Research Article Open Access

SNP Molecular Markers Development and Genetic Diversity Analysis of *Forsythia suspensa* Based on SLAF-seq Technology

Tao Jiang, Chunxiu Wen, Wei Tian, Xiaoliang Xie, Ruike Lu, Saiqun Wen, Lingdi Liu
Hebei Medicinal Plant & Technology Innovation Center, Institute of Cash Crops, Hebei Academy of Agricultural and Forestry Sciences, Shijiazhuang, 050051, China
Corresponding author email: nkyliulingdi@126.com

Medicinal Plant Research, 2022, Vol.12, No.2 doi: 10.5376/mpr.2022.12.0002
Received: 14 Apr., 2022
Accepted: 22 Apr., 2022
Published: 09 May, 2022

Copyright © 2022 Jiang et al., This article was first published in Molecular Plant Breeding in Chinese, and here was authorized to translate and publish the paper in English under the terms of Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Preferred citation for this article:
Jiang T., Wen C.X., Tian W., Xie X.L., Lu R.K., Wen S.Q., and Liu L.D., 2022, SNP Molecular markers development and genetic diversity analysis of *Forsythia suspensa* based on SLAF-seq technology, Medicinal Plant Research, 12(2): 1-10 (doi: 10.5376/mpr.2022.12.0002)

Abstract *Forsythia suspensa* (Thunb.) Vahl is an important medicinal plant that has great value to study in China. In this study, the 39 *Forsythia suspensa* materials were used specific loci amplified fragment sequencing technology (SLAF-seq) to develop SNP molecular markers. A total of 112.28 Mb reads data were obtained by sequencing. The reads data of each sample ranged from 1 324 860~5 911 565. The average sequencing quality value (Q30) and GC content of samples was 96.29% and 36.97%, respectively. Analysis of bioinformatics, there were 535 357 SLAF tags, in which 262 297 SLAF tags were polymorphic, and the average sequencing depth of the samples was 16.20 x. A total of 1 809 741 SNP molecular markers were obtained and 39 *Forsythia suspensa* materials were divided into 4 groups. This study could provide theoretical basis for germplasm resource identification and genetic diversity analysis of *Forsythia suspensa*.

Keywords *Forsythia suspensa* (Thunb.) Vahl; SNP; SLAF-seq; Molecular marker

Forsythia suspensa (Thunb.) Vahl is a deciduous shrub of *Forsythia* genus in the family of Oleaceae, mainly distributed in Hebei, Shanxi, Shaanxi, Shandong, Anhui, Henan and other places in China. *Forsythia suspensa* has high medicinal value, which is recorded in 'Divine Farmer’s Classic of Materia Medica' and 'Compendium of Materia Medica'. Its fruit can be used as medicine, which is divided into 'Qingqiao' and 'Laoqiao' in Chinese. *Forsythia suspensa* has the effect of clearing heat and removing toxicity, eliminating swelling and dispersing knots, which is the main component of Lianhuaqingwen capsules (Xie et al., 2010). Fruits of *Forsythia suspensa* contain a large number of flavonoids, lignin, alkaloids and acids, which are commonly used as bulk medicinal materials in China and are widely used in traditional Chinese medicine enterprises. Studies have shown that the fruits of *Forsythia suspensa* have obvious effects of anti-inflammatory, antibacterial, antiviral and hypolipidemic, which has important development value and application prospect in clinical research and drug development (Jiang, 2015). *Forsythia suspensa* is full of treasures, in which *Forsythia suspensa* leaves are rich in nutrients, such as total flavonoids, forsythiaside and phillyrin. And a kind of health care drink of *Forsythia suspensa* has been developed (Wang et al., 2016).

Forsythia suspensa is an important medicinal plant that has great value to study, the existing research on it is mainly focused on planting and cultivation techniques, determination of extract components and development and utilization of metabolite active components and so on. *Forsythia suspensa* resources are mainly wild, soil environment, climate conditions and varieties have great influence on the yield and quality of *Forsythia suspensa*. The wild *Forsythia suspensa* resources are limited, which are difficult to meet the needs of the market, and it is also easy to destroy the diversity of wild resources. However, the existing artificially cultivated *Forsythia suspense* is a wild resource without systematic artificial cultivation. It is mixed with various types, and its yield and quality are uneven. The lack of good *Forsythia suspensa* varieties has become the 'bottleneck' problem of large-scale and industrial production for *Forsythia suspensa*. Collecting and evaluating the existing *Forsythia suspensa* germplasm resources and screening high-yield and high-quality *Forsythia suspensa* germplasm are of great significance to the development of *Forsythia suspensa* industry. Therefore, the analysis

of genetic diversity of *Forsythia suspensa* germplasm resources and the development of unique molecular markers of *Forsythia suspensa* are beneficial to the collection and variety identification of *Forsythia suspensa* germplasm resources.

