Pak. J. Bot., 56(1): 259-275, 2024. DOI: http://dx.doi.org/10.30848/PJB2024-1(24)

GENETIC EVOLUTION ANALYSIS AND SNP DEVELOPMENT OF *TAXUS* SPP. BY SLAF-SEQ

BIHUA CHEN[1*], YONG LI[2], JIANMIN LI[1], HUIHUA FAN[1*], XINGHAO TANG[1], GUIBIN ZHOU[3] AND SHENGLIANG HUANG[3]

[1]Fujian Academy of Forestry Sciences; Fuzhou Fujian 350012, China
[2]Fujian South Pharmaceutical Co. Ltd, Mingxi Fujian 365200, China
[3]Mingxi County Hengfeng Forestry Co. Ltd, Mingxi Fujian 365200, China
**Corresponding author's email: jofhh@163.com; chenbihua.happy@163.com*

Abstract

The molecular markers within whole genome were obtained, evolutionary tree was constructed, and the genetic structure, PCA (principal co-ordinates analysis) and genetic diversity analysis were carried out to understand the genetic evolutionary relationship and specific SNP markers of *Taxus* species or provenances, The protocol was predicted by using the genome of *Taxus chinensis* var. *mairei* and digested with the *Hpy*166II restriction enzyme, the SLAF (specific-locus amplified fragment) tags of length 314-364 bp read length 126 bp×2 used for subsequent data assessment and analysis, a total of 148.78 Mb reads data were obtained from 62 *Taxus spp.* Samples including 7 species, with 94.46% of an average Q30 and 37.99% of an average GC content. 140,405 SLAF tags were developed with an average sequencing depth of 7.54×. SNPs were developed by genome analysis toolkits GATK and Sam Tools, and the intersection of SNPs acquired using the two methods was applied as the reliable SNP marker dataset, and a total of 7,795,093 population SNPs were obtained. The phylogenetic tree showed that the 62 *Taxus spp*. could be divided into two clades, *T. yunnanensis* was the first clade, and the others clustered as the second clade. The data amount obtained could be used for the verification and development of specific SNP markers and reveal the genetic relationship for the 7 *Taxus* species.

Key words: *Taxus* spp., genetic evolution; Phylogenetic tree; SLAF-seq (specific-locus amplified fragment sequencing); SNP.

Introduction

Taxus, a genus is divided from the common ancestor of Podocarpaceae and Cephalotaxaceae (Hao *et al*., 2008; Wu & Wang, 1983). There are 11 species worldwide mainly distributing on northern hemisphere (Cope, 1998; Miller, 1997; Price, 1990; Zheng & Fu, 1978) 1 variety and 4 species in China. *Taxus spp.* is considered as a rare and endangered plant with anticancer effect in the world. It is an ancient relict tree species left over from the Quaternary glacier and has 2.5-million-year history on the earth (Wu, 1986). *Taxus* has a complex population, so it is of great significance to study its genetic evolutionary relationship for the perspective of exploitation, utilization and conservation, as well as scientific research.

The population genetics is to apply mathematical and statistical methods for studying the gene and genotype frequency in a population, the selection and mutation effect which affect the frequencies, and the relationship between genetic and migration drift and its genetic structure, so as to explore their evolution. Using molecular markers to analyze genetic variation is a necessary means to study population genetics.

The previous studies showed applied the methods of SLAF-*seq* (Specific-locus Amplified Fragment Sequencing) to analyze genetic evolution and paternity identification of *Osmanthus fragrans* 'Pucheng Dangui' (Li *et al.*, 2020; Sun *et al.*, 2013). No study on genetic evolution of *Taxus* by SLAF-seq technique has been reported. In this study, by using this technique, a total of 62 *Taxus* spp., samples genomic DNA including 7 species, mainly cultivated in Fujian, were extracted for molecular marker development. The molecular markers within whole genome were obtained, evolutionary tree was constructed, and the genetic structure, PCA (principal co-ordinates analysis) and genetic diversity analysis were carried out, which is of great significance to understand more about the genetic evolutionary relationship and develop specific SNP markers of *Taxus* species or provenances.

Material and Methods

Design of enzyme digestion scheme: The genome of *T. chinensis* var. *mairei* was choosed as the reference genome for the prediction ofelectron enzyme digestion, and the *Hpy*166II was finally determined as the digestion enzyme. Fragment sequences of length 314-364 bp were defined as SLAF tags, and the SLAF tags were predicted.

Reference genome determination: Based on the genome size and GC content of *Taxus*, *T. chinensis* var. *maire* genome was choosed as the reference genome for theprediction of enzyme digestion.

Reference species information: Genome of *T. chinensis* var. *mairei*. The size of assembled genome was 10.23 Gb, and the GC content was 36.78%. The linkage address was as the following:https://www.ncbi.nlm.nih.gov/genome/81451

Digestion scheme: A BMK-developed enzyme digestion prediction software SLAF-Predict was used for the prediction of enzyme digestion based on the reference genome, and the best scheme of enzyme digestion was selected.

Experimental process: As per the optimal scheme of digestion, the qualified genomic DNA was separately digested. The enzyme digested fragment-SLAF tags was treated by tailling A to its 3' end and linking Dual-index (Kozich *et al.*, 2013). Sequencing joint, PCR amplification,

purification, sample mixing and gel cutting were used to select the target fragments. Illumina platform (Illumina, Inc., USA) was used for sequencing after the library passed quality inspection.

Information analysis process: Dual-index was apllied to identify these original sequencing data obtained, and each sample reads were obtained. After the connectors of sequencing readswere filtered, the sequencing quality and data volume were evaluated. The efficiency of enzyme digestion *Hpy*166II was assessed by the reference genome of of *T. chinensis* var. *mairei*, then to estimatethe accuracy and effectiveness of the experimental process. Genome-wide SNP (single nucleotide polymorphism) markers were developed in the population based on bioinformatics analysis, and population polymorphism was analyzed using high-quality SNPs with representativeness in the population.

SLAF tag development method: The sequencing reads generated were derived from the same length enzyme fragments generated by the digestionof same restriction enzyme on different samples. The reads of each sample were aligned according to sequence similarity, and the reads clustered together were derived from a SLAF fragment (SLAF tags). The same SLAF tag sequence similarity between different samples was higher than that of different SLAF tags. A SLAF tag defined as a polymorphic SLAF tag if its sequence was different (i.e., polymorphic) from sample to sample.The development flow chart of polymorphic SLAF tags was shown in (Fig. 1).

Results and Analysis

Evaluation of digestion scheme

Digestion scheme: The reference genome of *T. chinensis* var. *mairei* was predicted by electronic restriction digestion. According to the principle of selection of restriction digestion scheme, the restriction digestion combination was determined as *Hpy*166II. The sequence length of restriction digestion fragment between 314-364 bp was defined as SLAF tag, and 128,135 SLAF tags could be predicted, as shown in (Table 1).

Sequencing data statistics and assessment: In order to ensure the quality of this research analysis, this research adopts 126bp×2 read length as the subsequent data assessment and analysis.

Check the distribution of sequencing quality values: Sequencing quality value (Q) is an important index to evaluate the error rate of single base based on high-throughput sequencing. The higher the value of sequencing quality is, the lower the error rate of base sequencing is. The corresponding formula on the error rate of base sequencing P and thevalue of sequencing quality Q is as follows: $Q_{-score} = -10 \times \log_{10} P$ whilea base sequencing error probability is 0.001, the base quality value Q is considered as 30. Sample sequencing quality distribution of this research was shown in (Fig. 2).

Base distribution examination: LAF-seq sequencing reads were used as genomic DNA enzymatic fragments.the enzymatic digestion site and PCR amplification affected its base distribution. The first 2 sequencing reads bases showed the same base separation as the enzymatic digestion site, and the subsequent base distribution will showed varying degrees of fluctuation. Base distribution of the samples sequenced in this research was shown in (Fig. 3).

Sequencing data output and quality statistics: The sequencing data of each sample were counted, including reads quantity, Q30 and GC content, and the results were shown in (Table 2).

A total of 148.78 Mb reads data with an average Q30 of 94.46%and an average GC content of 37.99%were derived.

Evaluation of experimental database construction: The enzyme digestion effectiveness program was determined by evaluating and monitoring the normal experimental process of *T. chinensis* var. *mairei* sequencing data. The reference data for *T. chinensis* var. *mairei* with genome size 10.23 Gb in this research was downloaded from https://www.ncbi.nlm.nih.gov/genome/81451

Comparative efficiency statistics: Via SOAP (Li *et al*., 2008; Li *et al*., 2009b) software compared the sequenced reads of *Taxus* spp., with its reference genome of *T. chinensis* var. *mairei*. The results showed that the double-end comparison efficiency of this experiment was from 71.79% to 85.46%, which is basically normal.

SLAF marker development:

SLAF label statistics: Total of 222,921 SLAF tags were developed in this research, and the mean sequencing depth of the tags was 7.54×. The statistical results were shown in (Table 3).

SLAFTags distribution on the chromosomes: SLAF tags were located on the reference genome by BWA software, and SLAF tags and polymorphic SLAF tags were counted on different chromosomes. Detailed data were shown in the (Table 4).

According to the distribution of SLAF on chromosomes, the SLAF distribution map on chromosomes was shown below in (Fig. 4).

SNP statistics: Using the sequence type with the highest depth in each SLAF tagas a reference sequence, then BWA (Li & Rechard, 2009), GATK (McKenna *et al.*, 2010) and SAMtools (Li *et al.*, 2009a) compared to the reference sequence, SNPsmarkers were developed, and the SNPs intersection obtained was applied as the final reliable SNP marker dataset. Total of 7,795,093 population SNPs were derived. SNP nformation statistics were shown in (Table 5).

Summary of polymorphic SLAF tag development: A total of 222,921 SLAF tags were derived with an mean sequencing depth of 7.54×, and 140,405 polymorphic SLAF tags were derived with a total of 7,795,093 SNP markers.

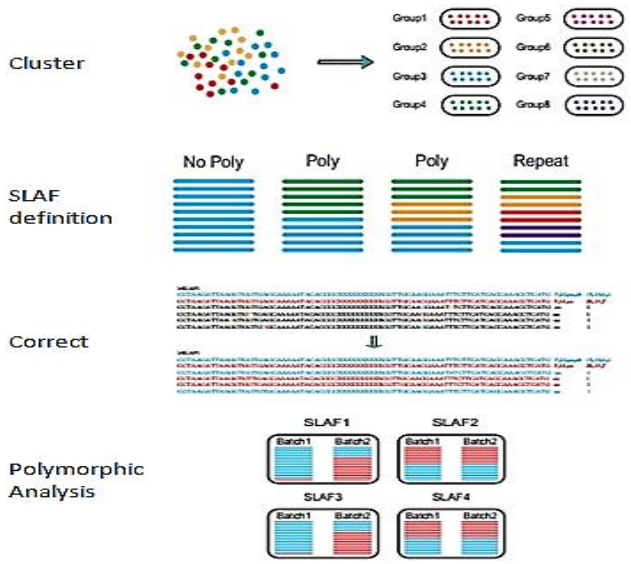

Fig. 1. SLAF tag development flowchart.

