



Qingsong Bai*, Qian Zhang, Yanling Cai, Huiming Lian,
Wanchuan Liu, Min Luo, Linghai Zeng*, Boxiang He*

Genome-wide association study of terpenoids in resin reveals candidate genes for resin yield in *Pinus massoniana*

Received: 22 July 2020; Accepted: 16 December 2020

Abstract: Masson pine (*Pinus massoniana*) is an important species for resin tapping in China. Molecular assisted selection (MAS) has been gradually introduced in to the selection and breeding of masson pine. Until now, the lack of connection between genome information and phenotypic data has hindered the breeding and selection of important chemical compounds of xylem resin.

In this study, gas chromatography mass spectrometer (GC-MS) was employed for qualitative and quantitative analysis of chemical compounds of resin collected from 69 masson pine germplasm resources. Then, genome-wide association study (GWAS) was conducted for the terpenoids with higher heritability.

A total of 15 primary compounds, including six diterpenes and nine monoterpenes/sesquiterpenes, were detected from xylem resin. By using five GWAS methods, 42 associated SNPs were identified from these masson pines. The *P. taeda* genome database was used for detecting candidate genes related to the associated SNPs. By using transcriptome data, several members of PKc_like super family, Cytochrome P450, AP2-ERF, F-box, and ABC transporter were found to be highly expressed in the masson pines with higher resin-yielding capacity.

The terpenoids associated SNPs can be directly used for MAS. These candidate genes, especially the two F-box members, will contribute to the study of terpenoid biosynthesis mechanisms in the future.

Keywords: *Pinus massoniana*, GC-MS, terpenoid, GWAS, candidate gene

Addresses: Q. Bai, Q. Zhang, Y. Cai, H. Lian, L. Zeng, B. He, Guangdong Provincial Key Laboratory of Silviculture, Protection and Utilization, Guangdong Academy of Forestry, Guangzhou, 510520, China, e-mail: baiqingsong1707@163.com, boxianghe@hotmail.com, zeng_linghai@163.com

W. Liu, M. Luo, Xinyi Forestry Research Institute, Maoming, 525300, China

*Corresponding authors

Introduction

Oleoresin is a natural product that is stored in various plant tissues or organs, including stems, roots, needles, and cones, in conifer species (Trapp & Croteau, 2001). Resin is naturally formed in response to insect and fungal pathogen invasion (Karanikas et al., 2010). The resin of coniferous trees is composed

of turpentine (a mixture of monoterpenes and sesquiterpenes) and rosin (diterpene resin acids). In addition to the role in the natural defending mechanism against pathogen invasion, terpenoids from resin are a renewable resources for industrial bioproducts, including solvents, flavors, coatings, adhesives, and inks (Bohlmann, 2011; Zerbe et al., 2012; Hall et al., 2013).

In China, masson pine (*Pinus massoniana*) is the most economically significant native resin producer (Wang et al., 2006). This species is distributed in dozens of provinces in southern China, which give rise to the high genetic diversity (Peng et al., 2003; Zhang et al., 2010). Some studies reported the qualitative and quantitative compositions of oleoresin in southern China, the highest concentration among the compounds in turpentine was α -pinene; palustric acid was the most concentrated compound in resin acid (An & Ding, 2012; Liu et al., 2016). Among the numerous kinds of terpenoids, several compositions, including pimaric acid, neoabietic acid, camphene, dipentene, and caryophyllene, were reported with high heritability (Liu et al., 2016). Moreover, the contents of chemical compounds showed diverse distributions among lines and families in the same tree species (Sjodin et al., 2000; Rezzi et al., 2005).

Due to the genetic stability of the chemical components of rosin, some terpenoids can also be used as phytochemical markers (Manninen et al., 2002). The terpenoids usually have different functions in various aspects, for example, α -pinene can be used as an insecticide; β -pinene can be used as an alternative biofuel as compared to limited natural resources (Peralta-Yahya et al., 2012). The quality of oleoresin is depended on the contents of different chemical compounds. Therefore, when certain components are required, the families which are abundant in those components should be selected and bred. Genetics is used to dissect phenotypes into their genetic components (Stinchcombe & Hoekstra, 2008; Eckert et al., 2012). And genetic studies have been conducted in many forestry species (Neale, 2007; Neale & Ingvarsson, 2008; Neale & Kremer, 2011). Genome-wide association studies (GWAS) can provide favorable DNA loci for different phenotypes, especially in those complex quantitative traits (Bartholome et al., 2017). Association analysis was conducted in loblolly pine to identify the favorable SNPs with metabolites (Eckert et al., 2012). Terpenoids also belong to the secondary metabolite. Therefore, GWAS of chemical compounds may also be feasible to provide us useful knowledge for masson pine.

Several reports on masson pine have focused on genetic relationships using random amplified polymorphic DNA (RAPD), sequence-related amplified polymorphism (SRAP), simple sequence repeat (SSR), and inter simple sequence repeat (ISSR) DNA markers; however, these were limited in scale (Peng et al., 2003; Zhang et al., 2010; Zhang et al., 2013; Bai et al., 2014; Chen et al., 2014; Feng et al., 2014). Conifer trees usually have large-scale genomes, whole-genome sequences are only available for a few conifer species (Neale et al., 2014). Next-generation sequencing (NGS) and the rapid advancement of DNA sequencing platforms have enabled the development

of large-scale DNA markers in large populations of nonmodel species, which typically have very large genomes (Andolfatto et al., 2011; Pavy et al., 2017; Liu et al., 2019). Specific-locus amplified fragment sequencing (SLAF-seq) is one of the advanced technologies that has increased the capacity for SNP discovery and genotyping at a large scale (Sun et al., 2013). By using SLAF-seq, we developed a large number of SNP markers and understood the population structure and genetic relationships of the conserved masson pine germplasm resources (Bai et al., 2019).

Until now, the lack of relationship between genetic information and chemical components lags behind the development of the resin industry. Germplasm resources with abundant genetic diversity are essential to the breeding work. Hence, the discovery of favorable molecular markers with chemical components is necessary to the marker-assisted selection (MAS) and collection of germplasm resources. In this study, qualitative and quantitative analyses of the chemical compounds in the xylem resin of 69 masson pines were carried out using gas chromatography mass spectrometry (GC-MS). The SNPs developed in our previous study were used for the favorable loci discovery of chemical compounds with high heritability using various GWAS methods (Bai et al., 2019). Several candidate genes were detected relating to the content of those chemical compounds. These favorable SNPs would benefit the selection and improvement of specific chemical compounds in masson pine. In addition, the candidate genes would provide us knowledge to better understand the mechanism of terpenoid biosynthesis.

Materials and Methods

Resin collection and chemical component analysis

A total of 69 masson pine germplasm resources were used for the collection of xylem resin (Table S1). Resin tapping was carried out using the narrow face system as described by Coppen and Hone (Coppen & Hone, 1995). The resin from each accession was collected by removing a sliver of wood from the stem without the application of chemical stimulants. GC-MS analysis was used to perform qualitative and quantitative analyses of the terpenoids in the collected resin according to the method described for *P. halepensis* (Karanikas et al., 2010). Turpentine and rosin were dissolved in methanol. Isobutylbenzene and heptadecanoic acid were used as the internal standards to quantify the contents of terpenoids in turpentine and rosin, respectively. For the process of gas chromatography (GC), the program of oven temperature was 60 °C for 2 min, increasing 2 °C/

Table 1. Summary statistics of different chemical compounds concentration in xylem resin of *P. massoniana*

Category	Compounds	Mean	Minimum	Maximum	SD	CV(%)	h^2
Oil percentage	–	20.08	8.6	30.45	4.33	21.56	unknown
Diterpene	Pimaric acid	8.81	3.3	11.4	1.46	16.56	0.47
	Sandaracopimaric acid	2.01	1.2	3.4	0.30	14.95	0.18
	Palustric acid	39.09	20.55	49.1	4.64	11.87	0.03
	Dehydroabietic acid	4.98	1.1	11.7	1.99	39.98	0.18
	Abietic acid	26.74	19.6	35.9	2.79	10.44	0.18
	Neoabietic acid	17.72	13.2	22.4	1.90	10.74	0.53
Monoterpene	α -pinene	74.97	54	86.6	6.99	9.33	0.18
	Camphene	1.91	0.9	3.9	0.52	27.34	0.59
	β -pinene	7.88	1.85	22.95	5.01	63.54	0.18
	Geraniolene	1.11	0.3	2.8	0.38	34.24	0.06
	Dipentene	1.85	0.8	3.55	0.55	29.55	0.47
Sesquiterpene	α -cedrene	0.64	0.2	1.65	0.32	49.56	unknown
	α -santalene	0.54	0.2	1.6	0.24	44.16	unknown
	Longifolene	6.67	1.8	20.3	3.11	46.67	0.24
	Caryophyllene	1.92	0.3	4.9	0.91	47.70	0.5