Forsythia suspensa has strong geographical adaptability and abundant genetic diversity. Fully exploiting genetic information can effectively improve the efficiency of germplasm resource identification and breeding of *Forsythia suspense*. Molecular markers such as randomly amplified polymorphic DNA (RAPD) and simple repeat sequences (SSR) have been developed for genetic relationship identification and polymorphism analysis of *Forsythia suspensa*. RAPD markers was used to analysis the *Forsythia suspensa* resources from different habitats and styles. It was considered that the differences of environmental factors would lead to the genetic diversity of *Forsythia suspensa* resources (Li et al., 2011, Proceedings of the 10th National Conference on Medicinal Plants and Phytomedicines, pp.118). Li et al. (2019) screened 13 ISSR primers to analyze 25 *Forsythia suspensa* germplasm resources. The results showed that the genetic diversity of *Forsythia suspensa* germplasm resources were related to ecological factors and growth environment. Meng (2014) used 11 pairs of polymorphic SSR primers to detect *Forsythia suspensa* germplasm resources in 16 regions of China, suggested that SSR molecular markers could be used for the genetic diversity analysis of *Forsythia suspensa* germplasm resources. Wang et al. (2015) developed SSR molecular markers for *Forsythia suspensa* based on RNA-seq transcriptome sequencing technology. A total of 3 199 SSR loci were detected, and 12 pairs of polymorphic SSR primers were screened to distinguish *Forsythia suspensa* resources from different habitats.

Specific-Locus Amplified Fragment Sequencing (SLAF-seq) is a new simplified genome sequencing technology, which has been widely used in germplasm collection, evaluation and genetic diversity analysis. At present, a large number of SNP markers have been developed based on SLAF-seq technology from *Manihot esculenta*, *Helianthus annuus*, *Camellia oleifera* and *Perilla frutescens*. However, there are few reports on the development of SNP molecular markers using SLAP-seq sequencing technology in *Forsythia suspensa*. In this study, the SLAF-seq database of *Forsythia suspensa* was constructed to screen the specific fragments from the database for sequencing. The bioinformatics software was used to analyze the polymorphism SLAF of *Forsythia suspensa*, so as to develop specific SNP molecular markers and provide reference for the identification and breeding of *Forsythia suspensa* germplasm resources.

1 Results and Analysis

1.1 Database construction and evaluation

SLAF-predict software was used to predict the reference genome of *Olea europaea* by electronic restriction. According to the principle of whether the enzyme segments were evenly distributed on the genome and whether the size of the enzyme segments was consistent with the experimental system (Davey et al., 2013), the restriction endonuclease combination was *Rsa*I+*Hpy*166II. Enzyme digestion efficiency is a key indicator to evaluate SLAF experiment. The experimental results showed that the efficiency of *Rsa*I+*Hpy*166II double digestion was 90.01%, and the proportion of residual enzyme digestion sites was 9.99%, indicating that the enzyme digestion reaction is normal. In this study, the sequences with the size of 364~414 bp in the enzyme section were identified as SLAF tags, and 129 643 SLAF tags were estimated. To evaluate the effectiveness of the SLAF experiment process and the implementation of the enzyme digestion program, this study used SOAP (Li et al., 2009b) software to compare the sequencing data of Nipponbare (rice) and the rice genome data. The results showed that the paired-end mapped was 98.09%, the singles-end mapped reads was 0.59%, and the unmapped reads was 1.32% (Table 1). The real fragment size of SLAF was calculated based on the position of the sequencing double-terminal comparison sequence on the reference genome (rice), and the distribution of insert fragments of rice sequences was drawn (Figure 1). The results showed that most of the inserted fragments were in the range of 364~414 bp, indicating that the sequencing method used in this study was highly reliable and the SLAF database was basically normal.

Table 1 The alignment of sequencing data with its Nipponbare (rice) genome sequences data

Sample	Paired-end mapped (%)	Singles-end mapped reads (%)	Unmapped reads (%)
Nipponbare (rice)	98.09	0.59	1.32