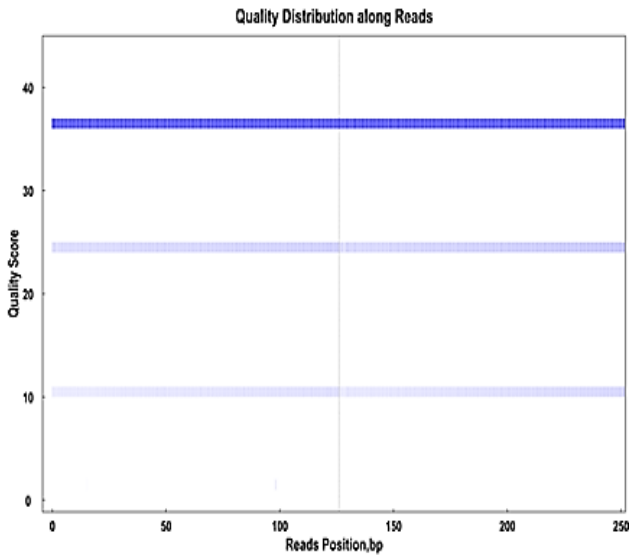

Fig. 2. Distribution of sequencing quality values.
Note: The horizontal coordinate is the base position of reads, and the vertical coordinate is the quality value of single base. The first 126bp was the quality value distribution of reads sequenced at the first end of the double-ended sequence, and the last 126bp was the quality value distribution of reads sequenced at the other end. Each bp represents each base of all reads sequenced, and the darker the color of each quality in the same position is, the higher the proportion of this quality value in the data is. For example, the first bp represents the distribution of the quality value of the first base of all sequenced.

Small in del detection and annotation:

Detection of small between sample and reference genome: As per the localization results of clean reads of the sample on the reference genome, the small fragment insertion and deletion (small) between the sample and reference genome were detected. The variation of small indel is generally less than thevariation ofSNP, reflecting the difference between the sample and the reference genome, and indel of coding region causes frameshift mutation, leading to changes gene function. (Table 6, Fig. 5).

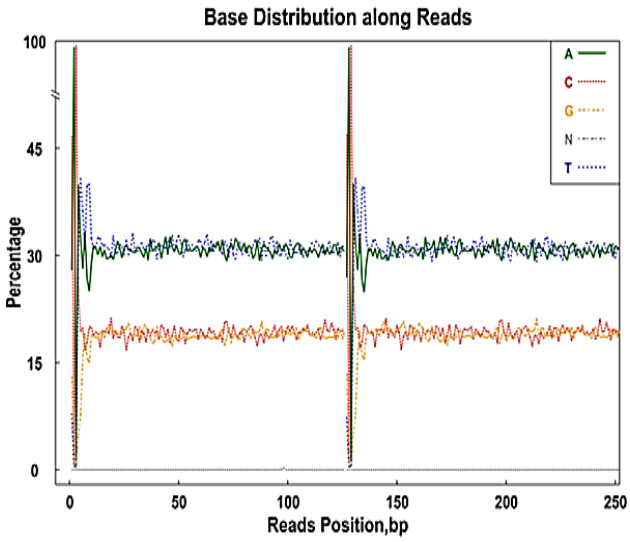

Fig. 3. Distribution of base content.
Note: The horizontal coordinate is the base position of reads, and the vertical coordinate is the proportion of bases; Different colors represent different base types: green represents base A, blue represents base T, red represents base C, orange represents base G, and gray represents base N that cannot be identified by sequencing. The first 126bp was the base distribution of Reads sequenced at the first end of the double-ended sequence, and the last 126bp was the base distribution of Reads sequenced at the other end. Each BP represents each base sequenced. For example, the first BP represents the distribution of A, T, G, C and N of all sequenced reads in the first base in the research.

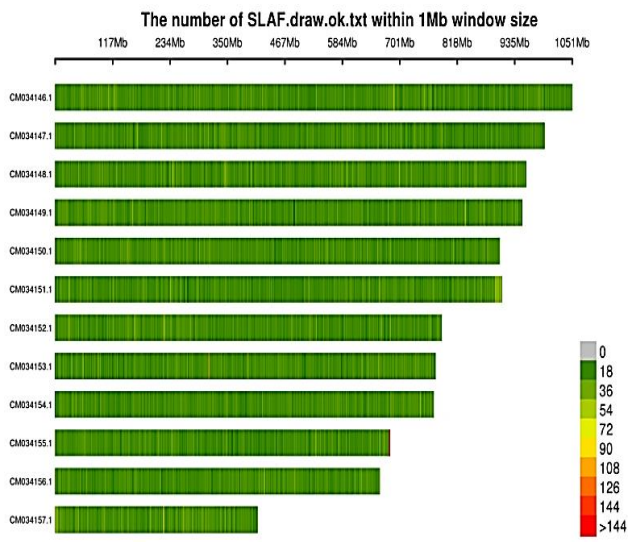

Fig. 4. The Distribution Map of SLAF on Chromosomes.
Note: The horizontal coordinate is the length of chromosome, each band represents a chromosome, and the genome is divided according to the size of 1Mb. The more SLAF tags isin each window, the darker the color is; The fewer SLAF tags is, the lighter the color is. The darker the color is, the more concentrated the SLAF tags are.

Genetic evolutionary analysis

Phylogenetic analysis: MEGA X (Kumar *et al.*, 2018) is used for constructing the phylogenetic tree based on 1,000 bootstrap repeats, Kimura 2-parameter model, neighbor-joining method. The phylogenetic tree was shown in (Fig. 6).

Table 1. Results of digestion scheme determined by digestion prediction.

Enzyme	Insert size	SLAF tag number
*Hpy*166II	314-364 bp	128,135

Genetic structure analysis: Population structure, known as population stratification, refers to the existence of subpopulations ondifferent gene frequencies in the related population. Materials in the same subgroup are closely related to each other, while subpopulations are far related to each other. Population structure analysis can determinethe number of ancestors of the related population and infer the genetic relationship of each sample. It is a widely used the method of cluster analysis at present, which helps to understand the evolutionary process of materials. Based on SNP, via Admixture (Alexander *et al.*, 2009) the group structure of the samples was analyzed by software, and the number of cluster (K value) of the samples was assumed as 1-10. As per the valley value of the error rate of cross-validation, the results of clustering are cross-verified and the optimal number of clustering is determined. See (Figs. 7 and 8) for clustering with K value 1-10 and the cross-validation error rate corresponding to each K value. The relationship between samples and populations was shown in (Table 7).

PCA analysis: In population genetics, based on SNP, by EIGENSOFT software (Price *et al.*, 2006), principal component analysis, principal component clustering of samples. PCA analysis can be used to determine which sample genetic relationship is relatively close and which sample genetic relationship are relatively distant, which can assist evolution analysis. PCA clustering and specific data were shown in (Table 8 and Figs. 9-10).

Genetic diversity analysis: GCTA software (Yang *et al.*, 2011)was used to estimate the relatedness of pairs of individuals in natural populations. The genetic relationship can be divided into A array and G array. A array is generally calculated by pedigree relationship, while G array is generally calculated by gene marker (referred to as SNP marker here). In this case, the mean value of marker expected variance is used to correct marker expected variance by default, namely, G array. Kinship value heat map wasshown in (Fig. 11).

Linkage disequilibrium analysis: When the probability of a particular allele in one locus co-occurring with an allele in another seat is greater than the probability of a random distribution of two alleles in the population, the two loci are said to be in linkage disequilibrium. Just conceptually, LD has nothing to do with chromosomes and nothing to do with linkage; But on the same chromosome, the stronger the LD at both loci, the closer the linkage is generally.

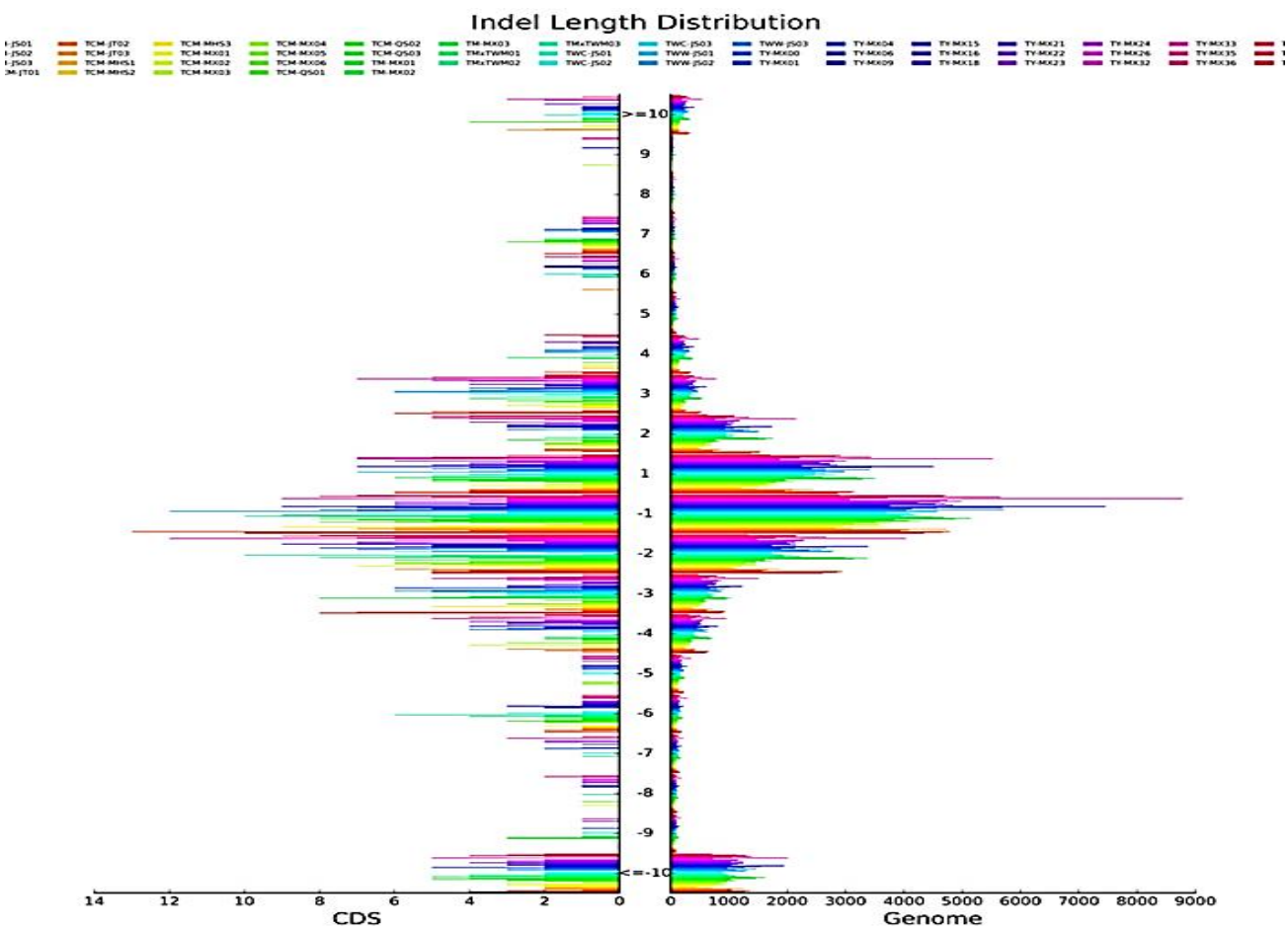

Fig. 5. InDel length distribution of the whole genome and coding region.
Note: The ordinate is InDel length distribution, greater than 0 is Insertion, less than 0 is Deletion, and the transverse coordinate is corresponding quantities.

Table 2. Statistical table of sequencing data of each sample.

