View metadata, citation and similar papers at core.ac.uk

brought to you by CORE

provided by Aberystwyth Research Portal

Aberystwyth University

Using MiddRAD-seq data to develop polymorphic microsatellite markers for an endangered yew species

Qin, Han-Tao; Yang, Guo-Qian; Provan, James; Liu, Jie; Gao, Lian-Ming

Published in:
Plant Diversity

DOI:
10.1016/j.pld.2017.05.008

Publication date:
2017

Citation for published version (APA):
Qin, H-T., Yang, G-Q., Provan, J., Liu, J., & Gao, L-M. (2017). Using *Midd*RAD-seq data to develop polymorphic microsatellite markers for an endangered yew species. *Plant Diversity*, *39*(5), 294-299. https://doi.org/10.1016/j.pld.2017.05.008

Document License
CC BY-NC-ND

General rights
Copyright and moral rights for the publications made accessible in the Aberystwyth Research Portal (the Institutional Repository) are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

• Users may download and print one copy of any publication from the Aberystwyth Research Portal for the purpose of private study or research.
• You may not further distribute the material or use it for any profit-making activity or commercial gain
• You may freely distribute the URL identifying the publication in the Aberystwyth Research Portal

Take down policy
If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

tel: +44 1970 62 2400
email: is@aber.ac.uk

Download date: 30. Aug. 2021

Plant Diversity 39 (2017) 294–299

Contents lists available at ScienceDirect

Plant Diversity

journal homepage: http://www.keaipublishing.com/en/journals/plant-diversity/
http://journal.kib.ac.cn

Using *Mi*ddRAD-seq data to develop polymorphic microsatellite markers for an endangered yew species

Hantao Qin [a, b], Guoqian Yang [b, c], Jim Provan [d], Jie Liu [a, **], Lianming Gao [a, *]

[a] Key Laboratory for Plant Diversity and Biogeography of East Asia, Kunming Institute of Botany, Chinese Academy of Sciences, Kunming, Yunnan, 650201, China
[b] University of Chinese Academy of Sciences, Beijing, 100049, China
[c] Germplasm Bank of Wild Species in Southwest China, Kunming Institute of Botany, Chinese Academy of Sciences, Kunming, 650201, China
[d] Institute of Biological, Environmental and Rural Sciences (IBERS), Aberystwyth University, Penglais, Aberystwyth, SY23 3DA, UK

ARTICLE INFO

Article history:
Received 19 January 2017
Received in revised form
26 May 2017
Accepted 27 May 2017
Available online 2 June 2017

(Editor: Xuewen Wang)

Keywords:
*Mi*ddRAD-seq
Endangered species
Microsatellite
Next-generation sequencing
Taxus florinii

ABSTRACT

Microsatellites are highly polymorphic markers which have been used in a wide range of genetic studies. In recent years, various sources of next-generation sequencing data have been used to develop new microsatellite loci, but compared with the more common shotgun genomic sequencing or transcriptome data, the potential utility of RAD-seq data for microsatellite ascertainment is comparatively under-used. In this study, we employed *Mi*ddRAD-seq data to develop polymorphic microsatellite loci for the endangered yew species *Taxus florinii*. Of 8,823,053 clean reads generated for ten individuals of a population, 94,851 (~1%) contained microsatellite motifs. These corresponded to 2993 unique loci, of which 526 (~18%) exhibited polymorphism. Of which, 237 were suitable for designing microsatellite primer pairs, and 128 loci were randomly selected for PCR validation and microsatellite screening. Out of the 128 primer pairs, 16 loci gave clear, reproducible patterns, and were then screened and characterized in 24 individuals from two populations. The total number of alleles per locus ranged from two to ten (mean = 4.875), and within-population expected heterozygosity from zero to 0.789 (mean = 0.530), indicating that these microsatellite loci will be useful for population genetics and speciation studies of *T. florinii*. This study represents one of few examples to mine polymorphic microsatellite loci from ddRAD data.

Copyright © 2017 Kunming Institute of Botany, Chinese Academy of Sciences. Publishing services by Elsevier B.V. on behalf of KeAi Communications Co., Ltd. This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/4.0/).

1. Introduction

Microsatellites (simple sequence repeats, SSR) are repeated motifs of 1–6 nucleotides which are found distributed throughout the genomes of eukaryotes (Jarne and Lagoda, 1996; Selkoe and Toonen, 2006). Compared with other PCR-based approaches such as RAPD, ISSR, and AFLP, SSRs represent a kind of cost-effective, neutral marker with the advantage of being codominant, genome wide, locus-specific and highly polymorphic (Li et al., 2002) SSRs have been widely used in genome mapping, parentage analysis, population genetics and phylogeography (Jarne and Lagoda, 1996; Selkoe and Toonen, 2006; Guichoux et al., 2011; Hodel et al., 2016). Traditionally, *de novo* SSR development for non-model species has involved construction of a genomic library followed by identifying SSR-containing clones by Sanger sequencing, which can be a tedious and costly process (Zalapa et al., 2012; Hodel et al., 2016). More recently, the development of SSR markers has been transformed with the advent of next-generation sequencing (NGS) technology (Davey et al., 2011; Zalapa et al., 2012), offering the ability to develop numerous microsatellite loci based on large amounts of sequence data with a reduced cost and effort (Guichoux et al., 2011; Zalapa et al., 2012; Hodel et al., 2016).

