

RESEARCH

Open Access



Identification of SNPs in a nonmodel macrofungus (*Lepista nuda*, Basidiomycota) through RAD sequencing

Fei Ye², Xiao-Dan Yu^{1*}, Qing Wang³ and Peng Zhao⁴

Abstract

Lepista nuda is a wild edible fungus that is valued for its odor and taste. Recent studies identified intraspecific morphological and genetic differences in *L. nuda*. Although single-nucleotide polymorphisms (SNPs) are useful for revealing intraspecific differences, the traditional methods used for investigating SNPs are time consuming and expensive, and they only locate a limited number of SNPs. This study used a “restriction-site associated DNA” (RAD) method combined with high throughput sequencing to efficiently identify a large number of SNPs in two samples of *L. nuda*. A total of 7 and 9 billion bp of raw data were obtained from the two collections. A total of 712 SNPs were found. These SNPs will be useful for the further analysis of the genetic variation within *L. nuda*. The study also confirms that the RAD method can be used to identify SNPs in a nonmodel macrofungus for which a reference genome is unavailable.

Keywords: Agaricales, Fungi, Single-nucleotide polymorphisms, Restriction-site associated DNA

Background

Lepista nuda (Bull.) Cooke, which is in the Agaricales (Basidiomycota) (Kirk et al. 2008), is a wild, edible mushroom that is common in many parts of the world (Singer 1986). This fungus is popular because of its taste, smell, and nutritional qualities and is considered edible in Europe (Singer 1986), China (Dai et al. 2010), and elsewhere. Consequently, the understanding and conserving of natural populations of *L. nuda* has attracted significant attention from mushroom collectors, government agencies, and conservation groups.

Descriptions of *L. nuda* morphology vary considerably. The pileus of *L. nuda*, for example, was described as grey brown or russet brown by some researchers but was described as purple brown by the other (Bon 1987; Hansen and Knudsen 1992; Mao 2000). Molecular data for the species also reveal substantial intraspecific differences. Thus, the analysis of ITS sequences showed that two collections of *L. nuda* did not group together

(Moncalvo et al. 2002). Using amplification polymorphism fragments (CAPS) and random amplified polymorphisms (RAPD), the results showed that the *L. nuda* from Greece and the United States differ from the *L. nuda* in France and Australia (Stott et al. 2005). To date, few reports have considered the intraspecific molecular variability of *L. nuda*, probably because the molecular methods available are time consuming and costly. Therefore, rapid and inexpensive methods are needed to clarify the morphological and genetic variation within *L. nuda*.

Single-nucleotide polymorphisms (SNPs) refer to variations in a single base pair of a DNA sequence. SNPs, which can function as useful markers for the study on population genetic, consist of unlinked loci that occur throughout the genome and that have relatively low mutation rates (Brumbfield et al. 2003). Generally, SNPs were developed mainly by DNA sequencing. To identify SNPs in *Tricholoma matsutake*, a genomic library was constructed and 73,065 bp were sequenced from random clones. Special primers from 20 sequenced fragments were then designed to amplify and analyzed more than 10,428 bp sequences from the two strains. Finally, a total of 178 SNPs were developed (Xu et al. 2007). There were only four SNPs from seven *Armillaria cepistipes* isolates

*Correspondence: yuxd126@126.com

¹ College of Biological Science and Technology, Shenyang Agricultural University, Shenyang 110866, Liaoning, People's Republic of China
Full list of author information is available at the end of the article

by sequencing the regions of ten single-copy protein-coding homologues and the housekeeping gene EF1- α (Heinzelmann et al. 2012). Therefore, the traditional method for detecting SNPs is time consuming and expensive and detects only a limited numbers of SNPs.

