Development of microsatellite markers for the critically endangered conifer *Glyptostrobus pensilis* (Cupressaceae) using transcriptome data

Xin-Yu Li[1], Xue-Ying Lin[1], Markus Ruhsam[2], Lu Chen[1], Xing-Tong Wu[1], Min-qiu Wang[1], Philip I. Thomas[2] and Ya-Feng Wen[1]*

[1] Central South University of Forestry and Technology, Changsha, Hunan, 410004, China.
[2] Royal Botanic Garden Edinburgh, 20A Inverleith Row, Edinburgh EH3 5LR, United Kingdom.

* Corresponding author: Ya-Feng Wen, E-mail: wenyafeng7107@163.com

Abstract

Glyptostrobus pensilis (Cupressaceae) is a critically endangered conifer which occurs in China, Laos, and Vietnam where it is only known from a few populations. Here we aim to develop microsatellite markers which can be used to study the genetic variation within this species. Using transcriptome data we tested 170 SSR loci for polymorphism in 16 samples. Twenty-three loci were polymorphic and selected for the genetic analysis of 83 individuals from three Chinese populations. The number of alleles per locus and population ranged from one to eight, the observed and expected heterozygosity from H_o = 0.00-1.00 and H_e = 0.00-0.83, respectively. Fifteen loci deviated from Hardy-Weinberg equilibrium in at least one of the three populations. The majority of loci could also be successfully amplified in four related species, namely *Cryptomeria fortunei, Taxodium distichum, Taxodium ascendens* and *Cunninghamia lanceolata*. These developed microsatellites are suitable for population genetic studies of *Glyptostrobus pensilis* and related species.

Keywords: *: Glyptostrobus pensilis; next-generation sequencing; Expressed Sequence Tag (EST); simple sequence repeat (SSR) markers*

Introduction

Glyptostrobus pensilis (Staunt.) Koch is a monotypic genus in the conifer family Cupressaceae. In China, its distribution is centered in the Pearl River delta (Guangdong Province), the central region of Fujian Province, and the northeast part of Jiangxi Province (Li and Xia, 2004). A few wild populations have also been found in Vietnam and Laos, extending its latitudinal distribution from 28° N to 13°N (Averyanov et al., 2009; Thomas and LePage, 2011). The fossil record shows that *G. pensilis* became extinct in northeastern China and adjoining areas possibly during the early Pleistocene (LePage, 2007). This species was widely distributed in Guangzhou, China until the late Holeocene (Xu and Li, 1959), however, due to widespread cultivation over several centuries, the true natural distribution remains uncertain (Yu, 1995; Li and Xia, 2005). During the Holocene, and especially over the last two millennia, the riparian and flood plain habitats of *G. pensilis* have been seriously impacted by human activities, especially as a result of the development of agriculture and rice cultivation. This has led to high individual mortality and the rapid decline of most *G. pensilis* populations (Li and Xia, 2004; Li and Xia, 2005; Nguyen et al., 2013). Natural regeneration is extremely rare, probably due to low seed viability and loss of suitable habitat. Currently, *G. pensilis* is evaluated as Critically Endangered (CR) on the IUCN Red List of Threatened Species (Thomas et al., 2011).

Simple sequence repeats (SSR) are co-dominant and highly polymorphic markers which can be developed cheaply and

DOI:10.2478/sg-2019-0007
edited by the Thünen Institute of Forest Genetics

efficiently using next-generation sequencing technology (Simon et al., 2009). Previous molecular studies on *Glyptostrobus* using chloroplast microsatellites (cpSSR) (Nguyen et al., 2013) and ISSR markers showed that genetic diversity was remarkably low (Li and Xia, 2005; Wu, 2011). However, these studies only focused on a limited part of the distribution range of this conifer and to date, there have been no range wide studies of its genetic diversity and structure or its phylogeographic history. Recently, Wang et al (2019) developed 10 polymorphic microsatellite markers for this species using restriction site-associated DNA sequencing (RAD-Seq). In this study, we developed 23 Expressed Sequence Tag-Simple Sequence Repeat (EST-SSR) markers for *G. pensilis* using transcriptome data, which can be used to investigate its range wide genetic structure and phylogeography. Additionally, we tested the transferability of the loci in four related species: *Cryptomeria fortunei, Taxodium distichum, Taxodium ascendens, Cunninghamia lanceolata*.

