Plant Systematics and Evolution (2021) 307:44
https://doi.org/10.1007/s00606-021-01764-1

ORIGINAL ARTICLE

Genetic variation and differentiation of *Quercus variabilis* populations at phosphate and non-phosphate rock sites in southwestern China

Huawei Ji[1,2] · Baoming Du[1] · Hongmei Du[3] · Ningxiao Sun[1] · Xu Yang[4] · Mingjun Peng[4] · Hongzhang Kang[1,5] · Shan Yin[1,5,6] · Chunjiang Liu[1,5,6]

Received: 16 August 2020 / Accepted: 4 May 2021 / Published online: 9 June 2021

Abstract

Subtropical soils are generally characterized by phosphorus (P), calcium (Ca) and magnesium (Mg) deficiency; extreme P-rich soils develop on phosphate rocks. How such contrasting geological-derived soils influence genetic diversity and structure of local plant populations have not been well documented, hindering our understanding of plant adaptive evolution in subtropical areas. In this study, we applied double digest restriction-site-associated DNA sequencing (ddRAD-seq) to investigate the genetic variation and differentiation of natural *Quercus variabilis* populations growing at geologically derived soil P-rich and P-deficient sites in subtropical China. Results showed that *Q. variabilis* populations had lower genetic diversity at P-rich sites than those at P-deficient sites, and genetic diversity was negatively correlated only with soil P content ($p < 0.05$) across these sites. The genetic variation mainly occurred within populations (95.76%) with a much smaller amount among populations (2.27%) and between the two P-type sites (1.97%). Moreover, significant genetic divergence ($p < 0.0001$) occurred between the two site-type populations. Among genes embedded in the selective sweep areas in genome for populations at P-rich sites and P-deficient sites, some were annotated as transcription factors, and some were involved in protein synthesis and degradation, involved in regulatory responses to P availability and other environmental stresses. These results suggest that the availability of soil P could be a key selective force driving adaptive genetic differentiation among *Q. variabilis* populations across variable P content soils in subtropical areas.

Keywords Adaptation · DdRAD-seq · Genetic variation and differentiation · Phosphate rocks · *Quercus variabilis* · Soil P content

Introduction

Phosphorus (P) is an essential macronutrient, and efficient acquisition of phosphate (Pi) is an important factor for plant growth and development (Raghothama and Karthikeyan 2005). Consequently, plants have evolved a wide range of morphological and molecular adaptations to increase remobilization, uptake and efficient use of Pi when availability is low (Raghothama and Karthikeyan 2005). These adaptations require the plants to sense the level of Pi and change the expression of a number of genes accordingly (Müller et al. 2007). The induction of high affinity phosphate transporters and the secretion of acid phosphatase and organic acids contribute to the mobilization and uptake of P from the rhizosphere (Wasaki et al. 2003, 2006; Fang et al. 2009). Wu et al. (2003) and Hammond et al. (2003) reported that the expression of various genes in *Arabidopsis*, such as the bypass pathways of

Handling editor: Hiroshi Kudoh.

✉ Chunjiang Liu
chjliu@sjtu.edu.cn

1 School of Agriculture and Biology, Shanghai Jiao Tong University, Shanghai 200240, China

2 Shaanxi Institute of Geological Survey, Xi'an 710054, China

3 School of Design, Shanghai Jiao Tong University, Dongchuan Rd. 800, Shanghai 200240, China

4 Yunnan Academy of Forestry, Lanan Rd. 2, Kunming 600000, Yunnan, China

5 Shanghai Urban Forest Research Station, State Forestry Administration, Shanghai 200240, China

6 Shanghai Yangtze River Delta Dense Population Region Research Station, Ministry of Education, People Republic of China, Shanghai 200240, China

carbon (C) metabolism and signal transduction, changes when plants are grown under low P conditions. Comprehensive examination of global gene expression in response to P deficiency revealed coordinated induction of 612 genes and suppression of 254 genes (Misson et al. 2005). Moreover, the regulated genes were involved in a wide range of functional groups related to metabolism, ion transport, signal transduction, transcriptional regulation, and growth and development (Misson et al. 2005).

Subtropical ecosystems are typically characterized as having P-, Ca- and Mg-deficient soils, while they are enriched with Fe and Al due to strong weathering and leaching under high temperatures and moisture levels (Walker and Syers 1976; Tiessen 2008; Cleveland et al. 2011). This could result in the low P stoichiometry in plants (Han et al. 2005). Despite the low supply of P in soils, good plant functionality might be maintained due to P conservation and utilization mechanisms in subtropical ecosystems (Vitousek 1982; Attiwill and Adams 1993). In terms of ecology, plants might have evolved and formed ecotypic populations that are adapted to P-deficient habitats (Snaydon and Bradshaw 1962; Silberbush et al. 1981; Yan et al. 1995; Xiao et al. 2009). In the Central Yunnan Plateau, located in subtropical China, some P-rich sites formed on P-rich ores (Yan et al. 2017), where soil resident P concentration is six-fold higher (> 1.2 mg g^{-1}) than those at P-deficient sites. The content of other chemical elements, such as N, S, Mg and Fe, is also significantly higher (Tao 2005; Yan et al. 2011; Zhou et al. 2015; Ji et al. 2017). At these P-rich sites, there were significantly P-rich based stoichiometry in plants (Xiao et al. 2009; Zhou et al. 2015) and in herbivorous insects due to the up-cascading transfer of the effects of P-rich soils (Ji et al. 2017). This likely meant that there were several P-rich ecosystems distributed across P-limiting subtropics, which resulted in P-rich ecotype populations.

Many studies show that differences in micro-habitat, such as slope (Owuor et al. 1997), soil fertility (Huff et al. 1998) and elevation (Owuor et al. 1997), can influence population genetic diversity and genetic differentiation among populations. Cao et al. (2014) suggested that the adaptation of dominant plants in phosphorus-enriched areas to local heterogeneous environments was characterized by certain genetic differentiation, and that edaphic factors, such as total nitrogen and available phosphorus, have had a significant impact on population differentiation of the dominant plants. Genetic differentiation is an important mechanism for the adaptation of biological populations to heterogeneous habitat conditions (Linhart and Grant 1996). The nutrient and metabolite compositions of *Quercus variabilis* significantly differed at P-rich and P-deficient sites in central Yunnan, suggesting that *Q. variabilis* might have formed ecotypes adapted to different local soil P supply (Ji et al. 2019). However, it is unclear how the patterns of population genetic variation and genetic differentiation of woody plants are related to soil P content.

Several types of molecular markers have been used to investigate population genetic variation in tree species (Catchen et al. 2013a). The emergence of next-generation sequencing (NGS) technologies has revolutionized genetic studies, providing a framework for detecting genome-wide highly informative markers in natural populations, including both model and non-model species (Allendorf et al. 2010; Zhang et al. 2016). For example, restriction-site-associated DNA tags sequencing (RAD-seq) uses Illumina next-generation sequencing platform to simultaneously discover and score tens to hundreds of thousands of SNP markers in hundreds of individuals (Miller et al. 2007; Etter et al. 2011). To date, RAD-seq approaches have been successfully used to identify SNPs of many species, such as temperate bamboo (Wang et al. 2013), American oak (Hipp et al. 2014) and threespine stickleback (Catchen et al. 2013a), opening the possibility for ecological, phylogenetic and population genetic studies at a genome-wide scale. RAD series mainly includes 2b-RAD (Wang et al. 2012), ddRAD (Peterson et al. 2012), and ezRAD (Toonen et al. 2013), of which, ddRAD can tune fragments number by employing two different enzymes and size selection, and the process of constructing a library is quite simple. Thus, ddRAD has become a more economical method to genotype thousands of individuals and has been proved to be suitable for population genetic and genomic studies in plants (Peterson et al. 2012; Yang et al. 2016). Selective sweep deals with the theory and practice of detection of recent adaptive evolution at the genomic level from the patterns of DNA polymorphism, and a selective sweep describes the reduction of diversity due to strong positive selection (Dmitry 2005). Therefore, in this study, the selective sweep method can be used to obtain possible selective site and candidate genes related to soil P variation.

The Central Yunnan Plateau of China is one of the richest phosphate rock regions in the world, resulting in many soil P-rich sites in subtropical areas where the soils are generally characterized by P, Ca and Mg deficiency (Zhou et al. 2015). *Q. variabilis* is one pan-Eastern-Asian distributed deciduous oak with high ecological, economic and cultural importance (Chen et al. 2012). It is also the dominant tree species in the natural forest communities at both P-rich and P-deficient sites in the Central Yunnan Plateau (Chen et al. 2012). Generally, genetic variation in *Q. variabilis* populations is high, with the majority of genetic variation within populations (Xu et al. 2004). Han (2004) reported that nitrogen, organic matters, soil thickness and rainfall played an important role in regulating genetic diversity of *Q. variabilis* populations. However, the distribution of *Q. variabilis* has greatly declined due to serious disturbances, such as cutting, grazing and agricultural activity, which have left the

stands highly fragmented (Chen et al. 2012). In this study, we employ the ddRAD-seq approaches to investigate the level of genetic diversity and differentiation of *Q. variabilis* populations growing at contrasting phosphorous levels in subtropical areas. The aim of this study was to investigate the extent to which soil P variation influences the genetic structure and variation of *Q. variabilis* populations in P-deficient subtropical areas.

Materials and methods

Study sites

The study area is in central Yunnan, subtropical China, where P-rich sites are surrounded by dominant P-deficient sites. P-rich sites (developed on P-rich phosphate rocks) (LT (Heilongtan), ZW (Zhiwuyuan), LKY (Linkeyuan) and AN (Anning)) are located near Kunming City where P-rich ores are distributed, and P-deficient sites (developed on non-phosphate rocks) (QR (Qingrengu), JB (Jiuba), ZT (Zhongtunshuiku) and FL (Fulongcun)) are located near Chuxiong City (Fig. 1). The chemical compositions of phosphate rocks on P-rich sites (Tao 2005) and non-phosphate rocks on P-deficient sites (Shi et al. 2011) can be seen in Online Resource 1. Climate is similar between the two site types, with a mean annual temperature of 15.4 °C and mean annual precipitation of 936.5 mm. In the study area, *Q. variabilis* occur in pure or mixed stands with other tree species, such as *Pinus yunnanensis* French.

Sampling

Four *Q. variabilis* sampling locations were selected for each of the P-rich and P-deficient site types (Fig. 1). Within each natural *Q. variabilis* sampling location, about ten well-growing trees were selected (distance from each tree was at least 30 m). Five of the youngest and well-developed leaves were collected from each tree and immediately frozen in situ in liquid nitrogen, and stored at −80° C pending analysis. Surface soil cores (0–10 cm, 3 cm Ø) collected from each stand were air-dried and sieved through a 60 mesh sieve (0.25 mm diameter) for chemical analysis. In this study, a total of 80 *Q. variabilis* individuals were used.

DdRAD sequencing

The DNA was extracted with a modified CTAB method (Doyle 1991). Library for ddRAD sequencing was prepared according to the double digest RAD-seq method described

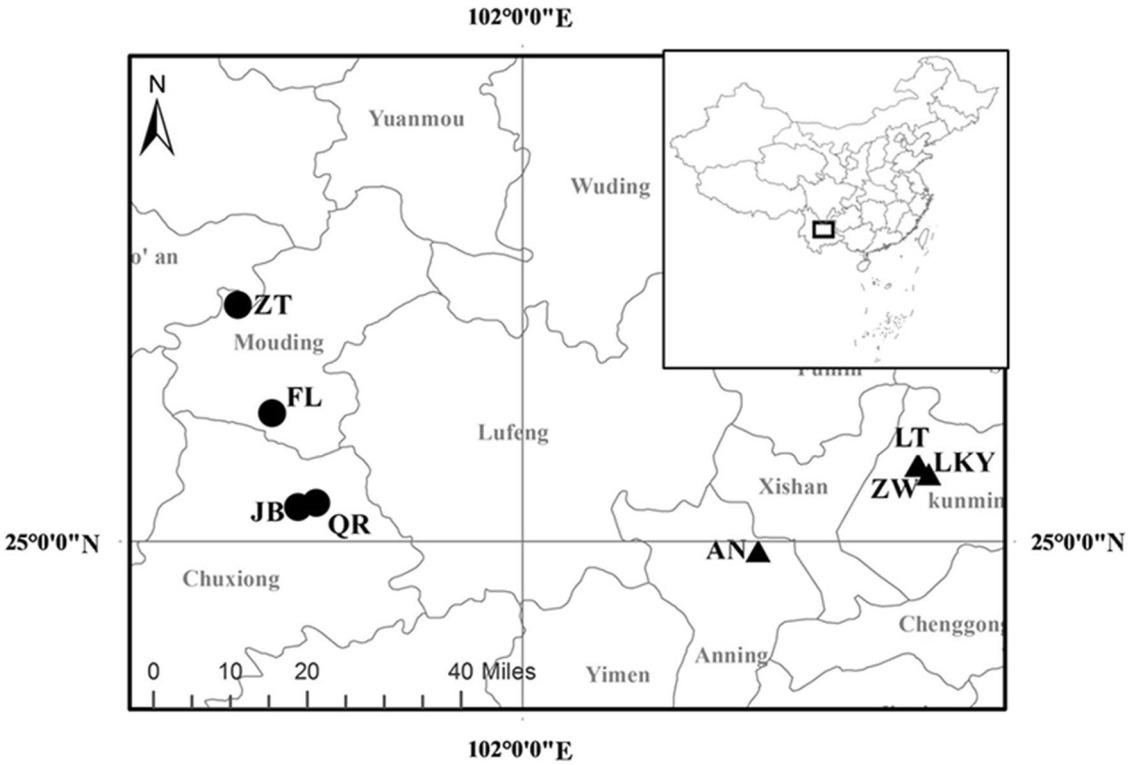

Fig. 1 The locations of sampling sites including 4 P-rich sites (LT, ZW, LKY and AN) and 4 P-deficient sites (QR, JB, ZT and FL) in Yunnan province. Circles represent P-deficient sites and triangles represent P-rich sites

