Research Article　　　　　　　　　　　　　　　　　　　　　　　　　　　　　　　**Open Access**

The Development of EST-SSR Markers from Transcriptome Sequencing and Genetic Diversity of Twenty Genotypes with High Yields of *Cornus wilsoniana*, an Important Wood Oil Plant

Zhenxiang He [1, 2], Minghuai Wang [3], Xiuyu Xu [3], Yunfei You [4], Yong Chen [5], Liangbo Zhang [6] ✉

1 Life Sciences College, Nanjing University, Nanjing, 210023, P.R. China
2 Nanjing University (Suzhou) High-tech Institute, Suzhou, 215123, P.R. China
3 Guangdong Academy of Forestry, Guangzhou, 510520, P.R. China
4 Yangkou National Forest Farm, Shunchang, 353211, P.R. China
5 Xishan Forest Farm, Chongqing, 404100, P.R. China
6 Hunan Academy of Forestry, Changsha, 410004, P.R. China
✉ Corresponding author Email: 40952257@qq.com

Abstract Oil plants are not only used as essential nutriments in food but also used broadly as ingredients of industrial products. *Cornus wilsoniana* Wanger is an important native woody oil plant with high-yield and high-oiliness characters in China. The fruit oil of *Cornus wilsoniana* not only could be used as equilibrated dietary oil, but also has hypolipidemic function and helps to overcome EFA deficiency. Genomic information is currently not available for *Cornus wilsoniana*, which will therefore affect its genetic improvement process. In this study, 8713 EST-SSRs were identified from the transcriptome sequencing of *Cornus wilsoniana*. Most of these EST-SSRs were composed of dimer and trimer repeats. The AG/CT motif is the most common dimeric EST-SSRs motif (53.5%), whereas the CG/CG (0.44%) microsatellites are present only at low abundances. Among the trimeric microsatellites, AAG/CTT (5.53%) and ATC/ATG (4.27%) are the most common motifs. There are no obvious dominant motifs among the tetra-, penta-, and hexa- nucleotide motifs. Fifteen pairs of EST-SSRs primers are developed with polymorphism and used for the genetic diversity analysis of twenty genotypes of *Cornus wilsoniana* with high fruit output. Allele number per locus ranges from 2 to 6 with an average of 2.87, and the PIC value ranges from 0.05 to 0.58 with an average of 0.36. Average genetic diversity overall SSR loci for the 20 genotypes was 0.427, ranging from 0.049 to 0.659. All the loci are polymorphic and clearly distinguish the genotypes. Cluster analyze (NJ tree, UPGMA) identifies a similar pattern of variation. 15developed EST-SSRs are informative, codominant and reliable, and could be applied in the improvement programs of *Cornus wilsoniana* in the future.

Keywords Woody oil plant; Genetic diversity; Polymorphism; *Cornus wilsoniana*

Background

Oil plants are widely spread in nature, which is not only used as essential nutriments in food, but also used broadly as ingredients of industrial products. There are approximately 300 million square kilometers of forest land throughout China, where oil plants cover 1 554 species belonging to 697 genera and 151 families (He et al., 2013a; He et al., 2013b). *Cornus wilsoniana* Wanger, a member of *Cornaceae*, is one of the important native woody oil plants with high-yield and high-oiliness characters in China. It distributes in the forest land with an elevation ranging from 100~1000 m, an average temperature of 18°C~25°C, and an annual precipitation of 1 000~1 570 mm. It can maintain maximum productivity for over 50 years and its life spans more than 200 years. An average of 50 kg dry fruits per year could be produced on a mature tree (Li et al., 2015). The oil content in dry fruits is 33%~36% and its oil has been taken as edible oil for over 100 years. Consumption of such oil is believed to prevent hyperlipidemia in Chinese folk recipe (Fu et al., 2012). Our previous study demonstrated the oil has a strong hypolipidemic effect, which is because of its very high content of PUFA (38.86%), especially linoleic acid and γ-linolenic acid. The activities of the two n-6 PUFA were possibly related to their high affinity to PPARδ. In addition, SFA/MUFA/PUFA ratio and EPA content are both good in the oil. Therefore, fruit oil of *Cornus*

wilsoniana could be used as equilibrated dietary oil, not only having a hypolipidemic function, but also helping to overcome EFA deficiency (Fu et al., 2012).

Cornus wilsoniana has been approved as a novel source of woody plant food oil by the Chinese government in 2013. The general office of the State Council of China promulgated "Opinion on accelerating the development of woody oil industry"(issued NO. 2014-68) at the end of 2014. Those have greatly raised the interest of private oil production companies and local government. According to our uncompleted calculations, the annual seed output of wild *Cornus wilsoniana* plants amounts to 600 000 kg in China. Apparently, this can't meet the demands of woody plant oil market. Magnanimous grafted seedlings with high seed yield are needed for afforestation. Unlike other oil crops such as soybean, maize, rapeseed and sunflower, there are less genetically improved varieties of *Cornus wilsoniana* available (Huang et al., 2016). Hence, selection and deployment of genetically improved materials are put on the agenda. Plus tree investigations on a large scale and selection of superior individuals have been done in the naturally distribution of *Cornus wilsoniana* based on the phenotype characteristics (Lin et al., 2010; Zhu et al., 2012; Zhang et al., 2016).

