CrossMark

Single nucleotide polymorphisms in two *GID1* orthologs associate with growth and wood property traits in *Populus tomentosa*

Haijiao Yang[1,2] · Xiaohui Yang[1,2] · Longxin Wang[1,2] · Chenrui Gong[1,2] · Beibei Chen[1,2] ·
Qingzhang Du[1,2] · Deqiang Zhang[1,2]

Abstract Gibberellic acids (GAs) have multiple functions in various stages of plant growth and wood development, and GA receptor *GIBBERELLIN INSENSITIVE DWARF 1* (*GID1*) is a key player in GA signaling. Here, we used association studies to examine the allelic variation within *PtoGID1a* and *PtoGID1b* associated with growth and wood properties in a natural population of *Populus tomentosa*. We isolated two full-length PtoGID1a and PtoGID1b cDNA by reverse transcription PCR, and both of these two genes had an open reading frame with the same length (1035 bp) and encoded a protein of 344 amino acids. Analysis of tissue-specific expression profiles indicated that these two genes had similar transcription patterns, being expressed predominantly in leaves. Nucleotide diversity and linkage disequilibrium (LD) analysis showed that *PtoGID1a* and *PtoGID1b* harbor high single nucleotide polymorphism (SNP) diversity ($\pi_T = 0.08237$ and 0.00921, respectively) and low LD ($r^2 > 0.1$, within 350 and 1000 bp, respectively). SNP- and haplotype-based association tests identified that six single SNPs ($Q < 0.05$) and six haplotypes ($P < 0.05$, $Q < 0.10$) were significantly associated with growth and wood properties, explaining 0.99–10.28 and 3.10–5.39 % of the phenotypic variance, respectively. Therefore, isolation of *PtoGID1a* and *PtoGID1b* and dissection of their allelic polymorphisms open an avenue to understand the regulation of *PtoGID1* in the growth and wood formation in trees.

Introduction

The gibberellin (GA) phytohormones function as important endogenous growth regulators in plants. The bioactive GAs (e.g., GA_1, GA_3, and GA_4) modulate diverse physiological and developmental processes including stem elongation, leaf expansion, seed germination, pollen maturation, and the induction of flowers and fruits (Olszewski et al. 2002; Sun and Gubler 2004). In some species, bioactive GAs also play a major role in stimulating early stages of xylem differentiation during wood formation and ultimately in determining wood quality (Israelsson et al. 2005). The accumulating evidence that GAs stimulate these processes indicates a close link between GA metabolism and GA response pathways to maintain GA homeostasis in plants. The precise mechanisms of GA perception and signal transduction have remained longstanding questions that are central to the regulation of gene expression and plant phenotype. Extensive studies showed that the components of GA signaling pathways include the transcriptional regulators DELLA proteins and the soluble GA receptor encoded by *GIBBERELLIN INSENSITIVE DWARF 1* (*GID1*). DELLAs act as repressors,

✉ Deqiang Zhang
  DeqiangZhang@bjfu.edu.cn

1   National Engineering Laboratory for Tree Breeding, College of Biological Sciences and Technology, Beijing Forestry University, No. 35, Qinghua East Road, Beijing 100083, People's Republic of China

2   Key Laboratory Genetics and Breeding in Forest Trees and Ornamental Plants, Ministry of Education, College of Biological Sciences and Technology, Beijing Forestry University, No. 35, Qinghua East Road, Beijing 100083, People's Republic of China

often negatively regulating GA signaling, and are degraded in response to GA treatments during growth (Peng et al. 1997; Ikeda et al. 2001; Chandler et al. 2002). *GID1* is essential for GA perception and was first isolated in rice (*Oryza sativa*, *OsGID1*) based on the *GA insensitive dwarf 1* (*gid1*) mutant (Ueguchi-Tanaka et al. 2005). Subsequently, three homologous *GID1s* (*AtGID1a*, *AtGID1b*, and *AtGID1c*) were cloned in *Arabidopsis thaliana*. Each *AtGID1* could complement the rice *gid1* mutant, restoring the wild phenotype, showing that GID1s can respond to GA and function as GA receptors (Nakajima et al. 2006; Griffiths et al. 2006).

Based on the identification of GID1s in rice and Arabidopsis, research in the past decade has identified *GID1* genes in fern (Hirano et al. 2007), cotton (Aleman et al. 2008), barley (Chandler et al. 2008), aspen (Mauriat and Moritz 2009), and wheat (Li et al. 2013). Previous studies have focused mainly on the identification and characterization of *GID1* loci in various species (Ueguchi-Tanaka et al. 2005; Nakajima et al. 2006; Chandler et al. 2008) including the functional expression (Iuchi et al. 2007), the crystal structure of GID1 bound with GA or a ternary complex of GA-GID1-DELLA (Murase et al. 2008; Shimada et al. 2008), and the molecular mechanisms of GA perception and signaling (Ueguchi-Tanaka et al. 2007a, b; Hirano et al. 2008; Sun 2011). However, many questions remain, such as the allelic variation of *GID1* associated with complex quantitative traits, especially the specific genetic variants in *GID1* related to growth and wood quality traits.

Linkage disequilibrium (LD)-based association studies (also known as LD mapping) can effectively bridge our understanding of complex quantitative traits and the underlying genetic variation in specific candidate genes in a natural or breeding population (Zhang and Zhang 2005; Eckert et al. 2009). For example, Wegrzyn et al. (2010) found 27 significant single marker associations across 40 candidate genes with three composite traits (lignin content, syringyl-to-guaiacyl ratio, and C6 sugars) in *Populus nigra*. In addition, a recent study identified eight significant single nucleotide polymorphisms (SNPs) from two *S*-adenosyl-L-homocysteine hydrolase (*SAHH*) genes associated with growth or wood traits in a natural population of *P. tomentosa* (Du et al. 2014a). These studies suggested that association mapping can effectively identify alleles associated with complex quantitative traits in tree breeding programs. Therefore, based on the SNP and haplotype association tests, the identification of new *GID1* genes and dissection of their allelic variation will provide essential information for improving growth and wood formation.

In this study, we identified *PtoGID1a* and *PtoGID1b*, which encode GID1 proteins, and characterized these genes using LD-based association studies in *P. tomentosa*. Quantitative real-time PCR (qRT-PCR) analysis showed that *PtoGID1a* and *PtoGID1b* were expressed in all tissues and organs, in accordance with GA being involved in all developmental stages of the plants, and may be inducible by GA treatment in trees. Association genetics identified six common SNPs and six haplotypes in *PtoGID1a* and *PtoGID1b* that were associated with growth and wood properties in 426 unrelated individuals from a natural population of *P. tomentosa*. To extend these observations, we investigated the genetic variance in *PtoGID1a* and *PtoGID1b*, including additive and dominant effects. This comprehensive study of *PtoGID1a* and *PtoGID1b* polymorphisms can provide important resources and information for marker-assisted selection and breeding in forest trees.

Materials and methods

Population materials

The association population consisted of 426 unrelated individuals growing in Guan Xian County, Shandong Province (36° 23′ N, 115° 47′ E), China, where root segments of 1047 native individuals collected throughout the entire natural distribution range of *P. tomentosa* (covering an area of 1 million km^2) were used to establish a clonal arboretum using a randomized complete block design with three replications in 1982 (Zhang et al. 2010). The total distribution of *P. tomentosa* can be divided into three climatic regions, southern, northwestern, and northeastern, on the basis of principal component analysis and ISODATA fuzzy clustering of 16 meteorological factors (Zhu 1992; Huang 1992). In the present study, these 426 uncorrelated individuals, representing almost the entire geographic distribution of *P. tomentosa* (148 from the southern region, 126 from the northwestern region, and 152 from the northeastern region), were used for association analysis. In addition, a panel of 45 individuals was randomly selected from the association population (15 individuals from each region) and directly sequenced to test for SNPs within *PtoGID1a* and *PtoGID1b*.

Phenotypic data

Ten phenotypic traits were measured for at least three ramets per genotype in all individuals of the association population. Of which, three growth characteristics including tree height, diameter at breast height (DBH), and stem volume were directly measured during field surveys in 2011 using standard procedures described in Zhang et al. (2006). For seven wood property measurement, each wood sample contains bark and pith (15 cm in length and 10 mm in diameter) were cored from the 426 trees at breast height (1.3 m above ground level) using instrument borers. Then, these wood cores were pulverized to wood meal, in which the contents of holocellulose, α-cellulose, hemicelluloses, and lignin were consulted to test using

near-infrared reflectance spectroscopy (NIRS) according to Schimleck et al. (2004) based on training sets and calibration models derived from wet chemistry analysis technique described by Porth et al. (2013) and Tian et al. (2012). Fiber length and width were determined by the Color CCTV Camera (Panasonic SDII) after macerated in Franklin solution at 70 °C for 48 h and washed with deionized water until the wood materials had been neutralized. An X-ray powder diffractometer (Philips, Eindhoven, the Netherlands) was used to measure the microfibril angle, and the X-ray diffraction profile was integrated at Chi between −180° and +180°. The distribution of values for each trait was approximately consistent with a normal distribution (Tian et al. 2012; Du et al. 2013, 2014b).

Isolation of *PtoGID1a* and *PtoGID1b* cDNAs

Total RNA was extracted from the leaves of 1-year-old *P. tomentosa* using the Plant Qiagen RNeasy kit and then reverse transcribed into cDNA with the Superscript First-Strand Synthesis system according to the manufacturer's instructions (Life Technologies, Carlsbad, CA, USA). The constructed cDNA library was composed of 5.0×10^6 pfu with insert size of 1.0–4.0 kb. Subsequently, random end sequencing of 5000 cDNA clones and comparison with all available *A. thaliana GID1* sequences and *P. trichocarpa GID1* sequences in NCBI (http://www.ncbi.nlm.nih.gov) identified two clones that were highly similar to *AtGID1s* and *PtGID1s*. The BLAST program was used to analyze these expressed sequence tag (EST) sequences. Finally, two cDNA clones encoding *GID1* members were detected and isolated from *P. tomentosa* by amplification using gene-specific primers and named *PtoGID1a* and *PtoGID1b*, respectively, according to their order of discovery.