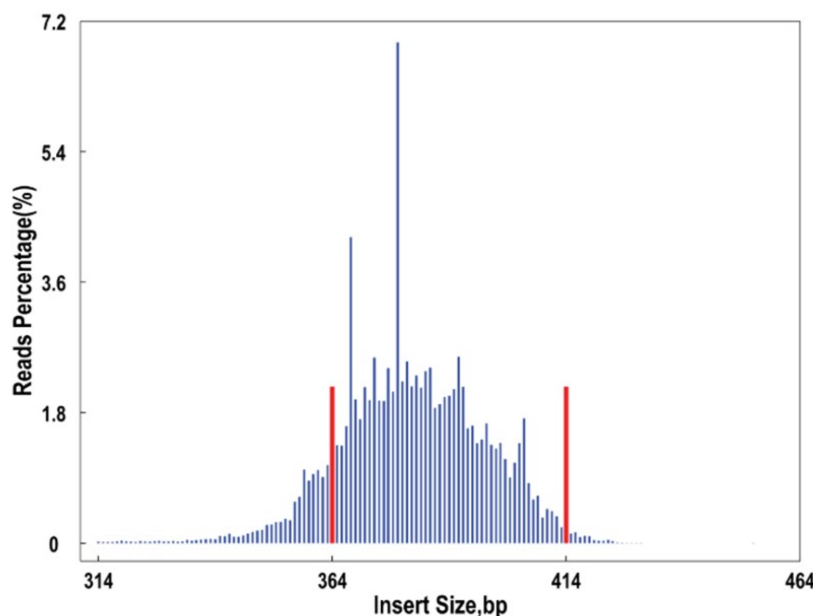

Figure 1 Distribution of insert fragments of rice sequences

1.2 Evaluation of sequencing data

To verify the accuracy of the experimental data analysis, the reads of 126 bp×2 was used as the analysis standard for subsequent data evaluation in this study. The sequencing quality value Q30 was used to evaluate the accuracy of sequencing single base. The base separation was checked by base distribution. The accuracy of the database was evaluated with Nipponbare (rice) as the control. After sequencing on the Illumina platform, the data amount of 0.46 Mb reads was obtained by the control, and a total of 112.28 Mb reads were obtained from 39 *Forsythia suspensa* samples. The reads data of each sample were between 1 324 860 and 5 911 565. Among them, No.6 (an) of Quzhou City Hebei Province obtained the largest amount of data of 5 911 565 reads. The data amount of Nipponbare (rice) reads was the smallest, which was 457 544 reads. The average Q30 percentage of this experiment was 96.29%, and the average GC percentage was 36.97% (Table 2). The sequencing results showed that the sequencing quality of this study was high, and the data were reliable to meet the sequencing requirements.

1.3 Development of SLAF tag and SNP marker

The results of *Forsythia suspensa* sequencing data analysis showed that a total of 535 357 SLAF tags were developed from 39 *Forsythia suspensa* germplasm resources, with an average of 162 683 SLAF tags per *Forsythia suspensa* sample. The average sequencing depth of tags was 16.20 x, and 262 297 polymorphic SLAF tags were developed. Based on the SLAF tag of *Forsythia suspensa*, 1 809 741 SNP molecular markers were developed. The number distribution of SNP in samples was 858 739~1 146 427, the hetloci ratio was 7.86%~13.36%, and the integrity ratio of SNP was 47.45%~63.34% (Table 3; Table 4).

Table 2 Statistics of sequencing data in different samples

Sample ID	ID	Total Reads	GC percentage (%)	Q30 percentage (%)
1	aa	3 384 290	37.42	96.45
2	aj	2 577 387	36.78	96.46
3	ak	5 084 443	36.46	96.33
4	al	3 777 474	36.93	96.46
5	am	2 474 741	36.72	96.2
6	an	5 911 565	35.72	96.15
7	ao	3 129 418	37.83	96.59
8	ap	3 532 885	36.06	96.17
9	aq	2 833 826	36.19	96.23
10	ar	2 439 553	37.45	96.23
11	as	2 818 605	37.85	96.09
12	ab	2 911 108	37.33	96.31
13	at	1 973 474	37.74	96.49
14	au	2 430 988	37.37	96.35
15	av	2 027 523	37.76	96.27
16	aw	2 478 635	37.45	96.36
17	ax	3 890 304	36.06	95.62
18	ay	3 769 908	36.43	95.63
19	az	2 680 306	37.48	96.42
20	ba	2 948 217	37.48	96.43
21	bb	2 626 893	37.56	96.55
22	bc	2 560 900	37.18	96.4
23	ac	1 814 866	38.4	96.24
24	bd	2 388 160	37.95	96.32
25	be	2 568 714	36.47	96.24
26	bf	3 252 986	36.49	96.33
27	bg	3 487 131	36.18	96.51
28	bh	1 324 860	37.13	96.26
29	bi	1 929 080	36.68	96.22
30	bj	3 638 878	36.72	96.27
31	bk	2 505 863	36.84	96.48
32	bl	2 561 008	36.9	96.29
33	bm	3 425 362	35.76	95.64
34	ad	2 041 422	36.49	96.5
35	ae	2 208 379	37.97	96.34
36	af	3 895 811	36.25	96.3
37	ag	2 479 782	36.8	96.16
38	ah	1 490 578	37.03	96.56
39	ai	3 010 172	36.7	96.63
SD	Rice	457 544	42.67	96.18

1.4 Genetic structure analysis

Genetic structure analysis can quantify the number of ancestors of the samples studied and infer the genetic relationship of each sample. It is a genetic structure analysis method that is widely used at present, which is helpful to study the evolution process of research materials. In this study, admixture software (Alexander et al., 2009) was used to analyze the genetic structure of the developed SNP molecular markers of *Forsythia suspensa*

germplasm resources. Based on the cross validation error rate of samples, the valley value of cross validation error rate is determined as the optimal grouping. The results of genetic structure analysis (Figure 2) showed that when K-value was 4, it was the best subgroup. Therefore, 39 *Forsythia suspensa* germplasm resources could be divided into 4 subgroups according to the developed SNP molecular markers (Table 5).