An efficient method for identifying SNP loci combines “restriction-site associated DNA” (RAD) with high throughput sequencing (Miller et al. 2007; van Tassel et al. 2008). The advantages of this method include: (1) the number of SNPs identified is ten-times greater than with the traditional technology; (2) the data utilization rate is high, and the cost of sequencing is relatively low; (3) the time and work required are less than with the traditional method; and (4) the method can be used for species that lack a reference genome. RAD techniques have been widely used to find SNP loci in animals and plants (Bourgeois et al. 2013; Lamer et al. 2014; Yu et al. 2015; Wang et al. 2013; Zhao et al. 2014; Xiao et al. 2015). For fungi, the RAD method has been used for *Laccaria bicolor* and for the plant-pathogenic fungi *Pyrenophora teres* and *Sphaerulina musiva* (Wilson et al. 2015; Leboldus et al. 2015). In this study, the RAD method was combined with Illumina sequencing to discover the SNPs in *L. nuda*. These SNPs could be used for further research concerning the population genetics within *L. nuda*.

Results and discussion

Sample identification

Two ITS sequences from the dried basidiomata of the two *L. nuda* specimens (HMAS 254481 and HMAS 254482) were obtained in this study. The sequences were submitted to GenBank: the GenBank accession numbers are KU215618 and KU215619. To assess the taxonomic status of these specimens, the ITS sequences obtained from this study were compared with the sequences in GenBank by a BLAST database search (Altschul et al. 1997). The results showed >99 % identity between the sequences obtained from this study and the sequences named “*L. nuda*” in GenBank. Given these molecular characteristics and the morphological characteristics of the specimens, we conclude that the two specimens were *L. nuda*.

Base calling

In this study, the RAD method was used to find SNPs in *L. nuda*. The distribution of bases is presented in Fig. 1, and the distribution of base quality is presented in Fig. 2. The base composition was similar among all of the reads, and the percentage of N was very low (Fig. 1). The base quality, which reflects the error rate of sequencing, was high for both samples (Fig. 2). Quality declines during sequencing as the activity of enzymes and the amount of reagent decline. Because the activity of enzymes and

the amount of reagent decline during sequencing, quality declines once a certain sequence length is attained (Wang et al. 2014). The base quality is also affected by the sequencing machine, the reagent, and the samples used. Statistical analysis indicated that the Q20 (indicating a 1 % sequencing error) of each sample was >92 %, and that the Q30 (indicating a 0.1 % sequencing error) of each sample was >84 % (Table 1). The mean GC content of the sequence in the two populations was about 46 %. A total of 7 billion bp of raw data was obtained from one sample and 9 billion from the other. Nine million or eight million clean reads were obtained by trimming the reads in the poor quality (quality value <Q20 and N > 10 %). The clean data was calculated by all of nucleotides of the clean reads without the adapter parts. The same reads cluster into a RAD-tag. There were about 170,000 RAD-tags in one sample and 140,000 in the other. The average depth was about 45 \times for one sample and 48 \times for the other. The overall sequencing depth in the current study was high. The redundancy of the two samples was relatively high (Table 1). There were two possible reasons why the redundancy was high. One possible reason was that all the reads were sequenced in forward direction. The other reason was that the size of genome of *L. nuda* is relatively small (about 70 Mb).

SNP calling

The RAD-tag depth distribution is shown in Fig. 3. In order to discharging the sequencing errors, the SNPs <6 \times were removed. A total of 712 SNPs were identified from the RAD-tags of the two samples. The SNPs distribution in the depth of reads is listed in Table 2. The numbers of SNPs were more than the numbers of SNPs obtained from the studies by the traditional methods (Xu et al. 2007; Heinzelmann et al. 2012). Although Wilson et al. (2015) obtained a high number of SNPs (17,854) from *L. bicolor* samples using the RAD method, the samples included both ingroup and outgroup specimens. Therefore, the number of informative SNP markers obtained for *L. bicolor* ranged from about 322 to 1000. In this paper, 712 SNP loci were obtained from two *L. nuda* samples using the RAD method. This number of SNP loci is sufficient to support further study of the genetic variation of *L. nuda*. The results of this study support that the RAD method was useful for identifying SNP loci in species for which genomic information is lacking.