Genetic improvement of forest tree species is usually based on recurrent selection for general combining ability. That is, the frequency of desirable genes in the population is increased progressively through cycles of selection and crossing (Shelbourne et al., 1989). Because this process involves sexual reproduction for progeny testing, seed has been the traditional output from genetic improvement programs. However, traditional genetic improvement of forest tree species is a time-, resource- and cost-intensive activity, which greatly restricts the genetic improvement process. Recently, El-Kassaby introduced a breeding strategy called "breeding without breeding" (BWB), which has been proven highly convenient for tree breeding. The efficiency of this strategy has been evaluated by progeny testing, parental selection, and construction of pedigrees. The BWB strategy was demonstrated in a number of tree species and it has been extensively used in different areas of forest tree breeding, including phylogenetic analysis, mating systems, estimation of genetic parameters and breeding value, and spatial variation (Yuan et al., 2016). Selecting superior individuals directly from the open-pollinated progeny implantations, coupled with identifying the phylogenetic relationship and genetic diversity of the selected materials based on molecular markers, could shorten the breeding cycle.

Microsatellites, also called simple sequence repeats (SSR), are usually short, 2 to 8 nucleotides in length, which widely exists in the genome of eukaryotes and prokaryotes. SSR markers are popular in population genetics, forensic studies, and paternity analyses because they are informative (multi-allelic, meaning that they have more than two forms of a gene), codominant (heterozygous individuals can be distinguished from homozygous individuals), and reliable (Smith et al., 1997; Jones and Ardren, 2003). The development of SSR marker primer is the prerequisite for the study of SSR markers. However, there isn't any genomic information available on the species of *Cornus wilsoniana*. In our previous study, 11 SSRs primers from *Cornus controversa* (Yoshihiro et al., 2010), which is the same genus, *Cornaceae*, with *Cornus wilsoniana* were used to amplify DNA samples from *Cornus wilsoniana*. Results showed that all 11 SSRs developed for *Cornus controversa* couldn't produce any polymorphic bands. To help bridge this knowledge gap, our objective was to develop a set of EST-SSR markers based on transcriptome sequencing of *Cornus wilsoniana* and to examine their polymorphisms. Then, it will allow us to make preparations for subsequent genetic diversity analysis and molecular phylogeny of selected twenty genotypes from naturally open-pollinated populations in relation to fruit output in this paper.

1 Materials and Methods
1.1 Plant materials
6 samples for high-throughput transcriptome sequencing were collected from fruits of two genotypes during three developmental stages in the germplasm bank of *Cornus wilsoniana* at Hunan Academy of Forestry. 20 samples of SSR amplification were collected from fresh leaves of grafted seedlings in the germplasm banks of *Cornus wilsoniana* at Longshan forest farm, Lechang city, Guangdong province. Xishan forest farm, Chongqing city and Yangkou national forest farm, Shunchang county, Fujian province.

1.2 DNA extraction
Genomic DNA of above samples was extracted using QIAamp DNA Mini Kit.

1.3 High-throughput transcriptome sequencing
6 samples were delivered to Shanghai GR Biotechnology Co. Ltd for transcriptome sequencing using RNA-Seq technique.

1.4 Identification of EST-SSRs
EST-SSRs were searched using the MISA (http://pgrc.ipk-gatersleben.de/misa/) with unit size of 3~5 and repeats >5 from the previous assembly of 271 Mb reads derived from the transcriptome sequences of *Cornus wilsoniana*.

1.5 Primer design and screening
Primers were designed to flank the EST-SSRs using software of Premier 3.0 (Premier Biosoft International, Palo Alto, CA). 84 pairs of primers met the demands of the expected product size at 100~300 bp and Tm value (minimum 55°C, optimum 57°C). Primer screening included two steps, unicity screening and polymorphism screening.

1.6 PCR amplification
The PCR reactions were carried out in a total volume of 20 μL, which contained 1 μL template DNA (approximately 30 ng), 0.2 μL of each primer, 10 μL Mix (Taq polymerase, dNTPs and PCR buffer), and 8.8 μL sterile distilled water. The PCR amplification was performed on a PCR 310A (Hangzhou, China) with a program of 94°C for 5 min; 35 cycles of 94°C for 30 s, 57°C for 30 s, 72°C for 20 s, and a final extension for 10 min at 72°C. The PCR products were separated and detected by 3730xl DNA Analyzer (ABI, U.S.A.).

1.7 Data analysis
Bands data of EST-SSR amplification were transferred from size to binary 0/1 and scored as present (1) or absent (0). The diversity level of the gene loci was evaluated with the polymorphic information content (PIC, Botstein et al., 1980). The PIC provides an estimate of the discriminatory power of a locus by taking into account the number and the relative frequencies of the alleles. PIC values vary from 0 (monomorphic) to 1 (highly discriminative, with many of the alleles present in equal frequencies). PIC, allele numbers and gene diversity of each EST-SSR were calculated using MolKin v3.0 (http://www.ucm.es/info/prodanim/html/JP_Web.htm).

Genetic distances of 20 genotypes of *Cornus wilsoniana* based on EST-SSRs were determined using POWER MARKER V3.25 (Liu and Muse, 2005). A dendrogram of UPGMA (un-weighted pair-group method with arithmetic means) cluster analysis was generated based on the neighbor-joining method as implemented in POWER MARKERV3.25with the tree viewed using FigTree v1.4.3.