Materials and Methods

Microsatellite detection

Fresh, young stems and leaf tissue of *G. pensilis* were collected from a plantation in Hunan Province (28°8′16.48″N, 112°59′28.36″E, 90m) and were immediately frozen in liquid nitrogen for RNA extraction. Total RNA was extracted using the RNAprep Pure plant kit (Tiangen) following the manufacturer's instructions. Qubit Flourometer and Agilent2100 were used to assess the concentration and the purity of the extracted RNA which was then used to construct a cDNA library and sequenced on a Illumina Genome analyzer HiSeq 2500 platform. Raw data were assembled de-novo using the software Trinity (http://trinityrnaseq.sourceforge.net/). The software MISA (http://pgrc.ipkgatersleben.de/misa/) was used to detect SSR motifs with 2, 3, 4, 5 and 6 bp nucleotide repeats. Primer3plus (http://www.bioinformatics.nl/cgi-bin/primer3plus/primer3plus.cgi) was used for primer design with the following parameters: primer length 18-27bp, annealing temperature 57-63°C, GC % content 20-80 %, and expected product length range of 100-300 bp; continuous A or T were not allowed in the 3′ end. Loci with a minimum number of 8 repeats were selected for amplification as they may have higher polymorphisms (Ueno et al., 2012). In total, 170 SSR loci were selected and evaluated for their suitability using 16 *G. pensilis* DNA samples (6 samples from DM, 5 from PNSL, and 5 from GZHN).

Microsatellite screening

Whole genomic DNA was extracted from dried leaves using the CTAB (cetyltrimethylammonium bromide) method (Tsumura et al., 1995). PCR amplification was carried out in volumes of 20 µl using the following protocol: 10 µL of 2×Taq PCR Master Mix (constituent: 20 mM Tris-HCl, 100 mM KCl, 3 mM $MgCl_2$, 0.1 U Taq Polymerase/µL, 500 µM dNTP each; Tiangen, Beijing, China), 1µl forward primer (10 µM) and 1µl reverse primer (10 µM), 3µl of 20-50ng/µl DNA template, and 5µl of ddH2O. The mixture was then cycled through the profile: 94°C for 4 min; 10 cycles of 94°C for 40 s, 65°C for 35 s (decreased by 1°C every cycle), 72°C for 30 s, followed by 32 cycles of 94°C for 40 s, 57°C for 35 s, 72°C for 30 s with a final extension at 60°C for 30 min. PCR products were visualized on a 2 % agarose gel. All loci which could be amplified successfully were tested individually using 16 *G. pensilis* samples to establish their polymorphic nature. For this second round of amplifications we used fluorescently labelled primers (FAM, HEX, TAMRA and ROX, Applied Biosystems) applying the same PCR protocol as above. PCR products were run on an Applied Biosystems 3130 Genetic Analyzer (ABI3130) adding a LIZ-500 labeled internal size standard (Applied Biosystems, Foster City, California, USA) to size fragments. The data were analyzed using the software Gene Marker V1.80. Among 170 SSR loci, 130 loci could be amplified successfully but only 23 loci were polymorphic and produced clear and interpretable peaks. The sequences of these 23 loci were deposited in the NCBI database (Table1) and used subsequently to investigate the genetic diversity of three Chinese *G. pensilis* populations (Appendix 1).