Sample ID	Species name	Total reads	GC Percentage (%)	Q30 Percentage (%)	Sampling site
TB-JS01	*T. baccata*	2,194,178	36.93	94.18	Jiangsu Province, China
TB-JS02	*T. baccata*	2,624,661	37.10	94.12	Jiangsu Province, China
TB-JS03	*T. baccata*	2,915,932	36.96	91.97	Jiangsu Province, China
TCM-JT01	*T. chinensis*var.*mairei*	2,732,228	38.20	94.73	Jitou Village, Jian'ou County, Fujian Privince, China
TCM-JT02	*T. chinensis*var.*mairei*	2,824,344	38.14	94.74	Jitou Village, Jian'ou County, Fujian Privince, China
TCM-JT03	*T. chinensis*var.*mairei*	3,372,320	38.10	95.06	Jitou Village, Jian'ou County, Fujian Privince, China
TCM-MHS1	*T. chinensis*var.*mairei*	2,142,252	36.84	93.30	Meihuashan, Longyan, Fujian, China
TCM-MHS2	*T. chinensis*var.*mairei*	2,809,651	36.72	93.49	Meihuashan, Longyan, Fujian, China
TCM-MHS3	*T. chinensis*var.*mairei*	3,017,080	36.96	94.22	Meihuashan, Longyan, Fujian, China
TCM-MX01	*T. chinensis*var.*mairei*	1,927,065	38.43	95.03	Mingxi County, Fujian Privince, China
TCM-MX02	*T. chinensis*var.*mairei*	2,475,684	38.07	94.31	Mingxi County, Fujian Privince, China
TCM-MX03	*T. chinensis*var.*mairei*	2,595,714	38.35	94.72	Mingxi County, Fujian Privince, China
TCM-MX04	*T. chinensis*var.*mairei*	2,295,160	38.00	94.25	Mingxi County, Fujian Privince, China
TCM-MX05	*T. chinensis*var.*mairei*	2,711,178	37.56	94.22	Mingxi County, Fujian Privince, China
TCM-MX06	*T. chinensis*var.*mairei*	2,977,591	38.38	94.84	Mingxi County, Fujian Privince, China
TCM-QS01	*T. chinensis*var.*mairei*	2,924,570	37.90	94.94	Qishan, Minhou County, Fujian Province, China
TCM-QS02	*T. chinensis*var.*mairei*	2,907,505	37.91	94.88	Qishan, Minhou County, Fujian Province, China
TCM-QS03	*T. chinensis*var.*mairei*	2,749,449	38.61	94.91	Qishan, Minhou County, Fujian Province, China
TM-MX01	*T.*×*media*	2,481,770	38.33	94.93	Mingxi County, Fujian Province, China
TM-MX02	*T.*×*media*	2,623,136	37.49	94.44	Mingxi County, Fujian Province, China
TM-MX03	*T.*×*media*	2,233,644	37.54	94.32	Mingxi County, Fujian Province, China
TMxTWM01	*T. media*×*T. wallichiana*var.*mairei*	2,515,935	37.11	94.05	Jiangsu Province, China
TMxTWM02	*T. media*×*T. wallichiana*var.*mairei*	2,116,155	37.14	94.34	Jiangsu Province, China
TMxTWM03	*T. media*×*T. wallichiana*var.*mairei*	2,264,319	37.12	94.15	Jiangsu Province, China
TWC-JS01	*T. media*×*T. wallichiana*var.*mairei*	3,112,007	36.85	94.02	Jiangsu Province, China
TWC-JS02	*T. media*×*T. wallichiana*var.*mairei*	2,686,668	37.00	93.98	Jiangsu Province, China
TWC-JS03	*T. media*×*T. wallichiana*var.*mairei*	2,381,478	36.94	93.96	Jiangsu Province, China
TWW-JS01	*T. media*×*T. wallichiana*var.*mairei*	2,241,839	37.02	94.17	Jiangsu Province, China
TWW-JS02	*T. media*×*T. wallichiana*var.*mairei*	2,504,381	37.37	93.17	Jiangsu Province, China
TWW-JS03	*T. media*×*T. wallichiana*var.*mairei*	1,848,360	36.73	93.76	Jiangsu Province, China
TY-MX00	*T. yunnanensis*	2,875,730	37.95	94.77	Mingxi County, Fujian Province, China
TY-MX01	*T. yunnanensis*	1,823,916	39.00	94.66	Mingxi County, Fujian Province, China
TY-MX04	*T. yunnanensis*	1,816,188	38.97	95.11	Mingxi County, Fujian Province, China

Table 2. (Cont'd.).

Sample ID	Species name	Total reads	GC Percentage (%)	Q30 Percentage (%)	Sampling site
TY-MX06	*T. yunnanensis*	2,432,950	38.19	95.11	Mingxi County, Fujian Province, China
TY-MX09	*T. yunnanensis*	3,538,889	38.49	94.77	Mingxi County, Fujian Province, China
TY-MX15	*T. yunnanensis*	1,470,914	39.00	95.17	Mingxi County, Fujian Province, China
TY-MX16	*T. yunnanensis*	2,243,322	38.06	93.92	Mingxi County, Fujian Province, China
TY-MX18	*T. yunnanensis*	1,947,182	38.59	95.14	Mingxi County, Fujian Province, China
TY-MX21	*T. yunnanensis*	2,062,946	39.65	95.09	Mingxi County, Fujian Province, China
TY-MX22	*T. yunnanensis*	2,081,495	38.22	94.32	Mingxi County, Fujian Province, China
TY-MX23	*T. yunnanensis*	1,781,149	37.98	95.07	Mingxi County, Fujian Province, China
TY-MX24	*T. yunnanensis*	2,200,366	37.93	94.16	Mingxi County, Fujian Province, China
TY-MX26	*T. yunnanensis*	2,531,190	39.06	94.36	Mingxi County, Fujian Province, China
TY-MX32	*T. yunnanensis*	1,258,216	38.36	94.39	Mingxi County, Fujian Province, China
TY-MX33	*T. yunnanensis*	5,589,264	37.70	93.82	Mingxi County, Fujian Province, China
TY-MX35	*T. yunnanensis*	1,634,839	39.22	95.32	Mingxi County, Fujian Province, China
TY-MX36	*T. yunnanensis*	3,164,821	38.23	94.08	Mingxi County, Fujian Province, China
TY-MX40	*T. yunnanensis*	1,857,774	38.62	94.74	Mingxi County, Fujian Province, China
TY-MX41	*T. yunnanensis*	2,231,915	37.97	94.44	Mingxi County, Fujian Province, China
TY-MX42	*T. yunnanensis*	1,078,106	37.95	94.48	Mingxi County, Fujian Province, China
TY-MX48	*T. yunnanensis*	1,833,788	38.73	94.74	Mingxi County, Fujian Province, China
TY-MX49	*T. yunnanensis*	2,394,589	38.55	94.36	Mingxi County, Fujian Province, China
TY-MX50	*T. yunnanensis*	2,304,330	38.32	95.18	Mingxi County, Fujian Province, China
TY-MX54	*T. yunnanensis*	2,286,174	38.66	95.07	Mingxi County, Fujian Province, China
TY-MX56	*T. yunnanensis*	2,127,665	38.35	94.84	Mingxi County, Fujian Province, China
TY-MX59	*T. yunnanensis*	2,073,483	38.04	94.21	Mingxi County, Fujian Province, China
TY-MX61	*T. yunnanensis*	2,047,885	38.15	95.03	Mingxi County, Fujian Province, China
TY-MX65	*T. yunnanensis*	2,381,633	38.55	93.80	Mingxi County, Fujian Province, China
TY-MX68	*T. yunnanensis*	2,471,225	38.22	95.26	Mingxi County, Fujian Province, China
TY-MX70	*T. yunnanensis*	2,487,998	38.34	94.84	Mingxi County, Fujian Province, China
TY-MX71	*T. yunnanensis*	1,974,124	38.81	95.36	Mingxi County, Fujian Province, China
TY-MX72	*T. yunnanensis*	1,570,278	38.02	93.88	Mingxi County, Fujian Province, China

Note: Sample ID: research sample number; Total reads: number of reads in each sample; GC percentage: the percentage of G and C bases in total bases in sequencing results; Q30 percentage: the percentage of bases whose sequencing quality value was greater than or equal to 30. For sample ID, TB: *T. baccata*; TCM: *T. chinensis* var. *mairei*; TM: *Taxus×media*; TWC: *T. wallichiana* var. *chinensis*; TWW: *T. wallichiana* var. *wallichiana*; TY: *Taxus yunnanensis*; TM×TWM: *T. media×T. wallichiana* var. *mairei;* TWW: *T. wallichiana* var. *wallichiana*

Table 3. SLAF label statistics.

Sample ID	SLAF number	Total depth	Average depth
TB-JS01	88,577	863,827	9.7523
TB-JS02	92,074	1,010,185	10.9714
TB-JS03	93,340	1,201,805	12.8756
TCM-JT01	120,465	947,417	7.8647
TCM-JT02	121,058	966,132	7.9807
TCM-JT03	126,439	1,089,260	8.6149
TCM-MHS1	110,526	863,813	7.8155
TCM-MHS2	89,128	1,044,261	11.7164
TCM-MHS3	111,942	1,244,706	11.1192
TCM-MX01	112,409	507,237	4.5124
TCM-MX02	117,389	848,423	7.2274
TCM-MX03	118,005	961,285	8.1461
TCM-MX04	114,509	821,013	7.1699
TCM-MX05	119,195	1,005,725	8.4376
TCM-MX06	118,739	1,079,718	9.0932
TCM-QS01	121,858	944,039	7.7470
TCM-QS02	120,487	1,026,888	8.5228
TCM-QS03	119,723	962,308	8.0378
TM-MX01	100,280	729,166	7.2713
TM-MX02	98,749	730,221	7.3947
TM-MX03	94,045	636,251	6.7654
TMxTWM01	120,908	1,091,590	9.0283
TMxTWM02	114,868	913,590	7.9534
TMxTWM03	117,946	979,250	8.3025
TWC-JS01	107,256	1,350,097	12.5876
TWC-JS02	101,814	1,084,062	10.6475
TWC-JS03	100,849	968,349	9.6020
TWW-JS01	92,917	920,172	9.9032
TWW-JS02	89,915	976,264	10.8576
TWW-JS03	86,624	718,575	8.2953
TY-MX00	110,203	1,038,778	9.4260
TY-MX01	102,012	506,341	4.9635
TY-MX04	102,273	489,245	4.7837
TY-MX06	113,902	746,380	6.5528
TY-MX09	117,201	898,178	7.6636
TY-MX15	95,245	371,848	3.9041
TY-MX16	110,124	764,242	6.9398
TY-MX18	106,620	564,018	5.2900
TY-MX21	101,510	462,651	4.5577
TY-MX22	105,270	675,110	6.4131
TY-MX23	105,222	560,209	5.3241
TY-MX24	104,461	659,062	6.3092
TY-MX26	109,463	852,975	7.7924
TY-MX32	88,447	385,466	4.3582
TY-MX33	123,400	1,679,327	13.6088
TY-MX35	96,683	391,331	4.0476
TY-MX36	113,821	1,028,684	9.0377
TY-MX40	103,964	505,624	4.8635
TY-MX41	107,169	760,901	7.1000
TY-MX42	86,462	366,421	4.2379
TY-MX48	102,647	475,740	4.6347
TY-MX49	104,447	683,491	6.5439
TY-MX50	110,996	625,441	5.6348
TY-MX54	109,278	610,341	5.5852
TY-MX56	108,788	615,050	5.6537
TY-MX59	104,386	685,707	6.5690
TY-MX61	109,074	571,092	5.2358
TY-MX65	107,835	696,248	6.4566
TY-MX68	114,121	750,302	6.5746
TY-MX70	112,112	682,818	6.0905
TY-MX71	106,133	530,881	5.0020
TY-MX72	98,754	488,495	4.9466

Note: Sample ID: research sample number; SLAF number: SLAF label number contained in the corresponding sample; Total depth: the total sequencing depth of the corresponding sample in SLAF label; Average depth: The average number of sequenced reads for each SLAF corresponding sample.

Table 4. The stastics of SLAF tags and polymorphic SLAF tags on different chromosomes.