Conclusions

This study used RAD method combined with high throughput sequencing to identify a total of 712 SNPs in two samples of *L. nuda*. These SNPs will be used to examine the population genetics of *L. nuda*. This paper also confirms that the RAD method can be used to

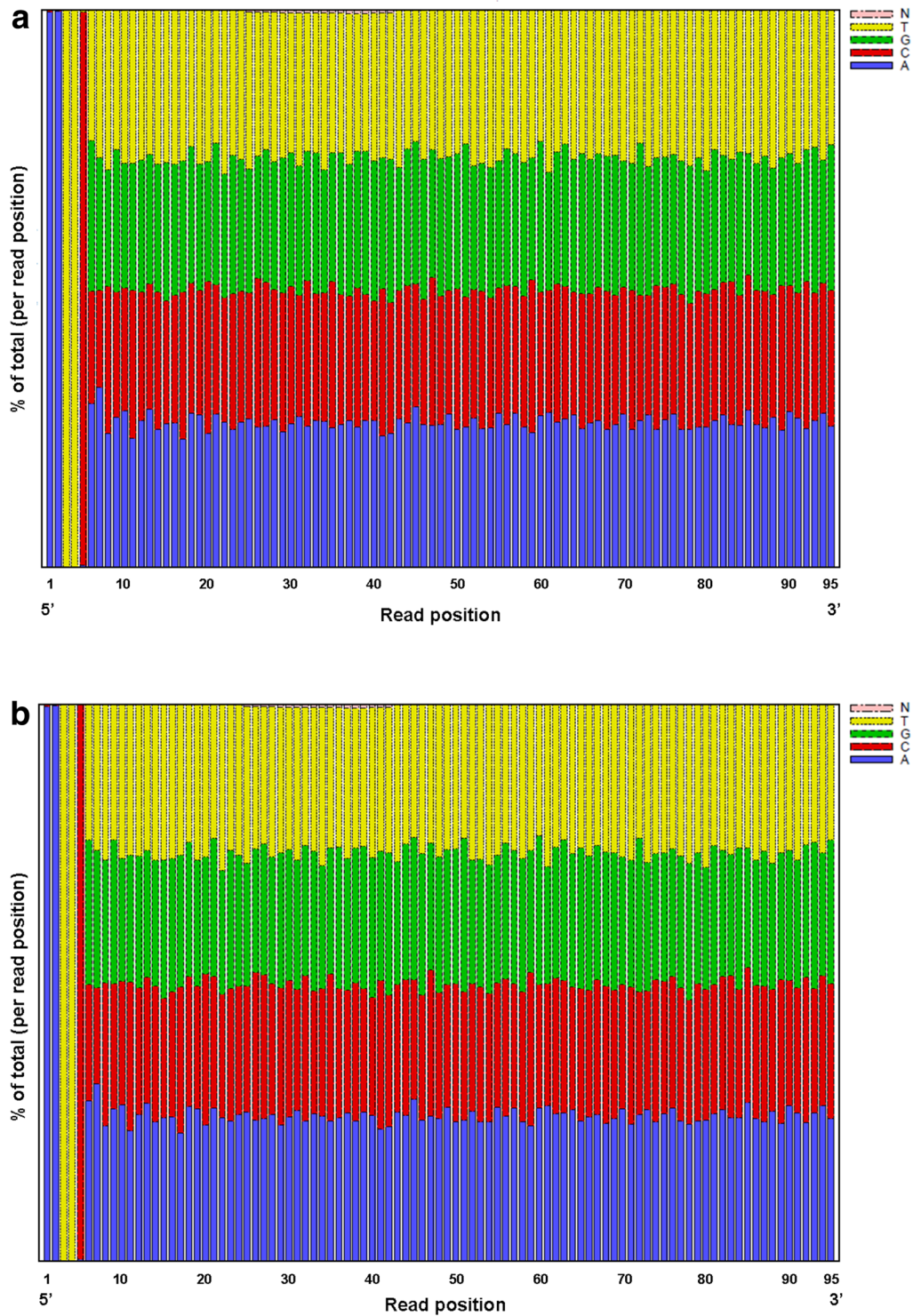


Fig. 1 a, b The distribution of bases in two *L. nuda* collections. The first five bases are the restriction enzymes loci. **a** HMAS 254481, **b** HMAS 254482

identify SNPs in a nonmodel macrofungus for which a reference genome is unavailable. Furthermore, the SNPs could provide the theoretical support for further research

about cultivation and breeding of *L. nuda*. It is important for the protection of resources and the conservation of the population genetics of *L. nuda* in the field.