Table 3 The statistics of SLAF and SNP

Sample ID	ID	SLAF number	Average depth	SNP number	Hetloci ratio (%)	Integrity ratio (%)
1	aa	169 975	19	1 049 141	9	58
2	ab	163 295	17	1 025 446	9	57
3	ac	155 430	10	981 621	8	54
4	ad	143 000	13	924 529	9	51
5	ae	151 834	13	965 501	9	53
6	af	174 192	21	1 057 089	10	58
7	ag	147 433	16	937 316	9	52
8	ah	133 555	10	864 029	8	48
9	ai	170 733	16	1 053 950	10	58
10	aj	155 948	15	988 828	9	55
11	ak	184 596	26	1 112 633	11	61
12	al	176 511	20	1 086 172	10	60
13	am	155 874	15	990 893	9	55
14	an	195 155	28	1 146 427	11	63
15	ao	169 739	17	1 056 161	9	58
16	ap	171 134	19	1 059 477	10	59
17	aq	162 852	16	1 023 975	10	57
18	ar	156 099	15	980 040	9	54
19	as	169 554	15	1 050 574	9	58
20	at	152 610	12	966 384	9	53
21	au	160 226	14	999 052	9	55
22	av	164 199	11	1 019 050	13	56
23	aw	158 413	15	1 008 801	9	56
24	ax	177 093	20	1 097 271	11	61
25	ay	180 930	19	1 109 397	10	61
26	az	168 507	15	1 048 006	9	58
27	ba	168 707	16	1 046 217	9	58
28	bb	170 532	14	1 053 139	10	58
29	bc	157 558	15	990 060	9	55
30	bd	165 000	13	1 022 117	9	56
31	be	168 097	14	1 054 124	10	58
32	bf	171 320	18	1 052 201	10	58
33	bg	171 488	19	1 066 954	10	59
34	bh	128 576	10	858 739	8	47
35	bi	142 505	13	925 649	8	51
36	bj	160 651	21	983 534	9	54
37	bk	150 649	16	953 261	9	53
38	bl	151 132	16	956 205	9	53
39	bm	169 570	19	1 047 302	10	58

Table 4 The statistics of SLAF and SNP of *Forsythia suspensa*

Total of SLAF	Polymorphic marker	Non-polymorphic marker	Total of SNP	Hetloci ratio (%)	Integrity ratio (%)
535 357	262 297	273 060	1 809 741	7.86~13.36	47.45~63.34