Chromosome ID	SLAF number	Polymorphic SLAF
CM034146.1	23,115	14,976
CM034147.1	21,385	13,783
CM034148.1	21,221	13,634
CM034149.1	21,044	13,758
CM034150.1	19,950	12,937
CM034151.1	19,707	12,279
CM034152.1	17,181	11,204
CM034153.1	17,163	10,930
CM034154.1	16,785	10,800
CM034155.1	14,555	8,912
CM034156.1	14,394	9,380
CM034157.1	9,839	5,191

Using PopLD decay (v3.41) software (Zhang *et al.*, 2019), the linkage disequilibrium between pairs of SNPs within a distance of 1000 Kb on the same chromosome was calculated. The strength of linkage disequilibrium was expressed as r2. The closer r2 is to 1, the stronger linkage disequilibrium is. The SNP spacing is fitted to r2, and the curve of r2 with distance is drawn. In general, more closer the SNP spacing is, more larger r2 is, and more farther the SNP spacing is, more smaller r2 is. The distance passed when the maximum r2 value drops to half is used to be the LD decay distance (LDD) in linkage disequilibrium. The longer the LDD, the smaller the probability of recombination within the same physical distance.The more shorter the LDD, the more greater the recombination probability within same physical distance. The LD decay of the study group in this project shown in the following (Fig. 12).

Analysis of genetic diversity: The variation of genes within a species can be explained by genetic diversity, including genetic variation among or betweenpopulations that differ significantly within a species and within a certainpopulation. There are often no completely consistent genotypes among individuals within a population, and a population is consisted of multiple individuals accompanying with these different genetic structures. This study of *Taxus spp*. genetic diversity coulddisclosethe evolutionary history of species andpopulations i.e., when and how these tested speciesoriginated, and provided important information for further studiesof their potential evolutionary or future fate. Genetic diversity is one of the core conservation biology study. The population genetic diversity was shown in (Table 9).

Discussion

SLAF-seq is a high-throughput sequencing technique, which can obtain specific length fragments according to the preset predigestion scheme, and adopt double-ended sequencing for specific enzyme digestion fragments. This sequencing method has high repeatability, and can obtain massive information throughout the whole genome within a short time, so as to achieve fine localization of candidate functional regions. Markers developed by this technique have the characteristics of high density, good consistency and low cost, and most SALF molecular markers are SNPs and a small number of InDel. At present, it has been successfully applied to many kinds of animals and plants.

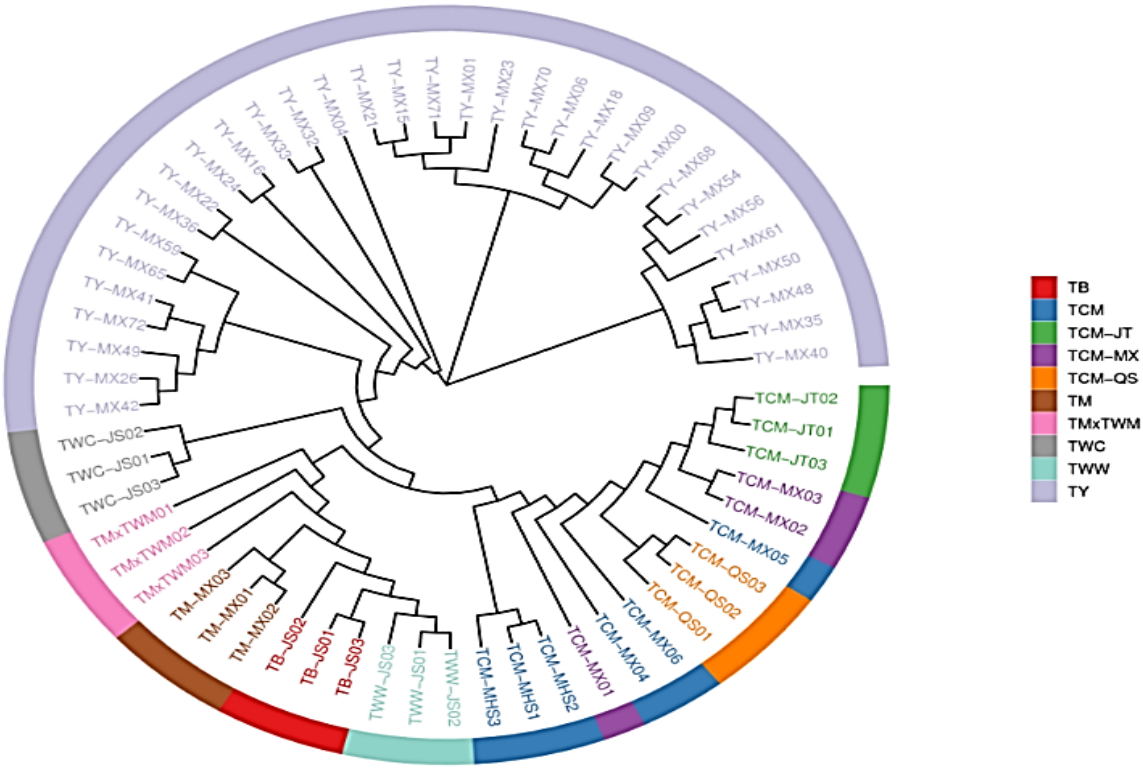

Fig. 6. Phylogenetic tree of sample system.

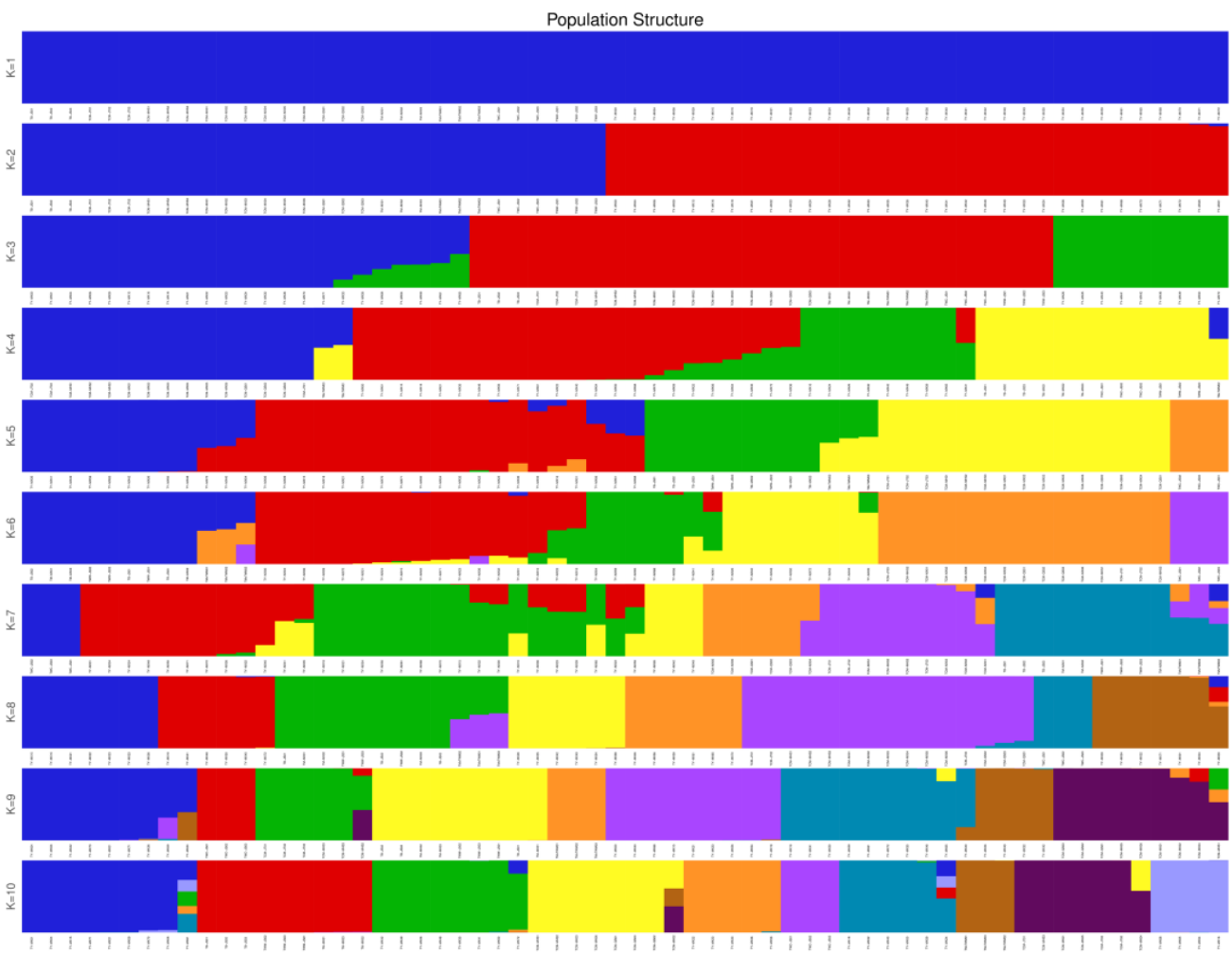

Fig. 7. Clustering results of admixture samples corresponding to each K value.

Table 5. SNP information statistics of samples.

Sample ID	SNP number	Hetloci ratio (%)	Integrity ratio (%)
TB-JS01	1,359,957	17.45%	5.60%
TB-JS02	1,471,054	18.87%	5.77%
TB-JS03	1,515,401	19.44%	6.54%
TCM-JT01	2,596,800	33.32%	5.80%
TCM-JT02	2,700,433	34.65%	5.70%
TCM-JT03	3,158,991	40.53%	6.12%
TCM-MHS1	2,032,895	26.08%	5.14%
TCM-MHS2	1,734,276	22.25%	7.18%
TCM-MHS3	2,289,541	29.37%	5.61%
TCM-MX01	2,133,440	27.37%	4.17%
TCM-MX02	2,477,937	31.79%	4.96%
TCM-MX03	2,407,353	30.89%	5.49%
TCM-MX04	2,234,410	28.67%	4.89%
TCM-MX05	2,518,492	32.31%	5.59%
TCM-MX06	2,547,006	32.68%	4.88%
TCM-QS01	2,854,666	36.62%	5.47%
TCM-QS02	2,620,522	33.62%	5.72%
TCM-QS03	2,542,438	32.62%	5.61%
TM-MX01	1,630,233	20.92%	7.38%
TM-MX02	1,715,443	22.01%	7.24%
TM-MX03	1,550,742	19.90%	6.68%
TMxTWM01	2,029,884	26.04%	7.53%
TMxTWM02	1,746,402	22.41%	6.89%
TMxTWM03	1,882,652	24.15%	7.26%
TWC-JS01	1,774,939	22.77%	6.00%
TWC-JS02	1,674,916	21.49%	5.40%
TWC-JS03	1,587,006	20.36%	5.39%
TWW-JS01	1,425,192	18.28%	5.86%
TWW-JS02	1,423,278	18.26%	6.13%
TWW-JS03	1,251,745	16.06%	5.40%
TY-MX00	2,509,930	32.20%	4.54%
TY-MX01	1,898,533	24.36%	3.96%
TY-MX04	2,096,317	26.89%	3.55%
TY-MX06	2,606,960	33.45%	4.01%
TY-MX09	3,355,283	43.05%	4.98%
TY-MX15	1,729,388	22.19%	3.14%
TY-MX16	2,172,734	27.88%	4.46%
TY-MX18	2,082,792	26.72%	3.43%
TY-MX21	2,169,880	27.84%	3.61%
TY-MX22	1,976,173	25.35%	4.20%
TY-MX23	2,093,852	26.86%	3.48%
TY-MX24	2,280,416	29.26%	4.22%
TY-MX26	2,256,230	28.95%	4.32%
TY-MX32	1,416,445	18.17%	3.58%
TY-MX33	3,920,328	50.30%	6.59%
TY-MX35	1,751,427	22.47%	3.23%
TY-MX36	2,631,116	33.76%	4.86%
TY-MX40	2,117,974	27.17%	3.49%
TY-MX41	2,155,632	27.66%	4.36%
TY-MX42	1,295,329	16.62%	3.09%
TY-MX48	2,093,410	26.86%	3.53%
TY-MX49	2,300,093	29.51%	4.15%
TY-MX50	2,534,908	32.52%	3.92%
TY-MX54	2,506,347	32.16%	3.98%
TY-MX56	2,382,678	30.57%	3.60%
TY-MX59	2,092,788	26.85%	4.08%
TY-MX61	2,344,930	30.08%	3.68%
TY-MX65	2,429,988	31.18%	4.40%
TY-MX68	2,637,677	33.84%	4.07%
TY-MX70	2,593,342	33.27%	4.00%
TY-MX71	2,236,607	28.69%	3.60%
TY-MX72	1,737,563	22.29%	3.68%
Total_SNP	7,795,093		

Table 5. (Cont'd.).