SD – standard deviation; CV – coefficient of variance; h^2 – broad-sense heritability.

min to 80 °C, subsequent increase rate of 4 °C/min until 280 °C. Helium was used as the carrier gas at a 1 mL/min flowing rate, the injector temperature was set up at 260 °C. The injection volume was 1 μ L with a split ratio of 1:50. Electron ionization mass spectrometry (EI-MS): the electron energy was 70 eV. Ion source was 230 °C, and the connection part temperature was 280 °C. Chemical components were identified by comparing the experimental retention indices (RI) with those of reference compounds run under identical conditions and by matching the fragmentation of mass spectra with the libraries NIST98, Wiley 275, and CNRS. The quantitative measurements were carried out under the same working conditions by connecting to a FID detector. Percentage of the total amount of components was served as the content of each chemical component. Pearson's correlation coefficients (r^2) were calculated to identify relationships among the different terpenoids. Statistical analysis was carried out using SPSS 13.0 (SPSS Inc., Chicago, USA).

Genome-wide association study

SLAF-seq technology was used to develop SNPs, and a total of 476,264 SNPs with MAF \geq 0.05 were identified from 69 masson pine accessions (Bai et al., 2019). A phylogenetic tree was constructed based on Nei's distance calculated by all the developed SNPs using neighbor-joining method in MEGA 6.0 (Tamura et al., 2013). Principal Component Analysis (PCA) was conducted for all the samples using Cluster software (de Hoon et al., 2004). The GWAS was performed using five methods, including EMMAX (Zhou & Stephens, 2012), FaST-LMM (Lippert et al., 2011), mrMLM (Wang et al., 2016), FASTmrEMMA (Wen et al., 2018), and ISIS EM-BLASSO (Tamba et

al., 2017). For the methods EMMAX and FaST-LMM, markers that met the criterion of $-\text{Log}_{10}(\text{P-value}) \geq 6.68$ were considered suggestively associated. Methods mrMLM, FASTmrEMMA, and ISIS EM-BLASSO were included in the mrMLM package (mrMLM.GUI v3.2, <https://cran.r-project.org/web/packages/mrMLM.GUI/index.html>), markers that met the criterion of LOD score ≥ 3 were considered significantly associated with the traits. Default values were used for all parameters.

Candidate gene identification and transcriptomic analysis

The SNPs located in SLAFs were used as queries, and a BLAST strategy was used to identify the genome region with the highest similarity to the *P. taeda* genome database (v1.0, <http://congenie.org/>) (Neale et al., 2014). Then, the genome region of 500 kb surrounding each SNP was used to search for the putative genes. The annotated genes that were near to the favorable SNPs were identified as candidate genes (McLaren et al., 2010). Gene annotation information from *P. taeda* genome database was also used to assign functions to the candidate genes.

Transcriptomic analysis was carried out in masson pine secondary xylem. The phenotype data of trait resin-tapping capacity (RYC) was calculated by formula described in previous study (Zeng et al., 2013; Bai et al., 2020). The masson pines carrying high (H), medium (M), and low (L) RYCs, three lines for each, were selected for RNA-seq. RNAprep Pure Plant Kit (Tiangen, China) was used to isolate the total RNA of each sample. After the assessment of RNA purity, concentration, and integrity, a total amount of 1.0 μ g RNA per sample was used to

prepare sequencing libraries by using the NEBNext® Ultra™ RNA Library Prep Kit for Illumina® (NEB, USA). Then, the library preparations were sequenced on an Illumina HiSeq 2500 platform. After removing the low quality reads, high quality clean reads were used for RNA *de novo* assembly. Five databases, including the Non-redundant (NR), Nucleotide (NT), EuKaryotic Orthologous Groups (KOG), PFAM, and Gene Ontology (GO), were used to annotate the function of transcripts using BLAST. The fragments per kilobase of exon model per million mapped reads (FPKM) values were used to identify the differentially expressed genes (DEGs). Differential expression analysis of the two groups was performed using the DESeq R package (1.10.1) (Wang et al., 2010).

Validation of expression levels of the putative genes

The method quantitative real-time PCR (qRT-PCR) was carried out to validate the expression levels of the candidate genes in the same materials used for transcriptomic analysis. Total RNAs were isolated from four tissues, including secondary xylem, mature needle leaves, one-year-old branches, and two-year-old branches of masson pine with high, medium, and low RYCs using an RNAPrep Pure Plant Kit (TIANGEN, China). Primer 3 (version 4.1.0) was used to design the specific primers (Untergasser et al., 2012). *Actin* was used as reference gene to normalize the relative expression profiles (Liu et al., 2015; Liu et al., 2017; Bai et al., 2020). The primers of all genes were listed in Table S2. The amplification reactions

performed for qRT-PCR were as below: 95 °C for 10 s, 57–60 °C for 30 s, and 72 °C for 20 s. The experiments were repeated three times with the cDNA samples. The expression levels of candidate genes are calculated using $2^{-\Delta\Delta C_t}$ method compared with L/H (Livak & Schmittgen, 2001). Fold changes were transformed and displayed by Tree View (Page, 1996).

Gene full-length cloning and phylogenetic analysis

The cDNA samples were used to clone the open reading frame (ORF) of selected candidate genes. Primer 3 (version 4.1.0) was used to design the specific primers for gene clones (Untergasser et al., 2012). The primers of all genes were listed in Table S2. The Mega 6.0 software was used to conduct the phylogenetic analysis with proteins from other species (Tamura et al., 2013).

Results

Genetic relationship analysis of masson pine individuals

A phylogenetic tree of those masson pines was constructed using the neighbor-joining method (Fig. 1A). As observed in our previous study, the masson pines did not group together based on their geographical distribution. PCA also could not clearly divide the population (Fig. 1B), and all masson pine accessions clustered into one major group, except for

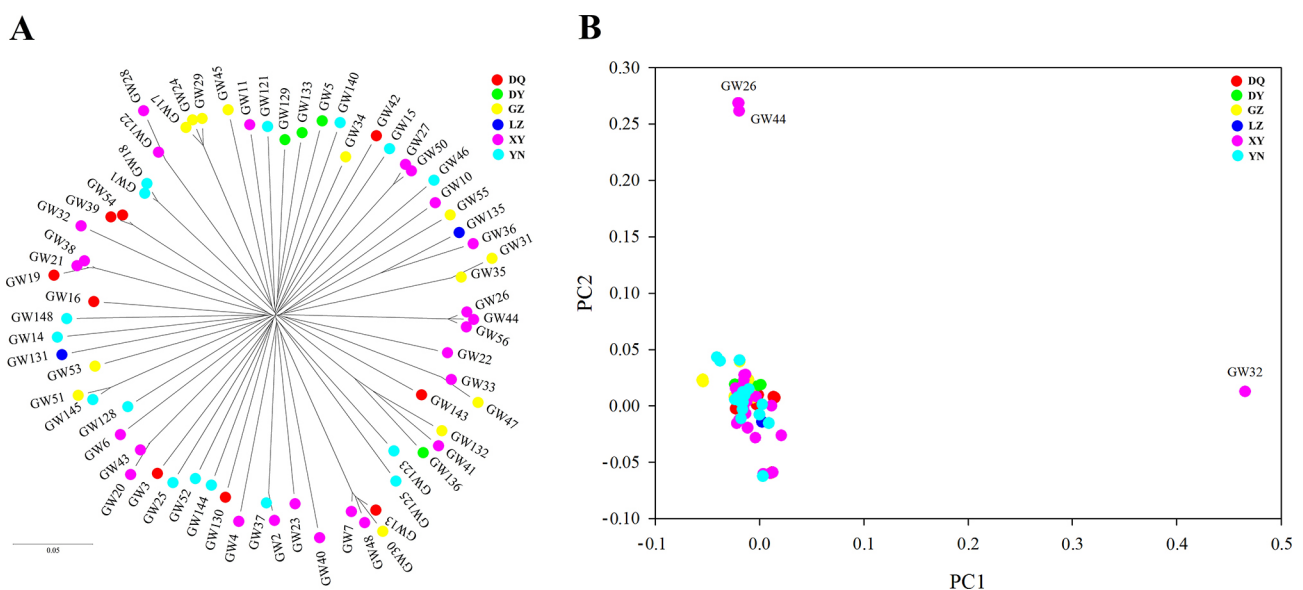


Fig. 1. Phylogenetic tree and PCA of 69 masson pines. (A) The phylogenetic tree was constructed with 1,000 bootstrap replications using Neighbor-joining method in MEGA 6.0 software. (B) Principal component analysis (PCA) of 69 masson pines. The two-dimensional PCA was built using the first principal component and the second principal component

three accessions (GW32, GW26, and GW44) from XY that were separated from the main group. The average genetic distance of these masson pines was 0.230, ranging from 0.009 to 0.289. GW32 had a high average genetic distance (0.249) in comparison to the other masson pines. The results indicated that there were no distinct genetic differentiations among these masson pines. A total of 476,264 SNPs with MAF ≥ 0.05 were used for GWAS analysis.

