in Plant Science* 7: 1–7.

Ritchie MD, Holzinger ER, Li R, Pendergrass SA, Kim D. 2015. Methods of integrating data to uncover genotype-phenotype interactions. *Nature Reviews Genetics* 16: 85–97.

Rubin EM. 2008. Genomics of cellulosic biofuels. *Nature* 454: 841–845.

Rui Y, Xiao C, Yi H, Kandemir B, Wang JZ, Puri VM, Anderson CT. 2017. POLYGALACTURONASE INVOLVED IN EXPANSION3 functions in seedling development, rosette growth, and stomatal dynamics in *Arabidopsis thaliana*. *Plant Cell* 29: 2413–2432.

Sanders D, Pelloux J, Brownlee C, Harper JF. 2002. Calcium at the crossroads of signaling. *Plant Cell* 14: S401–S417.

Schindelin J, Rueden CT, Hiner MC, Eliceiri KW. 2015. The ImageJ ecosystem: an open platform for biomedical image analysis. *Molecular Reproduction and Development* 82: 518–529.

Scholander APF, Hammel HT, Bradstreet ED, Hemmingsen EA. 2016. Sap pressure in vascular plants. *Science* 148: 339–346.

Sham PC, Purcell SM. 2014. Statistical power and significance testing in large-scale genetic studies. *Nature Reviews Genetics* 15: 335–346.

Shannon P, Markiel A, Owen O, Baliga NS, Wang JT, Ramage D, Amin N, Schwikowski B, Ideker T. 2003. Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome Research* 13: 2498–2504.

Shim H, Chasman DI, Smith JD, Mora S, Ridker PM, Nickerson DA, Krauss RM, Stephens M. 2015. A multivariate genome-wide association analysis of 10 LDL subfractions, and their response to statin treatment, in 1868 Caucasians. *PLoS ONE* 10: 1–20.

Sjödin A, Street NR, Sandberg G, Gustafsson P, Jansson S. 2009. The *Populus* genome integrative explorer (PopGenIE): a new resource for exploring the *Populus* genome. *New Phytologist* 182: 1013–1025.

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.

Solovieff N, Cotsapas C, Lee PH, Purcell SM, Smoller JW. 2013. Pleiotropy in complex traits: challenges and strategies. *Nature Reviews Genetics* 14: 483–495.

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, Street NR, Robinson KM, Silim SN, Albrectsen BR, Jansson S. 2015. Comparative physiology of allopatric *Populus* species: geographic clines in photosynthesis, height growth, and carbon isotope discrimination in common gardens. *Frontiers in Plant Science* 6: 528.

Stephens M. 2013. A unified framework for association analysis with multiple related phenotypes. *PLoS ONE* 8: e65245.

Sterling JD, Atmodjo MA, Inwood SE, Kumar Kolli VS, Quigley HF, Hahn MG, Mohnen D. 2006. Functional identification of an *Arabidopsis* pectin biosynthetic homogalacturonan galacturonosyltransferase. *Proceedings of the National Academy of Sciences, USA* 103: 5236–5241.

Tschaplinski TJ, Abraham PE, Jawdy SS, Gunter LE, Martin MZ, Engle NL, Yang X, Tuskan GA. 2019. The nature of the progression of drought stress drives differential metabolomic responses in *Populus deltoides*. *Annals of Botany*. doi: 10.1093/aob/mcz002

Tschaplinski TJ, Standaert RF, Engle NL, Martin MZ, Sangha AK, Parks JM, Smith JC, Samuel R, Jiang N, Pu Y *et al.* 2012. Down-regulation of the caffeic acid O-methyltransferase gene in switchgrass reveals a novel monolignol analog. *Biotechnology for Biofuels* 5: 71.

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.

Vining KJ, Pomraning KR, Wilhelm LJ, Priest HD, Pellegrini M, Mockler TC, Freitag M, Strauss SH. 2012. Dynamic DNA cytosine methylation in the *Populus trichocarpa* genome: tissue-level variation and relationship to gene expression. *BMC Genomics* 13: 27.

Visscher PM, Wray NR, Zhang Q, Sklar P, McCarthy MI, Brown MA, Yang J. 2017. 10 years of GWAS discovery: biology, function, and translation. *American Journal of Human Genetics* 101: 5–22.

Voiniciuc C, Engle KA, Günl M, Dieluweit S, Schmidt MH-W, Yang J-Y, Moremen KW, Mohnen D, Usadel B. 2018. Identification of key enzymes for pectin synthesis in seed mucilage. *Plant Physiology* 178: 1045–1064.

Weigel D, Nordborg M. 2015. Population genomics for understanding adaptation in wild plant species. *Annual Review of Genetics* 49: 315–338.

Weighill D, Jones P, Shah M, Ranjan P, Muchero W, Schmutz J, Sreedasyam A, Macaya-Sanz D, Sykes R, Zhao N *et al.* 2018. Pleiotropic and epistatic network-based discovery: integrated networks for target gene discovery. *Frontiers in Energy Research* 6: 30.

Wheeler NC, Steiner KC, Schlarbaum SE, Neale DB. 2015. The evolution of forest genetics and tree improvement research in the United States. *Journal of Forestry* 113: 500–510.