Genome-wide sequence data, generated mainly by 454 or Illumina platforms, have been increasingly used in the fields of ecology and evolutionary biology (Taylor et al., 2014; Vera et al., 2016). Several NGS-based methods, including restriction site-associated DNA sequencing (RAD-seq), genotyping by sequencing (GBS), and

* Corresponding author. Kunming Institute of Botany, Chinese Academy of Sciences, 132 Lanhei Road, Kunming, Yunnan, 650201, China. Fax: +86 0871 68123227.
** Corresponding author. Kunming Institute of Botany, Chinese Academy of Sciences, 132 Lanhei Road, Kunming, Yunnan, 650201, China. Fax: +86 0871 65225286.
E-mail addresses: liujie@mail.kib.ac.cn (J. Liu), gaolm@mail.kib.ac.cn (L. Gao).
Peer review under responsibility of Editorial Office of Plant Diversity.

http://dx.doi.org/10.1016/j.pld.2017.05.008
2468-2659/Copyright © 2017 Kunming Institute of Botany, Chinese Academy of Sciences. Publishing services by Elsevier B.V. on behalf of KeAi Communications Co., Ltd. This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/4.0/).

specific length/locus amplified fragment sequencing (SLAF-seq), have been used for large-scale discovery of SNPs (Baird et al., 2008; Elshire et al., 2011; Bonatelli et al., 2015; Ma et al., 2015). A variant of RAD-seq, double digest restriction-site associated DNA sequencing technology (ddRAD-seq) is becoming an increasingly popular approach for SNP marker development and genotyping in animals (Peterson et al., 2012).

Although NGS-based technologies are becoming widely utilized for SNP discovery and genotyping, they still cannot replace SSRs completely in population genetic studies, particularly in studies with large sample sizes and their associated expensive costs. Additionally, it has been shown that the results of genetic structure obtained by SSRs generally mirror those by SNPs of genomic data based on NGS (Elbers et al., 2016; Jansson et al., 2016; Jeffries et al., 2016). Therefore, microsatellites still have a pivotal role to play in population genetic studies, even in the "genomic era" (Hodel et al., 2016). Recently, a modified ddRAD-seq (*Mi*ddRAD-seq) approach was developed for SNP discovery and genotyping in angiosperm plants, which is quicker and cheaper than conventional ddRAD, and has a potential to be used for plant population genetics, phylogenetics, phylogeography (Yang et al., 2016), as well as for microsatellite isolation. To date, though, there has only been one reported study using ddRAD to isolate microsatellite markers for plant species by 454 sequencing (Bonatelli et al., 2015). In the present study, we have used the *Mi*ddRAD approach to develop microsatellite markers for a species of *Taxus* (yew), which are notorious for their large and complicated genomes (Leitch et al., 2001).

The yews are Tertiary relic species which have high economic and medicinal value, and are particularly well-known as the source of the cancer-inhibitory alkaloid Paclitaxel (Taxol; Kingston and Newman, 2007; Gao et al., 2007; Poudel et al., 2013). Several species of yew are listed as endangered by the International Union for the Conservation of Nature (IUCN) (https://www.iucn.org/), and all *Taxus* species in China were listed as first-class protection wild species in 1999 (http://www.forestry.gov.cn/portal/main/s/3094/minglu1.htm). Although several sets of microsatellite primers have been developed for different *Taxus* species using traditional methods (e.g. Dubreuil et al., 2008; Yang et al., 2009), cross-species amplification success has generally been very low (see Liu et al., 2011a). Thus, more species-specific polymorphic microsatellite loci of yews are needed to best facilitate further conservation genetic and evolutionary biological analyses.

Taxus florinii Spjut was described as a new species by Spjut (2007), supported by genetic and morphological evidence (Gao et al., 2007; Liu et al., 2011b, 2013; Möller et al., 2013). The species is confined to Southwestern China, with a narrow distribution in northwestern Yunnan and southwestern Sichuan (Liu et al., 2013; Möller et al., 2013). According to the categories and criteria of the International Union for the Conservation of Nature (IUCN) Red List of Threatened Species, the species should be classified as critically endangered due to sharp declines in natural populations over the past two decades. There are no polymorphic microsatellite markers reported for this endangered tree species so far, which hinders conservation genetic analysis and protection management. In this study, we aimed to develop and validate polymorphic microsatellite loci using the *Mi*ddRAD method for *T. florinii*.

2. Materials and methods

2.1. Plant material and DNA extraction

Ten individuals from *T. florinii* population TFP03 (Lanping County, Yunnan, China, 26°91′N, 99°42′E) were used to build a *Mi*ddRAD library. In addition, twelve individuals of this population and twelve from population TFP04 (Yulong xueshan, Yulong county, Yunnan, China, 27°14′N, 100°23′E) were selected for validating microsatellite polymorphism. Genomic DNA was extracted from silica-gel dried leaves using a modified CTAB method (Liu and Gao, 2011). The quality and concentration of DNA were assessed on 1% TAE agarose gels and using a NanoDrop® ND-1000 spectrophotometer (Thermo Fisher Scientific, Wilmington, DE, USA).