Qualitative and quantitative analysis of chemical compounds

The turpentine oil accounted for an average number of 20.08% of the total resin (Table 1). GC-MS was used for the chemical composition analysis of turpentine oil and rosin. A total of 15 primary chemical compounds were detected from masson pine resin, including six diterpenes in rosin; five monoterpenes and four sesquiterpenes in turpentine oil. For the quantitative analysis of compounds in rosin, the most concentrated component in rosin was palustric acid, which accounted for approximately 39.09% of the total content. For the quantitative analysis of turpentine oil, monoterpenes and sesquiterpenes accounted for 87.73% and 9.71% of the total resin. The most abundant constituent in the turpentine oil was α -pinene, which accounted for 74.97% of the total resin. Correlations between these traits were also identified, and several chemical components obtained correlation values greater than 0.5. For instance, significant positive correlations were detected in the chemical pairs of longifolene and α -santalene (0.799), α -santalene and α -cedrene (0.660), longifolene and

α -cedrene (0.516), and geraniolene and camphene (0.510), while significant negative correlations existed in the chemical pairs of abietic acid and palustric acid (-0.689), β -pinene and α -pinene (-0.574), and longifolene and α -pinene (-0.566).

All chemical compound contents showed approximately normal distribution, the distributions of pimaric acid, neoabietic acid, camphene, and dipentene were shown in Figure 2. Phenotypic evaluation revealed a broad range in chemical variation among the 69 masson pines. The coefficients of variation for the 15 chemical compounds ranged from 9.33~63.54% (Table 1). Among the 15 chemical compounds, only pimaric acid, neoabietic acid, camphene, dipentene, and caryophyllene have been reported with high heritability (Liu et al., 2016). Therefore, these compounds will be used for genome-wide associated SNP marker detection.

Association studies of five chemical compounds in masson pine resin

Five chemical compounds, pimaric acid, neoabietic acid, camphene, dipentene, and caryophyllene, were used to conduct GWAS. As a result, a total of 11, 7, 8, and 11 quantitative trait nucleotides (QTNs) across the masson pine genome were significantly associated with camphene, dipentene, neoabietic acid, and pimaric acid at the critical LOD ≥ 3 using multilocus GWAS methods (Table S3). In addition, five SNP markers were suggestively associated with four chemical compounds, including camphene, caryophyllene, dipentene, and pimaric acid (Table S4). The statistical analysis showed that several SNP markers were detected

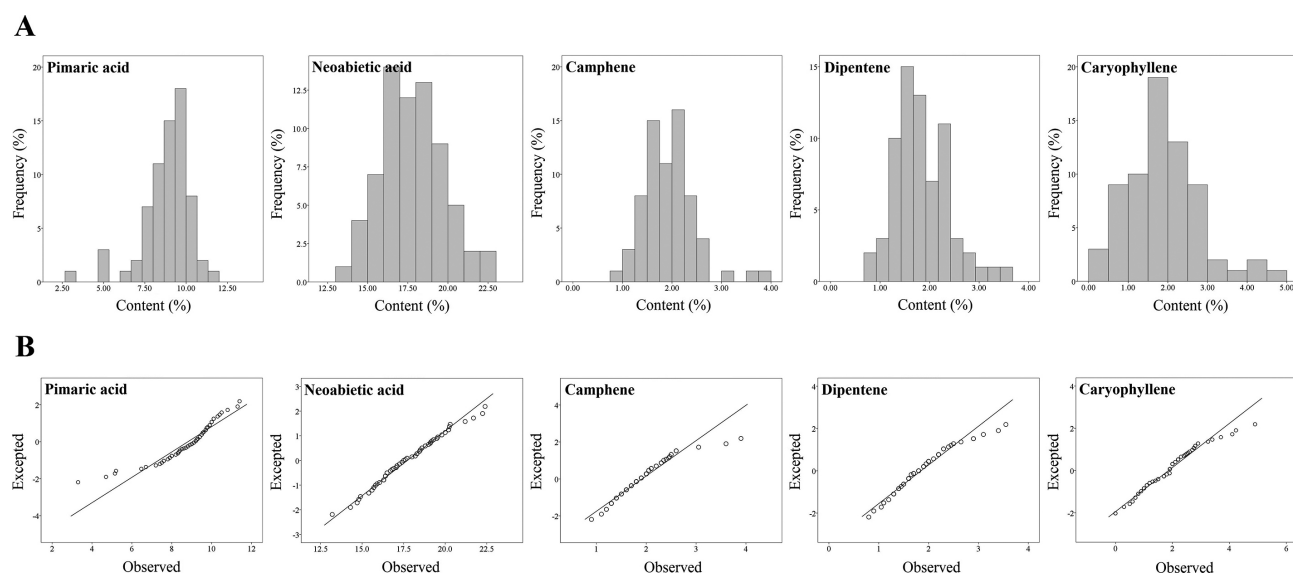


Fig. 2. Phenotypic analyses of five chemical compounds in masson pine xylem resin. (A) Frequency distribution of the content of pimaric acid, neoabietic acid, camphene, dipentene, and caryophyllene among 69 masson pine lines. (B) Quantile-quantile (Q-Q) plot of pimaric acid, neoabietic acid, camphene, dipentene, and caryophyllene