Population assignment analysis of the 20 genotypes was completed using GenAlEx 6.503 (http://biology.anu.edu.au/GenAlEx/).

2 Results
2.1 Characterization of the EST-SSR motifs distribution in the transcriptome of *Cornus wilsoniana*
In total, 126,173 transcripts and 78,496 singletons were obtained from a previous assembly of 217 Mb reads derived from the *Cornus wilsoniana* transcriptome sequences of the fruits.174, 326 sequences from the transcriptome sequencing were examined for EST-SSR development. The total size of examined sequences was 49,770,105 bp. A total number of identified EST-SSRs was 8713. The number of sequences containing more than 1 EST-SSR was 597. An average distance between EST-SSRs of approximately was 15.95 kb. The proportion of the EST-SSR unit size was not evenly distributed. Most of these EST-SSRs were composed of dimer and trimer repeats. The dinucleotides were the dominant motifs. Specifically, the abundance of di-, tri-, tetra-, penta-hexanucleotide motifs among these EST-SSRs were 6,635 (76.15%), 1,970 (22.61%), 78 (0.90%), 15 (0.17%), 15 (0.17%), respectively (Table 1).

Table 1 Characterization of 8,713 identified EST-SSR motifs

Repeat motif	Repeat number									Total	(%)
	5	6	7	8	9	10	11	12	13		
Dimer	-	2714	1550	1069	733	435	123	9	2	6635	76.15
Trimer	1242	481	212	30	2	3	0	0	0	1970	22.61
Tetramer	67	9	2	0	0	0	0	0	0	78	0.90
Pentamer	14	0	1	0	0	0	0	0	0	15	0.17
Hexamer	8	6	0	0	1	0	0	0	0	15	0.17
Total										8713	100

Considering sequence complementary, frequencies of classified repeat types of identified EST-SSR motifs were summarized in Table 2. The AG/CT motif was the most common dimeric EST-SSRs motif (53.5%), whereas the CG/CG (0.44%) microsatellites were present only at low abundances. Among the trimeric microsatellites, AAG/CTT (5.53%) and ATC/ATG (4.27%) were the most common motifs. There were no obvious dominant motifs among the tetra-, penta-, hexa- nucleotide motifs.

2.2 EST-SSR primer screening and validation

A total of 48 pairs of common primers and 1 fluorescent primer with tail were synthesized for unicity checking using 4 samples (#4, #5, #6, #7). 31 pairs of primers produced a single band or clear target bands. 15 pairs of primers were selected based on the EST-SSR amplification results of 8 samples (#4，#5，#6，#7, #16, #17, #18，#19) in polymorphism screening (Table 3).

2.3 Overall EST-SSR diversity

20 genotypes of *Cornus wilsoniana* from two provinces were evaluated using 15EST-SSR markers. A total of 43 alleles were detected at 15 loci (Table 3). A wide range of allelic variants were observed for each locus. The number of alleles per locus ranged from 2 alleles (SW24, SW33, and SW47) to 6 alleles (SW09), with an average of 2.87 alleles across the 15 loci. The parameter PIC refers to the value of a marker for detecting polymorphism within a population, depending on the number of detectable alleles and the distribution of their frequency, and has been proved to be a general measure of how informative a marker is (Guo and Elston, 1999); the higher the PIC value is, the more informative a marker is. The PIC values of the 15 primers exhibiting polymorphisms among the 20 individuals ranged from a minimum of 0.05 (SW47) to a maximum of 0.58 (SW48), with an average of 0.36,suggesting that the EST-SSR markers developed had a moderate level of polymorphisms .According to Nei's (Nei, 1973),the highest level of genetic diversity value (0.659) was observed with locus SW48, the lowest level of genetic diversity value (0.049) was observed with locus SW47, and a mean diversity of 0.427 was observed (Table 3).

2.4 Genetic relationship between20 genotypes of *Cornus wilsoniana* according to EST-SSR profiles

Genetic distance reflects the genetic relationships among materials. Simple sequence repeats (SSRs) are high-resolution markers that can identify different individuals within the same species. The combination of phenotypic selection, genetic distance-based phylogenetic analysis of selected individuals using SSR markers, and phylogenetic relationship-based field deployment would simplify breeding activities, decreasing inbreeding. A phylogenetic dendrogram of 20 genotypes of *Cornu swilsoniana* from two provinces with high fruit output was constructed based on their similarity matrix derived from the 15 SSR loci profiles. 20 individuals were mainly divided into four cluster groups (Figure 1). Group A included three genotypes (#1, #4, and #15), Group B included four genotypes (#2, #5, #6, and #7), Group C included six genotypes (#9, #10, #11, #12, #13, and #18), and Group D included seven genotypes (#3, #8, #14, #16, #17, #19, and #20). In the four cluster groups above, it was interesting to find that individuals came from the same collected site were tied to one clade as #1 and #15 in Group A, #2 and #7, #5 and #6 in Group B, #9 and #13, #10 and #11 in Group C, #16, #17, #19 and #20 in Group D, respectively. In this study, the SSRs clearly illustrated the well-documented similarities reflected by specific alleles at all the loci studied. The genetic distance-based results observed in the unrooted neighbor-joining tree revealed four major groups and agreed with genetic similarity analysis using UPGMA (Figure 2).