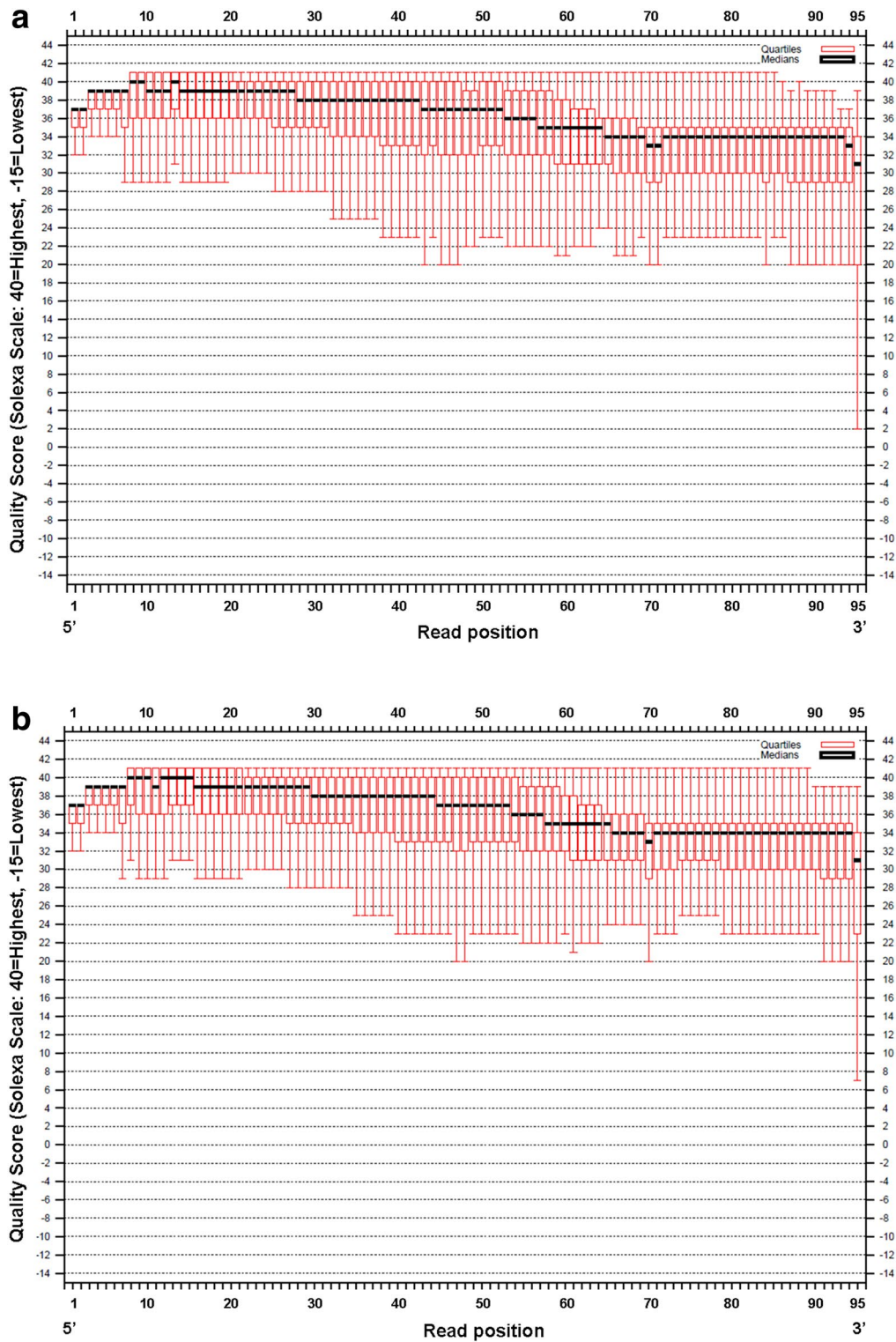


Fig. 2 a, b The distribution of base quality in the two *L. nuda* collections. The red vertical line represents the base quality of reads; the red vertical square represents the quartile of the base quality; the thick black line represents the median of the quartile. **a** HMAS 254481, **b** HMAS 254482