DNA extraction and full-length genomic DNA identification

Using the DNeasy Plant Mini kit (Qiagen China, Shanghai), total genomic DNA was extracted from young leaves of each *P. tomentosa* individual following the manufacturer's protocol. For sequencing the genomic DNA of *PtoGID1s*, a pair of specific primer (Table S1) was designed for amplification of *PtoGID1a* and *PtoGID1b*, respectively, based on their cDNA sequences and the available *P. trichocarpa GID1* sequences (Potri.013G028700 and Potri.002G213100, respectively) in NCBI, which contain the promoter regions and 3′ UTR-downstream regions. PCR amplification was performed according to the procedure described by Tian et al. (2014a), but the annealing temperature was replaced by 68 °C. Next, PCR products were resolved by agarose gel electrophoresis, excised, purified, ligated, transformed, and isolated. Ultimately, the BigDye Terminator Cycle Sequencing Kit,

version 3.1 (Applied Biosystems, Beijing, China), run on a Li-Cor 4300 genetic analyzer (Li-Cor Biosciences, Lincoln, NE), was used to directly sequence both strands of the plasmid DNA (three positive clones for each individual). For *PtoGID1a* and *PtoGID1b*, 5326 and 4912 bp of genomic DNA sequence were obtained, both including 2000 bp of the promoter region and 500 bp of 3′ UTR-downstream sequence. The 90 genomic clone sequences for *PtoGID1a* and *PtoGID1b* were deposited in GenBank under the accession numbers KU161176–KU161220 and KU161221–KU161265, respectively.

Gene structure and phylogenetic analysis

The gene structure diagrams of *PtoGID1a* and *PtoGID1b* were produced with Gene Structure Display Server (GSDS 2.0) (http://gsds.cbi.pku.edu.cn) after independently submitting their coding and genomic sequences. To investigate the phylogenetic relationship of *PtoGID1a* and *PtoGID1b* to the GID1 proteins from other species, 37 amino acid sequences of *GID1* from dicotyledons, monocotyledons, gymnospermous plants, as well as lower plants such as pteridophyte and bryophyte plants (Table S2) were obtained from NCBI (http://www.ncbi.nlm.nih.gov). Multiple sequence alignment, phylogenetic, and molecular evolutionary analyses were conducted using MEGA version 6.06, and the neighbor-joining method was used to build an unrooted phylogenetic tree (Tamura et al. 2007). Statistical confidence of tree nodes was based on 1000 bootstrap replicates.

Tissue-specific expression analysis

Plant materials and GA₃ treatment

The vegetatively propagated 1-year-old *P. tomentosa* "LM50" were grown under 16-h light and 8-h dark in a greenhouse at Beijing Forestry University, Beijing, China (40° 0′ N, 116° 20′ E). For GA treatment, 100 μM GA₃ (Sigma-Aldrich, St. Louis, MO, USA) was sprayed on seeding leaves until drops of liquid dripped down. The control plants were treated with water in the same manner. Mature leaves from the same position of control plants (at 0 h) and GA-treated plants were collected at 0, 3, 12, and 24 h after treatment, and three biological replicates were sampled per time points. Nine tissues (root, apex, bark, phloem, cambium, young leaf, mature leaf, developing xylem, and mature xylem) and four tissues (mature leaf, cambium, developing, and mature xylem) of three ramets were collected at 0 and 12 h, respectively, were used to examine the expression of *PtoGID1a* and *PtoGID1b* in different tissues and organs of *P. tomentosa*. All the harvested tissues were immediately frozen in liquid nitrogen after collection and stored at −80 °C until use.

RNA extraction

Total RNA was extracted using the Plant Qiagen RNeasy kit (Qiagen China, Shanghai) according to the manufacturer's instructions. During the RNA purification, additional on-column DNase digestions were performed three times using the RNase-Free DNase Set (Qiagen). Then, RNA was quantified based on absorption at 260 nm and reverse transcribed into cDNA using the Superscript First-Strand Synthesis system and the supplied polythymine primers (Invitrogen). All cDNA samples were used for tissue-specific expression analysis of *PtoGID1a* and *PtoGID1b* with qRT-PCR.

Quantitative real-time PCR

With the *PtoGID1a*- and *PtoGID1b*-specific and internal control (*Actin*) primer pairs (Table S3) designed with Primer Express 3.0 software (Applied Biosystems), qRT-PCR was performed on a 7500 Fast Real-Time PCR System (ABI). Two microliters of cDNA for all tissues was amplified in a reaction containing 10 μl of Quanti-Tect SYBR Green PCR reagent (Qiagen), 0.4 μl of ROX Reference Dye II solution (Invitrogen), 0.4 μl of 10 nM forward and reverse primers, and 6.8 μl of dH$_2$O. The PCR program included an initial denaturation at 95 °C for 30 s, 40 cycles of 5 s at 95 °C, 30 s at 60 °C, and a final melt curve of 72–95 °C. All reactions were carried out in triplicate for technical and biological repetitions with three plants. The specificity of the amplified fragments was checked through the melting curve, and the generated data was analyzed using the Opticon Monitor Analysis software 3.1 tools (Bio-Rad, USA) and standardized to the level of the *Actin* gene.

SNP identification and genotyping

The full-length genomic DNAs of *PtoGID1a* and *PtoGID1b*, both including 2000 bp of the promoter region and 500 bp of the 3′ UTR-downstream region, were resequenced and analyzed in 45 unrelated individuals from the association population to identify SNPs without consideration of insertions/deletions (InDels). To identify putative SNP variants, plasmid DNA was sequenced on both strands with conserved primers using the BigDye Terminator version 3.1 Cycle Sequencing Kit (Applied Biosystems, Beijing, China) and a Li-Cor 4300 genetic analyzer (Li-Cor Biosciences, Lincoln, NE). Then, BioEdit and DNA sequence polymorphism (DnaSP) software version 5.10 (Librado and Rozas 2009) were used for multiple-sequence alignment and polymorphism analysis. Ultimately, 159 common SNPs (minor allele frequencies ≥0.10, 78 SNPs for *PtoGID1a*, and 81 SNPs for *PtoGID1b*) were genotyped by single nucleotide primer extension using a Beckman Coulter sequencing system for all DNA samples.

Nucleotide diversity and linkage disequilibrium

DnaSP version 5.10 (Librado and Rozas 2009) was used to calculate summary statistics for nucleotide diversity and divergence. Nucleotide diversity was estimated using both the average number of pairwise differences per site between sequences, π (Nei 1987), and the average number of polymorphisms segregating per site, θ_w (Watterson 1975). In addition, the diversity statistics were also calculated separately for noncoding, synonymous, and nonsynonymous sites in *PtoGID1a* and *PtoGID1b*. To estimate whether natural selection (purifying selection or positive selection) was involved in the evolution of the two genes during speciation, we did d_N/d_S analysis on the DNA sequences of *PtoGID1a* and *PtoGID1b*.

To assess the pattern of LD in the sequenced candidate gene regions, the decay of LD within a specific physical distance (base pairs) between SNP sites within *PtoGID1a* and *PtoGID1b* was estimated by nonlinear regression analysis (Hill and Weir 1988) using DnaSP version 5.10 with sequence data. LD was measured as the squared correlation of allele frequencies (r^2), which is affected by both recombination and differences in allele frequencies between sites (Hill and Robertson 1968). The significance of the *r* value for each pair of common SNPs was calculated using the same software DnaSP version 5.10 (Librado and Rozas 2009).

Association analysis

Single SNP-based association All single SNP-based association tests between 159 SNP markers of *PtoGID1a* and *PtoGID1b* and 10 traits in the natural population were conducted via three statistical models: (1) a general linear model (GLM) only take population structure into account (*Q* model), (2) a mixed linear model (MLM1) only inferred relative kinship as a random covariate (*K* model), and (3) a unified mixed linear model (MLM2) uses both *Q* and *K* values (*Q* + *K* model) (Yu et al. 2006). These analyses with above three models were implemented in TASSEL version 2.1 (Yu et al. 2006; Bradbury et al. 2007). As described previously, the values of membership probability (*Q*) and pairwise kinship (*K*) were used to evaluate the effects of population structure and relatedness among individuals for marker-trait associations. The *Q* matrix was established based on the significant subpopulations (*K* = 11) by 20 genomic SSR markers (Du et al. 2012), as assessed according to the statistical model described by Evanno et al. (2005), using STRUCTURE version 2.3 (Pritchard et al. 2000). The relative kinship matrix (*K*) was obtained based on the same 20 SSR loci using the method proposed by Ritland (1996), which is built into the program SPAGeDi version 1.3 (Hardy and Vekemans 2002). The *K* matrix was calculated as described by Yu et al. (2006), and all negative values between individuals were set to 0. To correct for multiple testing, the false discovery rate (FDR) method

was additionally applied to identify significant SNPs using QVALUE in R (Storey and Tibshirani 2003; Storey et al. 2004).

Haplotype-based association All high-LD blocks of haplotypes ($r^2 \geq 0.75$, $P \leq 0.001$) were determined for each gene using the software Haploview version 4.1 (Barrett et al. 2012), and the haplotype frequencies were estimated based on the genotypic data for the 426 individuals. On the basis of the information from those significant SNPs in the haplotype blocks, haplotype-based association tests with growth and wood quality traits were performed using haplotype trend regression software (HTR) (Zaykin et al. 2002) with a three-marker sliding window. The significances of the haplotype associations were conducted based on 10^4 permutation tests. Singleton alleles and haplotypes with a frequency <5 % were discarded when constructing the haplotypes.