by Peterson et al. (2012). Briefly, for each individual 500 ng of genomic DNA was digested with HindIII and BfaI restriction enzymes (New England Biolabs). Then, 22.5 µL of 0.5X and 13.5 µL of 0.3X VAHTS™ DNA Clean Beads were used to get rid of big and small fragments, respectively, and we aimed at selecting the 220–450 bp fragments. After fragment selection, adapters, one with 7 bp barcode, were then ligated using Quick Ligase (NEB) in 50 µL reactions under room temperature. After ligation, purified adapter-ligated restriction fragments were used as templates of the PCR reaction. The PCR reactions (25 µL) contained: 5 µL of 5X Reaction buffer, 5 µL of 5X High GC enhancer, 2 µL of dNTP, 0.25 µL of Q5 polymerase, 1 µL of Index primer, 1 µL of Universal primer, 1/2/4 µL of adapter-ligated DNA fragments and 9.75/8.75/6.75 µL of ddH_2O. The PCR profile was as follows: 98 °C-30 s, 14 cycles: 98 °C-15 s, 65 °C-30 s, 72 °C-30 s and final extension 72 °C-5 min. The amplified samples were purified by 20 µL of 0.8X VAHTS™ DNA Clean Beads and followed quantification with ELISA and then were pooled, resulting in the library. The purified library was loaded on a 2.0% agarose gel for electrophoresing and we selected 350–600 bp gel as the target fragments. The library quality was detected by Agilent 2100 Bioanalyzer and concentration of the library was estimated by TBS-380 Mini-Fluorometer, and then, the library was sent for Illumina HiSeq 2500 sequencing (paired end, 2 × 150 bp) at Shanghai Personal Biotechnology Corporation.

Raw Illumina data (raw reads) of *Q. variabilis* can be seen in Online Resource 2. FastQC (http://www.bioinformatics.babraham.ac.uk/projects/fastqc/) was used to compute the per-base sequence quality. To avoid low-quality reads, data were filtered (Online Resource 3) according to the following criteria: (i) for paired raw sequencing reads, if the 5′-end of read1 (front 6 bp) did not match the HindIII recognition site sequence (AAGCTTT) or if the 5′-end of read2 (front 4 bp) did not match the BfaI recognition site sequence (CTAG), the reads were removed; (ii) adapter contamination of 3′-end of reads was removed by AdapterRemoval (version 2) (Schubert et al. 2016); (iii) quality filtering was achieved by using a sliding window (average quality per base within 5 bp windows met a minimum quality score of ≥ 20); (iv) paired raw sequencing reads were discarded when one of them was ≤ 50 bp in length.

Assembly of the reference genome

Due to the unavailability of the presently existing genomic information of the *Q. variabilis*, we specified one individual from the sampled populations for de novo genome sequencing, assembly and annotation. Then, the resulted assemblage reference genome was used as a reference for downstream alignment and variant calling. Two paired-end (PE) libraries with 250 bp and 500 bp inserts, respectively, and one 8 kb large insert mate-pair (MP) library were constructed. These libraries were sequenced on an Illumina HiSeq 2500 system using 2 × 150 bp, paired-end settings. After quality control of raw reads, total bases generated were 32.9 Gb and 38.0 Gb for the 250 bp and 500 bp PE libraries, respectively, and 3.3 Gb for the 8 kb MP library. Sequence coverage was determined using an estimated genome size of 0.9 Gb. Sequencing depth was approximately 36.6 × and 42.2 × for the 250 bp and 500 bp PE libraries, respectively, and 3.7 × for the 8 kb MP library. The assembly was performed using the SPAdes software (version 3.9.0), and assembly metrics are detailed in Online Resource 4. The SPAdes assembly generated scaffolds totaling 837.0 Mb with an N content of 20.72%. The GC content of the scaffolds was 35.7%. The assembled genome contained 169,545 scaffolds greater than 1,000 bp. The longest scaffold generated was 355.2 kb. The scaffold N50 and N90 was 19.3 Kb and 7.0 Kb, respectively.

Then, completeness of the assembled genome was assessed using BUSCO software (version 3.0.2), based on a set of 1440 conserved genes of 30 species in plant database. BUSCOs were classified as complete (C), complete and single-copy (S), complete and duplicated (D), fragmented (F) or missing (M). Assessment results of completeness of the assembled reference genome can be seen in Online Resource 5. In order to improve the accuracy of gene prediction, the following process was used to predict the gene model: (1) a combination of ab initio gene prediction methods with the software Augustus (version 3.03), glimmerHMM (version 3.0.1) and SNAP (version 2006-07-28) was used to obtain the ab initio prediction results of gene models; (2) gene prediction results including ab initio prediction results, homologous prediction results of related species and RNA-Seq prediction results were integrated in the software EvidenceModeler (EVM, version r2012-06-25) to predict all genes of the genome. The predicted results of the protein-coding genes in the assembled reference genome can be seen in Online Resource 6. The total gene number was 61,292 and average gene length was 2.1 Kb; average CDS (coding sequence) was 761.4 bp; average exon number per gene was 3.1 and average exon length was 241.3 bp; average intron length was 605.1 bp. All predicted protein-coding genes were annotated compared with various databases. Gene ontology (GO) annotation was completed by Blast2GO software. Kyoto Encyclopedia of Genes and Genomes (KEGG) ortholog and Pathway annotation was mainly completed by KEGG's automated annotation system and bidirectional best hit (BBH) method was used. The annotation of protein-coding genes based on the GO database and KEGG database can be seen in Online Resource 7 and Online Resource 8.

SNP calling

After generating a comprehensive reliable reference, all quality-filtered reads were then aligned to the de novo reference assembly using the BWA-MEM (Burrows–Wheeler aligner-maximum exact matches) algorithm [BWA (Burrows-Wheeler Alignment tool); Li and Durbin (2009)] and outputted as SAM files. BAM (Binary Alignment/Map format) files were generated from SAM (Sequence Alignment/Map format) files using the SAMtools toolkit (Li et al. 2009). The mapping ratios of sampled reads to the reference genome can be seen in Online Resource 9. Reads of all individuals were clustered with pstacks in Stacks software package (Catchen et al. 2013b) to obtain the number of loci. The major parameter m was set as 4. The loci number of each sample and the average sequencing depth of loci can be seen in Online Resource 10. Following the alignment, we performed SNP calling by using the Stacks software package. Several quality controls were considered for SNP calling: (i) a catalog across all individuals was constructed allowing two mismatches between catalog loci (n 2); (ii) loci present in all eight populations and genotyped in at least 75% of the individuals of each population (populations p 8, r 0.75) was used; (iii) minor allele frequencies (MAF) was set as 0.05. In this study, a total of 458,310 SNPs were detected in all individuals, and SNPs counts for each individual can be seen in Online Resource 11.

Statistical analysis

DdRAD-seq data were analyzed with the population genetics program Stacks. Population genetic statistics including observed heterozygosity, expected heterozygosity, number of haplotypes, haplotype diversity and nucleotide diversity were estimated, and the pairwise differentiation between populations (F_{ST}) also was estimated. Analysis of the molecular variance (AMOVA) was implemented in Arlequin (version 3.5.2.2) to measure the proportions of genetic variation within and among populations. The association between genetic distance (F_{ST} between populations) and geographic distance was assessed using Mantel's test with 10,000 random permutations implemented in the software package Vegan in R 3.2.0 (The R Foundation for Statistical Computing, 2015). Correlation analysis was performed by using SigmaPlot 10.0 (Systat Software, Inc., 2006). Phylogenetic trees were constructed with the software MEGA (available from www.megasoftware.net) using neighbor-joining (NJ) with p-distance model. A total of 7495 homozygous SNP loci (SNPs with MAF value less than 0.05 were removed) were used. Bootstrap (1,000 repetitions) was performed to assess tree reliability. Principle component analysis (PCA) was performed based on all genotyped SNPs using GCTA (Genome-wide Complex Trait Analysis) (http://cnsgenomics.com/software/gcta/).

In order to obtain independent loci with linkage disequilibrium, one SNP was selected for each tag, with a total of 68,625 SNP loci obtained for analysis of population genetic structure. The extent of population structuring was examined using a Bayesian clustering method implemented in software Structure (https://web.stanford.edu/group/pritchardlab/structure.html) (Pritchard et al. 2000). The number of clusters (K) was set from 2 to 10, and each value of K was determined over 8 runs with an admixture model using a burn-in length of 100,000 iterations and 1,000,000 Markov Chain Monte Carlo (MCMC) iterations. The optimal value of K was determined by examination of the Evanno's ΔK statistics (Evanno et al. 2005).

A sliding-window approach (20 kb windows sliding in 1 kb steps) was applied to quantify polymorphism levels ($\theta\pi$, pairwise nucleotide variation as a measure of variability) and genetic differentiation (F_{ST}) between *Q. variabilis* populations at P-rich sites (PR) and at P-deficient sites (PD) (Li et al. 2013). To detect regions with significant signatures of selective sweep (Li et al. 2013), we considered the distribution of the $\theta\pi$ ratios ($\theta\pi_{,PR}/\theta\pi_{,PD}$) and F_{ST} values. We used an empirical procedure and selected windows simultaneously with significant low and high $\theta\pi$ ratios (the 5% left and right tails) and significant high F_{ST} values (the 5% right tail) of the empirical distribution as regions with strong selective sweep signals along the genome, which should harbor genes that underwent a selective sweep. Then, genes in regions with strong selective sweep signals were annotated using the GO and KEGG databases, and each gene was classified according to GO categories and the KEGG pathways and KEGG Brite function databases.

Results

Genetic diversity and genetic structure of *Quercus variabilis* populations at P-rich and P-deficient sites

Among all the populations, soil P content of the four populations (LT, ZW, LKY and AN) at P-rich sites was significantly higher than that of the four populations (FL, ZT, QR and JB) at P-deficient sites; AN population had the highest soil P content, while JB and QR populations had much lower soil P content (Table 1). Genetic diversity (expected heterozygosity, nucleotide diversity and haplotype diversity) in the four *Q. variabilis* populations growing at P-rich sites was generally lower than those four *Q. variabilis* populations growing at P-deficient sites (Table 1). Among the eight *Q. variabilis* populations, AN population showed the lowest observed heterozygosity, expected heterozygosity, nucleotide diversity and haplotype diversity, while JB and QR

Table 1 Geographic locations, soil P content (mg g^{-1}) and genetic statistics of *Quercus variabilis* populations from P-rich sites (LT, ZW, LKY and AN) and P-deficient sites (QR, JB, ZT and FL) in central Yunnan, estimated from ddRAD-seq data

Population	Latitude (°)	Longitude (°)	Soil P content	Number of genotyped individuals	Observed heterozygosity	Expected heterozygosity	Nucleotide diversity	Number of haplotypes	Number of unique haplotypes	Haplotype diversity
FL	25.247	101.536	0.38	10.40	0.1216 ± 0.0003	0.1248 ± 0.0002	0.1311 ± 0.0002	1,549,638	241,023	0.6442
ZT	25.450	101.464	0.46	9.57	0.1228 ± 0.0003	0.1197 ± 0.0002	0.1264 ± 0.0002	1,747,036	226,055	0.6074
QR	25.075	101.621	0.22	11.23	0.1482 ± 0.0003	0.1457 ± 0.0003	0.1526 ± 0.0003	1,108,344	243,247	0.7371
JB	25.065	101.578	0.31	6.70	0.1460 ± 0.0003	0.1458 ± 0.0003	0.1576 ± 0.0003	1,590,184	224,540	0.7649
LT	25.148	102.759	1.26	9.55	0.1190 ± 0.0003	0.1200 ± 0.0002	0.1267 ± 0.0002	1,591,456	226,922	0.6217
ZW	25.147	102.749	1.33	9.71	0.1238 ± 0.0003	0.1183 ± 0.0002	0.1248 ± 0.0002	1,830,534	218,105	0.5944
LKY	25.084	102.768	1.31	9.52	0.1182 ± 0.0003	0.1180 ± 0.0002	0.1246 ± 0.0002	1,570,638	227,728	0.6079
AN	24.944	102.483	1.73	9.24	0.1147 ± 0.0003	0.1042 ± 0.0002	0.1102 ± 0.0002	1,588,150	198,367	0.5330

populations showed the highest observed heterozygosity, expected heterozygosity, nucleotide diversity and haplotype diversity (Table 1). Among all the measured chemical elements present in the soil, only P's concentration was significantly and negatively related to the expected heterozygosity, nucleotide diversity and haplotype diversity ($p < 0.05$) (Table 2; Online Resource 12).