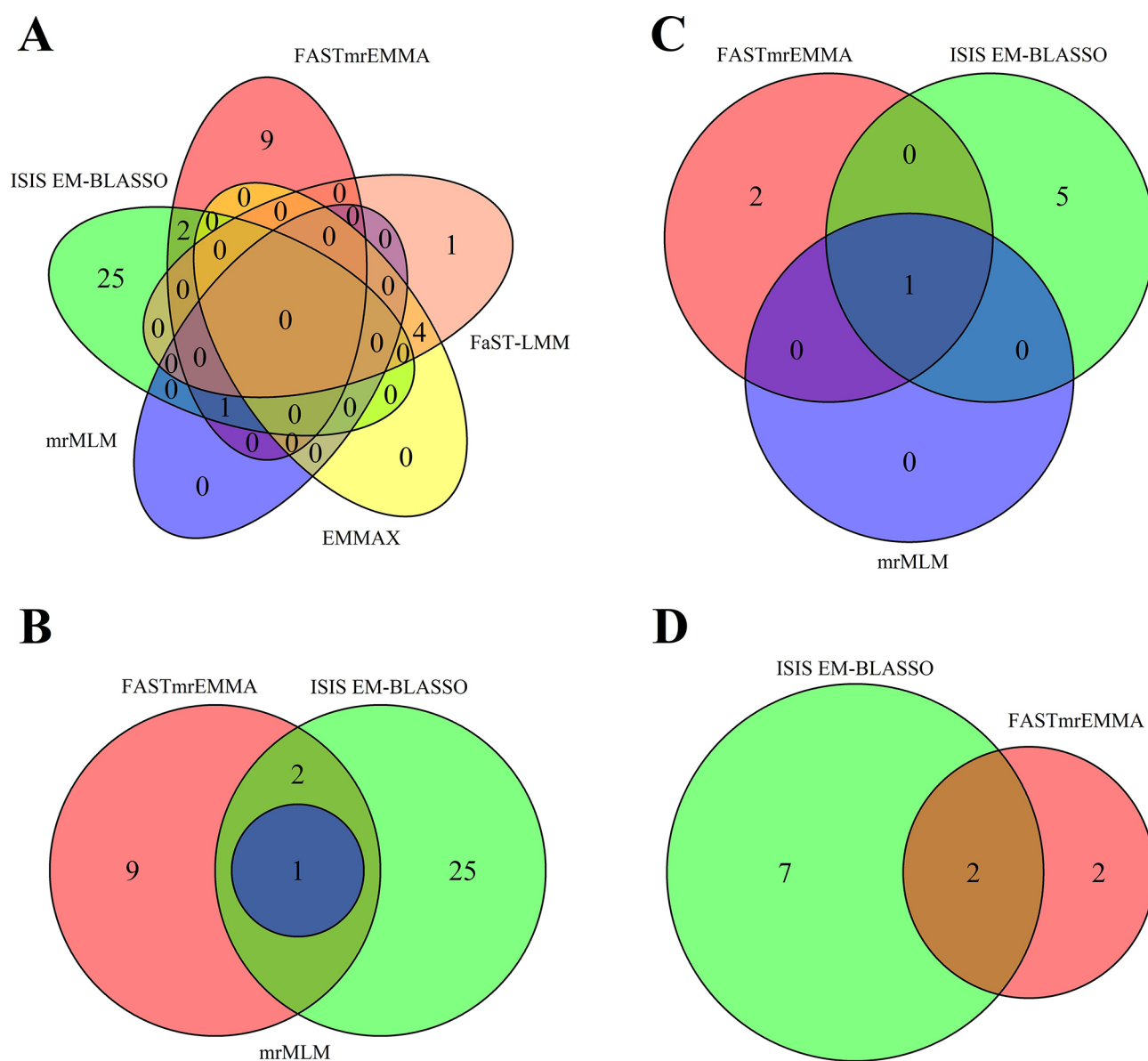


Fig. 3. Comparative analysis of five compound associated SNPs among five different GWAS methods. (A) Venn diagram of the number of five compounds associated SNPs among mrMLM, FASTmrEMMA, ISIS EM-BLASSO, EMMAX, and FaST-LMM. (B) Venn diagram of the number of the five compounds associated SNPs among mrMLM, FASTmrEMMA, and ISIS EM-BLASSO. (C) Venn diagram of the pimelic acid associated SNPs between FASTmrEMMA and ISIS EM-BLASSO. (D) Venn diagram of the neobietic acid associated SNPs among mrMLM, FASTmrEMMA, and ISIS EM-BLASSO

by multiple GWAS methods, and Venn diagrams were used to show the relationships among these methods (Fig. 3A). Among the associated SNPs detected by mrMLM, FASTmrEMMA, and ISIS EM-BLASSO, one SNP were detected by three methods and two SNPs can be detected by methods FASTmrEMMA and ISIS EM-BLASSO (Fig. 3B). The neobietic acid associated QTN Marker966639 was detected by mrMLM, FASTmrEMMA, and ISIS EM-BLASSO (Fig. 3C). The chemical compound pimelic acid associated QTNs Marker430961 and Marker612397 were detected by methods FASTmrEMMA and ISIS EM-BLASSO (Fig. 3D). With respect to the five suggestive SNP markers detected by methods EMMAX and FaST-LMM,

Marker312204, Marker399660, Marker835695, and Marker373310 were detected by two methods (Fig. S1). All associated SNP markers detected by EMMAX were also detected by FaST-LMM.

Identification and expression analysis of candidate genes

In this study, all associated SNP markers were used to identify putative functional genes. Totally, several categories of genes were identified from the SNP markers. The most enriched genes were used for statistical analysis. The result showed that the putative

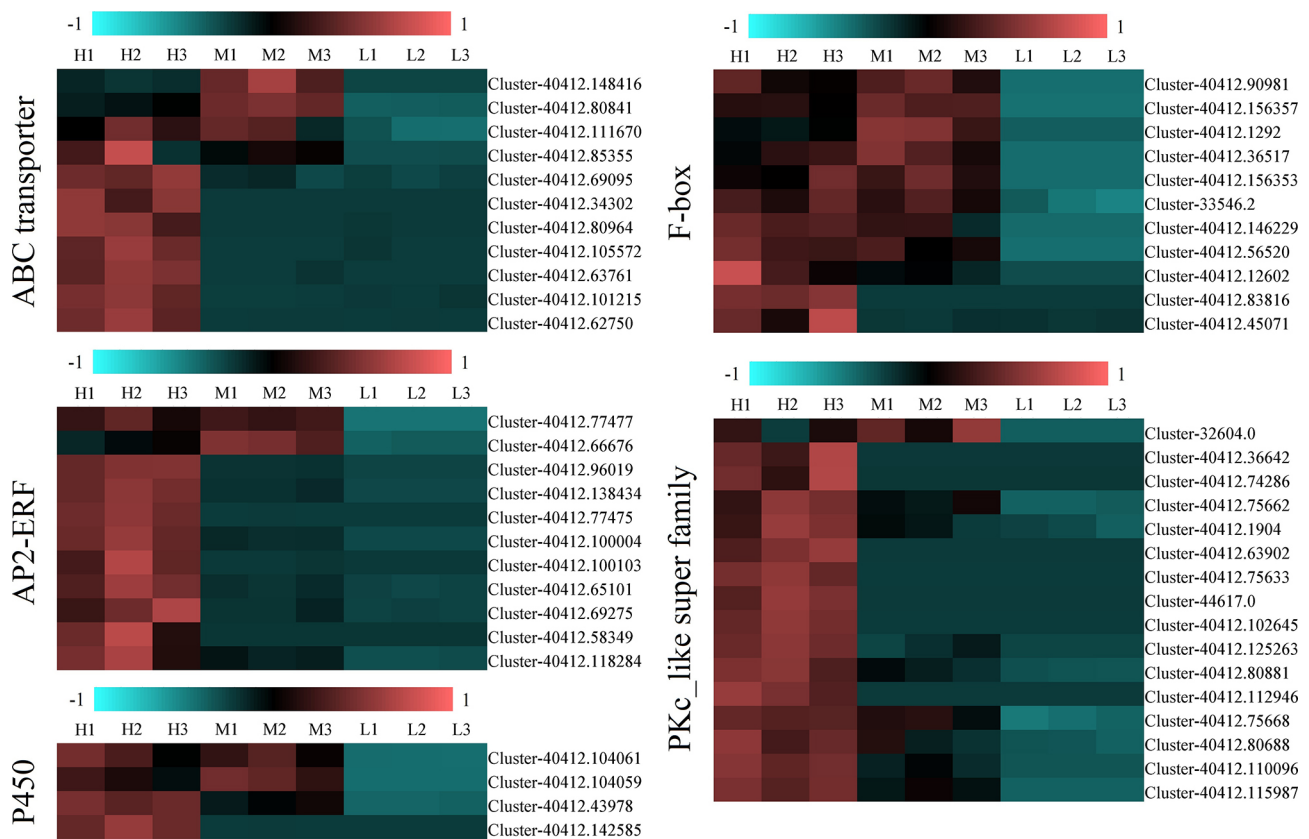


Fig. 4. Heatmap of the upregulated genes using transcriptome FPKM values. Relative expression analysis was conducted in five categories of genes, including ABC-transporter, APS-ERF, P450, F-box, and PKc_like super family. Relative expression levels among samples H, M, and L were conducted using FPKM from transcriptome sequencing. The FPKM values of these genes were normalized between 1 and -1

genes can be classified into five groups, which included PKc_like super family, Cytochrome P450, AP2-ERF, F-box, and ABC transporter. After gene expression analysis, the expression levels of several genes in the high (H) resin-yielding masson pines were significantly higher than those in low (L) resin-yielding masson pines. After searching for this type of genes, a total of 14, 4, 11, 11, and 11 members were identified from PKc_like super family, Cytochrome P450, AP2-ERF, F-box, and ABC transporter, respectively (Fig. 4).

Expression analyses, gene clone and phylogenetic analysis of candidate genes

Ten candidate genes were randomly selected from ABC-transporter, AP2-ERF, P450, F-box, and PKc_like super families. The transcription levels of these genes were assessed by qRT-PCR using samples from the secondary xylem of masson pines with different RYCs (Fig. 5A). All putative genes were differentially expressed among these samples with statistical significance from $P < 0.05$ to $P < 0.0000001$ and displayed similar expression patterns with FPKM values. All genes from five families were highly expressed in the samples H / M.