Modes of gene action

The modes of gene action were quantified using the ratio of dominant (*d*) to additive (*a*) effects ($|d/a|$) estimated from least square means for each genotypic class. When the ratio of dominance to additive effects was in the range $0.5 < |d/a| \leq 1.25$, gene action was defined as partial or complete dominance. Additive effects were defined as value of $|d/a| \leq 0.5$; under- or over-dominance was defined as values of $|d/a| > 1.25$. Details of the algorithm and formulas for calculating gene action were previously described (Eckert et al. 2009; Wegrzyn et al. 2010).

Results

Isolation and sequence analysis of *PtoGID1a* and *PtoGID1b* from *P. tomentosa*

Two full-length cDNAs of *PtoGID1a* and *PtoGID1b* were isolated from a cDNA library prepared from the leaves of *P. tomentosa* by reverse transcription PCR. The cDNAs of *PtoGID1a* and *PtoGID1b* are 2153 and 1852 bp in length, respectively, including 333 and 261 bp of the 5′ untranslated region (UTR), 785 and 556 bp of 3′ UTR, and have open reading frames of equal length 1035 bp. Alignment of the full-length cDNA sequence with the genomic sequence showed that *PtoGID1a* and *PtoGID1b* each have one intron (673 and 560 bp, respectively), located between the two exons (Fig. 1a). The deduced protein sequences of *PtoGID1a* and *PtoGID1b* revealed these genes each encode 344 amino acid polypeptides with estimated molecular weights of 38.54 and 39.12 kDa and theoretical pI values of 6.15 and 6.22, respectively.

We compared the amino acid sequences of PtoGID1a and PtoGID1b with those of known AtGID1s (AtGID1a, AtGID1b, AtGID1ac) (Fig. Fig. S1). The similarity between the two PtoGID1 sequences was 79.65 %, and both had 70.35–84.59 and 73.26–79.65 % similarities with the AtGID1s from *A. thaliana*, respectively. PtoGID1a and PtoGID1b share sequence similarity with the consensus sequence of the hormone-sensitive lipases (HSL) family, including the conserved motifs HGG and GXSXG. Moreover, the PtoGID1a and PtoGID1b amino acid sequences also retain the first two conserved residues, serine (S) and aspartic acid (D), which form the HSL catalytic triad. The third histidine (H) residue was replaced with isoleucine (I) at the corresponding positions in both PtoGID1a and PtoGID1b compared with OsGID1 and the AtGID1s (Fig. Fig. S1). In addition, there were 13 functional domains (TWVLIS, LDR, FFHGGSF, HS, IYD, YRR, DGW, GDSSGGNI, GNI, MF, LDGKYF, WYW, and GFY) important for GA binding activity or GID1-DELLA protein interaction (Ueguchi-Tanaka et al. 2007a, b) from the N-terminal to C-terminal, and these domains also occurred in PtoGID1a and PtoGID1b (Fig. Fig. S1), except for the LDR domain replaced by LER in PtoGID1b. It is therefore proposed that the *PtoGID1a* and *PtoGID1b* genes encode fully functional GA receptors.

Phylogenetic analysis of GID1 proteins in plant

To investigate the evolutionary relationship of PtoGID1a and PtoGID1b with the GID1 proteins from other plant species, an unrooted neighbor-joining (NJ) tree was constructed using 37 complete amino acid sequences (Table S2). Alignment of all predicted proteins showed that these GID1 proteins can be divided into five distinct groups, and the sequences in each group follow the same classification as the species (Fig. 1b). This suggests that *GID1* differentiation events occurred before the divergences of gymnosperm–angiosperm, moss–fern, and monocot–dicot plants. Further subdivisions of the eudicot groups revealed that GID1a and GID1c could be classified into the same groups, except for PhGID1c and MdGID1c, while GID1b was located in an independent subgroup, indicating that these individual GID1s may have distinct biological functions (Nakajima et al. 2006; Griffiths et al. 2006). The phylogenetic tree indicated that *PtoGID1a* is an ortholog of *AtGID1a/c* and that *PtoGID1b* should be classified with *AtGID1b* and other *GID1b* genes.

Ubiquitous expression of *PtoGID1a* and *PtoGID1b* in *P. tomentosa*

To examine the expression of *PtoGID1a* and *PtoGID1b* in different tissues and organs of *P. tomentosa*, we compared the steady-state mRNA levels of the two genes by qRT-PCR analysis with gene-specific primers and *Actin* as the internal control (Fig. 2a, b). Expression of the two genes

Fig. 1 Overview of *PtoGID1a*
and *PtoGID1b*. **a** Gene structure
of *PtoGID1a* and *PtoGID1b* in
P. tomentosa. Exons and
upstream/downstream regions,
which included untranslated
regions (UTRs), are shown as
black and gray boxes,
respectively, and the *gray lines
between boxes* indicate introns.
Positions of the translational start
(*ATG*) and termination codons
(*TAG/TAA*) of the predicted
coding sequences are shown. **b**
Neighbor-joining phylogenetic
tree of PtoGID1a, PtoGID1b, and
GID1 proteins from other plants.
Detailed information on 37
protein sequences is presented in
Additional file Table S2

(a)

(b)

Legend:
■ Exon   ━ Upstream/Downstream   — Intron
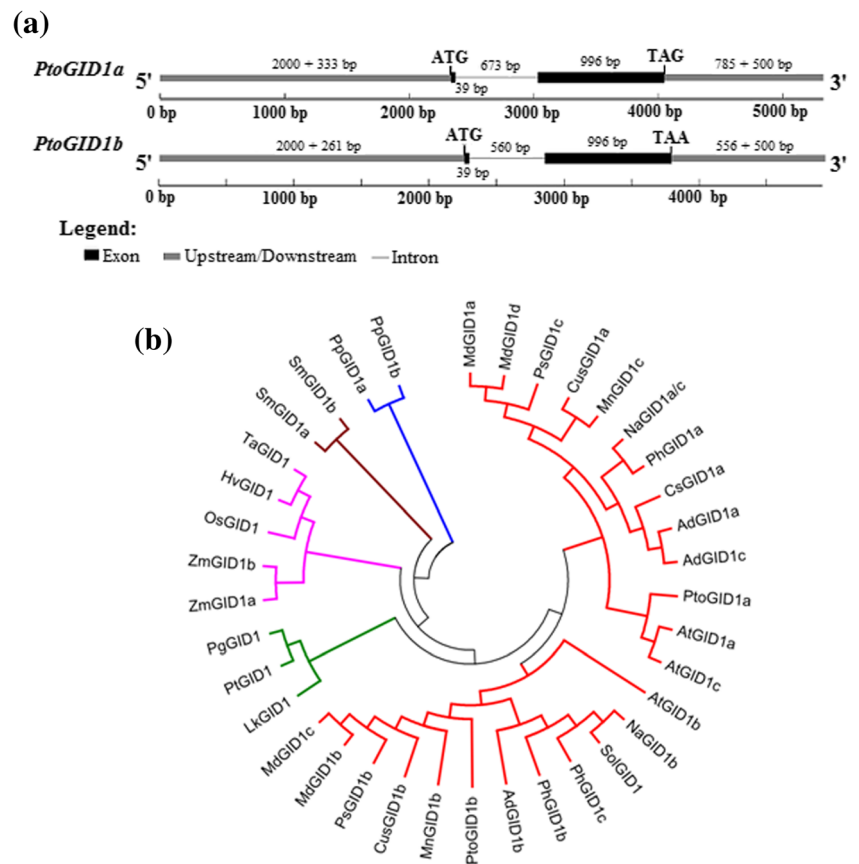
was detected in all the tissues and organs at varied levels
before or after GA$_3$ treatments, consistent with that GA is
active throughout the plants and during many develop-
mental processes. The transcripts of both *PtoGID1a* and
PtoGID1b were most abundant in the leaf and showed
similar overall profiles. The differences were that
PtoGID1a transcript levels were highest in mature leaf
tissue (4.24 ± 0.1695, in arbitrary units, relative to the
Actin control), followed by young leaf tissue
(2.35 ± 0.2919), but *PtoGID1b* mRNA transcript levels
were highest in young leaf tissue (2.74 ± 0.4208), follow-
ed by mature leaf tissue (0.35 ± 0.0506) at 0 h. By con-
trast, in the secondary tissues of bark, phloem, and xylem,
the expression levels of *PtoGID1a* and *PtoGID1b* showed
the lowest mRNA abundance. However, in comparison to
untreated tissues, transcript abundances of these two
genes were increased in mature xylem (0.184 ± 0.0035
vs. 0.027 ± 0.001 and 0.061 ± 0.0041 vs.
0.002 ± 0.0006) and developing xylem (0.017 ± 0.0009
vs. 0.003 ± 0.0003 and 0.058 ± 0.0030 vs.
0.0039 ± 0.0004) at the time points of 12 h after GA$_3$
treatment (Fig. 2b). Additionally, the transcript levels of
PtoGID1a and *PtoGID1b* were also increased in cambium
(0.886 ± 0.0391 vs. 0.115 ± 0.0529 and 0.0495 ± 0.0022
vs. 0.027 ± 0.0065) relative to the original, although

moderate expressions of the two genes were found in root,
apex, and cambium tissue before GA$_3$ treatment.

We further explored whether the expression of *PtoGID1a*
and *PtoGID1b* responds to GA by measure their transcript
levels at different time points after GA$_3$ treatment (Fig. 2c).
PtoGID1a expression first increased and then decreased fol-
lowing GA$_3$ treatment, peaking at 12 h (25.8 ± 2.114 times
higher than the *Actin* control), suggesting that GA$_3$
upregulates *PtoGID1a*. Conversely, *PtoGID1b* expression de-
creased following GA$_3$ treatment up to 12 h but then increased
gradually to a peak at 24 h (1.1 ± 0.222 times higher than the
Actin control) after treatment.