2.2. MiddRAD library preparation and NGS data processing

The *Mi*ddRAD library preparation protocol of Yang et al. (2016) was followed. Briefly, all DNA samples were double-digested with enzymes AvaII (NEB, Cat#: R0153S) + MspI (NEB, Cat#: R0106S). Barcoded adapters were annealed and ligated. Digested and ligated DNAs from the ten samples were pooled together and fragments of 600–700 base pairs (bp) retrieved from agarose gel. The fragments were enriched by PCR amplification using Phusion Master Mix with HF Buffer (NEB, Cat#: M0531S). Finally, the PCR products were purified with the QIAquick PCR purification kit (Qiagen, Cat#: 28106) and quantified with the Qubit® dsDNA High Sensitivity (HS) Assay kit (Invitrogen, Cat#: Q32851). Mixed samples were sequenced in a single lane of Illumina HiSeq X Ten (Illumina, Inc., San Diego, CA, USA) (150 bp, double-end). We used the script *process_radtags* from Stacks v1.39 (Catchen et al., 2013) to check the barcode of raw reads, to demultiplex the data and to discard any reads for which scores dropped below 90% probability of being correct (http://catchenlab.life.illinois.edu/stacks/). The software FASTQC (http://www.bioinformatics.babraham.ac.uk/projects/fastqc/) was employed to assess the sequencing quality.

2.3. Microsatellites mining and primer design

Because the Illumina sequence reads are relatively short and numerous, we used the *ustacks* script in Stacks v1.39 to create consensus sequences for each individual. These consensus sequences were obtained from aligning a set of short-read sequences into exactly-matching stacks. Next, the *pipe1.pl* scripts in QDD_v3 (Meglécz et al., 2010) were used to detect the sequences including microsatellites with di-to hexanucleotide motifs. The sequences with microsatellites of the ten individuals sequenced were aligned using the QDD *pipe2.pl* script. This step was used to identify polymorphic microsatellite loci which were then used to design primers. Polymorphic microsatellites were defined as polymorphisms present in the consensus sequences of the ten individuals from the TFP03 population.

The Perl script MIcroSAtelitte (MISA, http://pgrc.ipk-gatersleben.de/misa/) was used to screen di-to hexanucleotide motifs. The lowest thresholds of repeats for dinucleotides and trinucleotides were set to six, and others (tetra-, penta- and hexanucleotides) were set to three. Two neighboring SSRs with the maximum interruption less than 100 bp were considered as one compound microsatellite. The program Primer v3 was used to design SSR primers (http://bioinfo.ut.ee/primer3-0.4.0/). The expected products ranged from 100 to 280 bp. The primer size ranged from 18 to 23 bp with the optimal size of 20 bp. The optimum GC content was 50%, and the optimum melting temperature was 60 °C (ranged from 57 °C to 62 °C).

2.4. PCR validation

A total of 128 SSR primer pairs were selected for preliminary PCR screening with three individuals from population TFP03. Only the primer pairs which successfully amplified and produced the appropriate size with clear peaks were used for further polymorphism verification using the ten samples for *Mi*ddRAD-seq.

Table 1
Statistics of microsatellite markers developed from *Mi*ddRAD-seq data for ten samples of *Taxus florinii*.

Total no. of filtered reads	214,968,105
Total no. of unique reads (reduced by *ustacks*)	8,823,053
No. of reads containing microsatellites	94,851
No. of consensus sequences among individuals containing microsatellites	2993
No. of consensus sequences among individuals containing polymorphic microsatellites	526
Primer pairs designed	237
Primer pairs screened	128
Microsatellites selected for characterization	16

Finally, the selected validated microsatellite loci were characterized in 24 samples from two populations of *T. florinii*. The PCR procedure included two rounds. Forward primers were modified with the addition of an universal labeled M13 tail (TGTAAAACGACGGCCAGT; Schuelke, 2000), and the first round was carried out using the modified forward primer and unmodified reverse primer. Using 1 μL PCR product from the first round as template, the second round was carried out using a fluorescently-labeled M13 primer and the unmodified reversed primer. PCR amplification was carried out on a Veriti® 96-Well Thermal Cycler (Applied Biosystems, Foster City, USA). The PCR amplification conditions were as follows: an initial denaturation for 3 min at 95 °C, followed by 37 cycles of denaturation at 94 °C for 30 s, annealing temperature at 58 °C and/or 60 °C according to the Ta value (Table 2) for 30 s, extension at 72 °C for 30 s, and a final extension step at 72 °C for 4 min. The products were separated and visualized using an ABI3730 XL automated DNA fluorescent sequencer (Applied Biosystems, Foster City, CA, USA).

2.5. Data analysis

The SSR data were visualized and quantified using automated allele calling implemented in GENEMARKER v2.2.0 (SoftGenetics, State College, PA, USA). Population genetic statistics and Hardy–Weinberg equilibrium were calculated with GENALEX v6.41 (Peakall and Smouse, 2006). GENEPOP v4.0 was used to detect linkage disequilibrium between pairs of loci (Rousset, 2008).

3. Results

3.1. NGS MiddRAD dataset

The raw sequence data from the ten individuals of population TFP03 contained 214,968,105 reads, with an average size of 141 bp, of which 425,146 were removed due to low quality. After aligning and comparing using *ustacks*, the clean data reduced to 8,823,053 reads (Table 1).