Note: Sample ID: research sample number; Total SNP: the total number of SNPs detected. SNP num: number of SNP detected in the corresponding sample; HetLoci ratio: heterozygosity rate of SNP in the sample; Integrity ratio: SNP integrity detected in samples

Table 6. In Del statistics between whole genome and coding region.

Sample	CDS-Insertion	CDS-Deletion	CDS-Het	CDS-Homo	CDS-Total	Genome-Insertion	Genome-Deletion	Genome-Het	Genome-Homo	Genome-Total
TB-JS01	17	28	34	11	45	8,130	12,249	18,502	1,877	20,379
TB-JS02	14	31	38	7	45	8,790	13,058	19,639	2,209	21,848
TB-JS03	19	22	33	8	41	8,937	13,508	19,882	2,563	22,445
TCM-JT01	11	16	16	11	27	5,279	9,620	11,772	3,127	14,899
TCM-JT02	11	20	21	10	31	5,694	9,895	12,498	3,091	15,589
TCM-JT03	15	23	28	10	38	6,403	11,455	13,835	4,023	17,858
TCM-MHS1	7	16	16	7	23	3,966	7,250	9,060	2,156	11,216
TCM-MHS2	6	16	17	5	22	3,342	6,275	7,088	2,529	9,617
TCM-MHS3	7	20	20	7	27	4,570	8,321	10,274	2,617	12,891
TCM-MX01	9	22	28	3	31	4,080	7,143	9,715	1,508	11,223
TCM-MX02	11	16	18	9	27	4,953	8,757	11,419	2,291	13,710
TCM-MX03	12	16	19	9	28	4,807	8,696	10,948	2,555	13,503
TCM-MX04	7	13	14	6	20	4,377	7,802	10,012	2,167	12,179
TCM-MX05	9	20	17	12	29	5,164	9,089	11,392	2,861	14,253
TCM-MX06	9	22	24	7	31	5,043	8,653	11,436	2,260	13,696
TCM-QS01	16	20	32	4	36	5,754	10,398	12,996	3,156	16,152
TCM-QS02	12	18	26	4	30	5,455	9,821	12,172	3,104	15,276
TCM-QS03	11	22	19	14	33	5,231	9,367	11,766	2,832	14,598
TM-MX01	11	24	30	5	35	9,070	13,959	19,929	3,100	23,029
TM-MX02	17	25	34	8	42	9,732	14,951	21,382	3,301	24,683
TM-MX03	15	29	32	12	44	8,781	13,360	19,535	2,606	22,141
TMxTWM01	13	30	31	12	43	7,126	11,555	14,754	3,927	18,681
TMxTWM02	12	25	26	11	37	6,430	10,159	13,587	3,002	16,589
TMxTWM03	10	34	33	11	44	6,836	11,065	14,449	3,452	17,901
TWC-JS01	15	16	25	6	31	7,805	10,706	16,065	2,446	18,511
TWC-JS02	14	21	27	8	35	7,239	10,232	15,509	1,962	17,471
TWC-JS03	9	17	24	2	26	6,881	9,597	14,564	1,914	16,478
TWW-JS01	18	15	29	4	33	8,503	12,825	19,176	2,152	21,328
TWW-JS02	14	28	32	10	42	8,360	12,825	19,023	2,162	21,185
TWW-JS03	13	18	25	6	31	7,465	11,241	17,029	1,677	18,706
TY-MX00	15	25	32	8	40	9,032	14,172	20,873	2,331	23,204
TY-MX01	13	19	28	4	32	6,421	10,179	15,318	1,282	16,600
TY-MX04	14	31	38	7	45	7,137	11,592	17,554	1,175	18,729

Table 6. (Cont'd.).

Sample	CDS-Insertion	CDS-Deletion	CDS-Het	CDS-Homo	CDS-Total	Genome-Insertion	Genome-Deletion	Genome-Het	Genome-Homo	Genome-Total
TY-MX06	19	24	37	6	43	9,028	14,314	21,390	1,952	23,342
TY-MX09	19	27	38	8	46	11,595	18,449	26,779	3,265	30,044
TY-MX15	12	19	27	4	31	5,992	9,479	14,697	774	15,471
TY-MX16	11	21	28	4	32	7,456	11,767	17,462	1,761	19,223
TY-MX18	12	21	30	3	33	7,140	11,426	17,399	1,167	18,566
TY-MX21	11	25	30	6	36	7,247	11,674	17,680	1,241	18,921
TY-MX22	9	16	20	5	25	6,718	10,589	15,732	1,575	17,307
TY-MX23	14	26	35	5	40	7,251	11,663	17,696	1,218	18,914
TY-MX24	15	19	31	3	34	7,910	12,331	18,491	1,750	20,241
TY-MX26	13	20	26	7	33	7,522	11,908	17,757	1,673	19,430
TY-MX32	9	13	19	3	22	4,939	7,818	11,872	885	12,757
TY-MX33	27	48	53	22	75	14,148	21,601	30,034	5,715	35,749
TY-MX35	16	14	27	3	30	5,906	9,494	14,588	812	15,400
TY-MX36	17	26	38	5	43	8,993	14,206	20,672	2,527	23,199
TY-MX40	15	27	32	10	42	7,311	11,720	17,845	1,186	19,031
TY-MX41	13	24	34	3	37	7,612	11,697	17,514	1,795	19,309
TY-MX42	6	10	15	1	16	4,395	7,152	10,991	556	11,547
TY-MX48	12	22	28	6	34	7,258	11,638	17,640	1,256	18,896
TY-MX49	19	29	43	5	48	7,953	12,371	18,695	1,629	20,324
TY-MX50	14	28	38	4	42	8,674	13,758	20,689	1,743	22,432
TY-MX54	20	37	45	12	57	8,507	13,557	20,281	1,783	22,064
TY-MX56	11	21	29	3	32	8,198	12,933	19,766	1,365	21,131
TY-MX59	13	16	25	4	29	7,210	11,302	16,959	1,553	18,512
TY-MX61	18	20	36	2	38	8,169	12,907	19,554	1,522	21,076
TY-MX65	15	23	29	9	38	8,366	12,807	19,259	1,914	21,173
TY-MX68	22	31	44	9	53	9,351	14,500	21,939	1,912	23,851
TY-MX70	18	22	34	6	40	8,892	14,322	21,319	1,895	23,214
TY-MX71	12	24	32	4	36	7,503	12,241	18,428	1,316	19,744
TY-MX72	6	15	18	3	21	5,970	9,434	14,348	1,056	15,404
Total	**121**	**315**	**--**	**--**	**436**	**58,879**	**127,028**	**--**	**--**	**185,907**

Note: CDS: InDel statistics of coding area; InDel statistics of the whole Genome; Insertion: detected number of inserts; Deletion: the number of detected deletions. Heterozygosity: the number of heterozygous indels; Homozygosity: the number of homozygous indels; Total: indicates the total number of detected indels (excluding duplicates). According to the InDel length of samples in the CDS region and the whole genome, the length distribution diagram isshown in the following figure (if there are multiple samples, 50 are displayed by default)

Table 7. Sample group correspondence table.

Sample ID	Q1	Q2	Group
TB-JS01	0.999990	0.000010	Q1
TB-JS02	0.999990	0.000010	Q1
TB-JS03	0.999990	0.000010	Q1
TCM-JT01	0.999990	0.000010	Q1
TCM-JT02	0.999990	0.000010	Q1
TCM-JT03	0.999990	0.000010	Q1
TCM-MHS1	0.999990	0.000010	Q1
TCM-MHS2	0.999990	0.000010	Q1
TCM-MHS3	0.999990	0.000010	Q1
TCM-MX01	0.999990	0.000010	Q1
TCM-MX02	0.999990	0.000010	Q1
TCM-MX03	0.999990	0.000010	Q1
TCM-MX04	0.999990	0.000010	Q1
TCM-MX05	0.999990	0.000010	Q1
TCM-MX06	0.999990	0.000010	Q1
TCM-QS01	0.999990	0.000010	Q1
TCM-QS02	0.999990	0.000010	Q1
TCM-QS03	0.999990	0.000010	Q1
TM-MX01	0.999990	0.000010	Q1
TM-MX02	0.999990	0.000010	Q1
TM-MX03	0.999990	0.000010	Q1
TMxTWM01	0.999990	0.000010	Q1
TMxTWM02	0.999990	0.000010	Q1
TMxTWM03	0.999990	0.000010	Q1
TWC-JS01	0.999990	0.000010	Q1
TWC-JS02	0.999990	0.000010	Q1
TWC-JS03	0.999990	0.000010	Q1
TWW-JS01	0.999990	0.000010	Q1
TWW-JS02	0.999990	0.000010	Q1
TWW-JS03	0.999990	0.000010	Q1
TY-MX00	0.000010	0.999990	Q2
TY-MX01	0.039231	0.960769	Q2
TY-MX04	0.000010	0.999990	Q2
TY-MX06	0.000010	0.999990	Q2
TY-MX09	0.000010	0.999990	Q2
TY-MX15	0.000010	0.999990	Q2
TY-MX16	0.000010	0.999990	Q2
TY-MX18	0.000010	0.999990	Q2
TY-MX21	0.000010	0.999990	Q2
TY-MX22	0.000010	0.999990	Q2
TY-MX23	0.000010	0.999990	Q2
TY-MX24	0.000010	0.999990	Q2
TY-MX26	0.000010	0.999990	Q2
TY-MX32	0.000010	0.999990	Q2
TY-MX33	0.000010	0.999990	Q2
TY-MX35	0.000010	0.999990	Q2
TY-MX36	0.000010	0.999990	Q2
TY-MX40	0.000010	0.999990	Q2
TY-MX41	0.000010	0.999990	Q2
TY-MX42	0.000010	0.999990	Q2
TY-MX48	0.000010	0.999990	Q2
TY-MX49	0.000010	0.999990	Q2
TY-MX50	0.000010	0.999990	Q2
TY-MX54	0.000010	0.999990	Q2
TY-MX56	0.000010	0.999990	Q2
TY-MX59	0.000010	0.999990	Q2
TY-MX61	0.000010	0.999990	Q2
TY-MX65	0.012807	0.987193	Q2
TY-MX68	0.000010	0.999990	Q2
TY-MX70	0.000010	0.999990	Q2
TY-MX71	0.000010	0.999990	Q2
TY-MX72	0.000010	0.999990	Q2

Note: Sample ID: Sample ID; Q1: The possibility that samples come from the first primitive ancestor; Q2: The likelihood that the sample came from a second primitive ancestor; Group: Group of samples

In plants, 15,260,000 polymorphic SLAF tags were obtained by SLAF-Seq and 795,794 high-quality SNP markers were developed in sweet potato (Su *et al.*, 2016). A total of 164,874 SLAF tags including 21,031 polymorphic SLAF tags and having a polymorphism rate of 12.76% , were procured from the parents and two progeny pools in Phalaenopsis (Xiao *et al.*, 2021). In arbor willow, 277,333 high 99,526 polymorphism SLAFs were acquired including 35.89% polymorphism according to the difference gene sequence. 99,526 polymorphism SLAFs were coded by genetics of common alleles of encoding rules, having 58,763 labels coding success (Li *et al.*, 2018). 23,597,049 SLAF tags were obtained, including 370,659 polymorphic SLAF tags and 1,291,290 SNPs of *Pinus bungeana* population were developed (Tian *et al.*, 2021). A total of 857 589 SLAF labels of two *Eucalyptus urophylla* × *E. grandis* clones were developed, and the proportions of A1, A2 and A3 with the same genotype were 35.81% and 77.27% respectively, which reached the expectation of the experimental design to distinguish different clonal varieties (Pang *et al.*, 2019).