The full-length ORF sequences of Cluster-40412.12602 (*PmFBX1*) and Cluster-40412.56520 (*PmFBX2*) were cloned using the cDNA samples. Interestingly, the ORF lengths of *PmFBX1* and *PmFBX2* were both 1,086 bp. Except for the secondary xylem, needle leaves, one-year-old branches, and two-year-old branches were collected from masson pines with high, medium, and low RYCs and then used for qRT-PCR. *PmFBX1* was highly expressed in the needle leaves (OL) and one-year-old branches (OB) of masson pine with medium and low RYC but lowly expressed in the two-year-old branches of masson pine with medium and low resin yield (Fig. 5B). The expression of *PmFBX2* showed gradual declined trends from masson pine with high RYCs to low RYCs in all three tissues. Then, several F-box family members from *Arabidopsis thaliana*, *Oryza sativa*, and coniferous species *Picea abies* were used for multiple sequence alignment and phylogenetic evolution analysis with *PmFBX1* and *PmFBX2* in this study (Fig. 5C). *PmFBX1* was clustered with MA 50439g0010, a protein from *P. abies*, and then clustered with another protein (MA 71913G0010) from *P. abies*. *PmFBX2* was clustered with LOC Os01g07160.1, a protein from *O. sativa*.

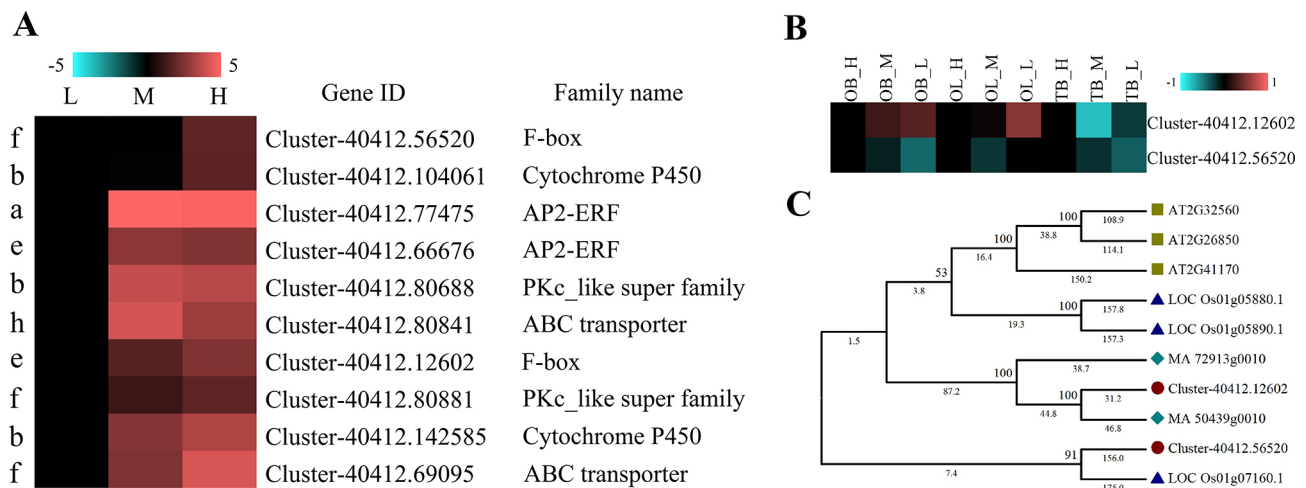


Fig. 5. Relative expression analyses and phylogenetic analysis of candidate genes. (A) The relative expression levels of ten selected candidate genes from five families. The relative expression levels of these genes were presented using fold change values transformed to log₂ format compared with L. Expression levels of each gene in sample L were unified to 1. T test was used to validate the significance among H, M, and L. English letters in the left side of heatmaps mean the grades of P-values. a: $P < 0.05$; b: $P < 0.01$; d: $P < 0.001$; e: $P < 0.0001$; f: $P < 0.00001$; h: $P < 0.0000001$. (B) Relative expression levels of PmFBX1 and PmFBX2 in needle leaves, one-year-old branches and two-year-old branches. (C) Phylogenetic analysis of PmFBX1 and PmFBX2 with the F-box family members from other plant species

Discussion

The biochemical components determine the quality of the pine resin and the utility value of masson pine germplasm resources, including the economic benefits. Masson pine is the primary resin-tapping species in China and contains more rosin than other pine species (Karanikas et al., 2010; Liu et al., 2016). In the present study, 69 *P. massoniana* accessions distributed across Guangdong province were analyzed for the qualitative and quantitative composition of terpenoid types of stems. Previous analyses of resins from Fujian and Guangxi have reported that α -pinene is the most abundant compound in turpentine, palustric acid is the most abundant compound in rosin (An & Ding, 2012; Liu et al., 2016). By calculating the turpentine percentage of each accession, the content of turpentine in masson pine resin showed a higher level than the coniferous tree species in Europe and America regions (Karanikas et al., 2010; Li et al., 2012). A total of 15 primary compounds were identified from the resin. The compounds are composed of five monoterpenes, four sesquiterpenes, and six diterpenes. In total, 13 out of 15 terpenoids were reported in the previous study of masson pine (Liu et al., 2016). In contrast, the terpenoids α -cedrene and α -santalene were not yet described in xylem resin. Generally, the most abundant compounds in our study are palustric acid and abietic acid in rosin, α -pinene in turpentine. Most terpenoids were also reported in other coniferous trees, but showed different concentration levels (Karanikas et al., 2010). Our results indicated that the category and content

of chemical compounds in masson pine in Guangdong were different from other regions and differs to the other *Pinus* species. The differences may be due to the provenance disparities and species differences, which indicated that abundant variations existed among masson pine families and pine species. No matter what, the masson pine in Guangdong can provide us different chemical quality of resin production. Although there were a limited number of correlations among the chemical components in our study, the correlation distributions showed some similar characteristics with *P. massoniana* in Fujian and *P. halepensis* (Karanikas et al., 2010; Liu et al., 2016).

The chemical type and quality have a significant influence on the industrial application of resin, the purposive selection of a certain chemical constituents in resin is a meaningful work for coniferous tree breeding (Li et al., 2012; Liu et al., 2016). The randomness of traditional breeding methods, which have typically neglected genetic information, has led to blindness in genome selection. Therefore, the mining of favorable SNP alleles is necessary for improving chemical selection in masson pine via molecular assisted selection (MAS). Due to their genome-wide abundance in genome DNA (Neale et al., 2014), SNP markers are playing an increasingly important role in plant research, such as in genetic map construction, novel gene discovery, evolutionary analysis, and MAS in breeding programs (Liu et al., 2012). As the availability of genomic DNA sequences has increased and the sequences have become more cost-effective to generate, the practicality of GWASs has increased

in the contemporary era. GWASs based on large-scale genome sequencing data provide a powerful platform for elucidating the genetic variants that underlie phenotypic diversity (Huang & Han, 2014). SLAF-seq is a recent and enhanced RRGs solution for large-scale SNP discovery and genotyping. This method has provided researchers with useful information regarding various types of traits (Sun et al., 2013; Su et al., 2016a; Su et al., 2016b). In our previous study, SLAF-seq identified 476,264 SNPs across 69 masson pine genomes and could largely support the association studies (Bai et al., 2019). The population structure and linkage disequilibrium analysis indicated that the masson pine population in Guangdong could be used for GWASs. Among the 15 primary chemical compounds in xylem resin, only pimaric acid, neoabietic acid, camphene, dipentene, and caryophyllene were reported with high heritability (Liu et al., 2016). The formation of complex traits was usually resulted from the combination of genes and environmental factors, which can easily lead to loss of heritability (Mackay et al., 2009). Hence, only those traits with high heritability can be stably detected. Stable favorable markers of target traits would be useful for masson pine breeding. In this study, five methods were used for association studies and gave us different numbers of associated SNP markers. Several markers were identified by at least two methods, which indicated that these SNPs have important application values in MAS. Using genetic data to guide breeding through MAS and genome selection has great potential for accelerating plant breeding while also improving its effectiveness (Varshney et al., 2014). The detected SNP markers can be PCR-amplified and are thus particularly useful for MAS (Scheben et al., 2017). The compounds, especially pimaric acid and neoabietic acid, have accounted 26% weight of the total resin acids. Hence, the pimaric acid and neoabietic acid associated SNPs may have important values in the future.