Table1
Characterization of 23 microsatellite loci developed in *Glyptostrobus pensilis*.[a]

Locus[a]	Repeat motif	Primer sequences (5'-3')	Size (bp)	GenBank accession no.	Putative function	E-value
Unigene_6272	$(AG)_{10}$	*F:*(FAM)TGGAGACCATTACGCGTTCA *R:* AAGAGGGGCATGTACGCTTC	287	MH061285	mRNA,clone:CFFL040_E19[*Cryptomeria japonica*]	0.000
Unigene_24035	$(AG)_8$	*F:* (HEX)CATTTCCCCCGGCAGATCAT *R:* GCCAAACATACAGCAGGTGC	253	MH061287	mRNA,clone:CLFL043_N20[*Cryptomeria japonica*]	0.000
Unigene_15398	$(AT)_{11}$	*F:* (TAMRA)CTGTTCCCCTGTGCATCATA *R:* AGATCGTGAGGGTTCGAGTC	307	MH158506	tRNA-Leu (trnL) gene and trnL-trnF intergenic spacer [*Glyptostrobus pensilis*]	1e-154
Unigene_62277	$(TA)_9$	*F:* (FAM)AGTATCCCGAGGGTGAATACA *R:* TAGCCGATGATCCTCGCATG	163	MH061297	CloneGQ04107_D16 mRNA sequence [*Picea glauca*]	4e-78
Unigene_35432	$(TG)_8$	*F:* (HEX)TCGCGATTCTGAGATCCAGC *R:* CGGCCACCAAATCCTTCTCT	209	MH061291	clone c11784_g1_i1 microsatellite sequence [*Chamaecyparis hodginsii*]	6e-105
Unigene_3959	$(AT)_{10}$	*F:* (TAMRA)TTCCTGATACCCACAAGCAT *R:* CAATTGAAATCAACCAAGTGGA	111	MH158507	uncharacterized LOC104229476 [*Nicotianasylvestris*]	2e-05
Unigene_51014	$(AG)_9$	*F:* (FAM)AGAGATGTGGGTATGGAGTTCT *R:* TCCATCCCTCTCTATCTCCCTAAC	143	MH061296	genome assembly S_erinaceieuropaei, scaffoldSPER_scaffold0054895 [*Spirometra erinaceieuropaei*]	2e-33
Unigene_36687	$(TCG)_8$	*F:* (HEX)TGCGCATCTCTCAGCATCAT *R:* GATGGAGAGCTCTGAGGCAC	254	MH061292	nucleolin-like LOC103497559 [*Cucumis melo*]	9e-22
Unigene_30801	$(AG)_{14}$	*F:*(TAMRA)TCATAATGCAACTTGACTCGATG *R:* TCCTGGACGGTTCAGTGC	172	MH158508	Uncharacterized LOC106298374 [*Brassica oleracea var. oleracea*]	3e-07
Unigene_88986	$(TC)_9$	*F:* (FAM)GCATTTGCTGGTGTTGCTTG *R:* GGTGGTGGGTGAAAGGGAAT	182	MH061298	clone CM43 microsatellite sequence [*Chilina dombeiana*]	3e-04
Unigene_16056	$(AT)_{10}$	*F:* (TAMRA)CGTCTGAGAGAGAAAGGAAAGC *R:* GTCCATTTCTGCTTTGGAGC	231	MH158509	mRNA,clone: CSFL017_C07 [*Cryptomeria japonica*]	7e-81
Unigene_8850	$(AT)_9$	*F:* (FAM)ACAAAAGGAAAAACCAGACGTAA *R:* CCTCACAAGGTAGCGGTTGT	200	MH061286	clone WS0461_N23 unknown mRNA [*Picea sitchensis*]	4e-10
Unigene_26046	$(AT)_8$	*F:* (HEX)CCGCTCCGATTGTTGATGTG *R:* ACCTGCCTGCTTGGTAAAATG	245	MH061289	mRNA, clone: CMFL023_M20 [*Cryptomeria japonica*]	0.000
Unigene_20393	$(AG)_{11}$	*F:* (TAMRA)TGCTCTCCAACAACAACAGC *R:* ATTTCATGGCATCTGGTGGT	91	MH158510	chromogranin-A-like (LOC106455924) [*Pundamilia nyererei*]	2e-06
Unigene_942	$(AT)_8$	*F:* (FAM)CTCAGACACAGCTGGCTTCA *R:* GGTGAAGATATCTGGAACTTGACG	236	MH061284	protein phosphatase 1 regulatory subunit 14B(LOC100709168) [*Oreochromis niloticus*]	0.001
Unigene_50154	$(AT)_8$	*F:* (HEX)ACTGCAAGAAAATTCCAGCGG *R:* ACCCTTCTGCTGAAAAGGCA	125	MH061295	mRNA, clone: CFFL004_I22 [*Cryptomeria*	0.000
Unigene_2343	$(AG)_{10}$	*F:* (TAMRA)CTTTAAGCGTTTCCAGGACAG *R:* CTTGGCATGCAGTCTTCTGA	261	MH158511	mRNA, clone: CLFL031_D03 [*Cryptomeria japonica*]	6e-19
Unigene_24759	$(CT)_9$	*F:* (FAM)TGTTCCGATGTCTACTGCCG *R:* TTTCACTCTCACCAGGTCGC	225	MH061288	putative exopolyphosphatase (LOC113346861) [*Papaver somniferum*]	1e-19
Unigene_37063	$(AT)_8$	*F:* (HEX)TCCATTGAGCACAACCCACT *R:* AGCTTTTAGTGGACATGCTTGA	284	MH061293	cultivar NCNSP0306 chromosome 3 [*Ipomoea trifida*]	0.001
Unigene_106035	$(AG)_8$	*F:* (TAMRA)GCAAGGCTCTGCAAGATGAAG *R:* CACCTTCCGTGCTGCTCTAT	187	MH061299	strain HSOK chromosome 15 [*Oryzias latipes*]	0.011
Unigene_16778	$(AG)_{11}$	*F:* (ROX)AGATGAGATCGAGGCGAAGA *R:* CGCATTTATGGATCCCGTTA	280	MH158512	genome assembly AptMant0, scaffold scaffold27 [*Apteryx australis mantelli*]	4e-08
Unigene_34713	$(GA)_8$	*F:* (HEX)GCAATCTTTGGGCCATGTGG *R:* CTAGACCAGCACGCGTAGAG	245	MH061290	putative ammonium transporter [*Apteryx australis mantelli*]	0.000
Unigene_45772	$(AT)_8$	*F:* (HEX)GTGGGTCCAGTGGTTCCAAT *R:* GGAGATGTTTGTCACACCCCA	273	MH061294	CloneDKEYP-98C11 in linkage group 5[*Danio rerio*]	0.019