Table 1 Statistics for reads of two *L. nuda* samples

Collections	HMAS 254481	HMAS 254482
Read length	95 SE	95 SE
Raw data	913,113,400	777,371,035
Clean data	910,348,330	775,132,075
Number of clean reads	9,582,614	8,159,285
Number of RAD-tags	176,688	141,777
Average depth	45.8628	48.8587
Q20 (%)	92.92	93.38
Q30 (%)	83.64	84.49
GC content (%)	46.1	45.87
Redundancy (%)	26.8	22.1

SE single end

Methods

Sample identification

Fungi used in this study were collected from Tianzhu Mountain, Shenyang City, Liaoning Province and Muleng town, Mudanjiang City, Heilongjiang Province of China (The locations are public areas. Therefore, there are no specific permissions were required for the locations. The authors confirm that the field studies did not involve endangered or protected species.). The specimens were dried with an electric air-ventilation drier and were deposited in the Mycological Herbarium of the Chinese Academy of Sciences (HMAS) with accession numbers of HMAS 254481 and HMAS 254482. Genomic DNA was extracted from the dried blocks of tissue of the herbarium specimens (Table 3) using the Plant DNA Extraction Kit (Sunbiotech Co., Ltd., Beijing, China) and following the manufacturer's instructions. The crude DNA extracts were used as templates for PCR. Primers ITS5/ITS4

(GGAAGTAAAAGTCGTAACAAGG/TCCTCCGCT-TATTGATATGC) were used for amplification of the ITS region including ITS1, 5.8S, and ITS2 (White et al. 1990). Reaction mixtures and PCR conditions were as described by the previous study (Yu et al. 2014). At the same time, the 16S region was amplified by the primers 27F/1492R (AGAGTTTGTATCMTGGCTCAG/TACGGYTAC-CTTGTTACGACTT) for confirming the absence of non-fungal DNA (Lane 1991). The PCR products were checked on a 1 % agarose gel and visualized by staining with ethidium bromide. Sequencing was performed on an ABI Prism[®] 3730 Genetic Analyzer (PE Applied Biosystems, Foster, CA, USA). Nucleotide sequences of the ITS regions that were amplified from the collections (HMAS 254481 and HMAS 254482) were aligned with the sequences of *L. nuda* ITS regions retrieved from GenBank using BioEdit 5.0.6 (Hall 1999) and Clustal X (Thompson et al. 1997).

Creation and sequencing of the RAD library

For construction of the RAD library, the DNA from the two samples was pooled. The sample indexing and pooling were followed by the published method (Baird et al. 2008). Because a reference genome for *L. nuda* is unavailable, the genomic DNA was first digested with *EcoRI* (NEB Company). The digestion was performed at 37 °C overnight; digestion ended with a 20-min deactivation step at 65 °C followed by cooling to 4 °C and the enzyme-digested product was pooled. The fragments of digested DNA were connected to the P1 adapter using T4 DNA ligase (NEB company). The sequences of the P1 adapter include the *EcoRI* restriction site, a reverse amplification site, and the Illumina sequencing primer