In this study, a total of 140,405 SLAF tags were developed, and a total of 7,795,093 population SNPs were obtained for the 62 *Taxus spp.* samples. The phylogenetic tree showed the overall tree divided into two clades, TY (*T. yunnanensis*) was the first calde, and TCM (*T. chinensis* var. *mairei*), TWW (*T. wallichiana* var. *wallichiana*), TB (*T. baccata*), TM (*Taxus*×*media*), TM x TWM (*T. media* × *T. wallichiana* var. *mairei*), TWC (*T. wallichiana* var. *chinensis*) clustered as the second clade. The data amount obtained could be used for the verification and development of specific SNP markers and reveal the genetic relationship for the 7 *Taxus* species. The developed SNP markers can be used for association study to the phenotypes of different individuals in the population and the molecular markers closely linked to a certain trait, so as to achieve fine mapping of good genes.

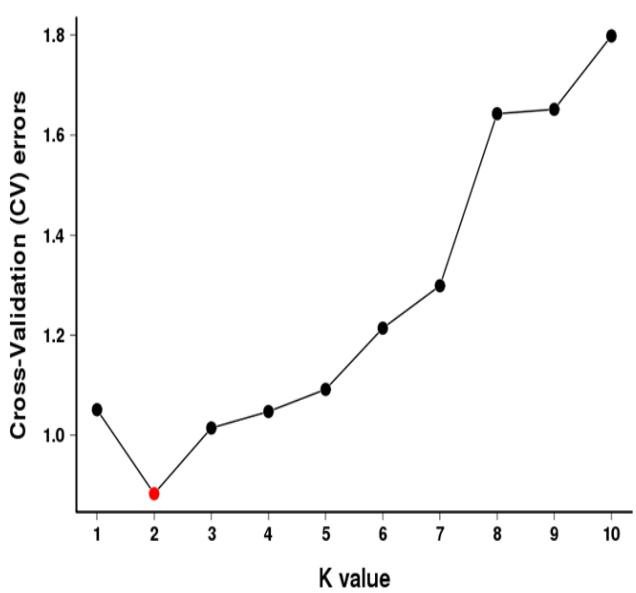

Fig. 8. Cross validation error rate of admixture K values.

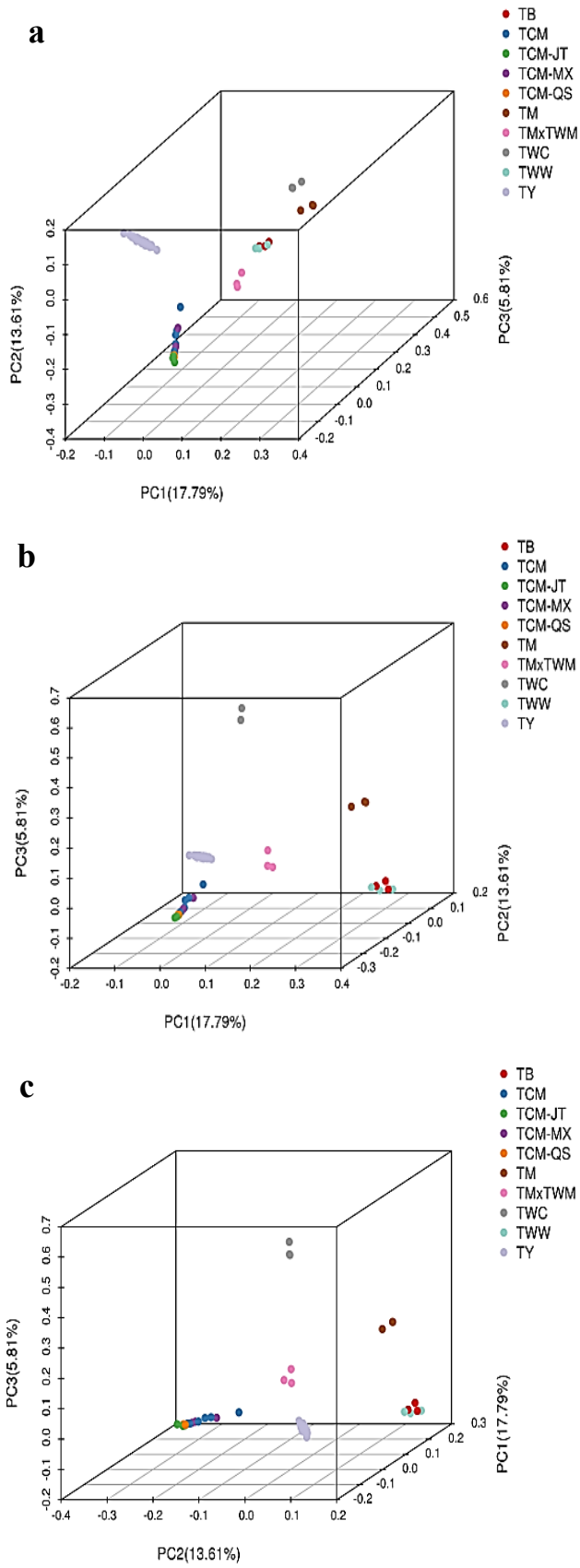

Fig. 9. Sample PCA three-dimensional clustering diagram.
Note: In the figure, the samples were clustered into three dimensions by PCA analysis, where PC1 represented the first principal component and PC2 represented the second principal component.PC3 represents the third principal component. A dot represents a sample and a color represents a group.

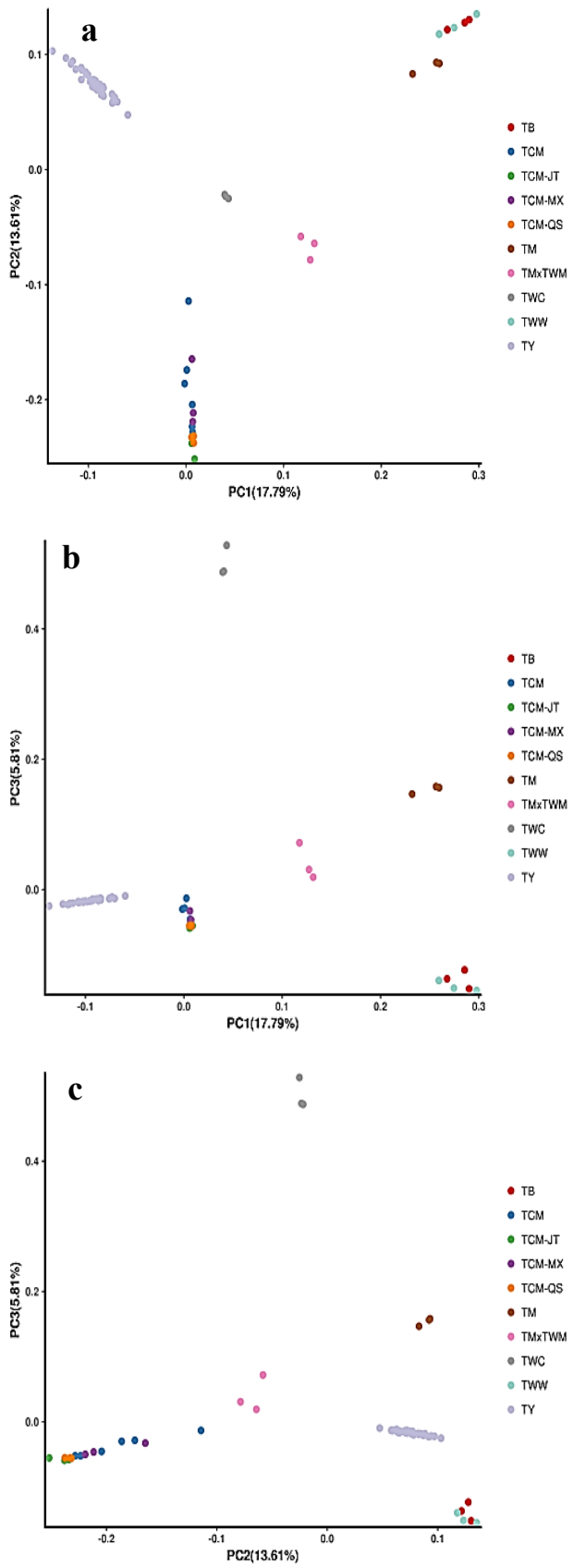

Fig. 10. Sample PCA two-dimensional clustering diagram.
Note: In the figure, the samples were clustered into three dimensions by PCA analysis, where PC1 represented the first principal component and PC2 represented the second principal component.PC3 represents the third principal component. A dot represents a sample and a color represents a group.

Table 8. Sample group correspondence table.