Nucleotide diversity and linkage disequilibrium of *PtoGID1a* and *PtoGID1b*

To identify nucleotide polymorphisms, we sequenced 5326
and 4912 bp genomic regions of *PtoGID1a* and *PtoGID1b*,
respectively, from 45 unrelated individuals of *P. tomentosa*.
All of the 90 sequences from *PtoGID1a* and *PtoGID1b*
contained the 5′ UTR, introns, 3′ UTR, 1035-bp exons,
2000-bp promoter regions, and 500-bp 3′ UTR-downstream
regions. Across the samples, a total of 251 SNPs (123 from
PtoGID1a and 128 from *PtoGID1b*) were detected in these
two genes at a frequency of approximately one SNP every 43

Fig. 2 Relative transcript levels of *PtoGID1a* and *PtoGID1b*. **a** Relative expression levels of *PtoGID1a* and *PtoGID1b* in various tissues and organs from *P. tomentosa* at 0 h. **b** Relative expression levels of *PtoGID1a* and *PtoGID1b* in various tissues at the time points of 12 h after GA$_3$ treatment. **c** Relative expression levels of *PtoGID1a* and *PtoGID1b* before and after GA treatment at different time points in leaves tissue. All expression levels were normalized to the mean value of *Actin*. The error bars represent ±SD
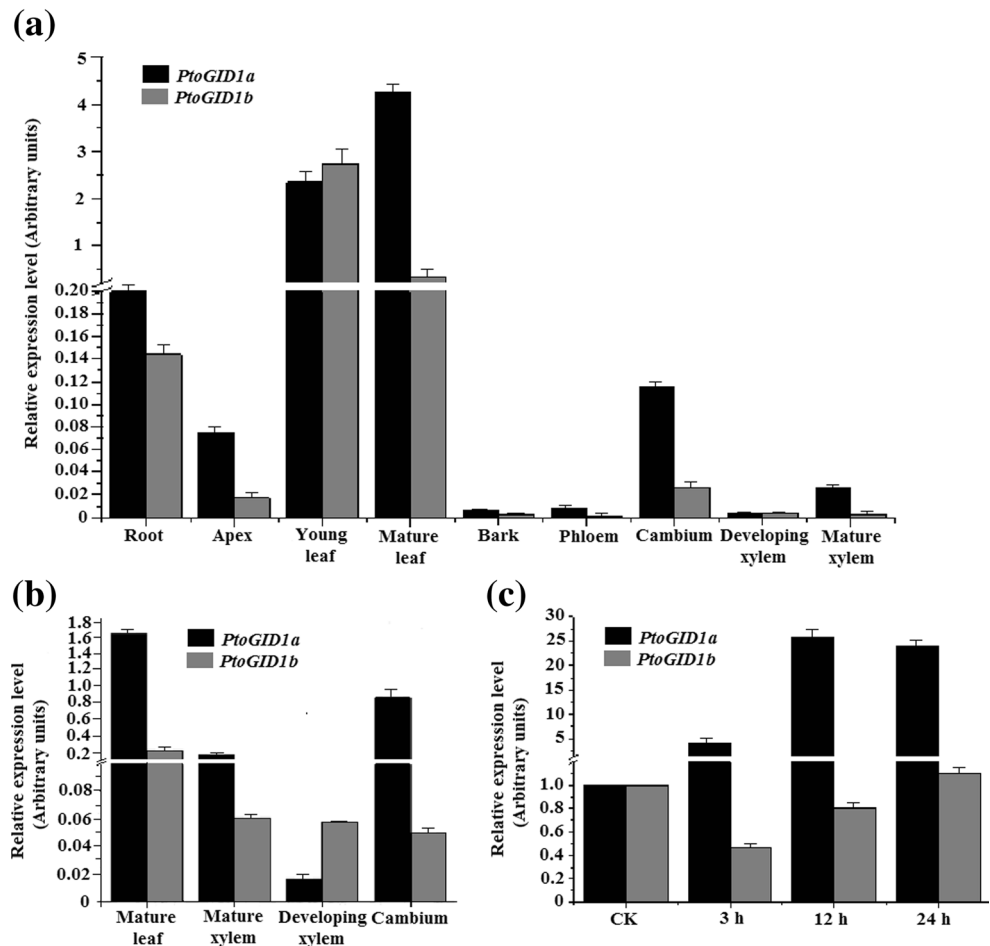
and 38 bp, respectively. Forty-three and eight of these SNPs occurred in coding regions of *PtoGID1a* and *PtoGID1b*, respectively, of which 33 SNPs were categorized as silent mutations and 18 were missense sites (Table 1). Altogether, 78 and 81 (63.41 and 63.28 %) were considered common SNPs (minor allelic frequency, MAF ≥ 0.10) in *PtoGID1a* and *PtoGID1b*, respectively. In general, both of these genes showed high nucleotide diversity (Table 1). The average levels of nucleotide diversity (π_T) for the different regions varied significantly ranging from 0.00336 (3′ UTR) to 0.01766 (exon 2) in *PtoGID1a* and from 0.00174 (intron) to 0.01365 (promoter) in *PtoGID1b* (Table 1). Similarly, estimates of the population mutation parameter (θ_w) varied between 0.00239 (intron) and 0.01012 (exon 2) in *PtoGID1a* and 0.00123 (intron) and 0.00924 (exon 2) in *PtoGID1b*. In coding regions of *PtoGID1a* and *PtoGID1b*, the values of nonsynonymous nucleotide diversity (π_{nonsyn} = 0.00847 and 0.00021, respectively) were markedly lower than the synonymous nucleotide diversity (π_{syn} = 0.04375 and 0.01429, respectively) (Table 1). The π_{nonsyn}/π_{syn} ratios for the exons were 0.1936 and 0.0147, respectively, reflecting the strong action of purifying selection involved in nonsynonymous sites of exon regions during speciation.

The pattern of the squared allelic correlation coefficient (r^2) with base pairs within *PtoGID1a* and *PtoGID1b* (including the promoter and the 3′ UTR-downstream regions) was used to assess the LD level. The nonlinear regression model showed that LD decayed quite rapidly with r^2 dropping to 0.1 within ~350 and ~1000 bp in *PtoGID1a* and *PtoGID1b*, respectively (Fig. 3), indicating that LD did not extend over the entire gene region. We then genotyped 78 and 81 common SNPs from 426 individuals, and 13 and four high-LD distinct haplotype blocks across the sequenced regions differed in *PtoGID1a* and *PtoGID1b*, respectively. There was a higher level of LD within each block ($r^2 > 0.8$), whereas LD levels were low between one block to the other haplotype block ($r^2 < 0.3$) (Fig. S2).

Single SNP-trait and haplotype-based associations

To explore the effects of SNPs in *PtoGID1a* and *PtoGID1b*, we conducted 1590 association tests (159 SNPs × 10 traits) using three statistical models (*Q*, *K*, and *Q* + *K*) for comparing (Table S4). For *PtoGID1a*, the numbers of significant markers ($P < 0.05$) were 33 using *K* model and 25 using *Q* model but fell to 14 with *Q* + *K* model. In *PtoGID1b*, we found 34 SNPs were associated with tree traits at $P < 0.05$ by the *K* model, 29

Table 1 Nucleotide polymorphisms in *PtoGID1a* and *PtoGID1b* locus

Locus	Region	Length (bp^{-1})	No. of polymorphic sites	Frequency (bp^{-1})	Nucleotide diversity	
					π_{T}	θ_{w}
PtoGID1a	Promoter	2000	43	47	0.00639	0.00596
	5′ UTR	333	9	37	0.00792	0.00628
	Exon 1	39	0	0	0	0
	Synonymous	6.66	0	0	0	0
	Nonsynonymous	32.33	0	0	0	0
	Intron	673	8	84	0.00428	0.00239
	Exon 2	996	43	23	0.01766	0.01012
	Synonymous	232.70	26	9	0.04500	0.04658
	Nonsynonymous	760.31	17	45	0.00883	0.00889
	3′ UTR	785	11	71	0.00336	0.00293
	3′ UTR-downstream	500	9	56	0.00364	0.00369
	Total silent[a]	4129.36	106	39	0.0739	0.00743
	Synonymous	239.36	26	9	0.04375	0.04524
	Nonsynonymous	792.64	17	47	0.00847	0.00852
	Total[b]	5326	123	43	0.08237	0.01595
PtoGID1b	Promoter	2000	82	24	0.01365	0.00924
	5′ UTR	261	11	24	0.00366	0.00441
	Exon 1	39	1	39	0.00427	0.00595
	Synonymous	7.61	0	0	0	0
	Nonsynonymous	31.39	1	31	0.0523	0.00728
	Intron	560	4	140	0.00174	0.00123
	Exon 2	996	7	125	0.00349	0.00161
	Synonymous	234.84	7	34	0.01475	0.00779
	Nonsynonymous	758.16	0	0	0	0
	3′ UTR	556	12	46	0.00493	0.00374
	3′ UTR-downstream	500	10	50	0.00440	0.00276
	Total silent[a]	4122.45	126	33	0.00900	0.00666
	Synonymous	242.45	7	35	0.01429	0.00755
	Nonsynonymous	789.55	1	790	0.00021	0.00029
	Total[b]	4912	128	38	0.00921	0.00695

Region containing indels were excluded from the calculation. The standard deviations (SD) of π and θ are not shown in this table

[a] Synonymous sites and noncoding position

[b] Silent sites plus nonsynonymous sites

by the Q model, and only 13 by the $Q + K$ model. After correction for multiple testing with a FDR test ($Q < 0.05$), the numbers of significant associations of SNPs with growth and wood property traits were reduced to four and two in *PtoGID1a* and *PtoGID1b*, respectively (Table 2). These six significant associations represented six unique sites from the promoter, 5′ UTR, and exon of *PtoGID1a* and *PtoGID1b*. These SNPs were significantly associated with five phenotypic traits: tree height, stem volume, fiber width, α-cellulose contents, and microfiber angle (Table 2). Of these markers, SNP52 and SNP55 from exon 2 of *PtoGID1a* were synonymous substitutions, and all associated markers from

PtoGID1b occurred in noncoding regions. Altogether, the percentage of phenotypic variance (R^2) explained by each SNP ranged from 0.99 to 10.28 %, in accordance with polygenic quantitative models of wood traits.