Except for discovering favorable SNP markers correlated with quantitative traits, excavating putative functional genes was also a general work for the breeding study. Many categories of genes, such as P450 and AP2, have been reported to be involved in the biosynthesis process of terpenoids (Bathe & Tissier, 2019; Chen et al., 2019; Zhang et al., 2019). Although the sequencing cost has been decreasing in recent years, it is still difficult in assembling the coniferous tree species which usually have large-scale genomes (Neale et al., 2014). Until now, the genome of *P. massoniana* has not been sequenced and *P. taeda* genome was used as the reference genome. In this study, a BLAST strategy was used to search for the related regions on *P. taeda* genome. After searching for 500 kb DNA information surrounding the SNP locations, several classifications of functional genes

were identified from these genome regions. Five gene categories, including PKc_like super family, Cytochrome P450, AP2-ERF, F-box, and ABC transporter, were obtained. Some differentially expressed genes were identified and treated as putative genes. The members of Cytochrome P450 gene family have been studied and reported in many model and non-model plants (Baek et al., 2019; Bathe & Tissier, 2019; Chen et al., 2019; Stagnati et al., 2019). The functions of several genes like terpene synthases (TPS) and Cytochrome P450 have also been reported in some coniferous trees (Hall et al., 2013; Geisler et al., 2016). CYP720B subfamily plays a central role in the biosynthesis of diterpene resin acids in *Sitka spruce*, *P. contorta* and *P. banksiana*. It has been reported that some members of the AP2 family are involved in the biosynthesis of various secondary metabolites including terpenoids (Udomsom et al., 2016; Paul et al., 2017; Zhang et al., 2019). ABC transporters are involved in the transportation of secondary metabolites in plants (Yazaki, 2006; Crouzet et al., 2013; Fu et al., 2017; Mao et al., 2019). It has been reported that ABC transporter is one of the important differentially expressed genes in *P. massoniana* (Liu et al., 2015). Some members of F-box were also involved in the biosynthesis of terpenoids, while PKc_like super family genes may be involved in the resistance process (Lee et al., 2010; Wu et al., 2018). Therefore, these genes identified in masson pine genome can be regarded as putative genes that may be involved in the biosynthesis of terpenoids. In this study, we analyzed the relative expression levels of five categories of genes. Several genes in each category were differentially expressed in masson pines with different resin yield. The relative expression levels of these genes were validated using qRT-PCR, which indicated that these differentially expressed genes could be treated as candidate genes that may play critical roles in the biosynthesis of terpenoids. Acquisition of the full-length of two F-box ORFs will be conducive to the molecular regulation mechanism study of terpenoids biosynthesis in *P. massoniana*.

Conclusions

The xylem resins of 69 masson pine germplasm resources were used for GC-MS analysis and 15 main terpenoids were detected from these masson pines. GWAS analyses were carried out on five chemical compounds, including pimaric acid, neoabietic acid, camphene, dipentene, and caryophyllene. A total of 42 associated SNPs were detected in these phenotypes using five GWAS methods. Five categories of candidate genes, including PKc_like super family, Cytochrome P450, AP2-ERF, F-box, and ABC transporter, were identified from these favorable SNP markers.

Two F-box family genes were full-length cloned. Our findings and results provide useful information in the selection of germplasm resources and resolve the molecular mechanism of terpenoid biosynthesis in *P. massoniana*.

Acknowledgements

This research was funded by the National Natural Science Foundation of China (No. 31800553), the Planning Subject of Guangdong Science and Technology Project (No. 2017B020232001), and ‘the 13th Five-Year-Plan’ in National Science and Technology for the Rural Development in China (No. 2017YFD0600303-5 and 2017YFD0600301).

Authors’ contributions

Qingsong Bai carried out the experimental, analytical work, wrote, and reviewed the manuscript. Qian Zhang, Yanling Cai, and Huiming Lian investigated the resources. Wanchuan Liu and Min Luo collected the materials and carried out experiment. Boxiang He and Linghai Zeng reviewed the manuscript. All authors read and approved the manuscript.

Availability of data and materials

All data generated in the present research is contained in this manuscript.

Competing interests

The authors declare that they have no competing interests.

Data archiving statement

Sequence data obtained in this study has been submitted to GenBank. The accession numbers of Cluster-40412.12602 and Cluster-40412.56520 were MT668424 and MT668425, respectively.

References

- An N & Ding G (2012) Study on chemical constituents of oleoresin from *Pinus massoniana* in Guangxi. *Journal of Central South University of Forestry & Technology* 32: 59–62. doi:10.14067/j.cnki.1673-923x.2012.03.026.
- Andolfatto P, Davison D, Erezylmaz D, Hu TT, Mast J, Sunayama-Morita T & Stern DL (2011) Multiplexed shotgun genotyping for rapid and efficient genetic mapping. *Genome Research* 21: 610–617. doi:10.1101/gr.115402.110.
- Baek YS, Goodrich LV, Brown PJ, James BT, Moose SP, Lambert KN & Riechers DE (2019) Transcriptome profiling and genome-wide association studies reveal GSTs and other defense genes involved in multiple signaling pathways induced by herbicide safener in grain sorghum. *Frontiers in Plant Science* 10: 192. doi:10.3389/fpls.2019.00192.
- Bai Q, He B, Cai Y, Lian H & Zhang Q (2020) Transcriptomic and metabolomic analyses reveal several critical metabolic pathways and candidate genes involved in resin biosynthesis in *Pinus massoniana*. *Molecular Genetics and Genomics* 295: 327–341. doi:10.1007/s00438-019-01624-1.
- Bai Q, Cai Y, He B, Liu W, Pan Q & Zhang Q (2019) Core set construction and association analysis of *Pinus massoniana* from Guangdong province in southern China using SLAF-seq. *Scientific Reports* 9: 13157. doi:10.1038/s41598-019-49737-2.
- Bai T-D, Xu L-A, Xu M & Wang Z-R (2014) Characterization of masson pine (*Pinus massoniana* Lamb.) microsatellite DNA by 454 genome shotgun sequencing. *Tree Genetics & Genomes* 10: 429–437. doi:10.1007/s11295-013-0684-y.
- Bartholome J, Bink MCAM, van Heerwaarden J, Chancerel E, Boury C, Lesur I, Isik F, Bouffier L & Plomion C (2017) Linkage and association mapping for two major traits used in the maritime pine breeding program: height growth and stem straightness. *PloS One* 12: e0171439. doi:10.1371/journal.pone.0171439.
- Bathe U & Tissier A (2019) Cytochrome P450 enzymes: A driving force of plant diterpene diversity. *Phytochemistry* 161: 149–162. doi:10.1016/j.phytochem.2018.12.003.
- Bohlmann J (2011) Terpenoid synthases—from chemical ecology and forest fires to biofuels and bio-products. *Structure* 19: 1730–1731. doi:10.1016/j.str.2011.11.009.
- Chen W, Cao M, Wang Y, Zhou Z & Xu L-A (2014) A genetic linkage map of *Pinus massoniana* based on SRAP, SSR and ESTP markers. *Silvae Genetica* 63: 1–9. doi:10.1515/sg-2014-0001.
- Chen X, Luck K, Rabe P, Dinh CQD, Shaulsky G, Nelson DR, Gershenzon J, Dickschat JS, Koellner TG & Chen F (2019) A terpene synthase-cytochrome P450 cluster in *Dictyostelium discoideum* produces a novel trisnorsesquiterpene. *Elife* 8: e44352. doi:10.7554/eLife.44352.
- Coppen JJW & Hone GA (1995) Gum naval stores: turpentine and rosin from pine resin. Natural Resources Institute, Food and Agriculture Organization of the United Nations.
- Crouzet J, Roland J, Peeters E, Trombik T, Ducos E, Nader J & Boutry M (2013) NtPDR1, a plasma membrane ABC transporter from *Nicotiana tabacum*, is involved in diterpene transport. *Plant*