Sample ID	PC1	PC2	PC3	PC4	PC5	PC6	PC7	PC8	PC9	PC10
TB-JS01	0.2677	0.1216	-0.1370	-0.1860	0.0105	0.0513	0.0149	-0.0042	0.5230	0.1302
TB-JS02	0.2855	0.1277	-0.1234	-0.1465	-0.0063	-0.0277	0.0603	-0.0204	0.2285	-0.3283
TB-JS03	0.2901	0.1303	-0.1523	-0.2086	-0.0012	0.0629	0.0090	-0.0102	0.4391	0.0826
TCM-JT01	0.0059	-0.2379	-0.0590	-0.0129	0.1019	-0.0269	0.1929	0.3148	0.0133	-0.0111
TCM-JT02	0.0068	-0.2339	-0.0579	-0.0102	0.0922	-0.0396	0.2039	0.3305	0.0136	-0.0012
TCM-JT03	0.0088	-0.2518	-0.0555	-0.0111	0.0873	-0.0368	0.2170	0.3412	0.0129	-0.0014
TCM-MHS1	0.0008	-0.1742	-0.0285	0.0120	-0.0796	0.0129	-0.0196	0.0141	-0.0123	0.0019
TCM-MHS2	0.0026	-0.1143	-0.0134	0.0115	-0.0617	0.0143	-0.0145	0.0110	-0.0069	0.0058
TCM-MHS3	-0.0014	-0.1861	-0.0300	0.0113	-0.0844	0.0065	-0.0191	0.0052	-0.0235	0.0067
TCM-MX01	0.0061	-0.1647	-0.0328	-0.0045	0.0159	-0.0001	0.0017	0.0028	-0.0063	0.0061
TCM-MX02	0.0074	-0.2115	-0.0464	-0.0100	0.0372	0.0033	0.0069	0.0148	-0.0019	0.0103
TCM-MX03	0.0069	-0.2191	-0.0501	-0.0126	0.0431	-0.0070	0.0062	0.0249	0.0011	0.0119
TCM-MX04	0.0064	-0.2043	-0.0455	-0.0111	0.0521	-0.0040	0.0118	0.0001	-0.0009	0.0072
TCM-MX05	0.0062	-0.2234	-0.0523	-0.0097	0.0525	0.0103	0.0331	0.0422	-0.0028	-0.0011
TCM-MX06	0.0067	-0.2282	-0.0522	-0.0098	0.0474	0.0038	0.0128	0.0128	-0.0071	0.0106
TCM-QS01	0.0056	-0.2326	-0.0554	-0.0127	0.1118	0.0023	-0.2037	-0.3677	0.0093	-0.0178
TCM-QS02	0.0079	-0.2375	-0.0553	-0.0123	0.1213	-0.0050	-0.1891	-0.3805	0.0128	-0.0125
TCM-QS03	0.0079	-0.2317	-0.0559	-0.0155	0.1299	-0.0031	-0.1904	-0.3873	0.0134	-0.0185
TM-MX01	0.2596	0.0922	0.1564	0.4249	0.2264	-0.0703	-0.0033	-0.0019	0.0168	0.2648
TM-MX02	0.2565	0.0931	0.1580	0.4344	0.2310	-0.0465	-0.0126	0.0141	-0.0014	0.4415
TM-MX03	0.2321	0.0831	0.1466	0.3924	0.1634	0.0016	0.1047	-0.0389	-0.0147	-0.7626
TMxTWM01	0.1273	-0.0785	0.0307	0.1699	-0.5615	-0.5280	-0.4993	0.2302	0.0726	-0.0515
TMxTWM02	0.1176	-0.0582	0.0718	0.1489	-0.3107	0.8352	-0.2882	0.1815	-0.0504	-0.0048
TMxTWM03	0.1315	-0.0642	0.0191	0.1419	-0.5900	-0.0090	0.6369	-0.3903	-0.0764	0.1136
TWC-JS01	0.0435	-0.0252	0.5284	-0.2626	0.0049	-0.0082	0.0001	0.0026	0.0004	-0.0043
TWC-JS02	0.0406	-0.0233	0.4885	-0.2417	0.0137	-0.0287	0.0157	-0.0097	0.0039	0.0028
TWC-JS03	0.0396	-0.0218	0.4872	-0.2435	0.0068	-0.0278	0.0058	-0.0027	-0.0080	0.0081
TWW-JS01	0.2977	0.1352	-0.1550	-0.2022	0.0343	-0.0332	-0.0370	0.0208	-0.4147	0.0447
TWW-JS02	0.2746	0.1233	-0.1515	-0.2010	0.0285	-0.0355	-0.0325	0.0181	-0.3220	0.0483
TWW-JS03	0.2593	0.1177	-0.1396	-0.1814	0.0291	-0.0218	-0.0427	0.0257	-0.4352	-0.0069
TY-MX00	-0.1178	0.0919	-0.0230	-0.0005	-0.0157	-0.0027	-0.0073	-0.0001	-0.0018	0.0024
TY-MX01	-0.0754	0.0579	-0.0132	0.0141	0.0177	-0.0005	0.0038	-0.0017	0.0004	-0.0048
TY-MX04	-0.0851	0.0721	-0.0134	0.0039	-0.0062	0.0033	-0.0029	0.0001	-0.0012	-0.0012
TY-MX06	-0.1129	0.0871	-0.0208	0.0038	-0.0090	0.0009	-0.0006	0.0011	-0.0035	0.0023
TY-MX09	-0.1230	0.0970	-0.0223	0.0072	-0.0059	-0.0024	-0.0040	-0.0037	0.0029	-0.0006
TY-MX15	-0.0730	0.0627	-0.0117	0.0027	-0.0097	-0.0020	-0.0006	0.0009	-0.0028	0.0010
TY-MX16	-0.0957	0.0720	-0.0175	0.0143	0.0185	-0.0035	0.0067	0.0019	0.0011	-0.0007
TY-MX18	-0.0949	0.0782	-0.0171	0.0020	-0.0106	-0.0055	-0.0039	-0.0004	0.0015	0.0035
TY-MX21	-0.0845	0.0708	-0.0136	0.0032	-0.0111	0.0030	-0.0023	0.0012	-0.0020	-0.0008
TY-MX22	-0.0849	0.0650	-0.0146	0.0153	0.0208	-0.0029	0.0085	-0.0033	0.0022	-0.0014
TY-MX23	-0.0931	0.0774	-0.0169	0.0045	-0.0118	-0.0018	-0.0004	-0.0010	0.0018	0.0030
TY-MX24	-0.0845	0.0638	-0.0159	0.0142	0.0168	-0.0048	0.0073	-0.0012	0.0035	0.0037
TY-MX26	-0.0983	0.0765	-0.0185	0.0149	0.0193	-0.0029	0.0076	0.0026	0.0047	-0.0072
TY-MX32	-0.0703	0.0589	-0.0137	0.0022	-0.0048	-0.0029	-0.0001	0.0041	0.0016	0.0019
TY-MX33	-0.1370	0.1030	-0.0252	0.0171	0.0155	-0.0020	-0.0004	0.0030	0.0018	0.0073
TY-MX35	-0.0758	0.0655	-0.0142	0.0023	-0.0127	0.0020	0.0009	0.0003	0.0014	0.0063
TY-MX36	-0.1071	0.0782	-0.0200	0.0172	0.0231	-0.0068	0.0017	-0.0030	0.0015	0.0018
TY-MX40	-0.0885	0.0740	-0.0143	0.0034	-0.0113	0.0008	0.0035	0.0028	0.0018	0.0054
TY-MX41	-0.0936	0.0712	-0.0170	0.0162	0.0201	-0.0033	0.0027	0.0016	0.0037	0.0011
TY-MX42	-0.0596	0.0475	-0.0098	0.0112	0.0147	-0.0023	0.0019	-0.0013	0.0032	-0.0023
TY-MX48	-0.0860	0.0715	-0.0147	0.0029	-0.0125	-0.0019	-0.0035	-0.0003	-0.0018	0.0016
TY-MX49	-0.0860	0.0664	-0.0155	0.0136	0.0186	-0.0017	0.0055	-0.0017	0.0033	-0.0030
TY-MX50	-0.1073	0.0885	-0.0184	0.0047	-0.0096	0.0025	-0.0036	-0.0044	0.0007	-0.0003
TY-MX54	-0.1003	0.0821	-0.0179	0.0032	-0.0103	-0.0031	0.0005	0.0020	0.0007	-0.0022
TY-MX56	-0.1028	0.0847	-0.0183	0.0055	-0.0047	-0.0009	-0.0031	-0.0019	0.0002	0.0024
TY-MX59	-0.0869	0.0651	-0.0170	0.0156	0.0195	-0.0008	0.0046	0.0077	0.0011	0.0038
TY-MX61	-0.1023	0.0830	-0.0193	0.0039	-0.0089	-0.0003	-0.0011	0.0029	0.0006	0.0020
TY-MX65	-0.0912	0.0694	-0.0149	0.0159	0.0201	-0.0040	-0.0015	-0.0002	0.0021	0.0000
TY-MX68	-0.1162	0.0941	-0.0217	0.0045	-0.0114	-0.0027	-0.0016	0.0016	-0.0018	0.0043
TY-MX70	-0.1064	0.0864	-0.0196	0.0042	-0.0111	-0.0022	-0.0030	-0.0017	-0.0028	-0.0020
TY-MX71	-0.0926	0.0770	-0.0146	0.0035	-0.0101	0.0061	-0.0003	-0.0013	-0.0036	-0.0034
TY-MX72	-0.0749	0.0594	-0.0132	0.0121	0.0170	-0.0003	0.0033	-0.0001	0.0031	0.0010

Note: Sample ID: Sample ID; BMK ID: Uniform serial number of the research samples by Baimai; PC1: first principal component; PC2: second principal component; PC3: the third principal component; etc.

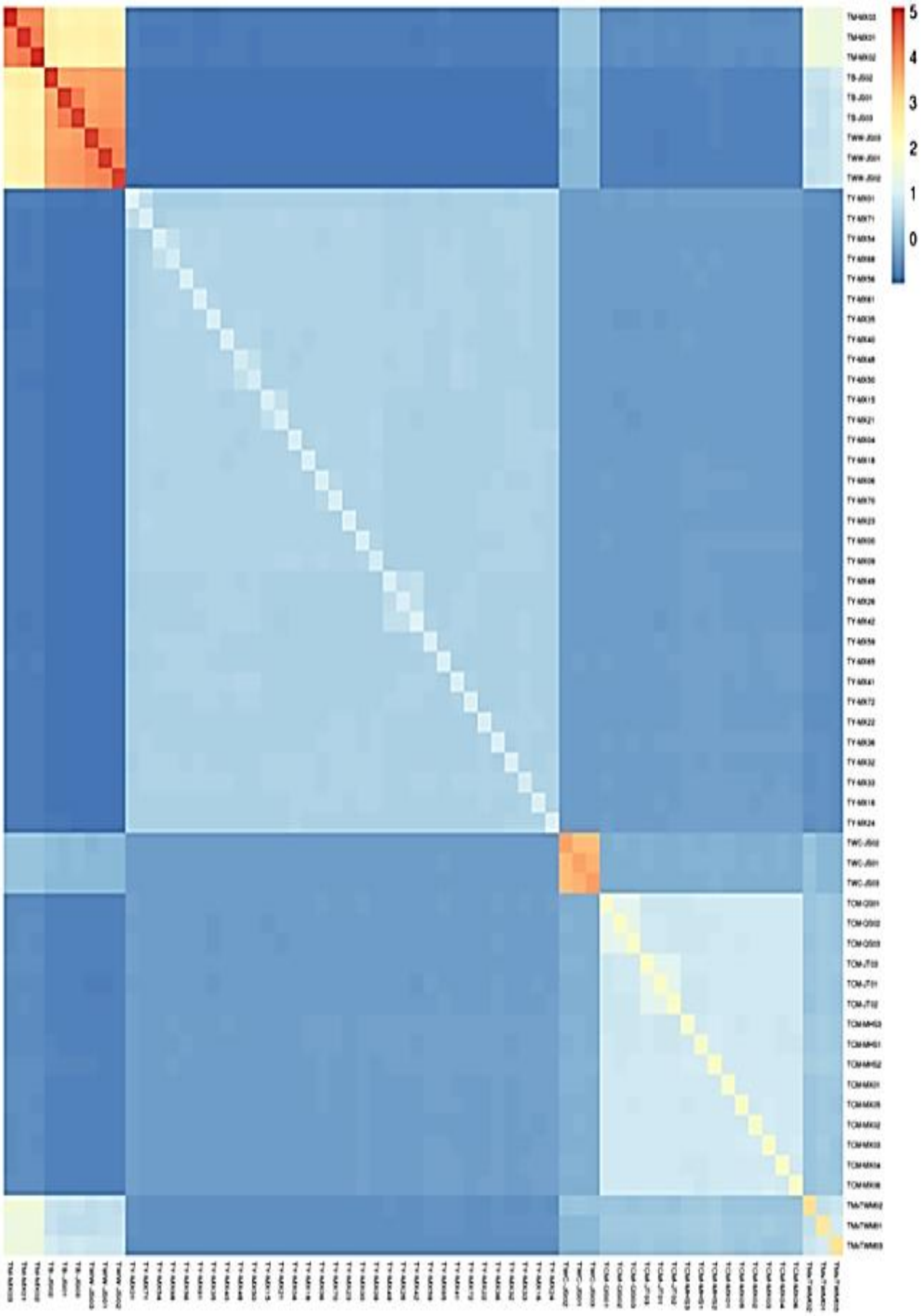

Fig. 11. Heatmap of kinship coefficient.

Table 9. Population genetic diversity.