Pairwise F_{ST} analyses indicated relatively low differentiation for the eight *Q. variabilis* populations ($F_{ST} = 0.028–0.073$) (Wright 1978); the greatest genetic differentiation was detected between the AN and JB populations ($F_{ST} = 0.073$); while the LT and LKY populations showed the smallest pairwise F_{ST} (0.028) (Table 3). F_{ST} values indicated very low levels ($F_{ST} < 0.05$) of genetic differentiation among ZW, LT and LKY populations and between FL and all other populations and indicated certain levels ($F_{ST} = 0.05–0.15$) of genetic differentiation between JB or QR and the four *Q. variabilis* populations growing at P-rich sites (AN, ZW, LT and LKY). Moreover, certain levels of genetic differentiation also were detected between AN and ZW or LT population (Table 3). Applying Mantel's test showed the significant positive relationship ($r = 0.46$, $p < 0.001$) between the genetic distance (F_{ST}) and geographical distance of *Q. variabilis* populations at P-rich and P-deficient sites (Online Resource 13).

When geographical population was not considered, AMOVA was used to measure the proportions of genetic variation within and among two groups, one including individuals from sampling locations occurring on P-rich soils and the other including individuals from sampling locations occurring on P-deficient soils (Table 4). AMOVA revealed that the majority of genetic variation occurred within populations (97.5%) and that genetic variation among populations constituted 2.5%. Moreover, significant genetic divergence ($p < 0.0001$) was found between the two P-type groups (P-rich and P-deficient), although the degree of genetic differentiation was rather low ($F_{ST} = 0.025$) (Table 4). When geographical population was taken into account, hierarchical AMOVA was used to measure the proportions of genetic variation within and among two groups, with one including individuals occurring at P-rich sampling locations separated by each of the four populations and another including individuals occurring at P-deficient sampling locations separated by each of the four populations. Hierarchical AMOVA revealed that the majority of genetic variation occurred within populations (95.76%, $p < 0.0001$), and that genetic variation among populations constituted 2.27% ($p < 0.05$), among two P-type groups constituted 1.97% ($p < 0.05$) (Table 4).

Based on the filtered 7495 homozygous SNP loci among individuals, 80 phylogenetic trees of *Q. variabilis* individuals from eight populations were obtained by merging branches with bootstrap value less than 50 (Fig. 2). The

Table 2 Pearson correlations of expected heterozygosity, nucleotide diversity and haplotype diversity for *Quercus variabilis* populations sampled at different field sites with concentrations of soil elements

Pearson correlation	Expected heterozygosity		Nucleotide diversity		Haplotype diversity	
	r	*p*	*r*	*p*	*r*	*p*
N	−0.553	0.156	−0.554	0.154	−0.538	0.169
P	−0.823	0.012	−0.807	0.015	−0.809	0.015
K	0.008	0.986	0.031	0.942	0.066	0.876
Ca	−0.474	0.235	−0.469	0.241	−0.450	0.263
Mg	−0.210	0.618	−0.204	0.629	−0.144	0.734
Fe	−0.487	0.221	−0.487	0.221	−0.488	0.220
Mn	−0.545	0.162	−0.547	0.160	−0.537	0.170
Zn	−0.363	0.377	−0.351	0.394	−0.331	0.423
Cu	−0.287	0.491	−0.287	0.491	−0.274	0.511
Al	−0.702	0.052	−0.691	0.058	−0.689	0.059
Na	−0.305	0.462	−0.304	0.465	−0.289	0.487

Table 3 Pairwise genetic differentiation between 8 *Quercus variabilis* populations from P-rich and P-deficient sites in central Yunnan

	ZT	QR	JB	LT	ZW	LKY	AN
FL	0.0335	0.0456	0.0480	0.0360	0.0391	0.0346	0.0450
ZT		0.0532	0.0581	0.0379	0.0401	0.0358	0.0487
QR			0.0407	0.0529	0.0556	0.0533	0.0649
JB				0.0584	0.0614	0.0582	0.0731
LT					0.0283	0.0280	0.0504
ZW						0.0305	0.0519
LKY							0.0479

Table 4 Results of AMOVA (analysis of molecular variance) with respect to *Quercus variabilis* populations within and among two groups, i.e., P-rich sites and P-deficient sites based on SNPs (with or without populations according to sites)

Source of variation	*df*	Sum of squares	Variance components	Percentage of variation (%)	*p* value
a. Without populations					
Among groups	1	99.75	1.26285	2.5	
Within groups	78	3840.4	49.2359	97.5	
Total	79	3940.15	50.49875		
F_{ST}	0.02501				0.00000
b. With populations					
Among groups	1	99.75	0.99289	1.97	0.02639
Among populations within groups	6	358.6	1.14752	2.27	0.01662
Within populations	72	3481.8	48.35834	95.76	0.00000
Total	79	3940.15	50.49875		

results showed that 80 individuals of *Q. variabilis* did not aggregate strictly according to the location of geographical populations, nor did they aggregate separately according to P-type of soil (-deficient vs -rich) (Fig. 2). There was a mixing phenomenon between individuals at P-deficient and P-rich sites. The four P-rich populations and FL population, occurring on a P-deficient sampling location, were basically mixed, but most individuals of JB, QR and ZT populations (all occurring on P-deficient sampling locations) were aggregated. Genetic distances between almost all individuals in the 8 populations of *Q. variabilis* indicate clustering, and only some individuals of P-deficient populations JB and QR are more isolated from the others (Fig. 2).

Principle component analysis for 80 *Q. variabilis* individuals from the eight populations was performed based on all genotyped 458,310 SNPs. The plot of the first and second component accounted for 10.09% and 3.51% of the variation, respectively, giving a cumulative variation of 13.6% (Fig. 3). JB and QR appeared obviously differentiated from the other 6 populations along the first principle component axis; AN was obviously distinguished from ZW and LT populations along the second principle component axis (Fig. 3).

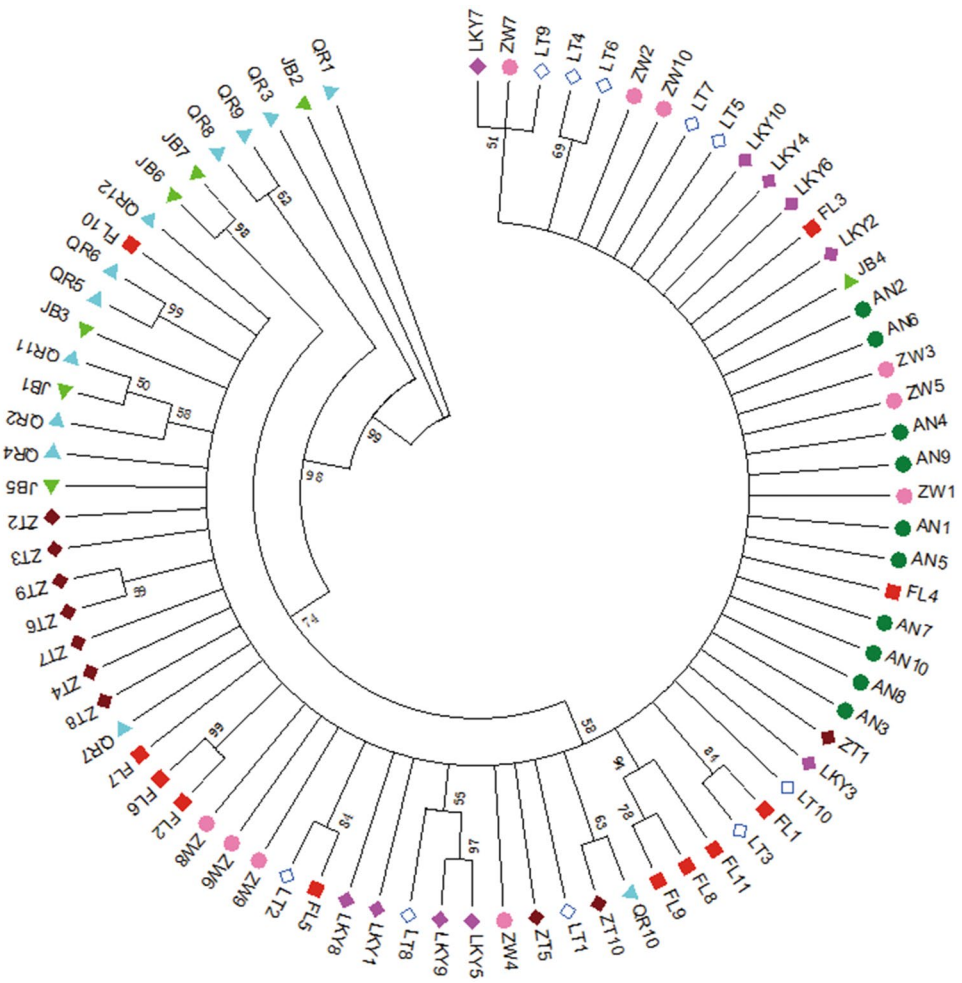

Fig. 2 The neighbor-joining phylogenetic tree of 80 *Quercus variabilis* individuals in 8 populations from P-rich (LT, ZW, LKY and AN) and P-deficient (QR, JB, ZT and FL) sites in central Yunnan. Numbers are percentage values over 1000 bootstrap replicates and only bootstrap values over 50% are shown

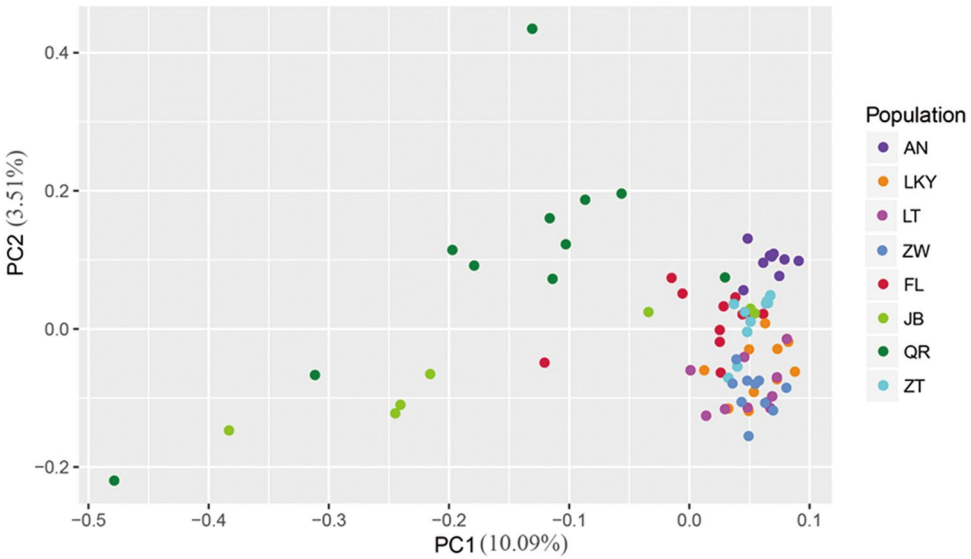

Fig. 3 Principle component analysis for 8 *Quercus variabilis* populations based on all genotyped SNPs

Structure model clustering analysis based on the filtered 68,625 SNPs was used for detection of genetic relationships among the eight *Q. variabilis* populations (Fig. 4). From $K=2$ to $K=10$, LKY, LT and ZW populations consistently

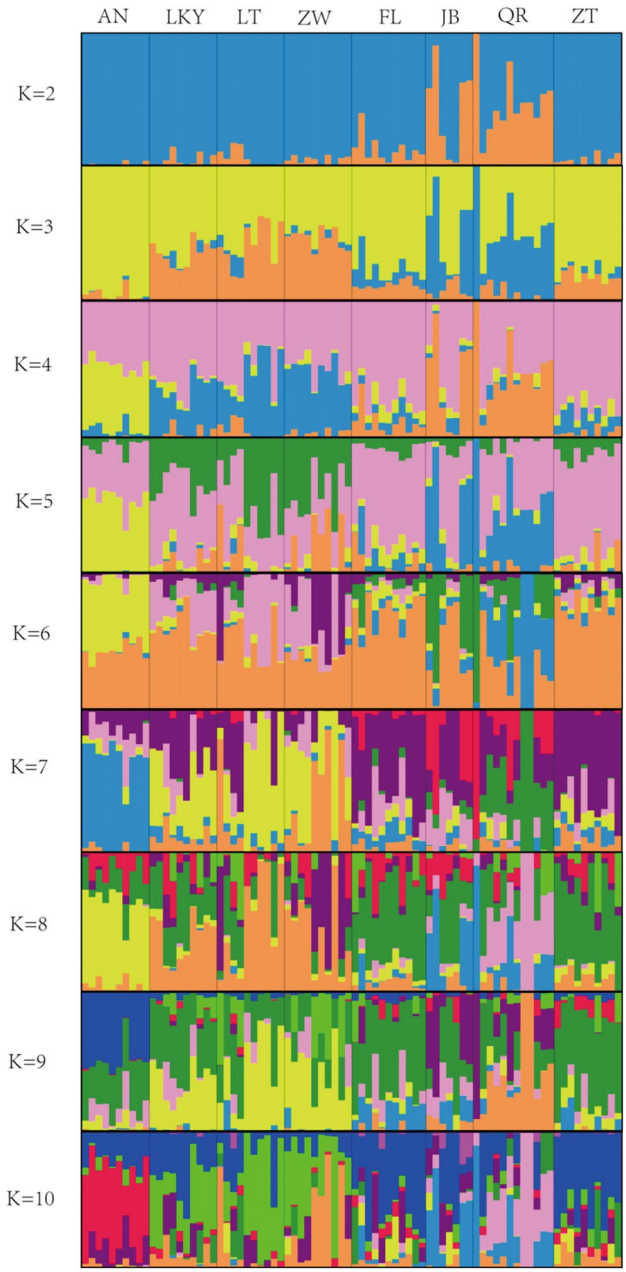

Fig. 4 Structure clustering of 8 *Quercus variabilis* populations

showed similar genetic composition proportion; ZT and FL also consistently showed similar genetic composition proportion; however, JB and QR populations showed similar genetic composition which differed considerably from other populations; and genetic composition of AN population was consistently distinct from all other populations (Fig. 4). Moreover, according to the ΔK statistics, the most likely K value was 3 (Online Resource 14), which meant that the *Q. variabilis* populations were structured into three genetic clusters; one cluster was composed of AN populations, a second cluster involved JB and QR populations, and a third cluster involved LKY, LT, ZW, FL and ZT populations (Fig. 4).