[a]Optimal annealing temperature was 60°C for all loci.

Data analysis

GenAlEx6.5 (Peakall and Smouse, 2012) was used to calculate the following parameters: allele number (*A*), effective number of alleles (A_e), observed heterozygosity (H_o) and expected heterozygosity (H_e). Microsatellite toolkit (Park, 2001) was used for the calculation of the polymorphic information content (*PIC*) and a Hardy-Weinberg equilibrium analysis was conducted in Genepop (http://www.genepop.curtin.edu.au/).

Results and Discussion

After cDNA library construction, transcriptome sequencing and quality filtering, a total of 106,901 unigenes were acquired with an average length of 872 bp. A total of 12,279 SSR motifs were identified in 9,597 unigenes. The SSR frequency, i. e. the ratio of unigenes with SSRs to the total number of unigenes was 8.98 %, including 1,892 unigenes that contained more than one SSR region. The distribution frequency of SSR motifs, i. e. the ratio of SSR loci to the total number of unigene was 11.48 %, which means that on average there is an SSR every 7.59 kb.

Among the motifs identified, trinucleotide repeats were the most abundant type with a frequency of 43.36 %, followed by tetranucleotide (30.60 %), hexanucleotide (10.45 %), dinucleotide (8.67 %) and pentanucleotide repeats (6.92 %). Among the trinucleotide loci, the AAG/CTT motif was the most abundant, followed by AGG/CCT and AGA/TCT.