Table 2 Frequency of classified repeat types of identified EST-SSR motifs considering sequence complementary

Repeat motif	Repeat number									Total
	5	6	7	8	9	10	11	12	13	
AC/GT	-	245	147	103	83	66	37	4	1	686
AG/CT	-	1925	1074	701	532	341	84	4	1	4662
AT/AT	-	519	326	256	117	28	2	1	-	1249
CG/CG	-	25	3	9	1	-	-	-	-	38
AAC/GTT	95	36	14	6	-	-	-	-	-	151
AAG/CTT	293	109	70	10	-	-	-	-	-	482
AAT/ATT	83	40	21	2	-	1	-	-	-	147
ACC/GGT	200	78	12	2	-	-	-	-	-	292
ACG/CGT	58	15	5	-	-	-	-	-	-	78
ACT/AGT	42	13	10	-	-	1	-	-	-	66
AGC/CTG	116	59	19	2	-	-	-	-	-	196
AGG/CCT	113	49	10	3	1	-	-	-	-	176
ATC/ATG	196	76	50	3	1	1	-	-	-	327
CCG/CGG	46	6	1	2	-	-	-	-	-	55
AAAC/GTTT	1	-	-	-	-	-	-	-	-	1
AAAG/CTTT	4	-	-	-	-	-	-	-	-	4
AAAT/ATTT	17	1	-	-	-	-	-	-	-	18
AACC/GGTT	2	1	1	-	-	-	-	-	-	4
AAGC/CTTG	3	1	-	-	-	-	-	-	-	4
AAGG/CCTT	3	-	-	-	-	-	-	-	-	3
AATC/ATTG	1	1	-	-	-	-	-	-	-	2
AATG/ATTC	3	-	-	-	-	-	-	-	-	3
ACAG/CTGT	4	-	-	-	-	-	-	-	-	4
ACAT/ATGT	11	3	-	-	-	-	-	-	-	14
ACTC/AGTG	1	-	1	-	-	-	-	-	-	2
ACTG/AGTC	3	-	-	-	-	-	-	-	-	3
AGAT/ATCT	8	1	-	-	-	-	-	-	-	9
AGCC/CTGG	1	1	-	-	-	-	-	-	-	2
ATCC/ATGG	4	-	-	-	-	-	-	-	-	4
ATGC/ATGC	1	-	-	-	-	-	-	-	-	1
AAAAG/CTTTT	1	-	-	-	-	-	-	-	-	1
AAAGC/CTTTG	1	-	-	-	-	-	-	-	-	1
AAAGG/CCTTT	1	-	-	-	-	-	-	-	-	1
AAATC/ATTTG	1	-	-	-	-	-	-	-	-	1
AAATG/ATTTC	1	-	-	-	-	-	-	-	-	1
AACAC/GTGTT	-	-	1	-	-	-	-	-	-	1
AAGAG/CTCTT	1	-	-	-	-	-	-	-	-	1
AAGCT/AGCTT	1	-	-	-	-	-	-	-	-	1
AAGGC/CCTTG	1	-	-	-	-	-	-	-	-	1
AATAC/ATTGT	1	-	-	-	-	-	-	-	-	1
AATAG/ATTCT	1	-	-	-	-	-	-	-	-	1
AATCC/ATTGG	1	-	-	-	-	-	-	-	-	1
AGATC/ATCTG	1	-	-	-	-	-	-	-	-	1
ATATC/ATATG	1	-	-	-	-	-	-	-	-	1
ATGCC/ATGGC	1	-	-	-	-	-	-	-	-	1
AAAATC/ATTTTG	-	2	-	-	-	-	-	-	-	2

Continued Table 2

Repeat motif	Repeat number									Total
	5	6	7	8	9	10	11	12	13	
AACACC/GGTGTT	1	-	-	-	-	-	-	-	-	1
AAGATG/ATCTTC	-	1	-	-	-	-	-	-	-	1
AAGGGC/CCCTTG	1	-	-	-	-	-	-	-	-	1
AATACC/ATTGGT	-	1	-	-	-	-	-	-	-	1
AATTCC/AATTGG	1	-	-	-	-	-	-	-	-	1
ACCACT/AGTGGT	-	2	-	-	-	-	-	-	-	2
ACCATC/ATGGTG	2	-	-	-	-	-	-	-	-	2
ACTAGC/AGTGCT	1	-	-	-	-	-	-	-	-	1
ACTATG/AGTCAT	1	-	-	-	-	-	-	-	-	1
AGCCTG/AGGCTC	-	-	-	-	1	-	-	-	-	1
AGCTCC/AGCTGG	1	-	-	-	-	-	-	-	-	1