3.2. The composition and characteristics of microsatellite loci

We obtained 2993 consensus sequences containing microsatellites by analyzing the *Mi*ddRAD data of ten individuals. A total of 3053 microsatellites were identified from these sequences with dinucleotides being the most abundant repeat motifs (41.87%, 1253 sequences), followed by tetranucleotides (706 sequences),

Table 2
Characteristics of the 16 validated polymorphic microsatellite loci developed for *Taxus florinii*.

Locus	GenBank accession no.	Repeat motif	Primer sequence (5′–3′)	*Ta* (°C)	Size (bp)
gr907	KY704069	$(CTT)_{6-7}$	**F:**CATTGCGCCTCTTTGGAGTC **R:**TGGCAGGCAGAATCAAAGGT	58	113–116
gr1114	KY704070	$(TG)_{7-10}$	**F:**AGACCCATCCAATATTTATAAAATGGT **R:**AGAGACAACTTGAAAAGACCAGA	58	106–112
gr1177	KY704071	$(AT)_{8-9}$	**F:**AGACCACACTTGCCCTTCAG **R:**AGCCAATGTATGCTAGGGAGTG	58	109–111
gr1258	KY704072	$(TA)_{9-10}$	**F:**GACCCCACATTGTTTCTTCAAG **R:**ACAACCTCTCAAAAACAACACAA	58	115–117
gr1439	KY704073	$(TC)_{7-8}$	**F:**TACCAATCACCCACCGTTCG **R:**GCCTCTTGCCATGACCGTTT	58	123–125
gr2591	KY704074	$(TC)_{6-7}$	**F:**GCCCTTAACGACGGAGTTGA **R:**CTCCTCTCTTGATAGCCGCG	58	111–113
gr2662	KY704075	$(AT)_{8-9}$	**F:**GTCCTATGTACCACTAAGCACA **R:**CGTTGTCTCAAGAATTTAATACATTCC	58	137–139
gr2680	KY704076	$(TA)_{8-9}$	**F:**AGTCAAAATGAGACCCATGACCA **R:**CGAGTTTAAGATGAATTTATAGCAAGC	58	126–128
gr4557	KY704077	$(CATTGA)_{9-11}$	**F:**CATCATGAAGGTCAGCCCGT **R:**AGATACCCTAATTAAGAGAGAAGGTG	58	106–118
gr4772	KY704078	$(GAGAGG)_{5-6}$	**F:**TACGGGGCGGTTAGGGATTA **R:**TGCTGTTCCAACAACCTCCT	58	108–114
gr6458	KY704079	$(CTC)_{9-11}$	**F:**GTGAGCGCGGACTTTTGC **R:**GCCACTACCGAACGCTTAGT	58	119–125
gr1031	KY704080	$(GAGAA)_{6-8}$	**F:**CATGGTTCCCCGTCTGGAAA **R:**TGTGGGGTGGGGAATGATTG	60	108–118
gr5502	KY704081	$(AGATA)_{5-6}$	**F:**AGCGGTGCAGAGTTTGATGA **R:**TTGTGGTCATTCGTTGGACA	60	104–109
gr2164	KY704082	$(TTAG)_{5-6}$	**F:**GTCCACAAGAAACCTCACCCT **R:**ATAACAAACTCTTGGCGGCG	60	130–134
gr1415	KY704083	$(CTG)_{6-7}$	**F:**ACCACCAGCTGCATCATGAA **R:**GGATTGGGAATGGCAATGGC	60	127–130
gr5371	KY704084	$(AGA)_{17-19}$	**F:**CCAAACGGGTTGCCTGAGAT **R:**CCCCCACACACCATCCTTTT	60	136–142

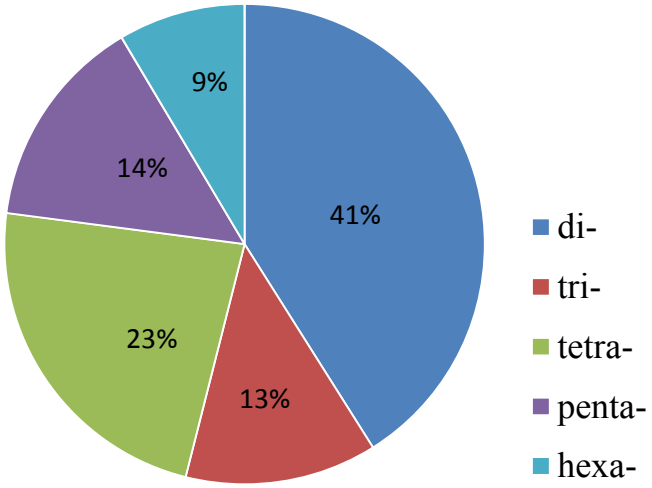

Fig. 1. The percentages of di-, tri-, tetra-, penta- and hexanucleotide repeats in SSR motif sequences of *Taxus florinii*.

pentanucleotides (439 sequences) and trinucleotides (394 sequences), while hexanucleotides represented the lowest frequency motifs (8.72%), found in only 261 sequences (Fig. 1). The dominant dinucleotide motifs were AG/CT repeats (40.15%) and the least common repeat was CG/CG (1.72%). Of the trinucleotides, the most frequent motifs were AGG/CCT repeats (19.80%), followed by AAG/CTT (19.54%); ACT/AGT had the lowest frequency (1.27%) (Fig. 2).