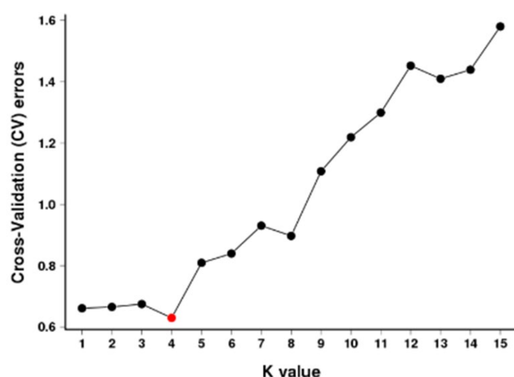

Figure 2 Analysis of genetic structure of *Forsythia suspensa*

Note: Cross validation error rates of K-values

Table 5 Classification of 39 *Forsythia suspensa*

Sample number	ID	Q1 value	Q2 value	Q3 value	Q4 value	Group
1	aa	0.019 847	0.980 133	0.000 01	0.000 01	Q2
2	ab	0.000 01	0.999 969	0.000 011	0.000 01	Q2
3	ac	0.000 01	0.999 97	0.000 01	0.000 01	Q2
4	ad	0.000 01	0.999 97	0.000 01	0.000 01	Q2
5	ae	0.0419 07	0.942 563	0.015 52	0.000 01	Q2
6	af	0.000 01	0.999 97	0.000 01	0.000 01	Q2
7	ag	0.000 01	0.999 97	0.000 01	0.000 01	Q2
8	ah	0.000 01	0.999 97	0.000 01	0.000 01	Q2
9	ai	0.065 498	0.842 459	0.038 871	0.053 172	Q2
10	aj	0.036 879	0.906 49	0.009 961	0.046 671	Q2
11	ak	0.065 087	0.866 224	0.035 919	0.032 77	Q2
12	al	0.045 215	0.859 967	0.030 008	0.064 811	Q2
13	am	0.061 318	0.874 547	0.040 245	0.023 89	Q2
14	an	0.055 101	0.881 333	0.017 363	0.046 202	Q2
15	ao	0.037 186	0.861 714	0.045 689	0.055 411	Q2
16	ap	0.999 97	0.000 01	0.000 01	0.000 01	Q1
17	aq	0.999 97	0.000 01	0.000 01	0.000 01	Q1
18	ar	0.999 97	0.000 01	0.000 01	0.000 01	Q1
19	as	0.040 05	0.886 588	0.042 386	0.030 976	Q2
20	at	0.051 247	0.894 918	0.029 685	0.024 151	Q2
21	au	0.062 863	0.872 266	0.042 678	0.022 194	Q2
22	av	0.023 938	0.876 323	0.071 499	0.028 241	Q2
23	aw	0.000 01	0.923 755	0.076 225	0.000 01	Q2
24	ax	0.000 01	0.937 466	0.062 14	0.000 01	Q2
25	ay	0.000 01	0.000 01	0.999 97	0.000 01	Q3
26	az	0.000 01	0.000 01	0.999 97	0.000 01	Q3
27	ba	0.000 01	0.000 01	0.999 97	0.000 01	Q3
28	bb	0.000 01	0.999 97	0.000 01	0.000 01	Q2
29	bc	0.000 01	0.999 97	0.000 01	0.000 01	Q2
30	bd	0.027 517	0.849 671	0.069 336	0.053 477	Q2
31	be	0.000 01	0.913 057	0.086 923	0.000 01	Q2
32	bf	0.000 01	0.928 986	0.070 994	0.000 01	Q2
33	bg	0.000 01	0.919 211	0.080 769	0.000 01	Q2
34	bh	0.000 01	0.908 652	0.049 332	0.042 006	Q2
35	bi	0.024 581	0.881 293	0.055 251	0.038 875	Q2
36	bj	0.003 581	0.913 624	0.034 516	0.048 279	Q2
37	bk	0.000 01	0.000 018	0.000 01	0.999 962	Q4
38	bl	0.000 02	0.000 01	0.000 01	0.999 96	Q4
39	bm	0.000 01	0.000 01	0.000 01	0.999 97	Q4

Note: Q1: The possibility that the sample came from the first primitive ancestor; Q2: The possibility that the sample came from the second primitive ancestor; Q3: The possibility that the sample came from the third primitive ancestor; Q4: The possibility that the sample came from the four primitive ancestor

2 Discussion

Forsythia suspensa has rich germplasm resources and long cultivation history in China. It has high medicinal development value and wide application prospect. With the increasing market demand of *Forsythia suspensa*, the phenomenon of preemption in *Forsythia suspensa* habitats is serious, *Forsythia suspensa* is picked before maturity, resulting in the quality of *Forsythia suspensa* medicinal materials failing to meet the requirements of Pharmacopoeia. In addition, *Forsythia suspensa* resources are mainly wild. Due to different climate, soil and other ecological environment factors, the yield and quality of *Forsythia suspensa* are uneven. Therefore, the collection and evaluation of *Forsythia suspensa* germplasm resources and the cultivation of excellent germplasm have become the key problems to be solved for the sustainable development of *Forsythia suspensa*.

In recent years, molecular markers such as RAPD, SRAP and SSRs have been gradually applied to the evaluation and genetic analysis of *Forsythia suspensa* germplasm resources. Wu et al. (2016) used RAPD technology to study the genetic diversity of 14 *Forsythia suspensa*. The results showed that the genetic diversity of *Forsythia suspensa* was rich, and there was a significant correlation with the habitats of *Forsythia suspensa*. Sequence-related amplified polymorphism (SRAP) is a new molecular marker developed by Li et al. (2010) from the University of California from *Brassica* crops. He (2013) used SRAP molecular marker technology to analyze the genetic diversity of 26 *Forsythia suspensa* germplasm resources, from which 12 pairs of primer combinations were selected and 153 polymorphic bands were obtained. Shen et al. (2019) used EST-SST molecular marker technology to study the genetic diversity of 77 *Forsythia suspensa* germplasm resources from 12 habitats in Henan and Shanxi. The results showed that the genetic similarity of *Forsythia suspensa* was related to growth environment and altitude, but not to geographical distance.