Table 3 EST-SSR primer sequences and repeat types of *Cornus wilsoniana* with polymorphism

| Name	Primer sequence	SSR motif	Product size (bp)	Allele no.	Gene diversity	PIC
SW05	F:CATCGATTAAGGCAGATAAGAA R:TCAAATTTCAGAGTCTCCACAA	(AAAT)5	113	2	0.495	0.37
SW09	F:GTGGTGGTGATTAGGAAGCTAT R: GTGACACTGGAATGAAGAGAGA	(TTTCC)5	289	6	0.486	0.46
SW10	F:AACGAAATTGAAACCCTAGAAA R:CAGAGACAAAATTCAAAAAGCA	(TGATT)5	112	3	0.301	0.28
SW17	F:ATGGCTATGATCCAAGTTCTTT R:TCAAACTGATCATAAAACAGGC	(TTGTA)5	225	4	0.341	0.32
SW19	F:GGCGGAAGAACTCTATCTTTGT R:TCAGAACAACGTACGTACAACA	(ATGGC)5	132	2	0.489	0.37
SW20	F:GGGGTTTATGTTCTTCAGTAGC R:TACCACCATCAATTTCTAACCC	(TGGT)5	108	3	0.571	0.50
SW23	F:TGTTTCAAGTTAAGGTCGATTG R:CTTCAGGTAGAAGTACCCAAGG	(TGAG)7	211	3	0.546	0.44
SW24	F:TAAACAAAACTGCTGCTGCTAT R:AGTCCAGAATAAAACCACTGCT	(GACT)5	181	2	0.180	0.16
SW28	F:CAGGCTGGTGTAACTCATAAAA R:CACTCTCACTACATTGCCCTAA	(TGGA)5	300	3	0.626	0.55
SW29	F:AGAAGTATTCCTGTCCTCATGC R:AACCCAAGTAAACAGAGGAAAC	(TTTA)5	201	2	0.289	0.25
SW30	F:ATAATGGCTAACCTACCCAAGA R:CGGTTTGTAATAGCACTCCAC	(ATAA)5	133	3	0.501	0.42
SW33	F:CTGGGACTTCACACAAGTAGAA R:AGGTAGAGCATTCTGGTTCCTA	(TCTG)5	151	2	0.455	0.35
SW34	F:GATGATGACGATTGATTGAGG R:GTTCAGAAAGAATACCACCCAT	(AGCTT)5	121	3	0.411	0.34
SW47	F:AATTTGAAAGAATACGTGATTGA R:ATGTCATTTGACTGTGGGTAGA	(CATA)5	244	2	0.049	0.05
SW48	F:ACAAGCACGGGCTTATTTATAC R:GTTTTCACAGTGATCAAACGAC	(TTAT)5	109	3	0.659	0.58
Mean		-	-	2.87	0.427	0.36
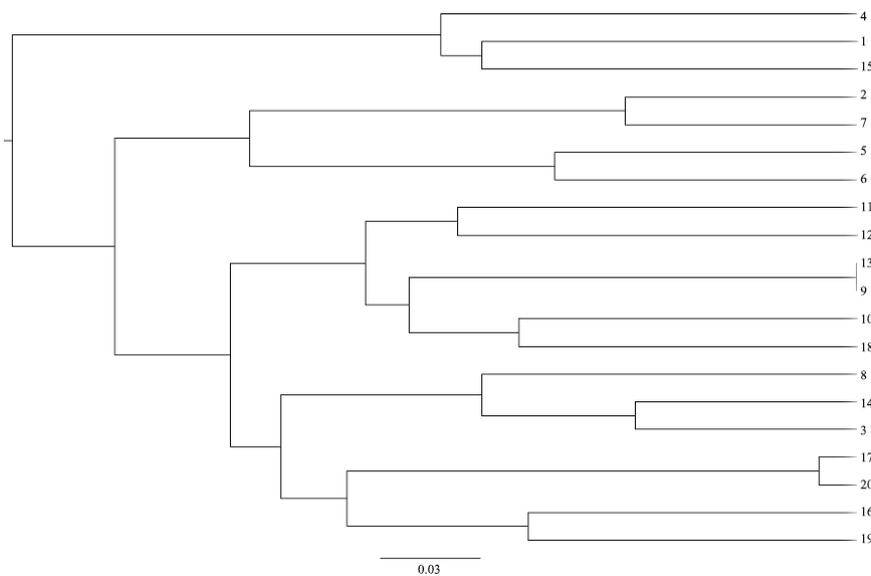
Figure 1 Dendrogram of 20 genotypes of *Cornus wilsoniana* showing the genetic similarity based on 43 alleles detected by 15 SSRs using UPGMA cluster analysis
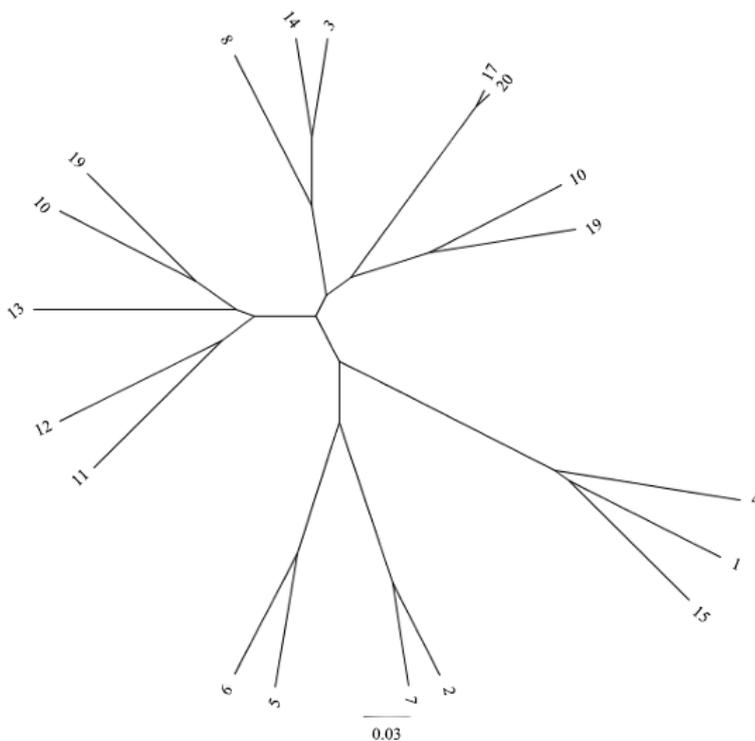

Figure 2 An un-rooted neighbor-joining tree showing the genetic relationship between the 20 genotypes of *Cornus wilsoniana* based on 15 SSRs.