The number of alleles (*A*) varied from one to eight in each population, the effective number of alleles (A_e) ranged from A_e = 1.00 to 5.97, the observed heterozygosity (H_o) from H_o = 0.00 to 1.00, and the expected heterozygosity (H_e) from H_e = 0.00 to 0.83. The polymorphic information content (*PIC*) varied between loci from *PIC* = 0.0120 to 0.7188, with an average of 0.4001 over all 23 loci (Table 2). Fifteen loci deviated from Hardy-Weinberg equilibrium (HWE) in at least one of the three populations and three (Unigene_24035, Unigene_51014, Unigene_2343) showed significant deviation in all assayed populations (HWE, P < 0.05; Table 2) which might be due to the presence of null alleles. However, two out of these three loci (Unigene_51014, Unigene_2343; Table 2) exhibited an excess of heterozygotes in all populations compared to what is expected under HWE which is unlikely to be caused by null alleles. A possible explanation could be associative overdominance or that these loci do not behave in a strictly neutral way (transcriptome data) and are under selective pressure. However, the deviations from HWE could also be due to chance.

Table 2.
Genetic diversity in threeChinese *Glyptostrobus pensilis* populations based on the 23 polymorphic microsatellite markers.[a]

Locus	DM(*N* = 33)				PNSL(*N* = 29)				GZHN(*N* = 21)				Total(*N* = 83)	
	A	A_e	H_o	H_e	*A*	A_e	H_o	H_e	*A*	A_e	H_o	H_e	*A*	*PIC*
unigene_6272	4	2.177	0.300	0.541*	3	1.110	0.103	0.099	3	1.706	0.333	0.414	5	0.4820
unigene_24035	2	1.502	0.000	0.334*	4	2.722	0.069	0.633*	3	2.014	0.000	0.503*	4	0.4789
unigene_15398	4	1.250	0.094	0.200*	3	1.319	0.276	0.242	3	2.228	0.200	0.551*	5	0.3206
unigene_62277	2	1.067	0.065	0.062	2	1.035	0.034	0.034	2	1.049	0.048	0.046	2	0.3576
unigene_35432	1	1.000	0.000	0.000	2	1.035	0.034	0.034	2	1.100	0.000	0.091*	3	0.0357
unigene_3959	5	2.580	0.448	0.612	2	1.039	0.038	0.038	5	3.653	0.650	0.726	7	0.7108
unigene_51014	5	3.372	0.970	0.703*	3	2.448	1.000	0.592*	4	3.000	0.952	0.667*	6	0.6929
unigene_36687	1	1.000	0.000	0.000	2	1.890	0.759	0.471*	2	1.049	0.048	0.046	2	0.2140
unigene_30801	8	4.787	0.576	0.791*	2	1.071	0.069	0.067	7	5.970	0.850	0.833*	9	0.7188
unigene_88986	2	1.031	0.030	0.030	1	1.000	0.000	0.000	1	1.000	0.000	0.000	2	0.0120
unigene_16056	4	1.213	0.125	0.175	3	1.764	0.500	0.433	4	2.427	0.667	0.588	5	0.4588
unigene_8850	1	1.000	0.000	0.000	4	2.184	0.862	0.542*	1	1.000	0.000	0.000	4	0.4706
unigene_26046	3	2.068	0.697	0.517	3	1.761	0.517	0.432*	3	1.537	0.429	0.349	3	0.4339
unigene_20393	2	1.695	0.333	0.410	2	1.109	0.103	0.098	3	1.800	0.476	0.444	3	0.3037
unigene_942	1	1.000	0.000	0.000	2	1.708	0.586	0.414	2	1.208	0.190	0.172	2	0.1966
unigene_50154	4	2.713	0.469	0.631*	2	1.035	0.034	0.034*	3	2.256	0.524	0.557*	4	0.5616
unigene_2343	5	3.290	0.970	0.696*	2	2.000	1.000	0.500*	4	3.065	1.000	0.674*	5	0.5873
unigene_24759	3	1.850	0.406	0.459*	1	1.000	0.000	0.000	2	1.630	0.333	0.387	3	0.2750
unigene_37063	5	2.510	0.438	0.602*	1	1.000	0.000	0.000	5	3.139	0.810	0.681*	5	0.5720
unigene_106035	2	1.330	0.161	0.248	2	1.035	0.034	0.034	2	1.724	0.300	0.420	2	0.3748
unigene_16778	1	1.000	0.000	0.000	2	1.665	0.552	0.400	3	2.172	0.857	0.540*	3	0.4978
unigene_34713	3	2.171	0.758	0.539	2	1.998	0.966	0.499*	3	2.146	0.857	0.534*	3	0.4157
unigene_45772	2	1.031	0.030	0.030	2	1.035	0.034	0.034	2	1.100	0.095	0.091	2	0.0459
Mean	3.043	1.854	0.299	0.330	2.261	1.477	0.329	0.245	3.000	2.086	0.418	0.405	3.869	0.4007