To quantify the modes of gene action, we calculated the ratio of dominant (d) to additive (a) effects ($|d/a|$). Two of the six marker-trait pairs showed additive effects ($|d/a| \leq 0.5$), one showed a partially to fully dominant effect ($0.5 < |d/a| \leq 1.25$), and the remaining associations were split between over- or under-dominance ($|d/a| > 1.25$) (Table 2). For example, the two genotypes of *PtoGID1a*-SNP16 and *PtoGID1b*-SNP3 showed difference in tree height (15.70 ± 1.162 m for AA,
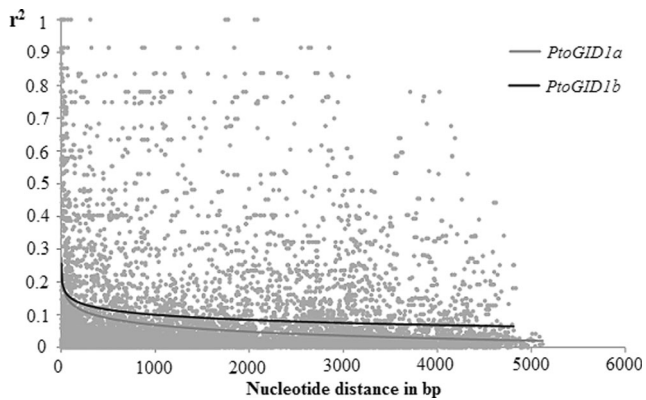
Fig. 3 Decay of linkage disequilibrium within *PtoGID1a* and *PtoGID1b*. Pairwise correlations between SNPs were plotted against the physical distance between the SNPs in base pairs. The *curve* describes the nonlinear regression of the least squares fit r^2 (Er2) onto the physical distance in base pairs

15.00 ± 0.296 m for GA, and 11.00 ± 0.335 m for GG) and stem volume (0.65 ± 0.049 m^3 for AA, 0.53 ± 0.100 m^3 for AC, and 0.35 ± 0.085 m^3 for CC), consistent with additive effects. The heterozygotes of *PtoGID1a*-SNP52 and *PtoGID1b*-SNP32 had higher stem volume (0.68 ± 0.070 m^3 for AG vs. 0.59 ± 0.056 m^3 for AA, 0.49 ± 0.012 m^3 for GG) and wider fibers (22.97 ± 0.474 mm for GT vs. 22.43 ± 0.269 mm for GG, 22.65 ± 0.733 mm for TT), respectively, than either homozygote class. The average values of microfiber angle across genotypic classes for *PtoGID1a*-SNP12 were suggestive of partial dominance ($18.21 \pm 1.512°$ for AA, $17.65 \pm 0.464°$ for AG, and $17.33 \pm 1.553°$ for GG), with the A allele being partially dominant (Fig. 4).

Moreover, we performed haplotype-based association tests to identify significant haplotypes with the same 10 growth and wood quality traits. Three obvious LD regions including eight common haplotypes were observed only in *PtoGID1a* with significant effect on these traits ($P < 0.05$, frequency > 1 %) (Table 3). Among these, two haplotypes from SNP1–4 were simultaneously associated with DBH and stem volume and three haplotypes from SNP61–64, which spans the 3′ UTR region, were associated with α-cellulose contents. SNP49–51 in exon 2 showed close association with fiber width. The proportion of phenotypic variation explained by each haplotype ranged from 3.10 to 5.39 % (Table 3). To additionally dissect the allelic variation in the haplotype identified in *PtoGID1a*, we also tested associations using haplotype trend regression (HTR) in *PtoGID1b* (Table S5). One haplotype (G-T-G) located in the promoter region extending from SNP35 to SNP37 was concurrently associated with fiber width, fiber length, and microfiber angle at a frequency of <1 %.

Discussion

Identification and characterization of *PtoGID1a* and *PtoGID1b* from *P. tomentosa*

The GID1-mediated GA perception system involves transcriptional regulation of GA-responsive genes in plant growth and development. In the present study, two GID1s encoded by *PtoGID1a* and *PtoGID1b* were determined to not only contain two conserved motifs (HGG and GXSXG) within the *HSL* family and retain the first two residues (S and D) of the catalytic triad at the corresponding position, but they also preserve all 13 important functional domains, in accordance with the expected conserved

Table 2 Significant SNP markers associated with growth and wood property traits and list of marker effects for significant marker-trait pairs

Trait	Locus	Region	Mutation	*P* value	*Q* value	R^2 (%)	$2a$[a]	d[b]	d/a	$2a/S_p$[c]	α[d]
PtoGID1a											
MFA	SNP12	Promoter	[A:G]	5.56E-11	8.27E-09	5.93	1.823	0.925	1.015	0.364	0.015
Height	SNP16	5′ UTR	[A:G]	1.84E-11	8.21E-09	0.99	0.100	−0.256	−0.256	0.027	1.056
Volume	SNP52	Exon 2	[A:G][s]	2.11E-06	2.35E-04	1.35	0.568	0.380	1.336	1.380	−0.019
α-Cellulose	SNP55	Exon 2	[A:C][s]	4.70E-11	8.27E-09	10.28	–	–	–	–	–
PtoGID1b											
Volume	SNP3	Promoter	[A:C]	8.44E-05	0.0187	7.92	0.328	−0.036	−0.219	0.795	0.018
Fiber width	SNP32	Promoter	[C:T]	4.18E-06	0.0021	5.41	−1.1708	−1.2931	−2.209	−0.567	0.549

P value significance level for association ($P \leq 0.05$), *Q* value a correction for multiple testing (false discovery rate (FDR) $Q \leq 0.05$), R^2 percentage of the phenotypic variance explained, *MFA* microfibril angel, *Volume* stem volume, *s* synonymous polymorphism, *ns* nonsynonymous polymorphism

[a] Calculated as the difference between the phenotypic means observed within each homozygous class ($2a = |G_{BB} - G_{bb}|$, where G_{ij} is the trait mean in the *ij*th genotypic class)

[b] Calculated as the difference between the phenotypic means observed within each heterozygous class and the average phenotypic mean across both homozygous classes ($d = G_{Bb} - 0.5 (G_{BB} + G_{bb})$, where G_{ij} is the trait mean in the *ij*th genotypic class)

[c] SD for the phenotypic trait under consideration

[d] The additive effects were calculated as $\alpha = p_B (G_{BB}) + p_b (G_{Bb}) - G$, where *G* is the overall trait mean, G_{ij} is the trait mean in the *ij*th genotypic class, and p_i is the frequency of the *i*th marker allele. These values were always calculated with respect to the minor allele
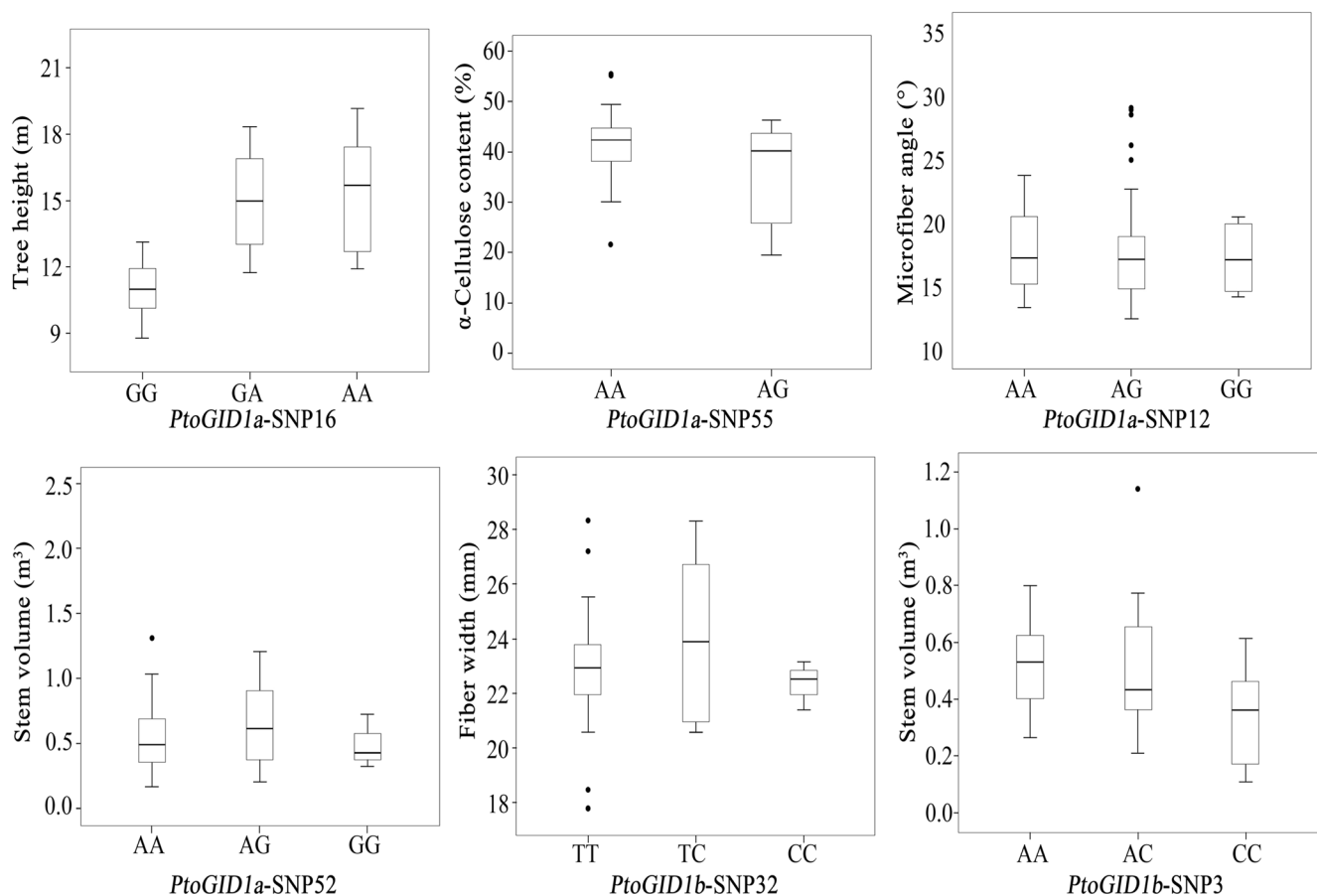
Fig. 4 Genotypic effects of the significant single nucleotide polymorphisms (SNPs) in *PtoGID1a* and *PtoGID1b* on different phenotypic traits in an association population. The *x* axes represent the genotypic classes. The left *y* axes represent the phenotypic data of the association population. *Upper* and *lower edges* of each *box* represent the first and third quartiles of the distribution, and the *central lines* indicate the average value. *Whiskers* in the *box plots* represent the maxima and minima