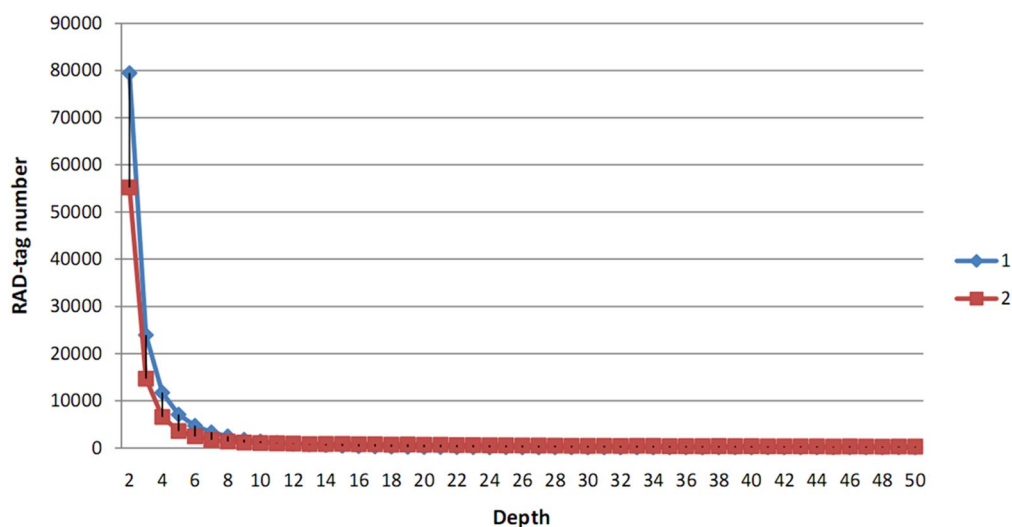
**Fig. 3** RAD-tag depth distribution in two *L. nuda* collections. 1 HMAS 254481, 2 HMAS 254482

Table 2 The correspondence between the depth of reads and the proportion of SNPs

The depth of reads (x)	The proportion of SNPs (%)
6–10	4.2
10–20	14.5
20–30	13.7
30–40	16.9
40–50	17.4
Above 50	33.3

Table 3 Collections of *L. nuda* used for the analysis

Taxa	Voucher collections	Origin	GenBank accession no.
<i>L. nuda</i>	HMAS 254481	Tianzhu Mountain, Shenyang City Liaoning Province, China	KU215618
<i>L. nuda</i>	HMAS 254482	Muleng Town, Mudanjiang City Heilongjiang Province, China	KU215619

sites. Subsequently, the fragments were physically broken using a Covaris ultrasonicator with 200 bursts of 90 s each on high power, after which the resulting fragments were confirmed to be 300–500 bp in length by agarose gel electrophoresis. Then, a P2 adapter was connected to the pieces (300–500 bp). The sequences of the P2 adapter include the forward amplification and Illumina sequencing primer sites. After PCR amplification, only the fragments that included the P1 and P2 adapters were screened. Single-end (101 bp, including 6 bp for barcode) sequencing was performed using the Illumina HiSeq2000 in a total throughput of 16 lanes (Shanghai Majorbio Bio-pharm Technology Co., Ltd.).

SNP calling

A total of 913,113,400 bp of raw data was obtained from one of the samples, and a total of 777,371,035 bp of raw data was obtained from the other. After processing, totals of 910,348,330 and 775,132,075 bp of clean data were derived from the processed raw data. The process was completed by SeqPrep and included removing the adapter parts, trimming the nucleotides that had a quality value <Q20, eliminating those reads in which N was >10 %. We distinguished the trimmed reads of the two samples according to the barcode, and the trimmed reads were clustered into read tags (hereafter referred to as RAD-tags) by sequence similarity using ustacks (Catchen et al. 2011) to produce unique candidate alleles for each RAD locus. A maximum base-pair mismatch of two was allowed in this step. RAD-tags were then collapsed into

clusters using ustacks under default parameters for SNP calling.

Abbreviations

CAPS: amplification polymorphism fragments; EF1- α : elongation factors 1- α ; HMAS: the Mycological Herbarium of the Chinese Academy of Sciences; RAD: restriction-site associated DNA; RAPD: random amplified polymorphisms; SNPs: single-nucleotide polymorphisms.

Authors' contributions

Conceived and designed the experiments: X-DY. Performed the experiments: FY. Analyzed the data: QW. Contributed reagents/materials/analysis tools: PZ. Contributed to the writing of the manuscript: FY, X-DY. All authors read and approved the final manuscript.