With the development of the third generation DNA sequencing technology, single nucleotide polymorphism (SNP) molecular marker technology has been applied to germplasm resources evolution and genetic relationship identification, animal and plant genetic linkage map construction, QTL mapping and genetic diversity analysis for its advantages of high genetic stability, rich distribution, high polymorphism and easy detection. SLAF-seq is a high-throughput sequencing technology, which can develop SNP molecular markers with high density, good stability and low cost. The technology has been widely used in genetic diversity analysis of plants such as *Theobroma cacao*, *Hevea brasiliensis*, *Citrus reticulata* and *Pinales* (Zhou et al., 2018). Zhou et al. (2017) used SLAF-seq technology to analyze the genetic evolution of 300 *Camellia oleifera* germplasm resources, a total of 238 771 polymorphic SLAF tags were obtained, and 1 197 282 SNP markers were developed. Zhao et al. (2016) used SNP molecular markers to analyze Exocarpium Citri Grandis, and screened 21 SNP loci with good polymorphism, which provided effective SNP molecular markers for identification of Exocarpium Citri Grandis. Sun et al. (2013) used SNP and EST-SSR molecular markers to identify the genetic relationship of 363 litchi germplasm (*Litchi chinensis* Sonn.). The results showed that SNP molecular markers could be used to identify the parental origin of new litchi germplasm.

In this study, genetic diversity analysis and SNP molecular marker development of 39 *Forsythia suspensa* germplasm resources were carried out by SLAF-seq sequencing technology. The results of sequencing data showed that the reads information of 112.28 Mb was obtained in this study. Based on the obtained reads data, 535 357 SLAF tags were developed, including 262 297 polymorphic SLAFs, and a total of 1 809 741 SNP markers were obtained. These SNP molecular markers were used to analyze the genetic structure of *Forsythia* suspensa, and 39 *Forsythia suspensa* germplasm resources were divided into four groups, providing new information for the classification of genetic structure of *Forsythia suspensa*. The development of SNP molecular markers can provide data support for the investigation of agronomic traits, population identification and genetic analysis of *Forsythia suspensa*.

3 Materials and Methods

3.1 Test materials

39 *Forsythia suspensa* (Thunb.) Vahl germplasm resources were collected from different counties and cities in Shaanxi Province, Shanxi Province, Henan Province and Hebei Province (Table 6). The healthy and mature

leaves of the above materials were collected and frozen in liquid nitrogen at the same time.

Table 6 Sources of 39 *Forsythia suspensa* accessions

Sample number	ID	Origin	Sample number	ID	Origin	Sample number	ID	Origin
1	aa	Jingxing City, Hebei Province	14	an	Pingshun City, Shanxi Province	27	ba	Luoyang City, Henan Province
2	ab	Jingxing City, Hebei Province	15	ao	Pingshun City, Shanxi Province	28	bb	Luanchuan City, Henan Province
3	ac	Jingxing City, Hebei Province	16	ap	Zhangzi City, Shanxi Province	29	bc	Luanchuan City, Henan Province
4	ad	Quzhou City Hebei Province	17	aq	Zhangzi City Shanxi Province	30	bd	Luanchuan City, Henan Province
5	ae	Quzhou City, Hebei Province	18	ar	Zhangzi City, Shanxi Province	31	be	Shangluo City, Shaanxi Province
6	af	Quzhou City, Hebei Province	19	as	Anze City, Shanxi Province	32	bf	Shangluo City, Shaanxi Province
7	ag	She City, Hebei Province	20	at	Anze City Shanxi Province	33	bg	Shangluo City, Shaanxi Province
8	ah	She City, Hebei Province	21	au	Anze City, Shanxi Province	34	bh	Tongchuan City, Shaanxi Province
9	ai	She City, Hebei Province	22	av	Zhengzhou City, Henan Province	35	bi	Tongchuan City, Shaanxi Province
10	aj	Qinyuan City, Shanxi Province	23	aw	Zhengzhou City, Henan Province	36	bj	Tongchuan City, Shaanxi Province
11	ak	Qinyuan City, Shanxi Province	24	ax	Zhengzhou City, Henan Province	37	bk	Huanglong City, Shaanxi Province
12	al	Qinyuan City, Shanxi Province	25	ay	Luoyang City, Henan Province	38	bl	Huanglong City, Shaanxi Province
13	am	Pingshun City, Shanxi Province	26	az	Luoyang City, Henan Province	39	bm	Huanglong City, Shaanxi Province

3.2 Acquisition of *Forsythia suspensa* DNA

Plant genomic DNA extraction kit (Tiangen Biotech (Beijing) Co., Ltd. DP305) was used to extract DNA from 39 *Forsythia suspensa* germplasm resources. The DNA concentration was detected by NanoDrop2000 spectrophotometer, and the extraction quality of DNA was detected by electrophoresis. Ensure that the extracted genomic DNA of *Forsythia suspensa* meets the requirements of database construction.