Note: *N* = sample size; *A* = number of alleles; A_e = effective number of alleles; H_o= observed heterozygosity; H_e = expected heterozygosity; *PIC* = polymorphic information content;
*Deviation from Hardy-Weinberg equilibrium (*P*<0.05)
[a]Locality and voucher information are provided in Appendix 1.

Additionally, cross-species amplification of the 23 loci developed in this study was tested in *Cryptomeria fortunei, Taxodium distichum, Taxodium ascendens* and *Cunninghamia lanceolata* (Appendix 1) using five individuals each. The number of loci that could be successfully amplified (n) and were polymorphic (n_{poly}) were as follows: *C. fortunei* (n = 22, n_{poly} = 15), *T. distichum* (n = 22, n_{poly} = 17), *Ta. ascendens* (n = 21, n_{poly} = 18) and *Cu. lanceolata* (n = 16, n_{poly} = 10) (Table 3). Interestingly, out of the four species *Cu. lanceolata* had the lowest number of polymorphic loci. This may be due to the more distant genetic relationship of *Cu. lanceolata* which belongs to another subfamily (Cunninghamioideae) compared to *Taxodium, Cryptomeria* and *Glyptostrobus* which are all in the same subfamily (Taxodioideae) (Mao et al. 2012).

Table 3
Characterization of 23 SSR markers developed for *Glyptostrobus pensilis* in four closely related species.[a]