Author details

¹ College of Biological Science and Technology, Shenyang Agricultural University, Shenyang 110866, Liaoning, People's Republic of China. ² College of Jilin Agricultural Science and Technology, Jilin, People's Republic of China. ³ Provincial Key Laboratory of Forest Protection, Liaoning Academy of Forestry, Shenyang, Liaoning, People's Republic of China. ⁴ Key Laboratory of Shandong Province for Edible Mushroom Technology, Institute of Mycological Science and Technology, Ludong University, Yantai, Shandong, People's Republic of China.

Acknowledgements

We are thankful to Prof. Yijian Yao for allowing the specimens cited to be kept in the Mycological Herbarium of the Chinese Academy of Sciences. We would also like to thank Shanghai Majorbio Bio-pharm Technology Co., Ltd. for assistance with Illumina sequencing. Prof. Bruce Jaffee is acknowledged for his work in kindly correcting the English sentences.

Competing interests

The authors declared that they have no competing interests.

Availability of data and materials

The datasets supporting the conclusions of this article are available in the GenBank.

Funding

This study was supported by the National Natural Science Foundation of China (Nos. 31200011, 31100299, 31300014 and 31400101).

Received: 1 June 2016 Accepted: 4 October 2016

Published online: 13 October 2016

References

- Altschul SF, Madden TL, Schäffer AA, Zhang J, Zhang Z (1997) Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res* 25:3389–3402
- Baird NA, Etter PD, Atwood TS, Currey MC, Shiver AL, Lewis ZA, Selker EU, Cresko WA, Johnson EA (2008) Rapid SNP discovery and genetic mapping using sequenced RAD markers. *PLoS One* 3:e3376
- Bon M (1987) The mushrooms and toadstools of Britain and North-western Europe. Hodder & Stoughton, London
- Bourgeois YX, Lhuillier E, Cézard T, Bertrand JA, Delahaie B, Cornuault J, Duval T, Bouchez O, Milá B, Thébaud C (2013) Mass production of SNP markers in a nonmodel passerine bird through RAD sequencing and contig mapping to the zebra finch genome. *Mol Ecol Resour* 13:899–907
- Brumfield RT, Beerli P, Nickerson DA, Edwards SV (2003) The utility of single nucleotide polymorphisms in inferences of population history. *Trends Ecol Evol* 18:249–256
- Catchen JM, Amores A, Hohenlohe P, Cresko W, Postlethwait JH (2011) Stacks: building and genotyping Loci de novo from short-read sequences. *G3* 1:171–182