features of GID1 identified in other plants. Ala-scanning muta-genesis of *OsGID1* showed that these conserved motifs and func-tional domains are crucial for the interaction of GID1 with GAs or DELLA proteins (Ueguchi-Tanaka et al. 2007a, b).

The known eudicot *GID1* genes fall into two distinct groups: *GID1a/c* and *GID1b* (Hirano et al. 2007; Voegele et al. 2011), which is in accordance with our findings that the two *PtoGID1* genes were classified into the "GID1a/c" and "GID1b" sub-groups of the phylogenetic tree (Fig. 1b). This supports the view that the different GID1 receptors seem to have distinct functions during plant development. The distinction between *PtoGID1a* and *PtoGID1b* is also supported by the tissue-specific expression patterns. In our study, although *PtoGID1a* and *PtoGID1b* were expressed in most of the tissues, their expression levels were different. For example, *PtoGID1a* was predominantly expressed in mature and young leaves, as was *PtoGID1b*. However, the *PtoGID1a* expression level in mature leaf was 12 times higher than that of *PtoGID1b* (Fig. 2a). Moreover, the hormonal regu-lation of expression of *PtoGID1a* and *PtoGID1b* at the same time points after GA$_3$ treatment also differed (Fig. 2c). The idea that the individual *GID1* genes have different expression profiles

and display partial redundancy has been postulated identified in *A. thaliana*, in which *AtGID1a* and *AtGID1c* transcripts were present at high and low levels in most of the analyzed tissues, while the level of the *AtGID1b* transcript was moderate (Griffiths et al. 2006). In our study, we found that the deduced amino acid sequences of the *PtoGID1* genes were similar. The transcription-al expression of *PtoGID1a* and *PtoGID1b* differed temporally and spatially and exhibited differential expression in various tis-sues and in response to GA$_3$ treatment. Therefore, by a combi-nation of structural analysis, functional expression, and phyloge-netic analysis of *PtoGID1a*, *PtoGID1b*, and *GID1s* in other plants, we infer that these two *PtoGID1* members as GA recep-tors are orthologs of *OsGID1* and *AtGID1s*.

Analysis of statistical models for association

The method of association mapping has the potential to pin-point genetic polymorphism underlying important traits (e.g., growth and wood properties) in highly heterozygous undo-mesticated forest trees. However, false discoveries (also known as false positive rates) are a major concern and can

Table 3 *PtoGID1a* haplotypes significantly associated with growth and wood property traits

Trait	P value	R² (%)	Haplotype	Frequency	P value (Ind)
DBH	0.014	5.30	SNPs 1–4		
			A-T-T-C	0.035	0.037
			G-T-T-T	0.103	0.019
Stem volume	0.022	3.56	SNPs 1–4		
			A-T-T-C	0.035	0.025
			G-T-T-T	0.103	0.047
Fiber width	0.020	5.39	SNPs 49–51		
			C-T-A	0.012	0.021
α-Cellulose	0.022	3.10	SNPs 61–64		
			C-G-G-A	0.454	0.028
			T-A-A-A	0.012	0.018
			T-A-G-A	0.012	0.006

P value significance level for the whole haplotype-based association ($P \leq 0.05$), *P* value (Ind) significance level for different individual haplotype associations, R^2 percentage of the phenotypic variance explained, *DBH* diameter at breast height

be partially attributed to spurious associations caused by population structure and unequal relatedness among individuals (Yu et al. 2006). Therefore, an appropriate statistical model is necessary to account for this point. Initially, the GLM model with *Q* matrix has attracted considerable attention as a powerful and effective approach for reducing the risk of false positive rates arising from population stratification. However, *Q* values alone are not adequate because the *Q* matrix only provides a rough dissection of population differentiation (Yang et al. 2010). Similarly, the simple model with kinship matrix alone would not be sufficient to correct genotypes among relatives in a population and thus can lead to spurious association in the similar way (Bourgain et al. 2003; Yu et al. 2006). To reduce the false positive rates on marker-trait associations while maintaining statistical power, a unified mixed-model approach with correction for population stratification and relatedness have been suggested. In the *Q* + *K* model, population structure is fit as a fixed effect, whereas kinship among individuals is incorporated as the variance–covariance of the random effect for the individuals (Pritchard et al. 2000; Zhao et al. 2007).

In the present study, we compared the classical GLM (*Q*) and MLM (*K*, *Q* + *K*) models on testing associations of two genes for growth and wood property traits carried out in a natural population of *P. tomentosa*. The numbers of detected significant markers from *PtoGID1a* and *PtoGID1b* and the explained phenotypic variance were varied with different models (Table S4). Of which, the *Q* + *K* model was even more conservative and detected fewer significant markers (14 vs. 33 or 25, and 13 vs. 34 or 29 in *PtoGID1a* and *PtoGID1b*, respectively) for growth and wood property traits than the simple *Q*- and *K*-based models. Also, almost all of those marker-trait association tests which are detected by the *Q* + *K* model appeared in other two models (Table S4). This result indicated

that the *Q* + *K* model is not prone to an enhanced false positive or false negative due to the model could sufficiently control both population structure and kinship among individuals in our data set. Therefore, we ultimately decided to use the MLM (*Q* + *K*) to test our phenotype–genotype associations. Similarly, a combined population structure and kinship approach has been successful in interpreting the association results. For example, Ehrenreich et al. (2009) used the MLM to conduct candidate gene association mapping of flowering time in *Arabidopsis*. Additionally, Zhao et al. (2007) also demonstrated that the MLM model was the most appropriate method for associations between flowering-related phenotypes and genotypes at candidate loci in *Arabidopsis*. Taken together, choosing of the proper statistical models can help to reduce the number of false positive rates and elevated the accuracy of associations.

Detection of significant single SNP and haplotypes associated with growth and wood properties

Populus is a widely used model organism for studies of tree molecular biology and biotechnology, having a rapid growth rate, modest genome size, and extensive polymorphic genetic markers. Poplar not only provides a complete genomics "toolkit" for the comparison of a long-lived perennial to short-lived model plants (e.g., Arabidopsis), but it also offers abundant genetic and adaptive variation for exploring the genetic basis of shoot growth, wood formation, and other important research topics (Bradshaw et al. 2000; Taylor 2002; Brunner et al. 2004). Given this, the identification of genes and allelic variants associated with growth and wood quality is important for practical application in tree breeding programs. The detection of specific marker-trait relationships using association genetics is a widely employed approach, and these

markers can be used for marker-assisted selection (MAS) breeding. In recent years, diverse groups of candidate genes affecting chemical wood properties have been identified in related *Populus* species using SNP-based associations (Tian et al. 2012; Du et al. 2014a).

In our study, we combined single marker association and haplotype-based tests jointly to dissect functional allelic polymorphisms within *PtoGID1a* and *PtoGID1b* for complex growth and wood property traits underlying LD-based association in *P. tomentosa*. Because of the rapid decay of LD in *P. tomentosa* (Fig. 3), once a marker-trait association has been discovered and validated, it is likely that the marker is located in close proximity to the causal polymorphisms or even is the functional variant itself (Neale and Kremer 2011). This rapid decay of LD could prevent researchers from focusing on linked but not relevant markers and missing potentially significant markers or functional variants. Therefore, LD-based methods will be a powerful approach for delineating complex functional genes. In all, we identified six single marker associations and six haplotypes within *PtoGID1a* and *PtoGID1b* that were significantly associated with wood and growth traits (Tables 2 and 3), which demonstrated that *PtoGID1a* and *PtoGID1b* may be involved in stem growth and wood formation. All of the significant SNPs identified by single marker association were synonymous mutations or in noncoding regions. Of these, two synonymous substitutions (SNP52 and SNP55) in exon 2 of *PtoGID1a* were strongly associated with stem volume and α-cellulose content, and the mode of gene action appeared to be over- or under-dominance, except for *PtoGID1a*-SNP55. Although the synonymous mutations did not cause an amino acid change, these silent SNPs can still affect the levels of gene expression and alter codon usage, thus affecting gene function via other mechanisms (Kimchi-Sarfati et al. 2007). A similar phenomenon has been discovered in many previous studies. For instance, a synonymous exonic SNP (SNP10) of *PtGA20ox* was associated with fiber width (Tian et al. 2012) and the same type of SNP (SNP45) in *PtoCesA4* of *P. tomentosa* was significantly associated with holocellulose content (Du et al. 2013).