Locus	*Taxodium distichum*(*N* = 5)				*Taxodium ascendens* (*N* = 5)				*Cryptomeria fortunei* (*N* = 5)				*Cunninghamia lanceolata*(*N* = 5)			
	A	A_e	H_o	H_e	*A*	A_e	H_o	H_e	*A*	A_e	H_o	H_e	*A*	A_e	H_o	H_e
unigene 6272	2	1.724	0.600	0.420	2	1.724	0.600	0.420	2	1.471	0.000	0.320	2	1.724	0.600	0.420
unigene 24035	1	1.000	0.000	0.000	1	1.000	0.000	0.000	1	1.000	0.000	0.000	—	—	—	—
unigene 15398	1	1.000	0.000	0.000	2	1.471	0.000	0.320	3	2.273	0.000	0.560	2	1.923	0.000	0.480
unigene 62277	2	1.471	0.400	0.320	1	1.000	0.000	0.000	1	1.000	0.000	0.000	—	—	—	—
unigene 35432	1	1.000	0.000	0.000	3	2.381	0.200	0.580	1	1.000	0.000	0.000	1	1.000	0.000	0.000
unigene 3959	4	2.941	0.200	0.660	—	—	—	—	2	1.471	0.400	0.320	—	—	—	—
unigene 51014	3	2.273	0.800	0.560	2	2.000	1.000	0.500	4	3.846	1.000	0.740	4	2.941	1.000	0.660
unigene 36687	2	1.923	0.800	0.480	3	1.852	0.600	0.460	2	2.000	1.000	0.500	2	1.471	0.400	0.320
unigene 30801	5	4.167	0.800	0.760	4	2.381	0.200	0.580	4	3.846	0.600	0.740	—	—	—	—
unigene 88986	2	1.923	0.800	0.480	3	1.852	0.600	0.460	2	1.724	0.600	0.420	1	1.000	0.000	0.000
unigene 34713	4	3.333	0.600	0.700	4	3.333	0.800	0.700	1	1.000	0.000	0.000	3	2.273	0.800	0.560
unigene 16056	—	—	—	—	—	—	—	—	1	1.000	0.000	0.000	—	—	—	—
unigene 16778	2	1.923	0.800	0.480	2	2.000	1.000	0.500	3	2.632	1.000	0.620	1	1.000	0.000	0.000
unigene 8850	1	1.000	0.000	0.000	2	1.471	0.000	0.320	2	1.923	0.000	0.480	1	1.000	0.000	0.000
unigene 26046	4	3.333	0.800	0.700	3	1.852	0.200	0.460	1	1.000	0.000	0.000	3	2.632	0.600	0.620
unigene 20393	8	7.143	1.000	0.860	7	5.556	1.000	0.820	5	4.545	0.800	0.780	3	2.632	0.800	0.620
unigene 942	6	4.545	0.800	0.780	5	3.125	0.600	0.680	3	2.381	1.000	0.580	4	3.333	0.200	0.700
unigene 50154	2	2.000	1.000	0.500	3	1.852	0.200	0.460	2	2.000	1.000	0.500	1	1.000	0.000	0.000
unigene 2343	5	4.545	0.200	0.780	3	2.778	0.000	0.640	—	—	—	—	—	—	—	—
unigene 24759	1	1.000	0.000	0.000	1	1.000	0.000	0.000	2	1.471	0.400	0.320	1	1.000	0.000	0.000
unigene 37063	2	1.220	0.200	0.180	2	1.923	0.800	0.480	1	1.000	0.000	0.000	—	—	—	—
unigene 106035	2	2.000	1.000	0.500	2	2.000	1.000	0.500	2	2.000	1.000	0.500	2	2.000	1.000	0.500
unigene 45772	3	2.381	1.000	0.580	2	1.923	0.800	0.480	6	5.556	1.000	0.820	5	3.846	0.400	0.740

Note:—= not available; *N* = sample size; *A* = number of alleles; A_e = effective number of alleles; H_o= observed heterozygosity; H_e = expected heterozygosity
[a]Locality and voucher information are provided in Appendix 1.

In this study we developed 23 polymorphic microsatellite loci which can be used to assess the genetic diversity of *G. pensilis* populations and underpin conservation efforts for this threatened species. One of the 23 loci (Unigene_24035) showed significant deviation from HWE due to an excess of homozygotes in all assayed populations which might be due to the presence of null alleles and should be used with caution. Two loci (Unigene_51014, Unigene_2343) had significantly more heterozygotes than expected. As these loci were developed from transcriptome data, they might linked to regions of the genome which are under selection. If this is the case then Unigene_51014 and Unigene_2343 could potentially be useful to study adaptation. Additionally, the result of our cross-amplification test suggests that the 23 loci might also be useful in genetic studies of related species in Cupressaceae.

Acknowledgements

This project was supported by forestry industry standard project of China (2014-LY-213). The Royal Botanic Garden Edinburgh is supported by the Scottish Government's Rural and Environment Science and Analytical Services Division.