Selection signal analysis of *Quercus variabilis* populations at P-rich and P-deficient sites

First, we grouped the four populations at P-rich sampling locations into a single population (PR) and grouped the four populations at P-deficient sampling locations into another population (PD). For the PR and PD populations, the selection signal analysis combined by F_{ST} and $\theta\pi$ ratios ($\theta\pi_{,PR}/\theta\pi_{,PD}$) showed that the threshold of regions with strong selective sweep signals along the genome was $F_{ST} > 0.089$ and $\theta\pi \leq 0.497/\theta\pi \geq 1.254$ (Fig. 5). Genomic regions with strong selective sweep signals in PR population were 888 Kb, or 0.11% of the genome and containing 38 genes, and that in PD population were 234 Kb, or 0.03% of the genome and containing eight genes.

Gene ontology database analysis showed that genes in regions with strong selective sweep signals in PR population were mainly involved in biological process and molecular function such as signal transduction, translation, phosphatase activity, DNA binding, structural molecule activity, kinase activity and so on (Table 5). GO analysis showed that genes contained in selected regions in PD population were mainly involved in biological process such as protein maturation, transport, small molecule metabolic process and so on (Table 5). KEGG analysis showed that genes contained in selected regions in PR population were mainly involved in metabolism and genetic information processing such as protein phosphatase and associated proteins, transcription factors and ubiquitin system (Table 6). KEGG analysis showed that genes contained in selected regions in PD population were involved in oxidative phosphorylation, pantothenate and CoA biosynthesis, terpenoid backbone biosynthesis and so on (Table 6).

Second, we found that the degree of genetic differentiation among all the populations was the highest between the AN and JB populations, respectively, occurring at P-rich and P-deficient sampling locations, and the soil P content between the two populations was very different. Thus, we also selected AN and JB populations for selection signal analysis, and through the F_{ST} and $\theta\pi$ ($\theta_{\pi,AN}/\theta\pi_{,JB}$) testing, the threshold of regions with strong selective sweep signals along the genome was $F_{ST} > 0.1919$ and $\theta\pi \leq 0.297/\theta\pi \geq 1.333$ (Fig. 6). Genomic regions with strong selective sweep signals in AN population were 2.464 Mb, or 0.29% of the genome and containing 91 genes, and that in JB population were 353 Kb, or 0.042% of the genome and containing 12 genes.

GO analysis showed that genes contained in selected regions in AN population were mainly involved in biological process and molecular function such as small molecule

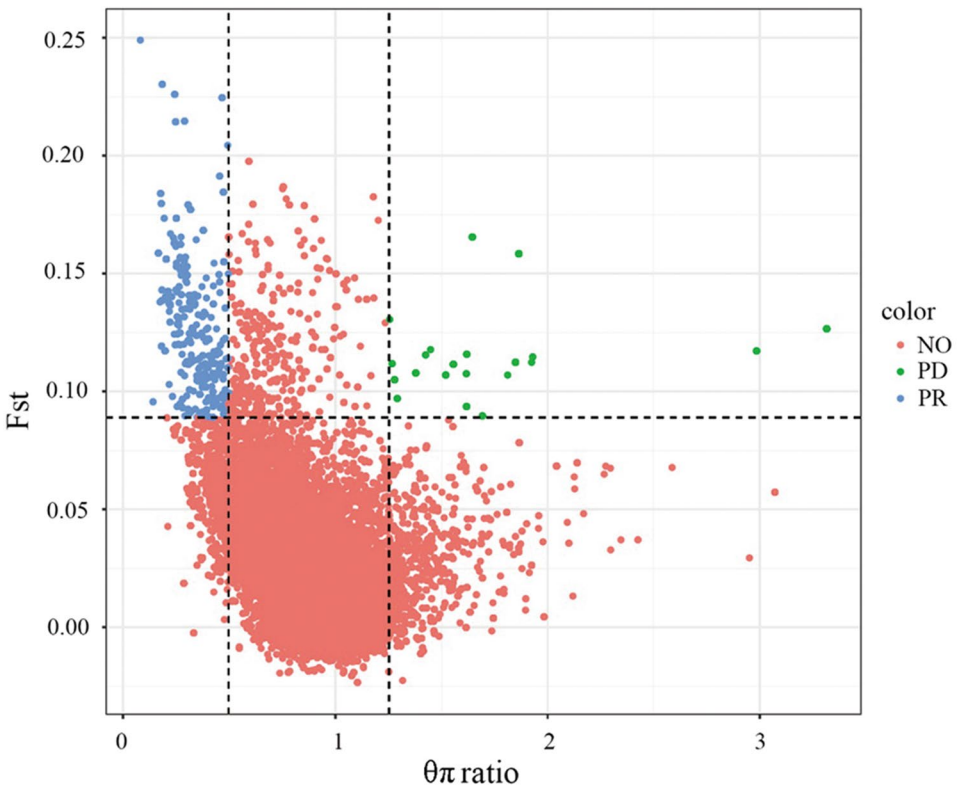

Fig. 5 Genomic regions with strong selective sweep signals in PR and PD populations of *Quercus variabilis*. Horizontal axis and vertical axis indicated $\theta\pi$ ratio ($\theta\pi$, $_{PR}/\theta\pi$, $_{PD}$) and F_{ST}, respectively. Data points located to the left and right of the left and right vertical dashed lines, respectively (corresponding to the 5% left and right tails of the empirical θπ ratio distribution, where the θπ ratios are 0.497 and 1.254, respectively), and above the horizontal dashed line (the 5% right tail of the empirical F_{ST} distribution, where F_{ST} is 0.089) were identified as selected regions for PR populations (blue points) and PD populations (green points), respectively

metabolic process, signal transduction, carbohydrate metabolic process, transport, transferase activity, RNA binding, kinase activity and so on (Table 7). GO analysis showed that genes contained in selected regions in JB population were involved in biosynthetic process, DNA binding, and so on (Table 7). KEGG analysis showed that genes contained in selected regions in AN population were mainly involved in metabolism and genetic information processing such as lysine degradation, transcription factors, ubiquitin mediated proteolysis, ubiquitin system and so on (Table 8). KEGG analysis showed that genes contained in selected regions in JB population were involved in transcription factors, RNA transport, ubiquitin system and so on (Table 8).

Discussion

The present study is the first effort to characterize the effects of available P at geological P-rich and P-deficient sites on genetics of plants using deciduous *Q. variabilis* in subtropical areas, showing novel insights into how plants adapt to variable nutrient sites in subtropical areas.

Lower genetic diversity of *Quercus variabilis* populations at P-rich sites

Our results showed that the haplotype diversity (*Hd*) of *Q. variabilis* across P-rich and P-deficient sites was 0.533–0.765, suggesting that *Q. variabilis* populations had a high level of genetic diversity. These results were in agreement with the previous studies. Chen et al. (2012) studied the genetic diversity and genetic structure of 528 samples from 50 populations of *Q. variabilis* across East Asia (including Mainland China, Zhoushan Islands, Taiwan Island as well as Korean Peninsula and Japan Islands) (*Hd* = 0.888). Xu et al. (2004) showed average expected heterozygosity (*He*) of 0.8059 using SSR markers for five *Q. variabilis* populations in central China. In the present study, ddRAD-seq analysis showed 0.1042–0.1458 of *He* and 0.1147–0.1482 of *Ho* (average observed heterozygosity) for 8 *Q. variabilis* populations across P variable content sites; these values are significantly lower than those measured using microsatellite markers (Xu et al. 2004; Wei 2015).

Furthermore, interestingly, our results showed that there was lower genetic diversity for *Q. varaibilis* populations at P-rich sites compared with those at P-deficient sites, and the genetic diversity was correlated with P, but not with other elements. It has been shown that genetic diversity is related to habitat complexity (Abeysinghe et al. 2000), rainfall (Jin et al. 1998), climate change (Pauls et al. 2013),

Table 5 Annotation of genes contained in selected regions in PR and PD populations of *Quercus variabilis* on the basis of the GO database

GO ID	Gene function	Gene number	Gene classification
PR population			
GO:0000003	Reproduction	1	Biological_process
GO:0007165	Signal transduction	1	
GO:0006412	Translation	1	
GO:0009058	Biosynthetic process	2	
GO:0034641	Cellular nitrogen compound metabolic process	2	
GO:0006457	Protein folding	2	
GO:0006464	Cellular protein modification process	3	
GO:0005576	Extracellular region	1	Cellular_component
GO:0005886	Plasma membrane	1	
GO:0005623	Cell	2	
GO:0005737	Cytoplasm	2	
GO:0005622	Intracellular	2	
GO:0043226	Organelle	2	
GO:0016853	Isomerase activity	1	Molecular_function
GO:0016791	Phosphatase activity	1	
GO:0003677	DNA binding	1	
GO:0019899	Enzyme binding	1	
GO:0030234	Enzyme regulator activity	1	
GO:0005198	Structural molecule activity	1	
GO:0051082	Unfolded protein binding	1	
GO:0016301	Kinase activity	2	
GO:0043167	Ion binding	5	
PD population			
GO:0009056	Catabolic process	1	Biological_process
GO:0051186	Cofactor metabolic process	1	
GO:0006091	Generation of precursor metabolites and energy	1	
GO:0051604	Protein maturation	1	
GO:0006950	Response to stress	1	
GO:0006810	Transport	1	
GO:0009058	Biosynthetic process	3	
GO:0034641	Cellular nitrogen compound metabolic process	3	
GO:0044281	Small molecule metabolic process	3	
GO:0005623	Cell	1	Cellular_component
GO:0005622	Intracellular	1	
GO:0004386	Helicase activity	1	Molecular_function
GO:0016301	Kinase activity	1	
GO:0016491	Oxidoreductase activity	1	
GO:0043167	Ion binding	2	

artificial disturbance (Farwig et al. 2008; Epps et al. 2010), slope direction (Jin et al. 1998) and soils (Li and Peng 2001; Xu et al. 2003). Chen and Song (1997) found that the clonal diversity of *Cyclobalanopsis glauca* populations was negatively correlated with soil total nitrogen (N). Li and Peng (2001) found that the genetic diversity of *Pinus massoniana* population was negatively correlated with the total N content in soils. Xu et al. (2003) observed that for *Reaumuria soongorica* populations in Fuliang desert, Xinjiang, the genetic diversity was negatively correlated with total P and Cl^{-1} content in the soils, and positively correlated with CO_3^{2-} content. In our study, P content in soils of JB and QR populations which located in P-deficient site was much lower compared with other populations, while the genetic diversity level of JB and QR population was much higher. Therefore, this might be due to the fact that *Q. variabilis* used restrictive nutrients (e.g., P) in the soil more fully (Raghothama and Karthikeyan 2005; Vance et al. 2003; Hammond et al. 2004; Zhang et al. 2013).