3.3. PCR validation of SSR primers

Among the 2993 consensus sequences containing microsatellites, a total of 526 contained polymorphic microsatellites in the ten individuals of the population TFP03 (Fig. 3). Of these, 237 were suitable for microsatellite primer design (listed in Appendix A.1 Supporting information).

A total of 128 primer pairs were randomly selected from the 237 designed primers for evaluating PCR amplification efficiency and polymorphism. Of the primers tested, 110 primers were excluded from further evaluation: 37 primers did not generate clear microsatellite peaks or failed to be amplified; 69 primers had a PCR amplification success rate of less than 80%; and according to the results of QDD *pipe2.pl*, four primers with length variations were not located in their microsatellite regions. The remaining 18 primers exhibited high amplification success, but two primer pairs (gr6734 and gr1679) did not generate the expected polymorphisms, with one (gr1679) being monomorphic in all 24 samples. Finally, 16 ideal polymorphic microsatellites were screened and characterized in 24 samples from two populations of *T. florinii*.

Of the 16 polymorphic SSR loci, seven were dinucleotides, with the remainder being trinucleotides (four), tetranucleotides (one), pentanucleotides (two) and hexanucleotides (two). All the 16 sequences have been uploaded to GenBank (KY704069–KY704084) (Table 3). The PCR amplification success rate for each locus ranged from 92% to 100%, and the number of alleles per locus (N_A) ranged from two to ten, with an average of 4.875 for the two studied populations (Table 3). The observed heterozygosity (H_O) and expected heterozygosity (H_E) varied from 0.083 to 1.000 (mean 0.578) and 0.080 to 0.778 (average 0.526) in population TFP03, and from zero to 1.000 (mean 0.547) and zero to 0.789 (average 0.534) in population TFP04, respectively. Three out of the 16 microsatellite loci (gr2662 in Population TFP03, gr6458 in Population TFP04 and gr4557 in both populations) exhibited significant deviation from Hardy–Weinberg equilibrium (Table 3). No significant pairwise linkage disequilibrium was detected between pairs of loci.

4. Discussion

In the present study, a total of 16 polymorphic microsatellite loci were validated and characterized for *T. florinii* using *Mi*ddRAD data. The levels of diversity observed at these microsatellite loci, measured as the number of alleles per locus (N_A), observed heterozygosity (H_O) and expected heterozygosity (H_E), are similar to those in previous studies (Dubreuil et al., 2008; Yang et al., 2009; Liu et al., 2011a). Two microsatellite loci significantly deviated from Hardy–Weinberg equilibrium, most likely due to the presence of null alleles. The 16 novel polymorphic microsatellite loci will be very useful for population genetics studies of the endangered species *T. florinii*. Comparisons between the results generated from SSRs and *Mi*ddRAD-seq will be very important for molecular marker selection in conservation genetics studies of yews in the future.

This is the first study to develop microsatellite loci using *Mi*ddRAD data. The results here indicate that *Mi*ddRAD technology can

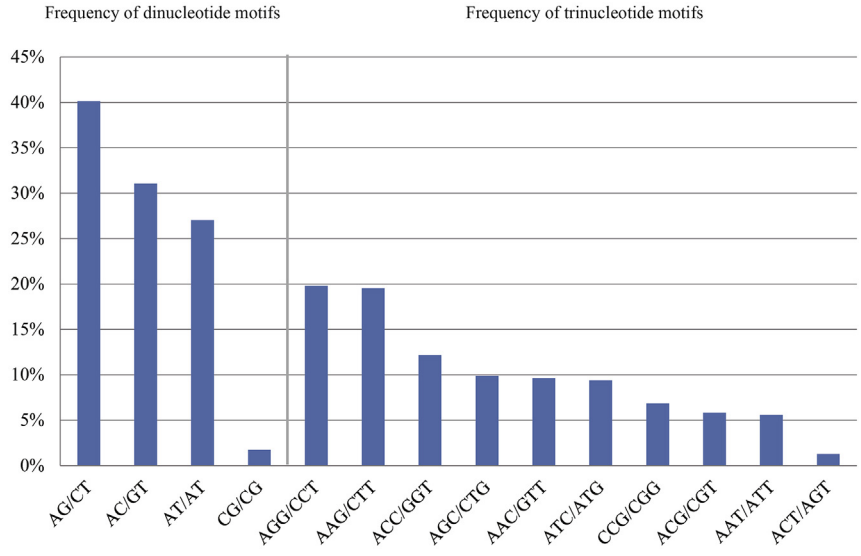

Fig. 2. Frequency of consensus sequences containing dinucleotide and trinucleotide motifs in *Taxus florinii*.