3 Discussions

SSRs have already been used broadly for a variety of purposes such as genetic diversity. SSRs can be found in either the coding or the non-coding regions of the genome. However, EST-SSRs score only the expressed region of the genome. Compared to SSR, EST-SSRs have some advantages, which can be developed at no cost from the transcriptome sequencing data and may be used across a number of related species. More importantly, EST-SSRs could illustrate marker-trait associations. In the present study, 8713 EST-SSRs were identified from the transcriptome sequencing of *Cornus wilsoniana*. 15 pairs of EST-SSRs primers were developed with polymorphism and successfully used for genetic diversity analysis of twenty genotypes of *Cornus wilsoniana* in

relation to fruit output. Those EST-SSR markers are informative, codominant and reliable, and could be applied in the improvement programs of *Cornus wilsoniana* in the future. However, SSRs primers from *Cornus controversa*, which is the same genus, *Cornaceae*, were failed in DNA samples from *Cornus wilsoniana*.

In this study, dimer and trimer repeats were found to be most abundant, which is in agreement with reports on other plants. The AG/CT motif was the most common dimeric EST-SSRs motif in *Cornus wilsoniana*.

On the highly efficient of EST-SSRs, 20 genotypes of *Cornus wilsoniana* has been produced clear distinctions using population assignment analysis. The genotypes clustered into two distinct groups according to the source of those individuals (Figure 3). Only #18 from the LC (Lechang, Guangdong province) group was much close to the group of JX (Jiangxi province). This showed #18 possibly originated from the JX. According to the historical record, farmers in Lechang city planted seedlings of *Cornus wilsoniana* from the JX province.

The values of pair-wise comparisons of Nei's genetic distance (D) between the genotypes analyzed were computed from the combined data for 15 primers. The pairwise distance ranged from 0.0000 to 0.9259 (Table 4). A comparatively higher genetic distance (0.9259) was observed between #7 and #15. This indicated that genetically they are diverse compared to those having a lower genetic distance value. Basically, this value is an indication of their genetic dissimilarity since #7 was collected at Yudu county in JX (N26°12′56.86″, E115°37′6.29″) and #15 was collected at LC (N25°06′51.44″, E113°19′24.93″). On the other hand, the lowest genetic distance (0.000) was found between #9 and #13 indicating that they are much closer to their genetic make-up. Both of them were collected from the same location with a close geographic distance, and their genetic distance revealed that there is a close genetic relationship between them.

The results of the present study may be finally tuned through utilization of more informative markers and more diverse set of genotypes. In cluster analysis, it is possible to both assess genetic diversity of elite genotypes and select genotypes with higher genetic diversity using EST-SSRs. This finding clearly demonstrated the reliability, usefulness, and efficiency of EST-SSRs in analyzing genomic diversity, genotype identification, plus tree screenings as parental breeding material and plant variety protection in the improvement programs of *Cornus wilsoniana*.
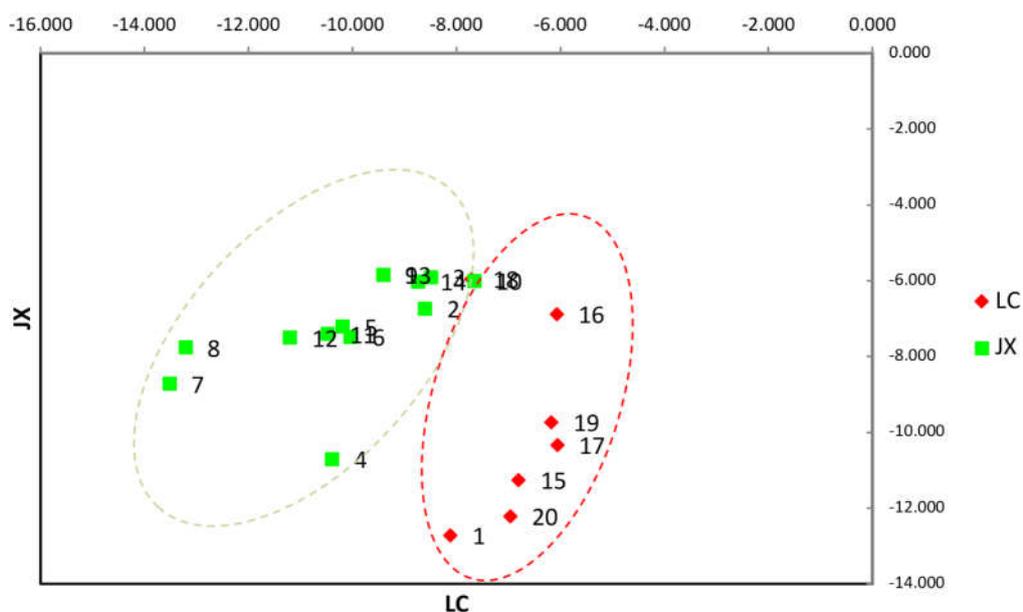