Table 6 Annotation of genes contained in selected regions in PR and PD populations of *Quercus variabilis* on the basis of the KEGG database

Gene definition	Gene function	Ko_ID	Gene classification
PR population			
tRNA dimethylallyltransferase	Prenyltransferases	ko01006	Protein families: metabolism
tRNA dimethylallyltransferase	Zeatin biosynthesis	ko00908	Metabolism of terpenoids and polyketides
Protein phosphatase inhibitor 2	Protein phosphatase and associated proteins	ko01009	Protein families: metabolism
Homeobox-leucine zipper protein	Transcription factors	ko03000	Protein families: genetic information processing
Large subunit ribosomal protein L17e	Ribosome	ko03010	Translation
tRNA dimethylallyltransferase	Transfer RNA biogenesis	ko03016	Protein families: genetic information processing
Calnexin	Chaperones and folding catalysts	ko03110	Protein families: genetic information processing
Calnexin	Protein processing in endoplasmic reticulum	ko04141	Folding, sorting and degradation
Calnexin	Membrane trafficking	ko04131	Protein families: genetic information processing
E3 ubiquitin-protein ligase CCNP1IP1	Ubiquitin system	ko04121	Protein families: genetic information processing
Protein phosphatase inhibitor 2	Chromosome and associated proteins	ko03036	Protein families: genetic information processing
WD repeat-containing protein 48	DNA repair and recombination proteins	ko03400	Protein families: genetic information processing
DNA cross-link repair 1B protein	DNA repair and recombination proteins	ko03400	Protein families: genetic information processing
Calnexin	Lectins	ko04091	Protein families: signaling and cellular processes
Calnexin	Phagosome	ko04145	Transport and catabolism
PD population			
NADH-ubiquinone oxidoreductase chain 5	Oxidative phosphorylation	ko00190	Energy metabolism
Type II pantothenate kinase	Pantothenate and CoA biosynthesis	ko00770	Metabolism of cofactors and vitamins
STE24 endopeptidase	Terpenoid backbone biosynthesis	ko00900	Metabolism of terpenoids and polyketides
STE24 endopeptidase	Peptidases	ko01002	Protein families: metabolism
NADH-ubiquinone oxidoreductase chain 5	Mitochondrial biogenesis	ko03029	Protein families: genetic information processing

Moreover, the AN population was located in P-rich site developed from P-rich phosphorus ores, and the P content in soil was the highest compared with other populations, while the genetic diversity level of AN population was the lowest. This seemed consistent with other findings, such as the reduced level of genetic diversity of plants grown in metal-contaminated soils in mining areas (Mengoni et al. 2001; Deng et al. 2007; Nkongolo et al. 2007). At the same time, it also indicated that *Q. variabilis* might be selected by the enrichment effect of geological P, which might affect the genetic structure of *Q. variabilis* populations.

Genetic structure and variation of *Quercus variabilis* populations across P-deficient and P-rich sites

It has been shown that there is higher genetic variation within populations rather than among populations for some oak species, e.g., *Q. petraea* (Bruschi et al. 2003), *Q. macrocarpa* (Craft and Ashley 2007) and *Q. aquifolioides* (Cheng et al. 2017). Similar results were obtained on *Q. variabilis* and *Q. acutissima* populations by both the means of allelic enzyme markers (Chung et al. 2002; Zhou et al. 2008) and DNA molecular marker techniques (such as SSR, ISSR and AFLP) (Lefort et al. 1998; Xu et al. 2004; Wang et al. 2005; Löpez-Aljorna et al. 2007). Consistent with the above results, our SNP marker study showed that the majority of genetic variation (95.76%) occurred within the eight *Q. variabilis* populations.

At the genome level, the overall genetic differentiation among eight populations of *Q. variabilis* was low: the average F_{ST} value was 0.047 (ranging from 0.028 to 0.073). Based on the island model hypothesis of Wright (1931) ($F_{ST}=1/(1+4Nm)$), for *Q. variabilis* populations in the present study, the inter-population genetic flow (Nm, migration

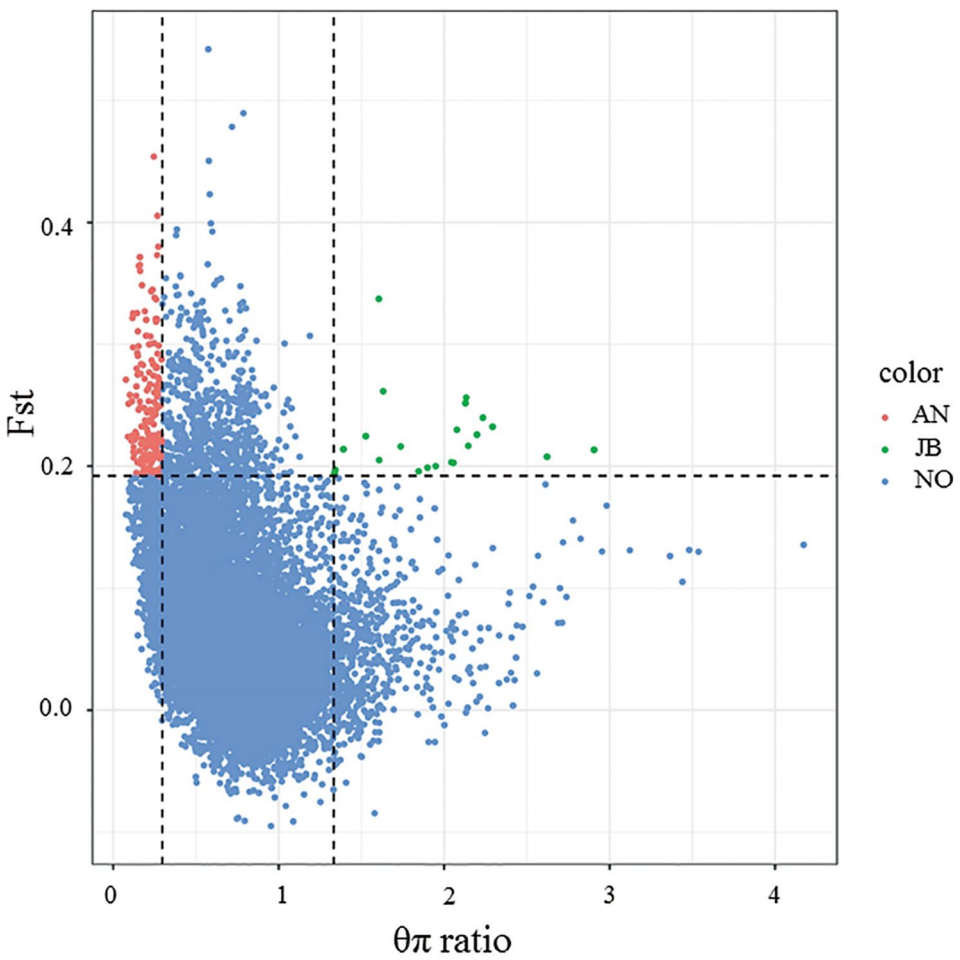

Fig. 6 Genomic regions with strong selective sweep signals in AN and JB populations of *Quercus variabilis*. Horizontal axis and vertical axis indicated θπ ratio ($\theta\pi_{,AN}/\theta\pi_{,JB}$) and F_{ST}, respectively. Data points located to the left and right of the left and right vertical dashed lines, respectively (corresponding to the 5% left and right tails of the empirical θπ ratio distribution, where the θπ ratios are 0.297 and 1.333, respectively), and above the horizontal dashed line (the 5% right tail of the empirical F_{ST} distribution, where F_{ST} is 0.1919) were identified as selected regions for AN populations (red points) and JB populations (green points), respectively

number per generation) was 5.089 (ranging from 3.170 to 8.679), which was similar in size to the general wind-pollinated plant gene flow (Nm = 5.24) (Hamrick et al. 1995). Although estimating Nm from F_{ST} should be carefully considered (Whitlock and Mccauley 1999), our calculation only shows that the migration rate between populations of *Q. variabilis* is higher than one effective migration per generation, and gene exchange is frequent. With the biological features of *Q. variabilis*, such as wide continuous distribution, long generation cycle, outcrossing, wind pollination and perennial woody plants (Hamrick and Godt 1989; Hamrick et al. 1992), as well as the frequent gene exchange among populations due to smaller pollen and longer transmission distance, may be the reason for the high genetic diversity and low genetic differentiation of *Q. variabilis* (Xu 2002).

Although the degree of inter-population differentiation of *Q. variabilis* was low, the analysis of molecular variance showed that the inter-population differentiation was significant, and the inter-population differentiation between P-rich and P-deficient sites was also significant. Mantel test showed that the genetic variation between *Q. variabilis* populations might be caused by geographical isolation, and there was a geographical isolation effect between populations. The cluster analysis of NJ phylogenetic tree, principal component analysis and genetic structure model based on SNP markers showed that eight populations of *Q. variabilis* could be divided into three clusters, one group of AN population, one group of JB and QR population, and one group of FL, ZT, LKY, LT and ZW population. According to Pairwise F_{ST} statistics of Wright (1978), the same results were found. The genetic divergence among the three clusters was large, but the genetic divergence within each cluster was small. Therefore, the population of *Q. variabilis* does not seem to be clustered strictly according to geographical location.

Moreover, in our study, the soil P content of AN population at P-rich site was the highest, while that of QR and JB population in P-deficient site was the lowest (Table 1). The soil P content of FL, ZT, LKY, LT and ZW populations was in the middle (Table 1). Therefore, the populations of *Q. variabilis* seemed to cluster into three clusters strictly according to the difference of soil P content, and the genetic differentiation level among the site populations with the greatest difference of soil P content was also the highest (Table 3). The difference of soil P content might be an

Table 7 Annotation of genes contained in selected regions in AN and JB populations of *Quercus variabilis* on the basis of the GO database

GO ID	Gene function	Gene number	Gene classification
AN population			
GO:0009056	Catabolic process	1	Biological_process
GO:0051276	Chromosome organization	1	
GO:0051604	Protein maturation	1	
GO:0044281	Small molecule metabolic process	1	
GO:0034641	Cellular nitrogen compound metabolic process	2	
GO:0007165	Signal transduction	2	
GO:0009058	Biosynthetic process	3	
GO:0005975	Carbohydrate metabolic process	3	
GO:0006810	Transport	3	
GO:0006464	Cellular protein modification process	4	
GO:0005654	Nucleoplasm	1	Cellular_component
GO:0043234	Protein complex	1	
GO:0005623	Cell	3	
GO:0005622	Intracellular	3	
GO:0043226	Organelle	3	
GO:0016798	Hydrolase activity, acting on glycosyl bonds	1	Molecular_function
GO:0016829	Lyase activity	1	
GO:0008168	Methyltransferase activity	1	
GO:0001071	Nucleic acid binding transcription factor activity	1	
GO:0016746	Transferase activity, transferring acyl groups	1	
GO:0016757	Transferase activity, transferring glycosyl groups	1	
GO:0003723	RNA binding	1	
GO:0016301	Kinase activity	3	
GO:0016491	Oxidoreductase activity	4	
GO:0003677	DNA binding	7	
GO:0043167	Ion binding	7	
JB population			
GO:0009058	Biosynthetic process	1	Biological_process
GO:0034641	Cellular nitrogen compound metabolic process	1	
GO:0001071	Nucleic acid binding transcription factor activity	1	Molecular_function
GO:0003677	DNA binding	2	

important factor in shaping the genetic structure of *Q. variabilis* population, and this might be related to the adaptive selection of *Q. variabilis* to P-rich and P-deficient habitats. In addition, the principal component analysis of *Q. variabilis* population also showed that soil P content was closely related to the genetic structure of *Q. variabilis* population.

Han (2004) reported that nitrogen, organic matters, soil thickness and rainfall played an important role in regulating genetic diversity of *Q. variabilis* populations. Artificial addition experiments also showed the effects of element enrichment on the genetic structure of population. Huff et al. (1998) found that fertilization had a significant effect on the genetic differentiation of little bluestem populations. Deng et al. (2007) also considered that metal pollution seemed to have a stronger effect on the genetic structure of *Sedum alfredii* populations than geographical distance. The plants growing in serpentine-developed soils (Brady et al. 2005), saline soils (Lowry et al. 2008; Baxter et al. 2010) and mine-affected soils (Macnair 1993) have formed a variety of life-history characteristics clines that are driven by genetic variation. Therefore, although there was geographic isolation effect, the genetic structure of *Q. variabilis* populations might show a certain spatial differentiation pattern along the P gradient in soils. Some studies suggested that the genetic basis and breeding system of plants themselves influenced the genetic structure of their populations, but natural selection, genetic drift and gene flow also affected the genetic structure of plant populations (Peterson et al. 2013). Therefore, habitat-based genetic differentiation, distance isolation and reproductive characteristics of *Q. variabilis* populations might together shape the genetic structure of *Q. variabilis* populations.