Data Archiving Statement

Raw sequencing reads were deposited in the National Center for Biotechnology Information (NCBI) Sequence Read Archive (accession no. PRJNA505868). All primer sequences were uploaded to NCBI (accession no. MH061284–MH1299; MH158506-MH158512; Table 1)

References

Averyanov LV, Phan KL, Nguyen TH, Nguyen SK, Nguyen TV, Pham TD (2009) Preliminary observation of native Glyptostrobus pensilis (Taxodiaceae) stands in Vietnam.Taiwania 54:191-212. https://doi.org/10.6165/tai.2009.54(3).191

Li FG, Xia NH (2004) The geographical distribution and cause of threat to Glyptostrobus pensilis (Taxodiaceae). J Trop Subrtop Bot 12:13-20. https://doi.org/10.3969/j.issn.1005-3395.2004.01.003

Li FG, Xia NH (2005) Population structure and genetic diversity of an endangered species, Glyptostrobus pensilis (Cupressaceae). Botanical Bulletin of Academia Sinica 46:155-162. https://doi.org/10.7016/BBAS.200504.0155

LePage BA (2007) The taxonomy and biogeographic history of Glyptostrobus Endlicher (Cupressaceae). Bulletin of the Peabody Museum of Natural History 48: 359-426. https://doi.org/10.3374/0079-032X(2007)48[359:TTABHO]2.0.CO;2

Mao K, Milne RI, Zhang L, Peng Y, Liu J, Thomas P, Mill RR, Renner SS (2012) Distribution of living cupressaceae reflects the breakup of pangea. Proceedings of the National Academy of Sciences, 109: 7793-7798. https://doi.org/10.1073/pnas.1114319109

Nguyen MT, Vu DD, Bui TTX, Nguyen MD (2013) Genetic variation and population structure in Chinese water pine (Glyptostrobus pensilis): a threatened species. Indian Journal of Biotechnology 12:499-503.

Park SDE (2001) The Excel microsatellite toolkit. Trypanotolerance in west African cattle and the population genetic effects of selection. [Ph.D. thesis]. University of Dublin, Dublin, Ireland.

Peakall R, Smouse PE (2012) GenAlEx version 6.5: Genetic analysis in Excel. Population genetic software for teaching and research-An update. Bioinformatics 28:2537-2539. https://doi.org/10.1093/bioinformatics/bts460

Simon SA, Zhai J, Nandety RS, Mccormick KP, Zeng J, Mejia D, Meyers BC (2009) Short-read sequencing technologies for transcriptional analyses. Annual Review of Plant Biology 60:305-333. https://doi.org/10.1146/annurev.arplant.043008.

Tsumura Y, Yoshimura K, Tomaru N, Ohba K (1995) Molecular phylogeny of conifers using RFLP analysis of PCR-amplified specific chloroplast genes. Theoretical and Applied Genetics 91:1222-1236. https://doi.org/10.1007/bf00220933

Thomas P, Lepage BA (2011) The end of an era?: the conservation status of redwoods and other members of the former Taxodiaceae in the 21st century. Japanese Journal of Historical Botany 19:89-100.

Thomas P, Yang Y, Farjon A, Nguyen D, Liao W (2011) Glyptostrobus pensilis. The IUCN Red List of Threatened Species 2011: e.T32312A9695181. https://doi.org/10.2305/iucn.uk.2011-2.rlts.t32312a9695181.en

Ueno S, Moriguchi Y, Uchiyama K, Ujino-Ihara T, Futamura N, Sakurai T, Shinohara K, Tsumura Y (2012) A second generation framework for the analysis of microsatellites in expressed sequence tags and the development of EST-SSR markers for a conifer, Cryptomeria japonica. BMC Genomics 13:1-16. https://doi.org/10.1186/1471-2164-13-136

Wu ZY (2011) Study on Conservation Biology and Restoration Technique of the Relict Plant Glyptostrobus pensilis [Ph.D.thesis]. Fujian Agriculture and Forestry University, Fujian Province, China.

Wang GT, Wang ZF, Wang RJ, Liang D, Jiang GB (2019) Development of microsatellite markers for a monotypic and globally endangered species, Glyptostrobus pensilis (Cupressaceae). Applications in Plant Sciences 7: e1217. https://doi.org/10.1002/aps3.1217

Xu XH, Li MP (1959) The ecology and geographical distribution of Glyptostrobus pensilis. J. South China Normal Univ. 3:84-99.

Yu YF (1995) Origin evolution and distribution of the Taxodiaceae. Acta Phytotaxon. Sin. 33:362-389.