- Molecular Biology 82: 181–192. doi:10.1007/s11103-013-0053-0.
- de Hoon MJL, Imoto S, Nolan J & Miyano S (2004) Open source clustering software. *Bioinformatics* 20: 1453–1454. doi:10.1093/bioinformatics/bth078.
- Eckert AJ, Wegrzyn JL, Cumbie WP, Goldfarb B, Huber DA, Tolstikov V, Fiehn O & Neale DB (2012) Association genetics of the loblolly pine (*Pinus taeda*, Pinaceae) metabolome. *New Phytologist* 193: 890–902. doi:10.1111/j.1469-8137.2011.03976.x.
- Feng YH, Yang ZQ, Wang J, Luo QF & Li HG (2014) Development and characterization of SSR markers from *Pinus massoniana* and their transferability to *P. elliottii*, *P. caribaea* and *P. yunnanensis*. *Genetics and Molecular Research* 13: 1508–1513. doi:10.4238/2014.March.12.2.
- Fu X, Shi P, He Q, Shen Q, Tang Y, Pan Q, Ma Y, Yan T, Chen M, Hao X, Liu P, Li L, Wang Y, Sun X & Tang K (2017) AaPDR3, a PDR transporter 3, is involved in sesquiterpene beta-caryophyllene transport in *Artemisia annua*. *Frontiers in Plant Science* 8: 723. doi:10.3389/fpls.2017.00723.
- Geisler K, Jensen NB, Yuen MMS, Madilao L & Bohlmann J (2016) Modularity of conifer diterpene resin acid biosynthesis: P450 enzymes of different CYP720B clades use alternative substrates and converge on the same products. *Plant Physiology* 171: 152–164. doi:10.1104/pp.16.00180.
- Hall DE, Zerbe P, Jancsik S, Quesada AL, Dullat H, Madilao LL, Yuen M & Bohlmann J (2013) Evolution of conifer diterpene synthases: diterpene resin acid biosynthesis in lodgepole pine and jack pine involves monofunctional and bifunctional diterpene synthases. *Plant Physiology* 161: 600–616. doi:10.1104/pp.112.208546.
- Huang X & Han B (2014) Natural variations and genome-wide association studies in crop plants. *Annual Review of Plant Biology* 65: 531–551. doi:10.1146/annurev-arplant-050213-035715.
- Karanikas C, Walker V, Scaltsoyiannes A, Comte G & Bertrand C (2010) High vs. low yielding oleoresin *Pinus halepensis* Mill. trees GC terpenoids profiling as diagnostic tool. *Annals of Forest Science* 67: 412. doi:10.1051/forest/2009132.
- Lee S, Badiyan S, Bevan DR, Herde M, Gatz C & Tholl D (2010) Herbivore-induced and floral homoterpene volatiles are biosynthesized by a single P450 enzyme (CYP82G1) in *Arabidopsis*. *Proceedings of the National Academy of Sciences of the United States of America* 107: 21205–21210. doi:10.1073/pnas.1009975107.
- Li Y, Luan Q, Shen D, Chen B, Jiang J & Fu Y (2012) Study on genetic variation of resin components among open-pollinated families of slash pine. *Forest Research* 25: 773–779. doi:10.13275/j.cnki.lykxyj.2012.06.007.
- Lippert C, Listgarten J, Liu Y, Kadie CM, Davidson DI & Heckerman D (2011) FaST linear mixed models for genome-wide association studies. *Nature Methods* 8: 833–835. doi:10.1038/nmeth.1681.
- Liu J, Huang S, Sun M, Liu S, Liu Y, Wang W, Zhang X, Wang H & Hua W (2012) An improved allele-specific PCR primer design method for SNP marker analysis and its application. *Plant Methods* 8: 34. doi:10.1186/1746-4811-8-34.
- Liu JJ, Schoettle AW, Sniezko RA, Yao F, Zamany A, Williams H & Rancourt B (2019) Limber pine (*Pinus flexilis* James) genetic map constructed by exome-seq provides insight into evolution of disease resistance and a genomic resource for genomics-based breeding. *The Plant Journal* 98: 745–758. doi:10.1111/tpj.14270.
- Liu Q, Zhou Z, Wei Y, Shen D, Feng Z & Hong S (2015) Genome-wide identification of differentially expressed genes associated with the high yielding of oleoresin in secondary xylem of Masson pine (*Pinus massoniana* Lamb) by transcriptomic analysis. *PloS One* 10: e0132624. doi:10.1371/journal.pone.0132624.
- Liu Q, Wei Y, Fan H, Shen D, Chen W & Zhou Z (2016) Family variation of oleoresin compounds and their correlation in *Pinus massoniana*. *Forest Research* 29: 324–330. doi:10.13275/j.cnki.lykxyj.2016.03.003.
- Liu Q, Wei Y, Xu L, Hao Y, Chen X & Zhou Z (2017) Transcriptomic profiling reveals differentially expressed genes associated with pine wood nematode resistance in masson Pine (*Pinus massoniana* Lamb.). *Scientific Reports* 7: 4693. doi:10.1038/s41598-017-04944-7.
- Livak KJ & Schmittgen TD (2001) Analysis of relative gene expression data using real-time quantitative PCR and the 2^{-CT} method. *Methods* 25: 402–408. doi:10.1006/meth.2001.1262.
- Mackay TFC, Stone EA & Ayroles JF (2009) The genetics of quantitative traits: challenges and prospects. *Nature Reviews Genetics* 10: 565–577. doi:10.1038/nrg2612.
- Manninen AM, Tarhanen S, Vuorinen M & Kainulainen P (2002) Comparing the variation of needle and wood terpenoids in Scots pine provenances. *Journal of Chemical Ecology* 28: 211–228. doi:10.1023/a:1013579222600.
- Mao J, He Z, Hao J, Liu T, Chen J & Huang S (2019) Identification, expression, and phylogenetic analyses of terpenoid biosynthesis-related genes in secondary xylem of loblolly pine (*Pinus taeda* L.) based on transcriptome analyses. *PeerJ* 7: e6124. doi:10.7717/peerj.6124.
- McLaren W, Pritchard B, Rios D, Chen Y, Flicek P & Cunningham F (2010) Deriving the consequenc-