3.3 Construction of SLAF library

The *Olea europaea* genome was selected as the reference genome in this study because the genome of *Forsythia suspensa* has not been published. The genome size of *Olea europaea* was 1.14 Gb, and the GC content was 35.40% (https://www.ncbi.nlm.nih.gov/genome/?term=Olea+europaea). Based on the genome size and GC content, the electronic enzyme digestion prediction of *Olea europaea* was carried out. The suitable restriction enzyme was selected to carry out the enzyme digestion experiment on *Forsythia suspensa*. The 3' end of the enzyme section was treated with A, and the Dual-index sequencing joint was connected. PCR amplification and electrophoresis gel cutting were carried out to recover the target fragment (Kozich et al., 2013). The SLAF sequencing of *Forsythia suspensa* was carried out on the Illumina platform.

3.4 Evaluation of sequencing data

In this study, the Dual-index method was used for statistical analysis of the sequencing data of *Forsythia suspensa*. The data information of 39 *Forsythia suspensa* samples was obtained by sequencing, and the sequencing data joints were filtered. The data volume and quality of the sequencing data were analyzed,

including the number of reads, Q30 and GC content. To analyze the quality of SLAF library construction, Nipponbare (rice) was selected as the control in this study, and the same enzyme was used for double enzyme digestion to construct the library. Nipponbare (rice) genome data reference was from http://rapdb.dna.affrc.go.jp.

3.5 SNP loci development

The sequencing data in this study were from the enzyme section of the same endonuclease, and the data with integrity less than 0.2 and depth less than 2 were filtered out. Then, according to the principle of sequence similarity, 39 *Forsythia suspensa* samples were clustered in this study. The data clustered together were defined as the same SLAF tags. If a SLAF tag had differences in sequences between different *Forsythia suspensa* samples, it could be considered as a polymorphic SLAF tag. The development of SNP loci is based on the SLAF tag, using the alignment software BWA (Li and Durbin, 2009) to compare the sequencing data with the reference genome, using GATK (McKenna et al., 2010) and SAM tools (Li et al., 2009a) method to detect SNP. Based on sequence consistency, the detected polymorphism SNP information is considered to be the most reliable SNP data set, and the SNP screening standard is secondary genotype frequency of (MAF)>0.05, integrity>0.8.

3.6 Analysis of genetic structure

In this study, admixture software (Alexander et al., 2009) was used to carry out the analysis of genetic population structure for the developed SNP molecular markers of *Forsythia suspensa*. Cluster analysis was carried out according to the cluster number of *Forsythia suspensa* samples (assuming that K value was 1~15), and the clustering results were verified by cross validation. According to the verification results, the valley value of cross validation error rate was determined as the optimal cluster number of *Forsythia suspensa* samples.

Authors' contributions

JT was the experimental designers and executor of this study. JT and WCX completed the data analysis and the first draft of the paper. TW, XXL, LRK and WSQ participated in some parts of the experiments. LLD was the designer and director of the project, guiding experimental design, data analysis, manuscript writing and revision. All authors read and approved the final manuscript.

Acknowledgments

This study was supported by the National Program on Key Research and Development Project of China (2017YFC1701702; 2017YFC1700702), Innovation Team Project of Traditional Chinese Medicine in the Technology system of Modern Agricultural Industry in Hebei Province (HBCT2018060201), Innovative Engineering Project of Hebei Academy of Agricultural and Forestry Sciences (2019-2-3) and Science and Technology Support Project in Hebei Province (19226345D).

Reference

Alexander D.H., Novembre J., and Lange K., 2009, Fast model-based estimation of ancestry in unrelated individuals, Genome Res., 19(9): 1655-1664
https://doi.org/10.1101/gr.094052.109
PMid:19648217 PMCid:PMC2752134

Davey J.W., Cezard T., Fuentes-Utrilla P., Eland C., Gharbi K., and Blaxter M.L., 2013, Special features of RAD sequencing data: implications for genotyping, Mol. Ecol., 22(11): 3151-3164
https://doi.org/10.1111/mec.12084
PMid:23110438 PMCid:PMC3712469

He M., 2013, Hight-throughput sequencing of *Hypericum perforatum* L. and research of germplasm resources in Qingling region, Dissertation for Ph.D., Shaanxi Normal University, Supervisor: Wang Z.Z., pp.10-12

Jiang N., Wang L.P., Zhang H.F., and Wang L., 2015, Evaluation of authentic Chinese herbal medicines Handan Forsythia, Yatai Chuantong Yiyao (Asia-Pacific Traditional Medicine), 11(1): 54-56

Kozich J.J., Westcott S.L., Baxter N.T., Highlander S.K., and Schloss P.D., 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(17): 5112-5120
https://doi.org/10.1128/AEM.01043-13
PMid:23793624 PMCid:PMC3753973

Li H., and Durbin R., 2009, Fast and accurate short read alignment with Burrows–Wheeler transform, Bioinformatics, 25(14): 1754-1760
https://doi.org/10.1093/bioinformatics/btp324