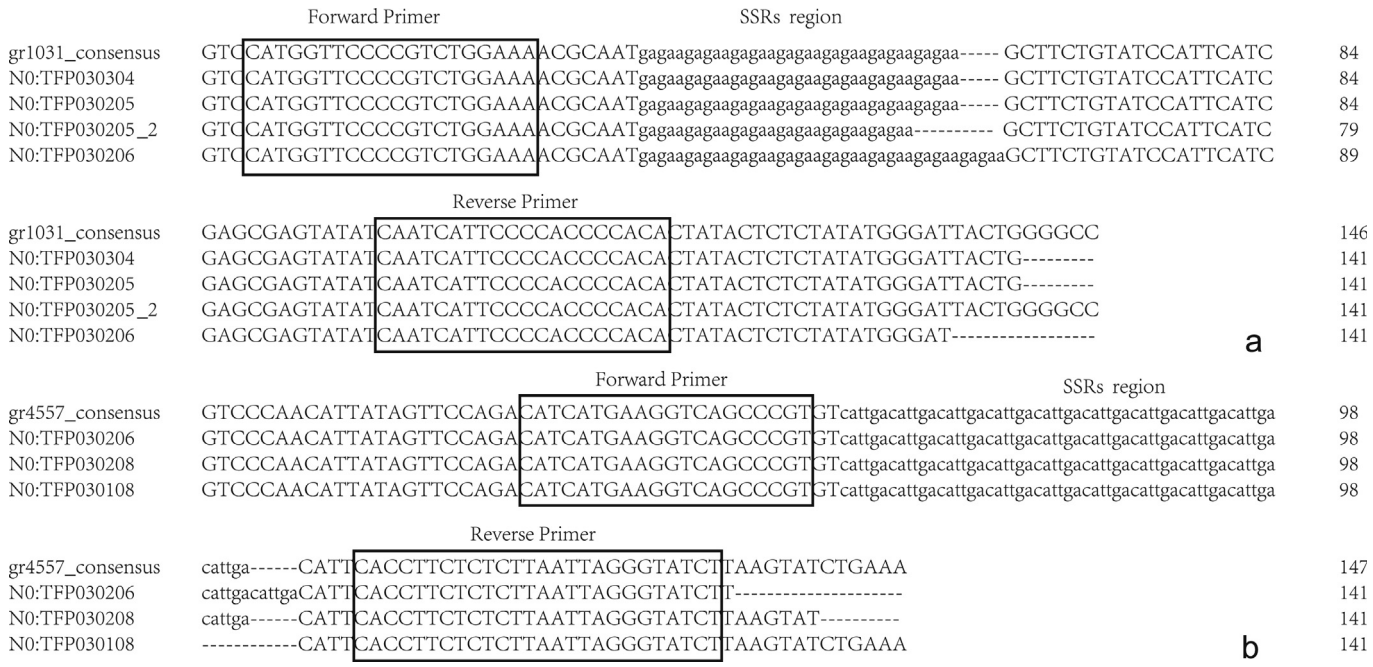

Fig. 3. The allelic diversity of microsatellite loci of *Taxus florinii* in silico, (a) allelic variation for primer pair gr1031, and (b) allelic variation for primer pair gr4557.

Table 3
Microsatellite loci analyzed on 24 individuals from two populations (TFP03 and TFP04) of *Taxus florinii*.

Locus	TFP03				TFP04				Total	
	N	N_A	H_O	H_E	N	N_A	H_O	H_E	N_A	PCR success
gr907	12	2	0.417	0.413	12	2	0.167	0.375	2	100%
gr1114	12	4	0.583	0.517	12	3	0.417	0.344	5	100%
gr1177	12	3	0.583	0.531	10	3	0.700	0.485	3	92%
gr1258	12	2	0.083	0.080	12	3	0.333	0.288	4	100%
gr1439	12	2	0.167	0.278	12	3	0.583	0.434	3	100%
gr2591	12	3	0.750	0.517	11	2	0.636	0.483	3	96%
gr2662	12	3	0.917	0.559[a]	12	3	0.833	0.569	3	100%
gr2680	12	5	0.917	0.684	11	6	1.000	0.789	6	96%
gr4557	12	6	0.333	0.778[a]	11	5	0.182	0.777[a]	7	96%
gr4772	12	4	0.667	0.656	12	5	0.750	0.733	5	100%
gr6458	12	5	0.667	0.628	12	8	0.167	0.743[a]	10	100%
gr1031	12	5	0.750	0.712	11	5	0.909	0.781	7	96%
gr5502	12	3	0.333	0.448	12	3	0.583	0.594	3	100%
gr2164	12	2	0.250	0.219	10	1	0.000	0.000	2	92%
gr1415	12	5	0.833	0.667	11	3	0.818	0.640	5	96%
gr5371	12	9	1.000	0.736	12	5	0.667	0.507	10	100%
Mean		3.938	0.578	0.526		3.75	0.547	0.534	4.875	98%

Number of individuals (N), number of alleles (N_A), and observed (H_O), expected heterozygosity (H_E), and PCR success. Total N_A indicates total number of alleles of the two populations.