- es of genomic variants with the Ensembl API and SNP Effect Predictor. *Bioinformatics* 26: 2069–2070. doi:10.1093/bioinformatics/btq330.
- Neale DB (2007) Genomics to tree breeding and forest health. *Current Opinion in Genetics & Development* 17: 539–544. doi:10.1016/j.gde.2007.10.002.
- Neale DB & Ingvarsson PK (2008) Population, quantitative and comparative genomics of adaptation in forest trees. *Current Opinion in Plant Biology* 11: 149–155. doi:10.1016/j.pbi.2007.12.004.
- Neale DB & Kremer A (2011) Forest tree genomics: growing resources and applications. *Nature Reviews Genetics* 12: 111–122. doi:10.1038/nrg2931.
- Neale DB, Wegrzyn JL, Stevens KA, Zimin AV, Puiu D, Crepeau MW, Cardeno C, Koriabine M, Holtz-Morris AE, Liechty JD, Martinez-Garcia PJ, Vasquez-Gross HA, Lin BY, Zieve JJ, Dougherty WM, Fuentes-Soriano S, Wu L-S, Gilbert D, Marcias G, Roberts M, Holt C, Yandell M, Davis JM, Smith KE, Dean JFD, Lorenz WW, Whetten RW, Sederoff R, Wheeler N, McGuire PE, Main D, Loopstra CA, Mockaitis K, deJong PJ, Yorke JA, Salzberg SL & Langley CH (2014) Decoding the massive genome of loblolly pine using haploid DNA and novel assembly strategies. *Genome Biology* 15: R59. doi:10.1186/gb-2014-15-3-r59.
- Page RDM (1996) TreeView: an application to display phylogenetic trees on personal computers. *Computer Applications in the Biosciences* 12: 357–358. doi:10.1093/bioinformatics/12.4.357.
- Paul P, Singh SK, Patra B, Sui X, Pattanaik S & Yuan L (2017) A differentially regulated AP2/ERF transcription factor gene cluster acts downstream of a MAP kinase cascade to modulate terpenoid indole alkaloid biosynthesis in *Catharanthus roseus*. *New Phytologist* 213: 1107–1123. doi:10.1111/nph.14252.
- Pavy N, Lamothe M, Pelgas B, Gagnon F, Birol I, Bohlmann J, Mackay J, Isabel N & Bousquet J (2017) A high resolution reference genetic map positioning 8.8K genes for the conifer white spruce: structural genomics implications and correspondence with physical distance. *Plant Journal* 90: 189–203. doi:10.1111/tbj.13478.
- Peng SL, Li QF, Li D, Wang ZF & Wang DP (2003) Genetic diversity of *Pinus massoniana* revealed by RAPD markers. *Silvae Genetica* 52: 60–63. doi:10.1007/s10722-004-5730-z.
- Peralta-Yahya PP, Zhang F, del Cardayre SB & Keasling JD (2012) Microbial engineering for the production of advanced biofuels. *Nature* 488: 320–328. doi:10.1038/nature11478.
- Rezzi S, Bighelli A, Castola V & Casanova J (2005) Composition and chemical variability of the oleoresin of *Pinus nigra* ssp *laricio* from Corsica. *Industrial Crops and Products* 21: 71–79. doi:10.1016/j.indcrop.2003.12.008.
- Scheben A, Batley J & Edwards D (2017) Genotyping-by-sequencing approaches to characterize crop genomes: choosing the right tool for the right application. *Plant Biotechnology Journal* 15: 149–161. doi:10.1111/pbi.12645.
- Sjodin K, Persson M, Faldt J, Ekberg I & Borg-Karlson AK (2000) Occurrence and correlations of monoterpene hydrocarbon enantiomers in *Pinus sylvestris* and *Picea abies*. *Journal of Chemical Ecology* 26: 1701–1720. doi:10.1023/a:1005547131427.
- Stagnati L, Lanubile A, Samayoa LF, Bragalanti M, Giorni P, Busconi M, Holland JB & Marocco A (2019) A genome wide association study reveals markers and genes associated with resistance to *Fusarium verticillioides* infection of seedlings in a maize diversity panel. *G3-Genes Genomes Genetics* 9: 571–579. doi:10.1534/g3.118.200916.
- Stinchcombe JR & Hoekstra HE (2008) Combining population genomics and quantitative genetics: finding the genes underlying ecologically important traits. *Heredity* 100: 158–170. doi:10.1038/sj.hdy.6800937.
- Su J, Fan S, Li L, Wei H, Wang C, Wang H, Song M, Zhang C, Gu L, Zhao S, Mao G, Wang C, Pang C & Yu S (2016a) Detection of favorable QTL alleles and candidate genes for lint percentage by GWAS in Chinese upland cotton. *Frontiers in Plant Science* 7: 1576. doi:10.3389/fpls.2016.01576.
- Su J, Pang C, Wei H, Li L, Liang B, Wang C, Song M, Wang H, Zhao S, Jia X, Mao G, Huang L, Geng D, Wang C, Fan S & Yu S (2016b) Identification of favorable SNP alleles and candidate genes for traits related to early maturity via GWAS in upland cotton. *BMC Genomics* 17: 687. doi:10.1186/s12864-016-2875-z.
- Sun X, Liu D, Zhang X, Li W, Liu H, Hong W, Jiang C, Guan N, Ma C, Zeng H, Xu C, Song J, Huang L, Wang C, Shi J, Wang R, Zheng X, Lu C, Wang X & Zheng H (2013) SLAF-seq: an efficient method of large-scale de novo SNP discovery and genotyping using high-throughput sequencing. *PloS One* 8: e58700. doi:10.1371/journal.pone.0058700.
- Tamba CL, Ni Y-L & Zhang Y-M (2017) Iterative sure independence screening EM-Bayesian LASO algorithm for multi-locus genome-wide association studies. *PLoS Computational Biology* 13: e1005357. doi:10.1371/journal.pcbi.1005357.
- Tamura K, Stecher G, Peterson D, Filipinski A & Kumar S (2013) MEGA6: molecular evolutionary genetics analysis version 6.0. *Molecular Biology and Evolution* 30: 2725–2729. doi:10.1093/molbev/mst197.
- Trapp S & Croteau R (2001) Defensive resin biosynthesis in conifers. *Annual Review of Plant Physiology*

- iology and Plant Molecular Biology 52: 689–724. doi:10.1146/annurev.arplant.52.1.689.
- Udomsom N, Rai A, Suzuki H, Okuyama J, Imai R, Mori T, Nakabayashi R, Saito K & Yamazaki M (2016) Function of AP2/ERF transcription factors involved in the regulation of specialized metabolism in *Ophiorrhiza pumila* revealed by transcriptomics and metabolomics. *Frontiers in Plant Science* 7: 1861. doi:10.3389/fpls.2016.01861.
- Untergasser A, Cutcutache I, Koressaar T, Ye J, Faircloth BC, Remm M & Rozen SG (2012) Primer3-new capabilities and interfaces. *Nucleic Acids Research* 40: e115. doi:10.1093/nar/gks596.
- Varshney RK, Terauchi R & McCouch SR (2014) Harvesting the promising fruits of genomics: applying genome sequencing technologies to crop breeding. *PLoS Biology* 12: e1001883. doi:10.1371/journal.pbio.1001883.
- Wang L, Feng Z, Wang X, Wang X & Zhang X (2010) DEGseq: an R package for identifying differentially expressed genes from RNA-seq data. *Bioinformatics* 26: 136–138. doi:10.1093/bioinformatics/btp612.
- Wang S-B, Feng J-Y, Ren W-L, Huang B, Zhou L, Wen Y-J, Zhang J, Dunwell JM, Xu S & Zhang Y-M (2016) Improving power and accuracy of genome-wide association studies via a multi-locus mixed linear model methodology. *Scientific Reports* 6: 19444. doi:10.1038/srep19444.
- Wang ZX, Calderon MM & Carandang MG (2006) Effects of resin tapping on optimal rotation age of pine plantation. *Journal of Forest Economics* 11: 245–260. doi:10.1016/j.jfe.2005.10.001.
- Wen Y-J, Zhang H, Ni Y-L, Huang B, Zhang J, Feng J-Y, Wang S-B, Dunwell JM, Zhang Y-M & Wu R (2018) Methodological implementation of mixed linear models in multi-locus genome-wide association studies. *Briefings in Bioinformatics* 19: 700–712. doi:10.1093/bib/bbw145.
- Wu P, Xie J, Hu J, Qiu D, Liu Z, Li J, Li M, Zhang H, Yang L, Liu H, Zhou Y, Zhang Z & Li H (2018) Development of molecular markers linked to powdery mildew resistance gene Pm4b by combining SNP discovery from transcriptome sequencing data with bulked segregant analysis (BSR-Seq) in wheat. *Frontiers in Plant Science* 9: 95. doi:10.3389/fpls.2018.00095.
- Yazaki K (2006) ABC transporters involved in the transport of plant secondary metabolites. *FEBS Letters* 580: 1183–1191. doi:10.1016/j.febslet.2005.12.009.
- Zeng LH, Zhang Q, He BX, Lian HM, Cai YL, Wang YS & Luo M (2013) Age trends in genetic parameters for growth and resin-yielding capacity in masson pine. *Silvae Genetica* 62: 7–18. doi:10.1515/sg-2013-0002.
- Zerbe P, Chiang A, Yuen M, Hamberger B, Hamberger B, Draper JA, Britton R & Bohlmann J (2012) Bifunctional cis-abienol synthase from *Abies balsamea* discovered by transcriptome sequencing and its implications for diterpenoid fragrance production. *Journal of Biological Chemistry* 287: 12121–12131. doi:10.1074/jbc.M111.317669.
- Zhang Y, Chu D-Y, Jin G-Q & Zhou Z-C (2010) Relationship between hybrid performance and genetic distances among *Pinus massoniana* clones based on ISSR molecular marker. *Forest Research* 23: 215–220. doi:10.3724/SPJ.1238.2010.00474.
- Zhang Y, Yang Q, Zhou Z & Jin G (2013) Divergence among masson pine parents revealed by geographical origins and SSR markers and their relationships with progeny performance. *New Forests* 44: 341–355. doi:10.1007/s11056-012-9340-x.
- Zhang Y, Ji A, Xu Z, Luo H & Song J (2019) The AP2/ERF transcription factor SmERF128 positively regulates diterpenoid biosynthesis in *Salvia miltiorrhiza*. *Plant Molecular Biology* 100: 83–93. doi:10.1007/s11103-019-00845-7.
- Zhou X & Stephens M (2012) Genome-wide efficient mixed-model analysis for association studies. *Nature Genetics* 44: 821–824. doi:10.1038/ng.2310.