- Dai YC, Zhou LW, Yang ZL, Wen HA, Bau T, Li TH (2010) A revised checklist of edible fungi in China. *Mycosystema* 29:1–29 **(in Chinese)**
- Hall TA (1999) BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. *Nucleic Acids Symp Ser* 41:95–98
- Hansen L, Knudsen H (1992) Nordic macromycetes. Polyporales, Boletales, Agaricales, Russulales, vol 2. Nordsvamp, Copenhagen
- Heinzelmann R, Rigling D, Prospero S (2012) Population genetics of the wood-rotting basidiomycete *Armillaria cepistipes* in a fragmented forest landscape. *Fungal Biol* 116:985–994
- Kirk PM, Cannon PF, Minter DW, Stalpers JA (2008) Ainsworth & Bisby's dictionary of the fungi, 10th edn. CAB International, Wallingford
- Lamer JT, Sass GG, Boone JQ, Arbieva ZH, Green SJ, Epifanio JM (2014) Restriction site-associated DNA sequencing generates high-quality single nucleotide polymorphisms for assessing hybridization between bighead and silver carp in the United States and China. *Mol Ecol Resour* 14:79–86
- Lane DJ (1991) 16S/23S rRNA sequencing. In: Stackebrandt E, Goodfellow M (eds) *Nucleic acid techniques in bacterial systematics*. Wiley, Chichester, pp 115–175
- Leboldus JM, Kinzer K, Richards J, Ya Z, Yan C, Friesen TL, Brueggeman R (2015) Genotype-by-sequencing of the plant-pathogenic fungi *Pyrenophora teres* and *Sphaerulina musiva* utilizing ion torrent sequence technology. *Mol Plant Pathol* 16:623–632
- Mao XL (2000) The macrofungi in China. Henan Science Technology Press, Zhengzhou **(in Chinese)**
- Miller MR, Dunham JP, Amores A, Cresko WA, Johnson EA (2007) Rapid and cost-effective polymorphism identification and genotyping using restriction site associated DNA (RAD) markers. *Genome Res* 17:240–248
- Moncalvo JM, Vilgalys R, Redhead SA, Johnson JE, James TY et al (2002) One hundred and seventeen clades of euagarics. *Mol Phylogenet Evol* 23:357–400
- Singer R (1986) The Agaricales in modern taxonomy, 4th edn. Koeltz Scientific Books, Koenigstein
- Stott K, Desmerger C, Holford P (2005) Relationship among *Lepista* species determined by CAPS and RAPD. *Mycol Res* 109:205–211
- Thompson JD, Gibson TJ, Plewniak F, Jeanmougin F, Higgins DG (1997) The ClustalX windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools. *Nucleic Acids Res* 24:4876–4882
- van Tassel CP, Smith TP, Matukumalli LK, Taylor JF, Schnabel RD, Lawley CT, Haudenschild CD, Moore SS, Warren WC, Sonstegard TS (2008) SNP discovery and allele frequency estimation by deep sequencing of reduced representation libraries. *Nat Methods* 5:247–252
- Wang XQ, Zhao L, Eaton DAR, Li DZ, Guo ZH (2013) Identification of SNP markers for inferring phylogeny in temperate bamboos (Poaceae: Bambusoideae) using RAD sequencing. *Mol Ecol Resour* 13:938–945
- Wang YK, Hu Y, Zhang TZ (2014) Current status and perspective of RAD-seq in genomic research. *Hereditas* 36:41–49
- White TJ, Bruns T, Lee S, Taylor J (1990) Amplification and direct sequencing of fungal ribosomal RNA genes from phylogenetics. In: Innes MA, Gelfand DH, Sninsky JS, White TJ (eds) *PCR protocols: methods and applications*. Academic Press, London, pp 315–322
- Wilson AW, Wickett NJ, Grabowski P, Fant J, Borevitz J, Mueller GM (2015) Examining the efficacy of a genotyping-by-sequencing technique for population genetic analysis of the mushroom *Laccaria bicolor* and evaluating whether a reference genome is necessary to assess homology. *Mycologia* 107:217–226
- Xiao B, Tan Y, Long N, Chen X, Tong Z, Dong Y, Li Y (2015) SNP-based genetic linkage map of tobacco (*Nicotiana tabacum* L.) using next-generation RAD sequencing. *J Biol Res* 22:11
- Xu JP, Guo H, Yang ZL (2007) Single nucleotide polymorphisms in the ectomycorrhizal mushroom *Tricholoma matsutake*. *Microbiology* 153:2002–2012
- Yu XD, Lv SX, Ma D, Li FF, Lin Y, Zhang L (2014) Two new species of *Melanoleuca* (Agaricales, Basidiomycota) from northeastern China, supported by morphological and molecular data. *Mycoscience* 55:456–461
- Yu S, Chu W, Zhang L, Han H, Zhao R, Wu W, Zhu J, Dodson MV, Wei W, Liu H, Chen J (2015) Identification of laying-related SNP markers in geese using RAD sequencing. *PLoS One* 10:e0131572
- Zhao J, Jian J, Liu G, Wang J, Lin M, Ming Y, Liu Z, Chen Y, Liu X, Liu M (2014) Rapid SNP discovery and a RAD-based high-density linkage map in jujube (*Ziziphus* Mill.). *PLoS One* 9:e109850

Submit your manuscript to a SpringerOpen® journal and benefit from:

- Convenient online submission
- Rigorous peer review
- Immediate publication on acceptance
- Open access: articles freely available online
- High visibility within the field
- Retaining the copyright to your article

Submit your next manuscript at ► springeropen.com