Table 8 Annotation of genes contained in selected regions in AN and JB populations of *Quercus variabilis* on the basis of the KEGG database

Gene definition	Gene function	Ko_ID	Gene classification
AN population			
Omega-hydroxypalmitate O-feruloyltransferase	Cutin, suberine and wax biosynthesis	Ko00073	Lipid metabolism
Sphinganine-1-phosphate aldolase	Sphingolipid metabolism	Ko00600	Lipid metabolism
Euchromatic histone-lysine N-methyltransferase	Lysine degradation	Ko00310	Amino acid metabolism
Histone-lysine N-methyltransferase SETD2	Lysine degradation	Ko00310	Amino acid metabolism
Cytochrome P450 family 714 subfamily A polypeptide 1	Cytochrome P450	Ko00199	Protein families: metabolism
Transcriptional activator Myb	Transcription factors	Ko03000	Protein families: genetic information processing
EREBP-like factor	Transcription factors	Ko03000	Protein families: genetic information processing
Translation initiation factor 4G	RNA transport	Ko03013	Translation
Translation initiation factor 4G	Messenger RNA biogenesis	Ko03019	Protein families: genetic information processing
Translation initiation factor 4G	Translation factors	Ko03012	Protein families: genetic information processing
Ubiquitin-conjugating enzyme E2 N	Ubiquitin mediated proteolysis	Ko04120	Folding, sorting and degradation
Ubiquitin-conjugating enzyme E2 N	Ubiquitin system	Ko04121	Protein families: genetic information processing
Euchromatic histone-lysine N-methyltransferase	Chromosome and associated proteins	Ko03036	Protein families: genetic information processing
Histone-lysine N-methyltransferase SETD2	Chromosome and associated proteins	Ko03036	Protein families: genetic information processing
Ubiquitin-conjugating enzyme E2 N	DNA repair and recombination proteins	Ko03400	Protein families: genetic information processing
JB population			
Homeobox-leucine zipper protein	Transcription factors	Ko03000	Protein families: genetic information processing
Small ubiquitin-related modifier	RNA transport	Ko03013	Translation
Small ubiquitin-related modifier	Messenger RNA biogenesis	Ko03019	Protein families: genetic information processing
Small ubiquitin-related modifier	Mitochondrial biogenesis	Ko03029	Protein families: genetic information processing
Small ubiquitin-related modifier	Ubiquitin system	Ko04121	Protein families: genetic information processing
Small ubiquitin-related modifier	Cytoskeleton proteins	Ko04812	Protein families: signaling and cellular processes

Population selection signals of *Quercus variabilis* at P-rich and P-deficient sites

Our results clearly showed that the size of selective sweep areas in the genome of PR population was more than three times larger than that of PD population. Meanwhile, the size of selective sweep areas in the genome of AN population was almost seven times larger than that of JB population. Compared with the typical P-deficient *Q. variabilis* populations in subtropical zone which are subjected to natural selection, the P-rich AN population with the highest P content in soil might also be artificially selected by factors such as phosphate mining; therefore, it seemed that artificial selection was more powerful than natural selection in shaping the genome of *Q. variabilis*, and this seemed to be consistent with the previous results, for instance, nucleotide diversity of P-rich AN population (0.1102) was significantly lower than the P-deficient JB population (0.1576) (Table 1).

According to GO gene function annotation, there were obvious differences in the selective sweep areas in the genome between PR and PD populations, and this was true for differences between AN and JB populations. Compared with the PD population, the genes in the selective sweep areas of PR population involve phosphatase activity, which was related to P activation and the separation of Pi from P-containing compounds (Müller et al. 2007). Moreover, the functions of the genes induced in the selective sweep areas of PR population were similar to the functions of the genes expressed in *Arabidopsis* leaves in response to carbohydrates sufficient, especially P sufficient (Müller

et al. 2007). The genes in the selective sweep areas of PR population also involved structural and molecular activities. Compared with JB population, genes in the selective sweep areas of AN population involved kinase, hydrolase, transferase and reductase activity, signal transduction, transport and carbohydrate metabolism, which were consistent with the functions of significant genes involved in P stress response regulation in Müller et al. (2007) and Hammond et al. (2011). Therefore, P-rich site populations might be selected by P content in P-rich site soils, forming the functional characteristics of genes related to the adaptation of P content in soils.

Many studies have shown that a series of transcription factors (TF) play a role in regulating plant response to environmental stress. For instance, the expression of transcription factors in *Arabidopsis thaliana* had tissue specificity (Chen et al. 2002), and the expression of transcription factors was induced in leaves of *A. thaliana* under P stress (Müller et al. 2007). Wu et al. (2003) found that seven homeobox genes were regulated in P-stressed leaves of *A. thaliana* (5 were induced and 2 were inhibited), of which the homeobox-leucine zipper protein HAT5 was significantly regulated, and four AP2/EREBP transcription factors were induced in P-stressed leaves. In addition, many EREBP and homeobox-leucine zipper protein transcription factors also responded to other environmental stresses (Dezar et al. 2005; Zhu et al. 2010). Many transcription factors belong to MYB family. MYB transcription factor genes were induced by different abiotic stresses (Chen et al. 2002), including low P stress (Müller et al. 2007). Wu et al. (2003) also found that *Arabidopsis* MYB-related protein CCA1 gene was significantly induced 24 h under low Pi stress. We found that both P-rich and P-deficient site populations contained transcription factors (e.g., EREBP and homeobox-leucine zipper protein transcription factors) in the selective sweep areas. Therefore, these genes might be related to the response of *Q. variabilis* to P availability or other stress factors in P-rich and P-deficient site soils.

Wu et al. (2003) and Qin et al. (2011) found that a considerable portion of plant regulatory genes responding to Pi stress showed distinct or even opposite expressions between roots and leaves. Based on the leaves of *Q. variabilis*, we found that the genes in the selective sweep areas of P-rich and P-deficient site populations also involved in protein synthesis and degradation. Morcuende et al. (2007) found that some potential regulatory genes responded to P availability had predictive functions of post-translational protein modification (e.g., protein kinase and phosphatase) and protein degradation (e.g., E3 ligase). For example, two tyrosine-specific protein phosphatase transcripts increased strongly during P deficiency and decreased significantly after 3 h of P resupply; four genes encoding E3 ubiquitin ligase-cyclic finger protein were induced and two were inhibited during P stress; and two E3 ubiquitin ligase U-box genes showed 6–10 times higher signal under P stress, and reversed rapidly after P resupply (Morcuende et al. 2007). Müller et al. (2007) also found that genes encoding tyrosine-specific protein phosphatase family proteins were induced by P stress. In our study, the genes in the selective sweep areas of PR population involved protein phosphatase and E3 ubiquitin-protein ligase, which might be related to the regulation of *Q. variabilis* in responding to P availability. In addition, similar to Müller et al. (2007), ribosomal protein genes were included in the selective sweep areas of PR population, which might be related to plant growth and development under P-rich nutrient conditions.

UBC encodes an ubiquitin-binding enzyme, which plays a role in the ubiquitination of proteins. UBC24-encoded ubiquitin-binding enzyme E2 has been proved to be the target gene of microRNA 399 (Kuo and Chiou 2011). The expression of microRNA399 was negatively correlated with UBC24. MicroRNA399 was induced by P starvation, and the expression of E2 binding enzyme gene UBC24 was inhibited, which promoted the expression of P transporter and promoted P uptake. Under P-sufficient conditions, the expression of microRNA399 decreased sharply, but UBC24 was highly abundant, which inhibited the expression of P transporter, thus avoiding P poisoning in plants (Chiou et al. 2006; Kuo and Chiou 2011). It has been observed that Fe deficiency response is mediated by ubiquitin-binding enzyme E2. For example, heterotopic expression of cucumber UBC13 gene in *Arabidopsis thaliana* resulted in the formation of lateral root hairs responded to Fe (Li and Schmidt 2010). Therefore, the genes in the selective sweep areas of AN population include E2 ubiquitin-binding enzyme gene, which may be related to the adaptation of *Q. variabilis* in AN population to nutrient availability (e.g., P or Fe). In Wu et al. (2003), a key cytoplasmic translation initiation factor, eIF2, which is related to protein synthesis, was down-regulated during Pi stress, suggesting that the mechanism of protein synthesis under Pi stress is usually inhibited. Therefore, the regulation of protein synthesis genes may be related to the adaptation of *Q. variabilis* to P-rich habitats.

Peptide modification has been proved to be a broad mechanism of biochemical pathways in plants and animals. Small ubiquitin-related modifier (SUMO) is an important member of the ubiquitin family (Kerscher et al. 2006). It has different effects compared with ubiquitination (Johnson 2004). SIZ1 is a SUMO E3 ligase in *Arabidopsis thaliana*. SIZ1 has been proved to be a key factor in regulating Pi stress response, and also involved in other stress responses (Miura et al. 2005, 2013; Colby et al. 2006; Yoo et al. 2006). Therefore, compared with AN population, the genes in the selective sweep areas of JB population include small ubiquitin-related modifier genes, which may be related to P stress

and other stress adaptation of *Q. variabilis* in JB population. Morcuende et al. (2007) found that most of the major amino acid contents of *Arabidopsis* seedlings remained unchanged or slightly increased under P deficiency. After 8 h of P resupply, some amino acid contents decreased, such as lysine, even lower than that of P-sufficient seedlings. The genes in the selective sweep areas of AN population involved lysine degradation function, so lysine content in leaves of *Q. variabilis* in P-rich sites may also be lower than that in P-deficient sites. Cytochrome P450 superfamily is the largest family of enzymes and proteins in plants. It closely participates in the regulation of plant hormone metabolism and directly acts on plant growth and development. Cytochrome P450 in some plants is involved in the pathway of gibberellin metabolism, such as CYP88A and CYP714A (Zhang et al. 2011; Jun et al. 2015). At the same time, some studies have found that Pi availability affects the expression of cytochrome P450-related genes (Misson et al. 2005; Wasaki et al. 2006). Compared with the JB population, the genes in the selective sweep areas of AN population include CYP714A1, suggesting that under P-rich conditions, the effect of gibberellin metabolic pathway may have a functional impact on the survival of *Q. variabilis* in AN population.

Conclusions

We examined the genetic characteristics of *Q. variablis* populations at P-rich and P-deficient sites, to demonstrate how plants have adapted to geological-derived variable P soils in subtropical areas. First, *Q. variabilis* populations had the high genetic diversity across geological-derived variable P sites, but the genetic diversity was higher for the populations at P-deficient sites compared with those at P-rich sites. Furthermore, the level of genetic diversity was only negatively correlated with soil P content. The degree of genetic differentiation is lower among *Q. variabilis* populations compared with that among individuals within the population, indicating that the gene exchange between populations of *Q. variabilis* is frequent. Second, *Q. variabilis* populations could be divided into three clusters according to soil P content at sites, and the genetic differentiation degree was the greatest between *Q. variabilis* populations at the richest P sites and the most deficient P sites. Third, there are differences in gene functions in the selective sweep areas between P-rich and P-deficient sites. Of genes in the selective sweep areas, some candidate genes involved transcription factors, protein synthesis and degradation, and might play a predominant role in regulating P availability and other stress factors of *Q. variabilis*. These results indicate that the availability of P in soils could be a key selection pressure leading to adaptive genetic differentiation among *Q. variabilis* populations.

Information on Electronic Supplementary Material

Online Resource 1. Means and standard deviations (SD) of chemical compositions (%) of phosphate rocks in Anning County (based on Tao (2005)) and non-phosphate rocks (the Matoushan Formation rocks) in Mouding County (based on Shi et al. (2011)), in Yunnan Province. The absence of the measurement data in the table is indicated by a single slash (/).
Online Resource 2. Raw Illumina data for each *Quercus variabilis* sample.
Online Resource 3. High-quality data for each *Quercus variabilis* sample. HQ reads %, percentage of high-quality reads on raw Illumina reads; HQ data (bp), bases of high-quality reads; HQ data %, percentage of bases of high-quality reads on bases of raw Illumina reads.
Online Resource 4. Assembly metrics of the assembled reference genome. N50, length of the smallest scaffold such that at least 50% of the total scaffold length is contained in scaffolds of that size or longer; N20, same as N50 but using 20% instead; N90, same as N50 but using 90% instead.
Online Resource 5. Assessment of completeness of the assembled reference genome using BUSCO.
Online Resource 6. Summary of predicted protein-coding genes in the assembled reference genome.
Online Resource 7. Annotation of protein-coding genes in the assembled reference genome on the basis of the GO database.
Online Resource 8. Annotation of protein-coding genes in the assembled reference genome on the basis of the KEGG database.
Online Resource 9. Sequence alignment results.
Online Resource 10. Loci number and the average sequencing depth of loci for each *Quercus variabilis* individual.
Online Resource 11. SNPs counts for each *Quercus variabilis* individual in 8 populations.
Online Resource 12. Pearson correlations of soil P concentrations with the expected heterozygosity (Exp Het), nucleotide diversity (Pi) and haplotype diversity (Hd) for *Quercus variabilis* populations sampled at different field sites.
Online Resource 13. Mantel test between the genetic distances (F_{ST} between populations) and log (geographic distances (Km)) of *Quercus variabilis* populations.
Online Resource 14. K = 3 had the highest ΔK vs. K peak height.

Supplementary Information The online version contains supplementary material available at https://doi.org/10.1007/s00606-021-01764-1.

Acknowledgements We are grateful for Dr. Washington Gapare for his comments and suggestions for improving the manuscript. We thank the Instrumental Analysis Center of Shanghai Jiao Tong University for the support on chemical analysis.

Funding This study was funded by the National Key R&D Program of China (2017YFC0505501) and the National Natural Science Foundation of China (31670626, 31270640 and 31070532).

Declarations

Conflict of interest The authors declare that there are no conflicts of interest.

