ISSN 1479-2621
Plant Genetic Resources: Characterization and Utilization (2018) **16**(4); 386–389
doi:10.1017/S1479262117000211

Short Communication

Genome-wide SNPs detection by genomic comparison between Dongxiang wild rice (*Oryza rufipogon*) and cultivated rice Nipponbare (*Oryza sativa* ssp. *japonica*)

Meng Zhang, Wenyang Huang, Zijun Xia, Jiahui Liu, Jiankun Xie* and Fantao Zhang*

College of Life Sciences, Jiangxi Normal University, Nanchang 330022, People's Republic of China

Received 22 January 2017; Accepted 16 June 2017 – First published online 26 July 2017

Abstract

Dongxiang wild rice (*Oryza rufipogon*, DXWR) exhibits valuable agronomic traits and represents a precious germplasm resource for rice breeding. The use of genetic markers can greatly speed up the breeding process and facilitate research on genetics and genomics. In our previous study, we identified insertion–deletion polymorphisms between DXWR and cultivated rice Nipponbare (*Oryza sativa* ssp. *japonica*), using whole-genome sequencing in DXWR. In this study, to further explore the genetic variations and enrich the available genetic markers of DXWR, we identified 1,089,478 single-nucleotide polymorphisms (SNPs) (corresponding to one SNP per 0.33 kb of the genome) by genomic comparison between DXWR and Nipponbare, using the genome sequencing data and bioinformatics approaches. Furthermore, the accuracy of the identified SNPs was also validated by polymerase chain reaction amplification and Sanger sequencing. This genome-wide SNPs identification greatly increases the number of genetic markers available for DXWR and provides new opportunities to exploit this valuable and endangered germplasm resource.

Keywords: genetic marker, germplasm resource, rice, single-nucleotide polymorphism, wild species

Introduction

Wild rice (*Oryza rufipogon*) is considered the ancestor of cultivated rice (*Oryza sativa*) and has been recognized as a precious germplasm resource for rice breeding (Sang and Ge, 2013; Atwell *et al*., 2014). Dongxiang wild rice (DXWR), a strain of *O. rufipogon* found in Dongxiang County, China, has the most northerly habitat (28°14′N) among wild rice populations worldwide (Xie *et al*., 2010; Mao *et al*., 2015). However, rapid urbanization around the habitat of DXWR has almost led to its extinction (Hu *et al*., 2012). Therefore, it should be effectively protected and exploited.

The discovery of genetic variations and development of genetic markers will facilitate research on rice genetics and genomics, and the exploration and utilization of elite gene resources in rice breeding. Among all genetic variations, single-nucleotide polymorphisms (SNPs) are the most abundant class of polymorphisms in the genome, and it is common for several SNPs to occur within a few hundred base pairs, which can enable a distinction to be made between highly similar cultivars (Ayed *et al*., 2014; Gurgul

*Corresponding author. E-mail: zhang84004@163.com; xiejiankun11@163.com

et al., 2014). Furthermore, compared with other common genetic markers, such as simple sequence repeats and insertion–deletion markers, SNP analysis can be automated in high-throughput assays without the need for DNA separation by size (Wang *et al.*, 2016). Therefore, SNPs have increasingly become the markers of choice for accurate genotype identification, diversity analysis, and genome-wide association studies in many species (Li *et al.*, 2015; Shavrukov, 2016). However, information available on SNPs in DXWR remains limited. The main objectives of this study were (1) to use genome sequencing data and bioinformatics approaches to detect genome-wide SNPs in DXWR and (2) to validate the reliability of the detected SNPs by polymerase chain reaction (PCR)-based Sanger sequencing.

Experimental

In this study, we used data from our previous study regarding whole-genome sequencing, reads filtering, and mapping to the published rice (Nipponbare) reference genome (http://rice.plantbiology.msu.edu/) (Kawahara *et al.*, 2013). The sequencing data of DXWR are deposited in NCBI with the accession number SRA167397 (Zhang *et al.*, 2015). SNP loci in the consensus sequence were detected by comparing with the reference sequence and then filtered using specific requirements (quality value >20 and the result supported by at least two reads) using SoapSNP (Chagné *et al.*, 2012). To compare the detected SNPs with annotated gene structures in the rice genome, physical positions of DXWR sequences were integrated into the annotated RAP2 Nipponbare full-length cDNA database (Ohyanagi *et al.*, 2006) using Generic Genome Browser software (http://gmod.org/wiki/GBrowse), as described by Stein *et al.* (2002).

Genomic DNA was extracted using the CTAB method (Porebski *et al.*, 1997). Based on the Nipponbare reference genome sequence, PCR primers specific for the sequences flanking the SNPs were designed. The details of the primers were presented in online Supplementary Table S1. PCR amplification was performed using the PrimeSTAR® Max DNA Polymerase Kit (Takara, Dalian, China). Target DNA fragments were cut from the agarose gel and purified using a DNA Gel Extraction Kit (Tiangen, Beijing, China), and then were sequenced at Shanghai Invitrogen Biotech (Shanghai, China). Sequence data were aligned using the publicly available programs Clustal X (http://www.clustal.org/) and BioEdit (http://www.mbio.ncsu.edu/bioedit/bioedit.html).