Many functional analyses of SNPs have emphasized coding regions and splice sites in candidate genes related to phenotypic traits. However, the vast majority (over 90 %) of trait-associated variants have been found in noncoding regulatory regions where they perturb binding sites of transcription factors, local chromatin structure, or co-factor recruitment, ultimately influencing important biological regulation (Thumma et al. 2009; Beaulieu et al. 2011; Maurano et al. 2012). This is especially true for promoters, which are usually assumed to be the key regulatory regions that control gene transcription (Roa-Rodrigues 2003). In the present study, we detected three significant markers located in the promoter region and one located in the 5′ UTR. For example, *PtoGID1a*-SNP12 and *PtoGID1a*-SNP16 in the promoter and 5′ UTR region,

respectively, were associated with microfiber angle and tree height, while *PtoGID1b*-SNP3 and *PtoGID1b*-SNP32 both resided in the promoter region and were associated with stem volume and fiber width (Table 2). These findings were consistent with previous studies that SNP loci in noncoding regions are significantly associated with wood traits. Southerton et al. (2010) detected a strong association between a SNP (HB16 located in the 5′ UTR) and microfiber angle in *Eucalyptus nitens*. Similarly, *PtoSAHHB*-1065, in the first intron of *SAHHB* associated with α-cellulose content, was identified in both discovery and validation populations of *P. tomentosa* (Du et al. 2014a). Marker-trait associations have also been reported for markers in promoter regions. Du et al. (2013) detected a significant marker 273 bp (T/A) upstream of the transcriptional start site of *CesA4* in *P. tomentosa* associated with α-cellulose content. These association results support the prevalent view that more SNPs occur in the promoter region or noncoding region than in the coding region of individual genes (Su et al. 2011).

Generally, the power of a single marker association test is often limited because LD information contained in the flanking markers is ignored, while haplotypes are intuitively regarded as more informative blocks since they preserve the joint LD structure among multiple adjacent markers (Akey et al. 2001). Recently, many association studies have stressed the necessity for a haplotype approach instead of merely relying on individual SNPs. Given the putative power of haplotype to assess the causal SNP loci in complex traits, we also conducted multi-locus haplotype LD tests in this mapping study. Through the comparison of single marker and haplotype-based associations (Tables 2 and 3), we found that the individual SNP loci within these haplotype blocks (SNP1–4, SNP49–51, and SNP61–64) have no significant correlation with growth and wood qualities. The marker trait-associated results of both methods demonstrated that the haplotype-based analysis may not be more powerful than single marker-based analysis in this low LD tree species, which is in accordance with a previous study on UDP-glucuronate decarboxylase (a key enzyme in synthesis of UDP-xylose) during cell wall biosynthesis in *P. tomentosa* by Du et al. (2013). However, contradictory studies suggest that haplotype-based tests can detect previously unknown quantitative traits and are well suited to unraveling the genetic basis of complex traits. Wegrzyn et al. (2010) examined individual genes and SNPs for complex traits controlling lignocellulosic biosynthesis in *P. trichocarpa* and found that the use of haplotypes can significantly increase the ability and robustness of mapping traits of interest. Therefore, we suggest that the best way is to combine single marker-based and haplotype-based methods, which will provide more insights than any one method into the

functional allelic variance underlying quantitative traits in an association population.

Additionally, most of the associations identified in our study explained a small proportion of the phenotypic variance (Tables 3 and S4) and the individual SNP loci within haplotype blocks have no significant correlation with traits, which suggested that each SNP of the haplotype blocks may has a minor effect, and the collective contribution of several minor sites can significantly affect phenotypic variance. These phenomena attributed to individual SNP alleles are consistent with a polygenic quantitative model of wood traits (Beaulieu et al. 2011), in accordance with earlier studies in a same nature population of *P. tomentosa*. For example, Du et al. (2013) and Tian et al. (2014a) identified cellulose synthase gene *PtoCesA4* and *PtoCesA7* allelic variants, respectively, which associated with growth and wood quality traits under the polygenic model, explaining 1.6–5.6 and 3.62–10.59 % of the phenotypic variability. Similarly, several significant associations were found and the phenotypic variances explained by a single SNP were also having a small proportion in *UDP-glucose dehydrogenase* (*PtUGDH*) (Tian et al. 2014b) and *PtGA20Ox* genes (Tian et al. 2012) in *P. tomentosa*. These significant associations were found with both tree growth (e.g., tree height, stem volume) and wood properties trait (e.g., fiber length, fiber width, microfibril angle, holocellulose, lignin, and α-cellulose contents), which we are usually interested in selecting for breeding. Therefore, we used a candidate gene-based association mapping approach to identify genes and allele variants controlling phenotypic variability in a nature population, which could provide more useful information and resources for forest tree breeding program.

Conclusions

In conclusion, we cloned two full-length cDNAs from *P. tomentosa*; these are similar to *OsGID1* and *AtGID1s*. The structural and phylogenetic analysis proved that these two clones, *PtoGID1a* and *PtoGID1b*, encode the likely GA receptors. Tissue-specific expression profiles of *PtoGID1a* and *PtoGID1b* revealed similar patterns, with either high or low expression levels in all tissues and organs, consistent with the participation of GA in growth processes throughout the life cycle of the plant. Subsequently, single marker and haplotype-based association analysis identified six single SNP loci and six haplotype blocks that were significantly associated with growth and wood properties. These findings implied that *PtoGID1a* and *PtoGID1b* may play an important role in tree growth and wood formation by mediating GA signaling. With the rapid development of next-generation sequencing technology and more sophisticated statistical software, there will be more opportunities to identify additional

GID1 genes in *P. tomentosa* because the genome of *Populus* is relatively larger than those of other plants (e.g., *A. thaliana*, which has three GID1-type GA receptors). The present study identified functional allelic variants of *PtoGID1a* and *PtoGID1b* with growth and wood quality traits, given new insights into genes affecting wood quality and quantity. From an applied perspective, the SNPs revealed in our study have potential applications in marker-assisted breeding.

References

Akey J, Jin L, Xiong M (2001) Haplotypes vs single marker linkage disequilibrium tests: what do we gain? Eur J Hum Genet 9:291–300

Aleman L, Kitamura J, Abdel-mageed H et al (2008) Functional analysis of cotton orthologs of GA signal transduction factors GID1 and SLR1. Plant Mol Biol 68:1–16

Barrett LW, Fletcher S, Wilton SD (2012) Regulation of eukaryotic gene expression by the untranslated gene regions and other non-coding elements. Cell Mol Life Sci 69:3613–3634

Beaulieu J, Doerksen T, Boyle B, Clement S, Deslauriers M et al (2011) Association genetics of wood physical traits in the conifer white spruce and relationships with gene expression. Genetics 188:197–214

Bourgain C, Hoffjan S, Nicolae R, Newman D et al (2003) Novel case-control test in a founder population identifies P-selectin as an atopy-susceptibility locus. Am J Hum Genet 73:612–626

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

Bradshaw HD, Ceulemans R, Davis J, Stettler R (2000) Emerging model systems in plant biology: poplar (*Populus*) as a model forest tree. J Plant Growth Regul 19:306–313

Brunner AM, Busov VB, Strauss SH (2004) Poplar genome sequence: functional genomics in an ecologically dominant plant species. Trends Plant Sci 9:49–56

Chandler PM, Marion-Poll A, Ellis M, Gubler F (2002) Mutants at the *Slender1* locus of 'Himalaya' barley: molecular and physiological characterization. Plant Physiol 129:181–190

Chandler PM, Harding CA, Ashton AR, Mulcair MD, Dixon NE, Mander LN (2008) Characterization of gibberellin receptor mutants of barley (*Hordeum vulgare* L.). Mol Plant 1:285–294

Du Q, Wang B, Wei Z, Zhang D, Li B (2012) Genetic diversity and population structure of Chinese white poplar (*Populus tomentosa*) revealed by SSR markers. J Hered 103:853–862

Du Q, Xu B, Pan W, Gong C (2013) Allelic variation in a cellulose synthase gene (*PtoCesA4*) associated with growth and wood properties in *Populus tomentosa*. G3: Genes Genomes Genetics 3:2069–2084

Du Q, Wang L, Zhou D, Yang H, Gong C et al (2014a) Allelic variation within the S-adenosyl-L-homocysteine hydrolase gene family is associated with wood properties in Chinese white poplar (*Populus tomentosa*). BMC Genet 15(Suppl 1):S4

Du Q, Xu B, Gong C, Yang X, Pan W, Tian J, Li B, Zhang D (2014b) Variation in growth, leaf and wood-property traits of Chinese white poplar (*Populus tomentosa* Carr.), a major industrial tree species in Northern China. Can J For Res 44:326–339

Eckert AJ, Bower AD, Wegrzyn JL, Pande B, Jermstad KD et al (2009) Association genetics of coastal Douglas fir (*Pseudotsuga menziesii var. menziesii*, Pinaceae). I. Cold hardiness related traits. Genetics 182:1289–1302

Ehrenreich IM, Hanzawa Y, Chou L, Roe JL, Kover PX et al (2009) Candidate gene association mapping of *Arabidopsis* flowering time. Genetics 183:325–335

Evanno G, Regnaut S, Goudet J (2005) Detecting the number of clusters of individuals using the software STRUCTURE: a simulation study. Mol Ecol 14:2611–2620