References

Abeysinghe PD, Triest L, Greef BD, Koedam N, Hettiarachi S (2000) Genetic and geographic variation of the mangrove tree Bruguierain Sri Lanka. Aquatic Bot 67:131–141. https://doi.org/10.1016/S0304-3770(99)00096-0

Allendorf FW, Hohenlohe PA, Luikart G (2010) Genomics and the future of conservation genetics. Nat Rev Genet 11:697–709. https://doi.org/10.1038/nrg2844

Attiwill PM, Adams MA (1993) Nutrient cycling in forests. New Phytol 124:561–582. https://doi.org/10.1111/j.1469-8137.1993.tb03847.x

Baxter I, Brazelton JN, Yu D, Huang YS, Lahner B, Yakubova E, Li Y, Bergelson J, Borevitz JO, Nordborg M (2010) A coastal cline in sodium accumulation in *Arabidopsis thaliana* is driven by natural variation of the sodium transporter athkt1;1. PLoS Genet 6:e1001193. https://doi.org/10.1371/journal.pgen.1001193

Brady KU, Kruckeberg AR, Bradshaw JHD (2005) Evolutionary ecology of plant adaptation to serpentine soils. Annual Rev Ecol Evol S 36:243–266. https://doi.org/10.1146/annurev.ecolsys.35.021103.105730

Bruschi P, Vendramin GG, Bussotti F, Grossoni P (2003) Morphological and molecular diversity among Italian populations of *Quercus petraea* (Fagaceae). Ann Bot (Oxford) 91:707–716. https://doi.org/10.1093/aob/mcg075

Cao J, Fu D, Yan K, Wang T, Zhang Y, Zhang Y, Zhang L, Duan C (2014) The ecological differentiation of dominant plants in phosphorus-enriched areas within the Dianchi Lake basin. Chin J Ecol 33:3230–3237

Catchen J, Bassham S, Wilson T, Currey M, O'Brien C, Yeates Q, Cresko WA (2013a) The population structure and recent colonization history of *Oregon threespine* stickleback determined using restriction-site associated DNA-sequencing. Molec Ecol 22:2864–2883. https://doi.org/10.1111/mec.12330

Catchen J, Hohenlohe PA, Bassham S, Amores A, Cresko WA (2013b) Stacks: an analysis tool set for population genomics. Molec Ecol 22:3124–3140. https://doi.org/10.1111/mec.12354

Chen D, Zhang X, Kang H, Sun X, Yin S, Du H, Yamanaka N, Gapare W, Wu HX, Liu C (2012) Phylogeography of *Quercus variabilis* based on chloroplast DNA sequence in East Asia: multiple glacial refugia and mainland-migrated island populations. PLoS ONE 7:e47268. https://doi.org/10.3732/apps.1500032

Chen W, Provart NJ, Glazebrook J, Katagiri F, Chang HS, Eulgem T, Mauch F, Luan S, Zou G, Whitham SA (2002) Expression profile matrix of Arabidopsis transcription factor genes suggests their putative functions in response to environmental stresses. Pl Cell 14:559–574. https://doi.org/10.1105/tpc.010410

Chen X, Song Y (1997) Clonal diversity of *Cyclobalanopsis glauca* populations and its relationship with environmental factors. Chin J Pl Ecol 21:342–348

Cheng X, Li X, Jiang Y, Huang X (2017) Analyzing genetic diversity of *Quercus aquifolioides* in different altitudes of Jade Dragon Snow Mountain in the northwest of Yunnan province with AFLPs. J Huazhong Agric Univ 36:22–27

Chiou TJ, Aung K, Lin SI, Wu CC, Chiang SF, Su CL (2006) Regulation of phosphate homeostasis by MicroRNA in *Arabidopsis*. Pl Cell 18:412–421. https://doi.org/10.1105/tpc.105.038943

Chung MY, Nason J, Chung MG, Kim KJ, Park CW, Sun BY, Pak JH (2002) Landscape-level spatial genetic structure in *Quercus acutissima* (Fagaceae). Amer J Bot 89:1229–1236. https://doi.org/10.3732/ajb.89.8.1229

Cleveland CC, Townsend AR, Taylor P, Alvarez-Clare S, Bustamante M, Chuyong G, Dobrowski SZ, Grierson P, Harms KE, Houlton BZ (2011) Relationships among net primary productivity, nutrients and climate in tropical rain forest: a pan-tropical analysis. Ecol Lett 14:939–947. https://doi.org/10.1111/j.1461-0248.2011.01658.x

Colby T, Matthäi A, Boeckelmann A, Stuible HP (2006) SUMO-conjugating and SUMO-deconjugating enzymes from Arabidopsis. Pl Physiol 142:318–332. https://doi.org/10.1104/pp.106.085415

Craft KJ, Ashley MV (2007) Landscape genetic structure of bur oak (*Quercus macrocarpa*) savannas in Illinois. Forest Ecol Managem 239:13–20. https://doi.org/10.1016/j.foreco.2006.11.005

Deng J, Liao B, Ye M, Deng D, Lan C, Shu W (2007) The effects of heavy metal pollution on genetic diversity in zinc/cadmium hyperaccumulator *Sedum alfredii* populations. Pl Soil 297:83–92. https://doi.org/10.1007/s11104-007-9322-5

Dezar CA, Fedrigo GV, Chan RL (2005) The promoter of the sunflower HD-Zip protein gene Hahb4 directs tissue-specific expression and is inducible by water stress, high salt concentrations and ABA. Pl Sci 169:447–456. https://doi.org/10.1016/j.plantsci.2005.04.008

Dmitry NPD (2005) Selective sweep. Springer, Berlin

Doyle J (1991) DNA protocols for plants. Springer, Berlin, Heidelberg. https://doi.org/10.1007/978-3-642-83962-7_18

Epps CW, Palsbøll PJ, Wehausen JD, Roderick GK, Ramey RR, Mccullough DR (2010) Highways block gene flow and cause a rapid decline in genetic diversity of desert bighorn sheep. Ecol Lett 8:1029–1038. https://doi.org/10.1111/j.1461-0248.2005.00804.x

Etter PD, Bassham S, Hohenlohe PA, Johnson EA, Cresko WA (2011) SNP discovery and genotyping for evolutionary genetics using RAD sequencing. Meth Molec Biol 772:157–178. https://doi.org/10.1007/978-1-61779-228-1_9

Evanno G, Regnaut S, Goudet J (2005) Detecting the number of clusters of individuals using the software structure: a simulation study. Molec Ecol 14:2611–2620. https://doi.org/10.1111/j.1365-294X.2005.02553.x

Fang Z, Shao C, Meng Y, Wu P, Chen M (2009) Phosphate signaling in Arabidopsis and *Oryza sativa*. Pl Sci 176:170–180. https://doi.org/10.1016/j.plantsci.2008.09.007

Farwig N, Braun C, Böhning-Gaese K (2008) Human disturbance reduces genetic diversity of an endangered tropical tree, *Prunus africana* (Rosaceae). Conservation Genet 9:317–326. https://doi.org/10.1007/s10592-007-9343-x

Hammond JP, Bennett MJ, Bowen HC, Broadley MR, Dan CE, May ST, Rahn C, Swarup R, Woolaway KE, White PJ (2003) Changes in gene expression in Arabidopsis shoots during phosphate starvation and the potential for developing smart plants. Pl Physiol 132:578–596. https://doi.org/10.2307/4281133

Hammond JP, Broadley MR, Bowen HC, Spracklen WP, Hayden RM, White PJ (2011) Gene expression changes in phosphorus deficient potato (*Solanum tuberosum* L.) leaves and the potential for diagnostic gene expression markers. PLoS ONE 6:e24606. https://doi.org/10.1371/journal.pone.0024606

Hammond JP, Broadley MR, White PJ (2004) Genetic responses to phosphorus deficiency. Ann Bot (Oxford) 94:323–332

Hamrick JL, Godt MJW (1989) Allozyme diversity in plant species. In: Brown AHD, Clegg MT, Kahler AL et al (eds) Plant population genetics, breeding and genetic resources. Sinauer Associates, Sunderland, pp 43–63

Hamrick JL, Godt MJW, Sherman-Broyles SL (1992) Factors influencing levels of genetic diversity in woody plant species. New Forests 6:95–124. https://doi.org/10.1007/BF00120641

Hamrick JL, Godt MJW, Sherman-Broyles SL (1995) Gene flow among plant populations: evidence from genetic markers. In: Hoch PC, Stephnon AG (eds) Experimental and molecular approaches to plant biosystematics. Missouri Botanical Garden, St. Louis, pp 215–232

Han W, Fang J, Guo D, Zhang Y (2005) Leaf nitrogen and phosphorus stoichiometry across 753 terrestrial plant species in China. New

Phytol 168:377–385. https://doi.org/10.1111/j.1469-8137.2005.01530.x

Han Z (2004) Study on area variability and genetic diversity in populations of Quercus variabilis from Shaanxi region. PhD Thesis, Northwest A&F University, Shaanxi

Hipp AL, Eaton DAR, Cavender-Bares J, Fitzek E, Nipper R, Manos PS (2014) A framework phylogeny of the American oak clade based on sequenced RAD data. PLoS ONE 9:e93975. https://doi.org/10.1371/journal.pone.0102272

Huff DR, Quinn JA, Higgins B, Palazzo AJ (1998) Random amplified polymorphic DNA (RAPD) variation among native little bluestem [*Schizachyrium scoparium* (Michx.) Nash] populations from sites of high and low fertility in forest and grassland biomes. Molec Ecol 7:1591–1597. https://doi.org/10.1046/j.1365-294x.1998.00473.x

Ji H, Du B, Liu C (2017) Elemental stoichiometry and compositions of weevil larvae and two acorn hosts under natural phosphorus variation. Sci Rep 7:45810. https://doi.org/10.1038/srep45810

Ji H, Ossipov V, Du B, Wen J, Liu C (2019) Differences in the relationship between metabolomic and ionomic traits of *Quercus variabilis* growing at contrasting geologic-phosphorus sites in subtropics. Pl Soil 439:339–355. https://doi.org/10.1007/s11104-019-04020-1

Jin L, Chen K, Li B (1998) The variation of genetic diversity of *Quercus aquifolioides* in different elevations. Acta Bot Sin 40:761–767. https://doi.org/10.1016/S1381-1177(97)00026-X

Johnson ES (2004) Protein modification by SUMO. Annual Rev Biochem 73:355–382. https://doi.org/10.1146/annurev.biochem.73.011303.074118

Jun XU, Wang X, Guo W (2015) The cytochrome P450 superfamily: key players in plant development and defense. J Integr Agric 14:1673–1686. https://doi.org/10.1016/S2095-3119(14)60980-1

Kerscher O, Felberbaum R, Hochstrasser M (2006) Modification of proteins by ubiquitin and ubiquitin-like proteins. Annual Rev Cell Developmental Biol 22:159–180. https://doi.org/10.1146/annurev.cellbio.22.010605.093503

Kuo HF, Chiou TJ (2011) The role of microRNAs in phosphorus deficiency signaling. Pl Physiol 156:1016–1024. https://doi.org/10.1104/pp.111.175265

Lefort F, Lally M, Thompson D, Douglas GC (1998) Morphological traits, microsatellite fingerprinting and genetic relatedness of a stand of elite oaks (*Q. robur* L.) at Tullynally. Ireland Silvae Genet 47:257–261. https://doi.org/10.1073/pnas.64.2.498

Li D, Peng S (2001) Genetic diversity in three *Pinus massoniana* populations in different elevations and its relationship with ecological factors. Acta Ecol Sin 21:415–421. https://doi.org/10.3321/j.issn:1000-0933.2001.03.011

Li H, Durbin R (2009) Fast and accurate short read alignment with Burrows-Wheeler transform. Bioinformatics 25:1754–1760. https://doi.org/10.1093/bioinformatics/btp324

Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, Marth G, Abecasis G, Durbin R (2009) The sequence alignment/map format and SAMtools. Bioinformatics 25:2078–2079. https://doi.org/10.1093/bioinformatics/btp352

Li M, Tian S, Jin L, Zhou G, Li Y, Zhang Y, Wang T, Yeung C, Chen L, Ma J et al (2013) Genomic analyses identify distinct patterns of selection in domesticated pigs and Tibetan wild boars. Nat Genet 45:1431–1438. https://doi.org/10.1038/ng.2811

Li W, Schmidt W (2010) A lysine-63-linked ubiquitin chain-forming conjugase, UBC13, promotes the developmental responses to iron deficiency in Arabidopsis roots. Plant J 62:330–343. https://doi.org/10.4161/psb.11424

Linhart YB, Grant MC (1996) Evolutionary significance of local genetic differentiation in plants. Annual Rev Ecol Syst 27:237–277. https://doi.org/10.1146/annurev.ecolsys.27.1.237

Löpez-Aljorna A, Bueno MA, Aguinagalde I, Martín JP (2007) Fingerprinting and genetic variability in cork oak (*Quercus suber* L.) elite trees using ISSR and SSR markers. Ann Forest Sci 64:773–779. https://doi.org/10.1051/forest:2007057

Lowry DB, Rockwood RC, Willis JH (2008) Ecological reproductive isolation of coast and inland races of *Mimulus guttatus*. Evolution 62:2196–2214. https://doi.org/10.1111/j.1558-5646.2008.00457.x