Yang J, Lee SH, Goddard ME, Visscher PM. 2011. GCTA: a tool for genome-wide complex trait analysis. *American Journal of Human Genetics* 88: 76–82.

Yang X, Li X, Li B, Zhang D. 2014. Identification of genes differentially expressed in shoot apical meristems and in mature xylem of *Populus tomentosa*. *Plant Molecular Biology Reporter* 32: 452–464.

Zeng H, Zhang Y, Zhang X, Pi E, Zhu Y. 2017. Analysis of EF-hand proteins in soybean genome suggests their potential roles in environmental and nutritional stress signaling. *Frontiers in Plant Science* 8: 1–15.

Zhang C, Du X, Tang K, Yang Z, Pan L, Zhu P, Luo J, Jiang Y, Zhang H, Wan H *et al.* 2018. *Arabidopsis* AGDP1 links H3K9me2 to DNA methylation in heterochromatin. *Nature Communications* 9: 4547.

Zhang J, Yang Y, Zheng K, Xie M, Feng K, Jawdy SS, Gunter LE, Ranjan P, Singan VR, Engle N *et al.* 2018. Genome-wide association studies and expression-based quantitative trait loci analyses reveal roles of HCT2 in caffeoylquinic acid biosynthesis and its regulation by defense-responsive transcription factors in *Populus*. *New Phytologist* 220: 502–516.

Zhou X, Stephens M. 2012. Genome-wide efficient mixed model analysis for association studies. *Nature Genetics* 44: 821–824.

Zhou X, Stephens M. 2014. Efficient algorithms for multivariate linear mixed models in genome-wide association studies. *Nature Methods* 11: 407–409.

Zhu JK. 2016. Abiotic stress signaling and responses in plants. *Cell* 167: 313–324.

Supporting Information

Additional Supporting Information may be found online in the Supporting Information section at the end of the article.

Fig. S1 Manhattan plots comparing GEMMA univariate and multivariate GWAS in *P. trichocarpa* for (a) stomatal density, carbon isotope, and pre-dawn leaf water potential; (b) height, leaf area, and petiole length; and (c) height, petiole diameter, and petiole length.

Fig. S2 Manhattan plots comparing GEMMA univariate and multivariate GWAS in *P. trichocarpa* for (a) leaf area, stomatal density, and SPAD; (b) leaf area, petiole length, height, and carbon isotope; and (c) leaf area, stomatal density, SPAD, and carbon isotope.

Fig. S3 Manhattan plots comparing GEMMA univariate and multivariate GWAS in *P. trichocarpa* for (a) carbon isotope, pre-dawn leaf water potential; (b) leaf dry weight, petiole diameter, SPAD; (c) petiole diameter, petiole length, specific leaf area.

Fig. S4 QQ-plot for single-trait GWAS in *P. trichocarpa* for carbon, leaf area, and stomatal density, and the corresponding multitrait GWAS with all three traits.

Fig. S5 QQ-plot for single-trait GWAS in *P. trichocarpa* for leaf area, leaf dry weight, leaf length and leaf width, and the corresponding multitrait GWAS with all four leaf traits.

Fig. S6 QQ-plot for single-trait GWAS in *P. trichocarpa* for leaf aspect ratio and specific leaf area and the corresponding multitrait GWAS with both traits.

Fig. S7 Pearson correlation of *GAUT9* (Potri.004G111000) gene expression in leaf and developing xylem of *P. trichocarpa* with genotype at locus Chr04_9996091.

Fig. S8 Allelic effects plots for single traits underlying the carbon isotope−leaf area−stomatal density multitrait association analysis in *P. trichocarpa*.

Fig. S9 Allelic effects plots for single traits underlying the leaf area−leaf diameter−leaf length−leaf width multitrait association analysis in *P. trichocarpa*.

Fig. S10 Allelic effects plots for single traits underlying the leaf aspect ratio−specific leaf area multitrait association analysis in *P. trichocarpa*.

Table S1 SNP PC covariates used in *P. trichocarpa* single- and multitrait GWAS analyses.

Table S2 Pearson pairwise correlation of morphological and physiological traits collected in *P. trichocarpa* association. Numbers below the diagonal represent correlations and numbers above the diagonal represent *P*-values. Red and blue colours indicate positive and negative correlations or *P*-values, respectively.

Table S3 PCA loadings of the traits of 13 morphological and physiological traits used in a PCA biplot (Fig. 3) collected in *P. trichocarpa* association plantation in Corvallis, Oregon, USA.

Table S4 Pearson correlation of morphological and physiological traits with latitude of origin in *P. trichocarpa*.

Table S5 Estimated percentage of variance explained (PVE) in *P. trichocarpa*.

Table S6 Significant SNPs detected by *P. trichocarpa* single- and multitrait GWAS based on a value of $P < 1 \times 10^{-7}$ with nearest genes and their expression levels, putative gene functions and closest *Arabidopsis thaliana* (AT) homologs.

Table S7 Gene models detected by *P. trichocarpa* single- and multitrait GWAS based on a value of $P < 1 \times 10^{-7}$, co-expressed genes, gene description and linked metabolites from network analysis using Cytoscape.

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.