Figure 3 Two-dimensional plot of population assignment analysis of 20genotypes from two provinces (LC and JX) based on the 15 EST-SSR markers

Table 4 Genetic distances of 20 genotypes of *Cornus wilsoniana* based on 15 EST-SSRs

OTU	01	02	03	04	05	06	07	08	09	10	11	12	13	14	15	16	17	18	19	20
01	*	0.6561	0.6154	0.2428	0.4147	0.5439	0.7536	0.8454	0.5543	0.3677	0.5044	0.4772	0.5543	0.5790	0.2264	0.4976	0.5719	0.5226	0.4053	0.5719
02	0.6561	*	0.1871	0.4722	0.3107	0.3648	0.1400	0.4305	0.3878	0.3774	0.7113	0.7702	0.3878	0.3929	0.6312	0.2310	0.3494	0.3118	0.4133	0.4083
03	0.6154	0.1871	*	0.5243	0.4700	0.4372	0.3653	0.1864	0.3107	0.4147	0.4644	0.3159	0.3107	0.1335	0.3236	0.2027	0.2168	0.3648	0.4259	0.2723
04	0.2428	0.4722	0.5243	*	0.3653	0.3363	0.6539	0.7793	0.6014	0.4351	0.5493	0.6335	0.6014	0.5594	0.2594	0.5447	0.5159	0.5030	0.3857	0.5523
05	0.4147	0.3107	0.4700	0.3653	*	0.1823	0.3951	0.3842	0.5471	0.2736	0.4961	0.3747	0.5471	0.4700	0.3534	0.3653	0.4604	0.2106	0.3653	0.4955
06	0.5439	0.3648	0.4372	0.3363	0.1823	*	0.3951	0.5541	0.6161	0.4455	0.3747	0.4372	0.6161	0.3747	0.4487	0.3653	0.5319	0.2106	0.2807	0.6088
07	0.7536	0.1400	0.3653	0.6539	0.3951	0.3951	*	0.5381	0.5030	0.4351	0.6547	0.8006	0.5030	0.3653	0.9259	0.4155	0.5159	0.3852	0.5447	0.5900
08	0.8454	0.4305	0.1864	0.7793	0.3842	0.5541	0.5381	*	0.2894	0.4242	0.5076	0.2390	0.2894	0.2664	0.4613	0.3440	0.4052	0.3161	0.6515	0.4391
09	0.5543	0.3878	0.3107	0.6014	0.5471	0.6161	0.5030	0.2894	*	0.2041	0.2695	0.3374	0.0000	0.3107	0.5948	0.2310	0.4083	0.3365	0.3852	0.4390
10	0.3677	0.3774	0.4147	0.4351	0.2736	0.4455	0.4351	0.4242	0.2041	*	0.2324	0.3003	0.2041	0.4455	0.4242	0.3207	0.4337	0.2041	0.3481	0.5004
11	0.5044	0.7113	0.4644	0.5493	0.4961	0.3747	0.6547	0.5076	0.2695	0.2324	*	0.2412	0.2695	0.3192	0.4431	0.3396	0.4526	0.3183	0.2616	0.5193
12	0.4772	0.7702	0.3159	0.6335	0.3747	0.4372	0.8006	0.2390	0.3374	0.3003	0.2412	*	0.3374	0.2336	0.2664	0.4576	0.3620	0.3107	0.4259	0.3937
13	0.5543	0.3878	0.3107	0.6014	0.5471	0.6161	0.5030	0.2894	0.0000	0.2041	0.2695	0.3374	*	0.3107	0.5948	0.2310	0.4083	0.3365	0.3852	0.4390
14	0.5790	0.3929	0.1335	0.5594	0.4700	0.3747	0.3653	0.2664	0.3107	0.4455	0.3192	0.2336	0.3107	*	0.3236	0.3081	0.2442	0.4518	0.3653	0.3013
15	0.2264	0.6312	0.3236	0.2594	0.3534	0.4487	0.9259	0.4613	0.5948	0.4242	0.4431	0.2664	0.5948	0.3236	*	0.4046	0.2511	0.4930	0.3440	0.2801
16	0.4976	0.2310	0.2027	0.5447	0.3653	0.3653	0.4155	0.3440	0.2310	0.3207	0.3396	0.4576	0.2310	0.3081	0.4046	*	0.2928	0.3051	0.1985	0.3516
17	0.5719	0.3494	0.2168	0.5159	0.4604	0.5319	0.5159	0.4052	0.4083	0.4337	0.4526	0.3620	0.4083	0.2442	0.2511	0.2928	*	0.5375	0.2646	0.0230
18	0.5226	0.3118	0.3648	0.5030	0.2106	0.2106	0.3852	0.3161	0.3365	0.2041	0.3183	0.3107	0.3365	0.4518	0.4930	0.3051	0.5375	*	0.3852	0.6089
19	0.4053	0.4133	0.4259	0.3857	0.3653	0.2807	0.5447	0.6515	0.3852	0.3481	0.2616	0.4259	0.3852	0.3653	0.3440	0.1985	0.2646	0.3852	*	0.3217
20	0.5719	0.4083	0.2723	0.5523	0.4955	0.6088	0.5900	0.4391	0.4390	0.5004	0.5193	0.3937	0.4390	0.3013	0.2801	0.3516	0.0230	0.6089	0.3217	*

Authors' contributions

Zhenxiang He carried out the molecular genetic studies and drafted the manuscript. Minghuai Wang, You Yunfei and Chen Yong participated in the plus tree collection and carried out the establishment of the germplasm banks of *Cornus wilsoniana*. Xiuyu Xu performed the statistical analysis. Liangbo Zhang conceived of the study and participated in its design and coordination. All authors read and approved the final manuscript.