Group	Average_MAF	Expected_allele_number	Expected_heterozygous_number	Nei_diversity_index	Number_of_poly_marker	Observed_allele_number	Observed_heterozygous_number	Polymorphysm_information_content	Shnnon_Wiener_index
TB	0.347	1.000-2.000(1.079)	0.278-0.500(0.044)	0.333-0.667(0.055)	7889	1.000-2.000(1.103)	0.333-1.000(0.028)	0.239-0.375(0.034)	0.451-0.693 (0.063)
TCM	0.317	1.000-2.000(1.171)	0.153-0.500(0.097)	0.167-0.667(0.114)	38984	1.000-2.000(1.240)	0.167-1.000(0.032)	0.141-0.375(0.076)	0.287-0.693 (0.141)
TCM-JT	0.347	1.000-2.000(1.126)	0.278-0.500(0.070)	0.333-0.667(0.088)	20873	1.000-2.000(1.166)	0.333-1.000(0.044)	0.239-0.375(0.055)	0.451-0.693 (0.101)
TCM-MX	0.372	1.000-2.000(1.153)	0.278-0.500(0.083)	0.333-0.667(0.106)	20241	1.000-2.000(1.190)	0.333-1.000(0.034)	0.239-0.375(0.064)	0.451-0.693 (0.119)
TCM-QS	0.345	1.000-2.000(1.121)	0.278-0.500(0.067)	0.333-0.667(0.084)	19465	1.000-2.000(1.159)	0.333-1.000(0.043)	0.239-0.375(0.052)	0.451-0.693 (0.097)
TM	0.351	1.000-2.000(1.129)	0.278-0.500(0.071)	0.333-0.667(0.090)	13642	1.000-2.000(1.168)	0.333-1.000(0.058)	0.239-0.375(0.056)	0.451-0.693 (0.103)
TMxTWM	0.368	1.000-2.000(1.335)	0.278-0.500(0.183)	0.333-0.667(0.233)	40477	1.000-2.000(1.420)	0.333-1.000(0.091)	0.239-0.375(0.142)	0.451-0.693 (0.262)
TWC	0.355	1.000-2.000(1.087)	0.278-0.500(0.048)	0.333-0.667(0.060)	8875	1.000-2.000(1.111)	0.333-1.000(0.021)	0.239-0.375(0.037)	0.451-0.693 (0.069)
TWW	0.349	1.000-2.000(1.056)	0.278-0.500(0.031)	0.333-0.667(0.039)	5482	1.000-2.000(1.074)	0.333-1.000(0.028)	0.239-0.375(0.024)	0.451-0.693 (0.045)
TY	0.215	1.000-2.000(1.163)	0.031-0.500(0.098)	0.031-0.571(0.101)	59365	1.000-2.000(1.334)	0.031-0.774(0.031)	0.030-0.375(0.079)	0.080-0.693 (0.151)

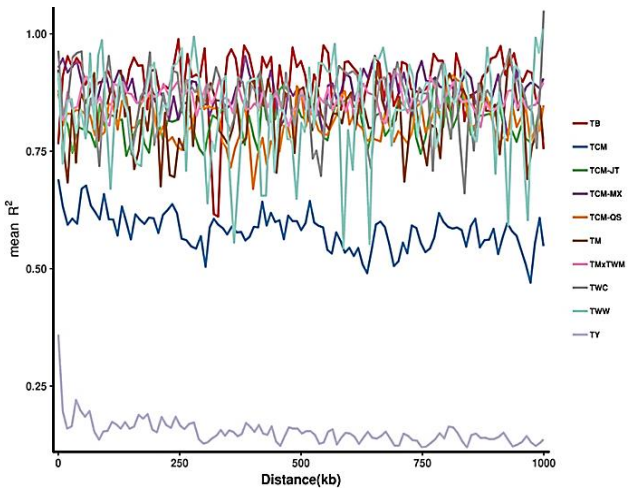

Fig. 12. LD decay.

Conclusions

The protocol was predicted by using the genome of *T. chinensis* var. *mairei* and digested with the *Hpy*166II restriction enzyme, the SLAF tags of length 314-364 bp were predicted. Read length 126bp x 2 was used for subsequent data evaluation and analysis. A total of 148.78 Mb reads data were obtained from 62 *Taxus spp.* Samples including 7 species, with 94.46% of an average Q30 and 37.99% of an average GC content. A total of 140,405 SLAF tags were developed with an average sequencing depth of 7.54 x. SNPs were developed by GATK and SamTools, and the intersection of SNPs acquired using the two methods was applied as the reliable SNP marker dataset, and a total of 7,795,093 population SNPs were obtained. It can be used to the development of subsequent specific SNP markers. From these results, the 7 species of *Taxus spp*. could be divided into two clades, *T. yunnanensis* was the first clade, and the others clustered as the second clade. The data amount obtained could be used for the verification and development of specific SNP markers and reveal the genetic relationship for the 7 *Taxus* species.

Acknowledgements

This research was funded by the Public Welfare Special Projects for Research Institutes from Fujian Department of Science and Technology, Project title: The Screening and High-efficient Cultivation Technique of High Paclitaxel *Taxus* (Grant number: 2016R1010-3); China Central Governmental Forestry Science and Technology Promotion Project, Project title: Demonstration Area Construction for Standardized Cultivation of Paclitaxel *Taxus* Raw Material Forest. [Grant number: MIN (TG18)]; Study on Key Technique of Breeding and Industrialization of High Paclitaxel *Taxus* (Grant number: 2015N618). The study was supported by the Key Laboratory of Timber Forest Breeding and Cultivation for Mountainous Areas in Southern China, China National Forestry and Grassland Administration and the Key Laboratory of Forest Culture and Forest Product Processing Utilization of Fujian Province. We acknowledge Dr. Zhong Wang, Institute of Botany, Jiangsu Province and Chinese Academy of Sciences, for his helps of collecting same *Taxus* samples. Thanks to Biomaker Technology Company Limited for sequencing.

References

Alexander, D.H., J. Novembre and K. Lange. 2009. Fast model-based estimation of ancestry in unrelated individuals. *Genom. Res.*, 19: 1655-1664.

Cope, E.A. 1998. Taxaceae: the genera and cultivated species. *Bot. Rev.*, 64: 291-322. DOI: https://doi.org/10.1007/BF02857621.

Hao, D.C., P.G. Xiao, B.L. Huang, G.B. Ge and L. Yang. 2008. Interspecific relationships and origins of Taxaceae and Cephalotaxaceae revealed by partitioned Bayesian analyses of chloroplast and nuclear DNA sequences. *Plant System. Evol.*, 276: 89-104. https://doi.org/10.1007/s00606-008-0069-0.

Kozich, J.J., S.L. Westcott, N.T. Baxter, S.K. Highlander and P.D. Schloss. 2013. Development of a dual-index sequencing strategy and curation pipeline for analyzing amplicon sequence data on the MiSeqIllumina sequencing platform. *Appl. Environ. Microbiol.*, 79: 5112-5120. DOI: 10.1128/AEM.01043-13.

Kumar S., G. Stecher, M. Li, C. Knyaz and K. Tamura. 2018. MEGA X: Molecular evolutionary genetics analysis across computing platforms. *Mol. Biol. Evol.*, 35: 1547-1549.

Li, H. and D.A. Rechard. 2009. Fast and accurate short read alignment with Burrows-Wheeler Transform. *Bioinformatics*, 25:1754-1760.

Li, H., B. Handsaker, A. Wysoker, T. Fennell, J. Ruan, N. Homer and G. Marth. 2009a. The sequence alignment/map format and SAMtools. *Bioinformatics*, 25: 2078-2079.

Li, J.M., B.H. Chen, S.H. Jiang, B.M. Ma and S.J. Trueman. 2020. Paternity analysis and SNP development of *Osmanthus fragrans* 'Pucheng Dangui' using Slaf-seq. *Pak. J. Bot.*, 52: 213-218. DOI: http://dx.doi.org/10.30848/PJB2020-1(15).

Li, M., C. Guo, Y. Wang, Y. Li, F. Tan and J. Han. 2018. SNP sites developed by SLAF-seq technology in arbor willow. *Southwest China J. Agri. Sci.*, 31: 891-895. Doi: 10.16213/j.cnki.scjas.2018. 5.003

Li, R., Y. Li, K. Kristiansen and J. Wang. 2008. SOAP: short oligonucleotide alignment program, *Bioinformatics*, 24(5): 713-714. https://doi.org/10.1093/bioinformatics/btn025.

Li, R.Q., C. Yu, Y.R. Li, T.W. Lam, S.M. Yiu, K. Kristiansen and J. Wang. 2009b. SOAP2: an improved ultrafast tool for short read alignment. *Bioinformatics*, 25: 1966-1967. DOI:10.1093/bioinformatics/btp336.

McKenna, A., M. Hanna, E. Banks, A. Sivachenko and K. Cibulskis. 2010. The genome analysis toolkit: A MapReduce framework for analyzing next-generation DNA sequencing data. *Genom. Res.*, 20: 1297-1303.

Miller, C.N. 1977. Mesozoic conifers. *Bot. Rev.*, 43: 217-280. DOI: https://doi.org/10.1007/BF02860718.

Pang, Z.W., X. Liu, X.Y. Sun, J. Lan and D.W. Cheng. 2019. Application of SLAF-seq technology in identification of *Eucalyptus* varieties. *Eucal. Sci. Technol.*, 36: 26-28. https://www.cnki.com.cn/Article/CJFDTotal-ASKJ201901005.htm.

Price, A.L., N.J. Patterson, R.M. Plenge, M.E. Weinblatt, N.A. Shadick and D. Reich. 2006. Principal components analysis corrects for stratification in genome-wide association studies. *Nature Gen.*, 38: 904. https://doi.org/10.1038/ng1847

Price, R.A. 1990. The genera of Taxaceae in the southeastern United States. *J. Arnold Arboretum*, 71(1): 69-91. DOI: https://doi.org/10.5962/bhl.part.24926.

Su, W., N. Zhao, J. Lei, L. Wang, S. Chai and X. Yang. 2016. SNP sites developed by specific length amplification fragment sequencing (SLAF-seq) in sweet potato. *Sci. Agri. Sinica*, 49: 27-34. doi: 10.3864/j.issn.0578-1752.2016.01.003.

Sun, X., D. Liu, X. Zhang, W. Li, H. Liu, W. Hong, C.B. Jiang, N. Guan, C.X. Ma, H.P. Zeng, C.H. Xu, J. Song, L. Huang, C.M. Wang, J.J. Shi, R. Wang, X.H. Zheng, C.Y. Lu, X.W. Wang and H.K. Zheng. 2013. SLAF-seq: an efficient method of large-scale *De novo* SNP discovery and genotyping using high-throughput sequencing. *PloS One*, 8(3): e58700. DOI:10.1371/journal.pone.0058700.

Tian, Q., S. Liu, S. Niu and W. Li. 2021. Development of SNP molecular markers of *Pinus bungeana* based on SLAF-seq technology. *J. Beijing Forest. Univ.*, 43: 1-8. DOI: 10.12171/j.1000-1522.20200211

Wu, Z.Y. 1986. Flora of Yunan, Vol. 4. Science Press: Beijing, China.

Wu, Z.Y. and H.S.Wang. 1983. Natural Geography of China (Phytogeography) Vol. 1. Science Press: Beijing, China.

Xiao, W.F., Z. Li, H.M. Chen and F.B. Lv. 2021. Identification and validation of single-nucleotide polymorphism markers linked to flower ground color in Phalaenopsis by using combined specific-locus amplified fragment sequencing and bulked segregant analysis. *J. China Agri. Univ.*, 26: 92-100. DOI: 10.11841/j.issn.1007-4333.2021.09.10

Yang, J., S.H. Lee, M.E. Godd ard and P.M. Visscher. 2011. GCTA: a tool for genome-wide complex trait analysis. *Amer. J. Hum. Genet.*, 88: 76-82. doi: 10.1016/j.ajhg.2010.11.011. pmid:21167468.

Zhang, C., S. Dong, J. Xu, W. He and T. Yang. 2019. PopLD decay: a fast and effective tool for linkage disequilibrium decay analysis based on variant call format files. *Bioinformatics*, 35: 1786-1788. https://doi.org/10.1093/bioinformatics/bty875.

Zheng, W.J. and L.G. Fu. 1978. Flora of China, vol. 7. Press: Beijing, China.

(Received for publication 22 February 2023)