[a] Indicates significant departure from Hardy–Weinberg equilibrium ($P < 0.05$).

offer great potential as an approach to isolate microsatellite markers. A total of 2993 consensus sequences were found containing microsatellites, which included various types of simple sequence repeat motifs, 526 (17.6%) of which were polymorphic among the ten individuals used in the ascertainment process. The relatively quick and simple ddRAD approach offers greater efficiency compared to traditional methods (e.g. Dubreuil et al., 2008; Yang et al., 2009), not only in terms of the isolation of the microsatellites themselves, but also in the identification of polymorphic loci, which traditionally would require additional fragment analysis (Fig. 3). This approach also allows discrimination between length polymorphism due to changes in repeat number within the microsatellite motif, as opposed to insertions and/or deletions in the flanking regions (Angers and Bernatchez, 1997; Anderson et al., 2000). Thus, the *Mi*ddRAD approach offers a quick, alternative method to develop and characterize microsatellite markers, particularly for non-model species.

Of the 128 primers tested, the PCR amplification success rate of 69 primers was less than 80%. This is probably due to the poor quality of primers. The short read lengths associated with the Illumina platform (average 141 bp) limits the amount of flanking sequence available for optimal primer design, which may be overcome to an extent by generating longer sequences. Therefore, the longer paired-end reads of the Illumina Miseq sequencing platform (2×300 bp) or the Roche 454 platform (350–600 bp) may offer greater potential in designing more optimal microsatellite primer pairs (Wei et al., 2014). A reference genome can provide homologous sequences upstream and downstream, which may also

promote the design efficiency of the microsatellite primers. For example, powerful software has been developed for microsatellite screening and primer design of genome-scale data, such as GMATA (Wang and Wang, 2016).

5. Conclusions

The *Mi*ddRAD data offers an effective approach to develop and characterize microsatellite markers for non-model species. The 16 novel polymorphic microsatellite loci for *T. florinii* characterized based on *Mi*ddRAD-seq data will be important for investigating genetic diversity and population structure, and these results in turn will provide crucial information for conservation and management of this endangered yew tree.

Acknowledgements

We are grateful to Jun-Bo Yang, Zhi-Rong Zhang and Lin-Jiang Ye to providing technical assistance of lab work. This research was funded by the National Natural Science Foundations of China (31370252, 41571059) and the National Key Basic Research Program of China (2014CB954100). Jie Liu was supported by the China Scholarship Council for one-year study at the Aberystwyth University, UK. Laboratory work was performed at the Laboratory of Molecular Biology at the Germplasm Bank of Wild Species, Kunming Institute of Botany, Chinese Academy of Sciences.

Appendix A. Supplementary data

Supplementary data related to this article can be found at http://dx.doi.org/10.1016/j.pld.2017.05.008.