Griffiths J, Murase K, Rieu I et al (2006) Genetic characterization and functional analysis of the GID1 gibberellin receptors in *Arabidopsis*. Plant Cell 18:3399–3414

Hardy OJ, Vekemans X (2002) SPAGeDi: a versatile computer program to analyse spatial genetic structure at the individual or population levels. Mol Ecol Notes 2:618–620

Hill W-G, Robertson A (1968) Linkage disequilibrium in finite population. Theor Appl Genet 38:226–231

Hill WG, Weir BS (1988) Variances and covariances of squared linkage disequilibria in finite populations. Theor Popul Biol 33:54–78

Hirano K, Nakajima M, Asano K et al (2007) The *GID1*-mediated gibberellin perception mechanism is conserved in the lycophyte *Selaginella moellendorffii* but not in the bryophyte *Physcomitrella patens*. Plant Cell 19:3058–3079

Hirano K, Ueguchi-Tanaka M, Matsuoka M (2008) *GID1*-mediated gibberellin signaling in plants. Plant Cell 13:192–199

Huang Z (1992) The study on the climatic regionalization of the distributional region of *Populus tomentosa*. J Beijing For Univ 14:26–32

Ikeda A, Ueguchi-Tanaka M, Sonoda Y, Kitano H, Koshioka M, Futsuhara Y, Matsuoka M, Yamaguchi J (2001) Slender rice, a constitutive gibberellin response mutant, is caused by a null mutation of the *SLR1* gene, an ortholog of the height regulating gene *GAI/RGA/RHT/D8*. Plant Cell 13:999–1010

Israelsson M, Sundberg B, Moritz T (2005) Tissue-specific localization of gibberellins and expression of gibberellin-biosynthetic and signaling genes in wood forming tissues in aspen. Plant J 44:494–504

Iuchi S, Suzuki H, Kim Y-C, Iuchi A et al (2007) Multiple loss-of-function of Arabidopsis gibberellin receptor *AtGID1s* completely shuts down a gibberellin signal. Plant J 50:958–966

Kimchi-Sarfati C, Oh JM, Kim IW, Sauna ZE, Calcagno AM et al (2007) A "silent" polymorphism in the *MDR1* gene changes substrate specificity. Science 315:525–528

Li A, Yang W, Li S et al (2013) Molecular characterization of three *GIBBERELLIN-INSENSITIVE DWARF1* homologous genes in hexaploid wheat. J Plant Physiol 170:432–443

Librado P, Rozas J (2009) DnaSP v5: a software for comprehensive analysis of DNA polymorphism data. Bioinformatics 25:1452–1452

Maurano M, Humbert R, Rynes E, Thurman R, Haugen E et al (2012) Systematic localization of common disease-associated variation in regulatory DNA. Science 337:1190–1195

Mauriat M, Moritz T (2009) Analyses of *GA20ox*- and *GID1*-over-expressing aspen suggest that gibberellins play two distinct roles in wood formation. Plant J 58:989–1003

Murase K, Hirano Y, Sun T, Hakoshima T (2008) Gibberellin-induced DELLA recognition by the gibberellin receptor GID1. Nature 456:459–464

Nakajima M, Shimada A, Takashi Y et al (2006) Identification and characterization of Arabidopsis gibberellin receptors. Plant J 46:880–889

Neale DB, Kremer A (2011) Forest tree genomics: growing resources and applications. Nat Rev Genet 12:111–122

Nei M (1987) Molecular evolutionary genetics. Columbia University Press, New York

Olszewski N, Sun T, Gubler F (2002) Gibberellin signaling biosynthesis, catabolism, and response pathways. Plant Cell 14:S61–S80

Peng J, Carol P, Richards DE, King KE, Cowling RJ, Murphy GP, Harberd NP (1997) The *Arabidopsis GAI* gene defines a signaling pathway that negatively. Gene Dev 11:3194–3205

Porth I, Klápště J, Skyba O, Lai BS, Geraldes A et al (2013) *Populus trichocarpa* cell wall chemistry and ultrastructure trait variation, genetic control, and genetic correlations. New Phytol 197:777–790

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

Ritland K (1996) Estimators for pairwise relatedness and individual inbreeding coefficients. Genet Res 67:175–185

Roa-Rodrigues C (2003) Promoters used to regulate gene expression. CAMBIA Intellectual Property, Canberra

Schimleck LR, Kube PS, Raymond CA (2004) Genetic improvement of Kraft pulp yield in *Eucalyptus nitens* using cellulose content determined by near infrared spectroscopy. Can J For Res 34:2363–2370

Shimada A, Ueguchi-Tanaka M, Nakatsu T, Nakajima M (2008) Structural basis for gibberellin recognition by its receptor GID1. Nature 456:520–523

Southerton SG, MacMillan CP, Bell JC, Bhuiyan N, Downes G et al (2010) Association of allelic variation in xylem genes with wood properties in *Eucalyptus nitens*. Austral For 73:259–264

Storey JD, Tibshirani R (2003) Statistical significance for genome wide studies. Proc Natl Acad Sci U S A 100:9440–9445

Storey JD, Taylor JE, Siegmund D (2004) Strong control, conservative point estimation and simultaneous conservative consistency of false discovery rates: a unified approach. J R Stat Soc Ser B 66:187–205

Su Z, Hao C, Wang L, Dong Y, Zhang X (2011) Identification and development of a functional marker of *TaGW2* associated with grain weight in bread wheat (*Triticum aestivum* L.). Theor Appl Genet 122:211–223

Sun T (2011) The molecular mechanism and evolution of the GA–GID1–DELLA signaling module in plants. Curr Biol 21:R338–R345

Sun T, Gubler F (2004) Molecular mechanism of gibberellins signaling in plants. Annu Rev Plant Biol 55:197–223

Tamura K, Dudley J, Nei M, Kumar S (2007) MEGA4: molecular evolutionary genetics analysis (MEGA) software version 4.0. Mol Biol Evol 24:1596–1599

Taylor G (2002) *Populus*: Arabidopsis for forestry. Do we need a model tree? Ann Bot 90:681–689

Thumma BR, Matheson BA, Zhang D, Meeske C, Meder R et al (2009) Identification of a *cis*-acting regulatory polymorphism in a Eucalypt *cobra*-like gene affecting cellulose content. Genetics 183:1153–1164

Tian J, Du Q, Chang M, Zhang D (2012) Allelic variation in *PtGA20Ox* associates with growth and wood properties in *Populus* spp. PLoS One 7(12):e53116

Tian J, Chang M, Du Q, Xu B, Zhang D (2014a) Single-nucleotide polymorphisms in *PtoCesA7* and their association with growth and wood properties in *Populus tomentosa*. Mol Gen Genomics 289:439–455

Tian J, Du Q, Xu B, Li B, Zhang D (2014b) Single-nucleotide polymorphisms in the 5′ UTR of *UDP-glucose dehydrogenase* (*PtUGDH*) associate with wood properties in *Populus tomentosa*. Tree Genet Genomes 10:339–354

Ueguchi-Tanaka M, Ashikari M, Nakajima M et al (2005) *GIBBERELLIN INSENSITIVE* DWARF1 encodes a soluble receptor for gibberellins. Nature 437:693–698

Ueguchi-Tanaka M, Nakajima M, Katoh E et al (2007a) Molecular interactions of a soluble gibberellin receptor, GID1, with a rice DELLA protein, SLR1, and gibberellin. Plant Cell 19:2140–2155

Ueguchi-Tanaka M, Nakajima M, Motoyuki A, Matsuoka M (2007b) Gibberellin receptor and its role in gibberellin signaling in plants. Annu Rev Plant Biol 58:183–198

Voegele A, Linkies A, Muller K, Leubner-Metzger G (2011) Members of the gibberellin receptor gene family GID1(*GIBBERELLIN INSENSITIVE DWARF1*) play distinct roles during *Lepidium sativum* and *Arabidopsis thaliana* seed germination. J Exp Bot 62:5131–5147

Watterson GA (1975) On the number of segregating sites in genetical models without recombination. Theor Popul Biol 7:188–193

Wegrzyn JL, Eckert AJ, Choi M, Lee JM, Stanton BJ et al (2010) Association genetics of traits controlling lignin and cellulose biosynthesis in black cottonwood (*Populus trichocarpa*, Salicaceae) secondary xylem. New Phytol 188:515–532

Yang XH, Yan JB, Shah T, Warburton ML, Li Q et al (2010) association mapping panel for quantitative trait loci dissection. Theor Appl Genet 121:417–431

Yu J, Pressoir G, Briggs WH, Vroh Bi I, Yamasaki M et al (2006) A unified mixed-model method for association mapping that accounts for multiple levels of relatedness. Nat Genet 38:203–208

Zaykin DV, Westfall PH, Young SS, Karnoub MA, Wagner MJ, Ehm MG (2002) Testing association of statistically inferred haplotypes with discrete and continuous traits in samples of unrelated individuals. Hum Hered 53:79–91

Zhang D, Zhang Z (2005) Single nucleotide polymorphisms discovery and linkage disequilibrium. For Studies China 7:1–14

Zhang DQ, Zhang ZY, Yang K (2006) QTL analysis of growth and wood chemical content traits in an interspecific backcross family of white poplar (*Populus tomentosa* × *P. bolleana*) × *P. tomentosa*. Can J For Res 36:2015–2023

Zhang D, Du Q, Xu B, Zhang Z, Li B (2010) The actin multigene family in *Populus*: organization, expression and phylogenetic analysis. Mol Gen Genomics 284:105–119

Zhao K, Aranzana MJ, Kim S, Lister C et al (2007) An *Arabidopsis* example of association mapping in structured samples. PLoS Genet 3:e4

Zhu Z (1992) Collection, conservation and utilization of plus tree resources of *Populus tomentosa* in China. J Beijing For Univ 14:1–25