Li H., Handsaker B., Wysoker A., Fennell T., and Ruan J., 2009a, The sequence alignment/map format and SAMtool, Bioinformatics, 25(16): 2078-2079
https://doi.org/10.1093/bioinformatics/btp352
PMid:19505943 PMCid:PMC2723002

Li H., Ruan C.J., Teixeirada S.J., and Liu B.Q., 2010, Associations of SRAP markers with dried-shrink disease resistance in a germplasm collection of sea buckthorn (*Hippophae* L.), Genome, 53(6): 447-457
https://doi.org/10.1139/G10-020
PMid:20555434

Li L., Dong C.M., Zhu Y.H., and Xia W., 2019, Genetic diversity of *Forsythia suspensa* based on ISSR, Zhongcaoyao (Chinese Traditional and Herbal Drugs), 50(15): 3673-3680

Li R., Yu C., Li Y.R., Lam T.W., Yiu S.M., Kristiansen K., and Wang J., 2009b, SOAP2: an improved ultrafast tool for short read alignment, Bioinformatics, 25(15): 1966- 1967
https://doi.org/10.1093/bioinformatics/btp336
PMid:19497933

McKenna A., Hanna M., Banks E., Sivachenko A., Cibulskis K., Kernytsky A., Garimella K., Altshuler D., Gabriel S., Daly M., and DePristo M., 2010, The genome analysis toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data, Genome Res., 20(9): 1297-1303
https://doi.org/10.1101/gr.107524.110
PMid:20644199 PMCid:PMC2928508

Meng L.Z., 2014, The development of forsythia SSR molecular markers and transcriptome analysis based on high-throughput sequencing technology, Thesis for M.S., Shanxi Agricultural University, Supervisor: Wang X.C., pp.14-20

Shen P.L., Dong C.M., Xia W., and Li M., 2019, Genetic diversity of EST-SSR markers in *Forsythia suspensa* (Oleaceae) from different habitats, Beifang Yuanyi (Northern Horticulture), 43(13): 140-147

Sun Q.M., Li Y.Z., Xiang X., Chen D.M., Yang X.Y., Fang J., Wu X.B., Zhou D.H., Ma S.P., and Ma W.Z., 2013, A novel litchi germplasm (*Litchi chinensis* Sonn.), Yujinqiu, identified by EST-SSR and SNP analysis, Fenzi Zhiwu Yuzhong (Molecular Plant Breeding), 11(3): 403-414

Wang X.C., Tan H.L., Chen Z., Meng L.Z., Wang W.B., and Fan S.C., 2015, Assembly and characterization of the transcriptome and development of SSR markers in *Forsythia suspensa* based on RNA-Seq technology, Zhongguo Kexue: Shengming Kexue (Scientia Sinica Vitae), 45(3): 301-310
https://doi.org/10.1360/N052014-00273

Wang X.L., Yang Y.Y., Ye S.H., Feng X.Z., and Gao X.Y., 2016, Determination of core nutrients and functional components of organic *Forsythia suspensa* leaves tea, Nongchanpin Jiagong (Academic Periodical of Farm Products Processing), (7): 55-58

Wu T., Wei S., Mi L.H., Li M., and Zhang S.R., 2016, DNA fingerprint library construction and cluster analysis of *Forsythia suspensa* from different origins, Zhongcaoyao (Chinese Traditional and Herbal Drugs), 47(5): 816-820

Xie B.X., Li Z.M., and Chen S.C., 2010, Determination of phillyrin and ephedrine hydrochloride in Lianhuaqingwen capsules by HPLC, Zhongguo Yiyao Gongye Zazhi (Chinese Journal of Pharmaceuticals), 41(9): 690-691

Zhao J.S., Yang X.Y., Ceng X.Y., Zhong S., Fang J., Luo J.B., Ceng Y.Y., and Xiang X., 2016, Analysis on germplasm resources of Exocarpium Citri Grandis using SNP molecular markers, Fenzi Zhiwu Yuzhong (Molecular Plant Breeding), 14(5): 1203-1211

Zhou L., Duan Y., Wen B., Ma Y.C., Zhu X.J., Wang Y.H., and Fang W.P., 2018, Research and application of SNP molecular markers in ligneous plants genetics and breeding, Yaredai Zhiwu Kexue (Subtropical Plant Science), 47(2): 187-193

Zhou Q.H., Zhou C., Zheng W., Mason A.S., Fan S.Y., Wu C.J., Fu D.H., and Huang Y.J., 2017, Genome-wide SNP markers based on SLAF-Seq uncover breeding traces in rapeseed (*Brassica napus* L.), Front. Plant Sci., 8: 648
https://doi.org/10.3389/fpls.2017.00648
PMid:28503182 PMCid:PMC5409215