References

Anderson, T.J.C., Su, X.Z., Roddam, A.W., et al., 2000. Complex mutations in a high proportion of microsatellite loci from the protozoan parasite *Plasmodium falciparum*. Mol. Ecol. 9, 1599–1608.
Angers, B., Bernatchez, L., 1997. Complex evolution of a salmonid microsatellite locus and its consequences in inferring allelic divergence from size information. Mol. Biol. Evol. 14, 230–238.
Baird, N.A., Etter, P.D., Atwood, T.S., et al., 2008. Rapid SNP discovery and genetic mapping using sequenced RAD markers. PLoS ONE 3, e3376.
Bonatelli, I.A., Carstens, B.C., Moraes, E.M., 2015. Using next generation RAD sequencing to isolate multispecies microsatellites for *Pilosocereus* (Cactaceae). PLoS ONE 10, e0142602.
Catchen, J., Hohenlohe, P.A., Bassham, S., et al., 2013. Stacks: an analysis tool set for population genomics. Mol. Ecol. 22, 3124–3140.
Davey, J.W., Hohenlohe, P.A., Etter, P.D., et al., 2011. Genome-wide genetic marker discovery and genotyping using next-generation sequencing. Nat. Rev. Genet. 12, 499–510.
Dubreuil, M., Sebastiani, F., Mayol, M., et al., 2008. Isolation and characterization of polymorphic nuclear microsatellite loci in *Taxus baccata* L. Conserv. Genet. 9, 1665–1668.
Elbers, J.P., Clostio, R.W., Taylor, S.S., 2016. Population genetic inferences using immune gene SNPs mirror patterns inferred by microsatellites. Mol. Ecol. Resour. http://dx.doi.org/10.1111/1755-0998.12591.
Elshire, R.J., Glaubitz, J.C., Sun, Q., et al., 2011. A robust, simple genotyping-by-sequencing (GBS) approach for high diversity species. PLoS ONE 6, e19379.
Gao, L.M., Möller, M., Zhang, X.M., et al., 2007. High variation and strong phylogeographic pattern among cpDNA haplotypes in *Taxus wallichiana* (Taxaceae) in China and North Vietnam. Mol. Ecol. 16, 4684–4698.
Guichoux, E., Lagache, L., Wagner, S., et al., 2011. Current trends in microsatellite genotyping. Mol. Ecol. Resour. 11, 591–611.
Hodel, R.G.J., Segovia-Salcedo, M.C., Landis, J.B., et al., 2016. The report of my death was an exaggeration: a review for researchers using microsatellites in the 21st century. Appl. Plant Sci. 4, 1600025.
Jansson, E., Taggart, J.B., Wehner, S., et al., 2016. Development of SNP and microsatellite markers for goldsinny wrasse (*Ctenolabrus rupestris*) from ddRAD sequencing data. Conserv. Genet. Resour. 8, 201–206.
Jarne, P., Lagoda, P.J., 1996. Microsatellites, from molecules to populations and back. Trends Ecol. Evol. 11, 424–429.
Jeffries, D.L., Copp, G.H., Lawson, H.L., et al., 2016. Comparing RADseq and microsatellites to infer complex phylogeographic patterns, an empirical perspective in the Crucian carp, *Carassius carassius*. L. Mol. Ecol. 25, 2997–3018.
Kingston, D.G., Newman, D.J., 2007. Taxoids: cancer-fighting compounds from nature. Curr. Opin. Drug Discov. Dev. 10, 130–144.
Leitch, I.J., Hanson, L., Winfield, M., et al., 2001. Nuclear DNA C-values complete familial representation in gymnosperms. Ann. Bot. 88, 843–849.
Li, Y.C., Korol, A.B., Fahima, T., et al., 2002. Microsatellites: genomic distribution, putative functions and mutational mechanisms: a review. Mol. Ecol. 11, 2453–2465.
Liu, J., Gao, L.M., 2011. Comparative analysis of three different methods of total DNA extraction used in *Taxus*. Guihaia 31, 244–249.
Liu, J., Gao, L.M., Li, D.Z., et al., 2011a. Cross-species amplification and development of new microsatellite loci for *Taxus wallichiana* (Taxaceae). Am. J. Bot. 98, e70–e73.
Liu, J., Moeller, M., GAO, L.M., et al., 2011b. DNA barcoding for the discrimination of Eurasian yews (*Taxus* L., Taxaceae) and the discovery of cryptic species. Mol. Ecol. Resour. 11, 89–100.
Liu, J., Moeller, M., Provan, J., et al., 2013. Geological and ecological factors drive cryptic speciation of yews in a biodiversity hotspot. New Phytol. 199, 1093–1108.
Ma, J.Q., Huang, L., Ma, C.L., et al., 2015. Large-scale SNP discovery and genotyping for constructing a high-density genetic map of tea plant using specific-locus amplified fragment sequencing (SLAF-seq). PLoS ONE 10, e0128798.
Meglécz, E., Costedoat, C., Dubut, V., et al., 2010. QDD: a user-friendly program to select microsatellite markers and design primers from large sequencing projects. Bioinformatics 26, 403–404.
Möller, M., Gao, L.M., Mill, R.R., et al., 2013. A multidisciplinary approach reveals hidden taxonomic diversity in the morphologically challenging *Taxus wallichiana* complex. Taxon 62, 1161–1177.
Peakall, R.O.D., Smouse, P.E., 2006. GENALEX 6: genetic analysis in Excel. Population genetic software for teaching and research. Mol. Ecol. Notes 6, 288–295.
Peterson, B.K., Weber, J.N., Kay, E.H., et al., 2012. Double digest RADseq: an inexpensive method for de novo SNP discovery and genotyping in model and non-model species. PLoS ONE 7, e37135.
Poudel, R.C., Gao, L.M., Möller, M., et al., 2013. Yews (*Taxus*) along the Hindu Kush-Himalayan region: exploring the ethnopharmacological relevance among communities of Mongol and Caucasian origins. J. Ethnopharmacol. 147, 190–203.
Rousset, F., 2008. genepop'007: a complete re-implementation of the genepop software for Windows and Linux. Mol. Ecol. Resour. 8, 103–106.
Schuelke, M., 2000. An economic method for the fluorescent labeling of PCR fragments. Nat. Biotechnol. 18, 233–234.
Selkoe, K.A., Toonen, R.J., 2006. Microsatellites for ecologists: a practical guide to using and evaluating microsatellite markers. Ecol. Lett. 9, 615–629.
Spjut, R.W., 2007. Taxonomy and nomenclature of *Taxus* (Taxaceae). J. Bot. Res. Inst. Tex. 1, 203–289.
Taylor, S.A., White, T.A., Hochachka, W.M., et al., 2014. Climate-mediated movement of an avian hybrid zone. Curr. Biol. 24, 671–676.
Vera, M., Díez-del-Molino, D., García-Marín, J.L., 2016. Genomic survey provides insights into the evolutionary changes that occurred during European expansion of the invasive mosquitofish (*Gambusia holbrooki*). Mol. Ecol. 25, 1089–1105.
Wang, X., Wang, L., 2016. GMATA: an integrated software package for genome-scale SSR mining, marker development and viewing. Front. Plant Sci. 7, 1350.
Wei, N., Bemmels, J.B., Dick, C.W., 2014. The effects of read length, quality and quantity on microsatellite discovery and primer development: from Illumina to PacBio. Mol. Ecol. Resour. 14, 953–965.
Yang, G.Q., Chen, Y.M., Wang, J.P., et al., 2016. Development of a universal and simplified ddRAD library preparation approach for SNP discovery and genotyping in angiosperm plants. Plant Meth. 12, 39.
Yang, J.B., Li, H.T., Li, D.Z., et al., 2009. Isolation and characterization of microsatellite markers in the endangered species *Taxus wallichiana* using the FIASCO method. HortScience 44, 2043–2045.
Zalapa, J.E., Cuevas, H., Zhu, H., et al., 2012. Using next-generation sequencing approaches to isolate simple sequence repeat (SSR) loci in the plant sciences. Am. J. Bot. 99, 193–208.