References

Botstein D., White R.L., Skolnick M., and Davis R.W., 1980, Construction of a genetic linkage map in man using restriction fragment length polymorphisms, Am J Hum Genet,32:314-331

Fu J., Zhang X.W., Liu K., Li Q.S., Zhang L.R., Yang X.H., Zhang Z.M., Li C.Z., Luo Y., He Z.X., and Zhu H.L., 2012, Hypolipidemic activity in sprague–dawley rats and constituents of a novel natural vegetable oil from *Cornus wilsoniana* Fruits, Journal of Food Science, 77(8):160-169
https://doi.org/10.1111/j.1750-3841.2012.02786.x
PMid:22747885

Guo X., and Elston R.C.,1999, Linkage information content of polymorphic genetic markers, Hum Hered, 49:112-118
https://doi.org/10.1159/000022855
PMid:10077733

He Z.X., Shi S.D., Hu Y.M., Zhang L.B., and Fu J., 2013a, The China Oil Plant Species (COPS) Database: A comprehensive web-based database and informatics system providing ecological and chemical traits of oil plants in China, Vegetos, 26(2):15-22
https://doi.org/10.5958/j.2229-4473.26.2.048

He Z.X., Yin Z.F., Hu Y.M., Zhang L.B., Li C.Z, and Fu J., 2013b, Assessment of biodiversity in China oil plant species-class, content and distribution characteristics of fatty acids, Advances in Food Sciences, 35(3):117-125

Huang J.H, Ye D.Q., Zhang Z.C., He Z.X, and Cheng X.R., 2016, Spatial heterogeneity and distribution patterns of soil nutrients in sloping field of *Cornus wilsoniana* Young Plantation, Journal of Southwest Forestry University, 36(4): 30-35

Jones A.G., and Ardren W.R., 2003, Methods of parentage analysis in natural populations, Mol Ecol.,12: 2511-2523
https://doi.org/10.1046/j.1365-294X.2003.01928.x

Li C.Z., Zhang L.B., Xiao Z.H., Li P.W., Liu R.K., Chen J.Z., He Z.X., Fu J., 2015, Production of biodiesel using a vegetable oil from *Cornus wilsoniana* fruits, Periodica Polytechnica Chemical Engineering, 59(4):283-287
https://doi.org/10.3311/PPch.7868

Lin J., Li P.S., Liang Q., Zhang F.Q., Zhang L.B., and He Z.X., 2010, Ecological investigation of germplasm resources of *Cornus wilsoniana* and its development and utilization in North Guangdong Area, Journal of Guangdong Forestry Science and Technology,26(6):45-48

Liu K., and Muse S.V., 2005, Power Marker: Integrated analysis environment for genetic marker data, Bioinformatics, (21): 2128-2129
https://doi.org/10.1093/bioinformatics/bti282

Nei M.,1973, Analyses of gene diversity in subdivided populations, Proc. Natl. Acad. Sci., USA70: 3321-3323
https://doi.org/10.1073/pnas.70.12.3321
PMid:4519626 PMCid:PMC427228

Shelbourne C.J.A., Carson M.J., and Wilcox M.D., 1989, New techniques in the genetic improvement of radiata pine, Commonwealth Forestry Review, (68): 191-201

Smith J.S.C., Chin E.C.L., Shu H., Smith O.S., Wall S.J., Senior M.L., Mitchell S., Kresovich S., and Ziegle J., 1997, An evaluation of the utility of SSR loci as molecular markers in maize (Zeamays L.): comparisons with data from RFLPs and pedigree, Theor Appl Genet, 95: 163-173
https://doi.org/10.1007/s001220050544

Yoshihiro Y., Shingo K., Shoji N., Takashi M., and Isagi Y.J., 2010 Isolation and characterization of 11 microsatellite loci in *Cornus controversa* (Cornaceae), Conservation Genet Resour, 2:145-147
https://doi.org/10.1007/s12686-010-9276-4

Yuan H.W., Niu S.H., Zhou X.Q., Du Q.P., and Li Y., 2016, Evaluation of seed production in a first-generation seed orchard of Chinese pine (*Pinus tabuliformis*), For Res, 27(5):1003-1008
https://doi.org/10.1007/s11676-016-0238-x

Zhang Z.C., Huang J.H., Ye D.Q., Li W.F., He Z.X., and Cheng X.R, 2016, Nitrogen and phosphorus stoichiometry of different families of *Cornus wilsoniana* and their relationships to nutrient availability, Journal of Northwest Forestry University,31(4):53-58

Zhu C.L, Zhang L.B., Li X.C., Nai X.E., and He Z.X., 2012, Study on the fruit oil content of *Cornus wilsoniana* plus trees by Nuclear Magnetic Resonance (NMR) approach, Journal of Guangdong Forestry Science and Technology, 28(4):21-24