Macnair MR (1993) The genetics of metal tolerance in vascular plants. New Phytol 124:541–559. https://doi.org/10.1111/j.1469-8137.1993.tb03846.x

Mengoni A, Barabesi C, Gonnelli C, Galardi F, Gabbrielli R, Bazzicalupo M (2001) Genetic diversity of heavy metal-tolerant populations in *Silene paradoxa* L. (Caryophyllaceae): a chloroplast microsatellite analysis. Molec Ecol 10:1909–1916. https://doi.org/10.1046/j.0962-1083.2001.01336.x

Miller MR, Dunham JP, Amores A, Cresko WA, Johnson EA (2007) Rapid and cost-effective polymorphism identification and genotyping using restriction site associated DNA (RAD) markers. Genome Res 17:240–248. https://doi.org/10.1101/gr.5681207

Misson J, Raghothama KG, Jain A, Jouhet J, Block MA, Bligny R, Ortet P, Creff A, Somerville S, Rolland N (2005) A genome-wide transcriptional analysis using Arabidopsis thaliana Affymetrix gene chips determined plant responses to phosphate deprivation. Proc Natl Acad Sci USA 102:11934–11939. https://doi.org/10.1073/pnas.0505266102

Miura K, Okamoto H, Okuma E, Shiba H, Kamada H, Hasegawa PM, Murata Y (2013) SIZ1 deficiency causes reduced stomatal aperture and enhanced drought tolerance via controlling salicylic acid-induced accumulation of reactive oxygen species in Arabidopsis. Pl J 73:91–104. https://doi.org/10.1111/tpj.12014

Miura K, Rus A, Sharkhuu A, Yokoi S, Karthikeyan AS, Raghothama KG, Baek D, Koo YD, Jin JB, Bressan RA (2005) The Arabidopsis SUMO E3 ligase SIZ1 controls phosphate deficiency responses. Proc Natl Acad Sci USA 102:7760–7765. https://doi.org/10.1073/pnas.0504375102

Morcuende R, Bari R, Gibon Y, Zheng W, Pant BD, BLÄSING O, Usadel B, Czechowski T, Udvardi MK, Stitt M, (2007) Genome-wide reprogramming of metabolism and regulatory networks of Arabidopsis in response to phosphorus. Plant Cell Environ 30:85–112. https://doi.org/10.1111/j.1365-3040.2006.01608.x

Müller R, Morant M, Jarmer H, Nilsson L, Nielsen TH (2007) Genome-wide analysis of the Arabidopsis leaf tanscriptome reveals interaction of phosphate and sugar metabolism. Pl Physiol 143:156–171. https://doi.org/10.1104/pp.106.090167

Nkongolo KK, Mehes M, Deck A, Michael P (2007) Metal content in soil and genetic variation in Deschampsia cespitosa populations from Northern Ontario (Canada): application of ISSR markers. Eur J Genet Toxicol 1:1–38

Owuor ED, Fahima T, Beiles A, Korol A, Nevo E (1997) Population genetic response to microsite ecological stress in wild barley, Hordeum spontaneum. Molec Ecol 6:1177–1187. https://doi.org/10.1046/j.1365-294X.1997.00296.x

Pauls SU, Nowak C, Bálint M, Pfenninger M (2013) The impact of global climate change on genetic diversity within populations and species. Molec Ecol 22:925–946. https://doi.org/10.1111/mec.12152

Peterson BJ, Bricker E, Brisbin SJ, Furman BT, Stubler AD, Carroll JM, Berry DL, Gobler CJ, Calladine A, Waycott M (2013) Genetic diversity and gene flow in Zostera marina populations surrounding Long Island, New York, USA: no evidence of inbreeding, genetic degradation or population isolation. Aquat Bot 110:61–66. https://doi.org/10.1016/j.aquabot.2013.05.003

Peterson BK, Weber JN, Kay EH, Fisher HS, Hoekstra HE (2012) Double digest RADseq: an inexpensive method for de novo SNP

discovery and genotyping in model and non-model species. PLoS ONE 7:e37135. https://doi.org/10.1371/journal.pone.0037135
Pritchard JK, Stephens M, Donnelly P (2000) Inference of population structure using multilocus genotype data. Genetics 155:945–959
Qin L, Zhang CL, Zhang B (2011) Differential gene expression in leaves and roots of winter rape in response to phosphorus starvation. Russian J Pl Physiol 58:142–148. https://doi.org/10.1134/S1021443710061032
Raghothama KG, Karthikeyan AS (2005) Phosphate acquisition. Pl Soil 274:37. https://doi.org/10.1007/s11104-004-2005-6
Schubert M, Lindgreen S, Orlando L (2016) AdapterRemoval v2: rapid adapter trimming, identification, and read merging. BMC Res Notes 9:88. https://doi.org/10.1186/s13104-016-1900-2
Shi C, Han R, Wu P, Fang M, Lai H, Shentu L (2011) Geochemistry and provenance of source rock for matoushan formation rocks (Late Cretaceous) in the Mouding area, Central Yunnan. Acta Sedimentol Sin 29:303–311. https://doi.org/10.1007/s11589-011-0776-4
Silberbush M, Waisel Y, Kafkafi U (1981) The role of soil phosphorus in differentiation of edaphic ecotypes in *Aegilops peregrina*. Oecologia 49:419–424. https://doi.org/10.1007/BF00347610
Snaydon RW, Bradshaw AD (1962) Differences between natural populations of *Trifolium repens* L. in response to mineral nutrients. J Exp Bot 13:422–434. https://doi.org/10.1093/jxb/13.3.422
Tao Y (2005) A proximate mathematical model of the content of main chemical constituents in an industrial phosphorus ore. Yunnan Geol 24:151–166
Tiessen H (2008) Phosphorus in the global environment. In: White PJ, Hammond JP (eds) The ecophysiology of plant-phosphorus interactions. Springer, Dordrecht, pp 1–7. https://doi.org/10.1007/978-1-4020-8435-5_1
Toonen RJ, Puritz JB, Forsman ZH, Whitney JL, Fernandezsilva I, Andrews KR, Bird CE (2013) ezRAD: a simplified method for genomic genotyping in non-model organisms. PeerJ 1:e203. https://doi.org/10.7717/peerj.203
Vance CP, Uhde-Stone C, Allan DL (2003) Phosphorus acquisition and use: critical adaptations by plants for securing a nonrenewable resource. New Phytol 157:423–447. https://doi.org/10.1046/j.1469-8137.2003.00695.x
Vitousek P (1982) Nutrient cycling and nutrient use efficiency. Amer Naturalist 119:553–572. https://doi.org/10.1086/283931
Walker TW, Syers JK (1976) The fate of phosphorus during pedogenesis. Geoderma 15:1–19. https://doi.org/10.1016/0016-7061(76)90066-5
Wang S, Hu Y, Zhou Z (2005) AFLP analysis of genetic variation and structure of natural populations of *Quercus senescens*. Acta Bot Yunnan 27:49–58. https://doi.org/10.1360/biodiv.050022
Wang S, Meyer E, Mckay JK, Matz MV (2012) 2b-rad: A simple and flexible method for genome-wide genotyping. Nature Methods 9:808–810. https://doi.org/10.1038/nmeth.2023
Wang XQ, Zhao L, Eaton DA, Li DZ, Guo ZH (2013) Identification of SNP markers for inferring phylogeny in temperate bamboos (Poaceae: Bambusoideae) using RAD sequencing. Molec Ecol Resources 13:938–945. https://doi.org/10.1111/1755-0998.12136
Wasaki J, Shinano T, Onishi K, Yonetani R, Yazaki J, Fujii F, Shimbo K, Ishikawa M, Shimatani Z, Nagata Y (2006) Transcriptomic analysis indicates putative metabolic changes caused by manipulation of phosphorus availability in rice leaves. J Exp Bot 57:2049–2059. https://doi.org/10.1093/jxb/erj158
Wasaki J, Yamamura T, Shinano T, Osaki M (2003) Secreted acid phosphatase is expressed in cluster roots of lupin in response to phosphorus deficiency. Pl Soil 248:129–136. https://doi.org/10.1023/A:1022332320384
Wei G (2015) Genetic variation of populations and introgression among four sympatric oaks in Jiangsu and Anhui provinces. PhD Thesis, Nanjing Forestry University, Nanjing
Whitlock MC, Mccauley DE (1999) Indirect measures of gene flow and migration: FST not equal to 1/(4Nm+1). Heredity 82:117–125. https://doi.org/10.1038/sj.hdy.6884960
Wright S (1931) Evolution in Mendelian populations. Genetics 16:97–159. https://doi.org/10.1007/BF02459575
Wright S (1978) Evolution and the genetics of populations, volume 4: variability within and among natural populations. University of Chicago Press, Chicago. https://doi.org/10.2307/2529965
Wu P, Ma L, Hou X, Wang M, Wu Y, Liu F, Deng XW (2003) Phosphate starvation triggers distinct alterations of genome expression in Arabidopsis roots and leaves. Pl Physiol 132:1260–1271. https://doi.org/10.1104/pp.103.021022
Xiao G, Li T, Zhang X, Yu H, Huang H, Gupta DK (2009) Uptake and accumulation of phosphorus by dominant plant species growing in a phosphorus mining area. J Hazard Mater 171:542–550. https://doi.org/10.1016/j.jhazmat.2009.06.034
Xu L (2002) Developments in population and evolutionary genetics within the genus *Quercus* L. J Nanjing Forest Univ, Nat Sci Ed 26:73–77
Xu L, Wang Y, Wang X, Zhang L, Yue M, Gu F, Pan X, Zhao G (2003) Genetic structure of Reaumuria soongorica population in Fukang Desert, Xinjiang and its relationship with ecological factors. Acta Bot Sin 45:787–794. https://doi.org/10.1117/12.466683
Xu X, Xu L, Huang M, Wang Z (2004) Genetic diversity of microsatellites (SSRs) of natural populations of *Quercus variabilis*. Hereditas 26:683–688. https://doi.org/10.1088/1009-0630/6/5/011
Yan K, Fu D, He F, Duan C (2011) Leaf nutrient stoichiometry of plants in the phosphorus-enriched soils of the Lake Dianchi watershed, southwestern China. Chin J Pl Ecol 35:353–361. https://doi.org/10.3724/SP.J.1258.2011.00353
Yan K, Ranjitkar S, Zhai D, Li Y, Xu J, Li B, Lu Y (2017) Current revegetation patterns and restoration issues in degraded geological phosphorus-rich mountain areas: a synthetic analysis of Central Yunnan, SW China. Pl Diversity 39:140–148. https://doi.org/10.1016/j.pld.2017.04.003
Yan X, Lynch JP, Beebe SE (1995) Genetic variation for phosphorus efficiency of common bean in contrasting soil types: I. Vegetative Response. Crop Sci 35:1086–1093. https://doi.org/10.2135/cropsci1995.0011183X003500040028x
Yang G, Chen Y, Wang J, Guo C, Zhao L, Wang X, Guo Y, Li L, Li D, Guo Z (2016) Development of a universal and simplified ddRAD library preparation approach for SNP discovery and genotyping in angiosperm plants. Pl Meth 12:39. https://doi.org/10.1186/s13007-016-0139-1
Yoo CY, Miura K, Jin JB, Lee J, Park HC, Salt DE, Yun DJ, Bressan RA, Hasegawa PM (2006) SIZ1 small ubiquitin-like modifier E3 ligase facilitates basal thermotolerance in Arabidopsis independent of salicylic acid. Pl Physiol 142:1548–1558. https://doi.org/10.1104/pp.106.088831
Zhang BD, Xue D-X, Wang J, Li Y-L, Liu B-J, Liu J-X (2016) Development and preliminary evaluation of a genomewide single nucleotide polymorphisms resource generated by RAD-seq for the small yellow croaker (*Larimichthys polyactis*). Molec Ecol Resources 16:755–768. https://doi.org/10.1111/1755-0998.12476
Zhang Y, Zhang B, Yan D, Dong W, Yang W, Li Q, Zeng L, Wang J, Wang L, Hicks LM (2011) Two Arabidopsis cytochrome P450 monooxygenases, CYP714A1 and CYP714A2, function redundantly in plant development through gibberellin deactivation. Pl J 67:342–353
Zhang Y, Zhou Z, Yang Q (2013) Genetic variations in root morphology and phosphorus efficiency of *Pinus massoniana* under

heterogeneous and homogeneous low phosphorus conditions. Pl Soil 364:93–104. https://doi.org/10.1007/s11104-012-1352-y

Zhou J, Lu Y, Zhang J, Zhang W (2008) Study on isozyme genetic diversity of *Quercus variabilis* natural populations in Shaanxi province. J SW Forest Coll 28:1–4

Zhou X, Sun X, Du B, Yin S, Liu C (2015) Multielement stoichiometry in *Quercus variabilis* under natural phosphorus variation in subtropical China. Sci Rep 5:7839. https://doi.org/10.1038/srep07839

Zhu Q, Zhang J, Gao X, Tong J, Xiao L, Li WB, Zhang H (2010) The Arabidopsis AP2/ERF transcription factor RAP2.6 participates in ABA, salt and osmotic stress responses. Gene 457:1–12. https://doi.org/10.1016/j.gene.2010.02.011