Discussion

In our previous study, whole-genome sequencing of DXWR was performed and 282,383,842 paired-end reads (90 bp per read) were obtained after filtering, and approximately 228 million reads were uniquely aligned with the reference genome of Nipponbare. An overall effective depth of 55× coverage was achieved, and the resulting consensus sequence was 357,407,193 bp, with 95.76% coverage of the Nipponbare genome (Zhang *et al.*, 2015).

In this study, 1,089,478 SNPs were detected by genomic comparison between DXWR and Nipponbare (Table 1 and online Supplementary Table S2). The genome-wide average SNP density was one SNP per 0.33 kb, whereas the density varied among the chromosomes, from one SNP per 0.28

Table 1. Densities of SNPs on individual chromosomes detected between DXWR and Nipponbare genomes

Chromosome	Reference size (bp)	DXWR aligned length (bp)	Number of SNPs	Average[a]
1	43,270,923	41,887,727	134,280	0.31
2	35,937,250	34,943,035	97,240	0.36
3	36,413,819	35,125,571	105,197	0.33
4	35,502,694	33,825,642	97,659	0.35
5	29,958,434	29,110,085	85,416	0.34
6	31,248,787	29,591,680	94,306	0.31
7	29,697,621	27,397,024	86,503	0.32
8	28,443,022	27,778,151	67,233	0.41
9	23,012,720	22,166,476	71,735	0.31
10	23,207,287	22,252,558	78,633	0.28
11	29,021,106	27,100,318	85,629	0.32
12	27,531,856	26,228,926	85,647	0.31
Total	373,245,519	357,407,193	1,089,478	0.33

[a]The 'Average' column is a SNP density in DXWR, and equals to aligned length (bp)/number of SNPs/1000.

https://doi.org/10.1017/S1479262117000211 Published online by Cambridge University Press

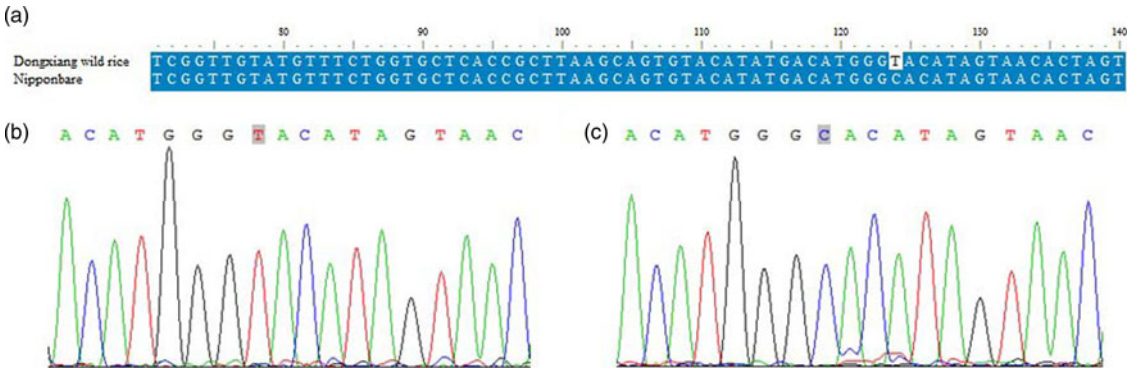

Fig. 1. Validation of the SNP locus (chromosome 5/19727118) and genotyping between DXWR and the cultivated rice Nipponbare by PCR-based Sanger sequencing. (a) SNP locus and its flanking sequences obtained from Sanger sequencing were aligned between DXWR and the cultivated rice Nipponbare. (b) Confirmation of the genotype (*T*) of the SNP locus in DXWR by Sanger sequencing. (c) Confirmation of the genotype (*C*) of the SNP locus in the cultivated rice Nipponbare by Sanger sequencing.

kb (chromosome 10) to one SNP per 0.41 kb (chromosome 8) (Table 1). Previously, Feltus *et al.* (2004) reported a genome-wide average SNP density of one SNP per 0.93 kb between 9311 (*O. sativa* ssp. *indica*) and Nipponbare, and Arai-Kichise *et al.* (2011) reported a genome-wide average SNP density of one SNP per 2.9 kb between Omachi (*O. sativa* ssp. *japonica*) and Nipponbare. The high density of SNPs between DXWR and Nipponbare may reflect high sequence divergence among the diverse combinations used in our study. Furthermore, 166,903 SNPs were found to occur in the coding regions of 39,283 genes. Among these SNPs, 63,338 SNPs in 25,863 genes were synonymous, whereas 103,565 SNPs in 32,797 genes were non-synonymous, and 19,377 genes exhibited both synonymous and non-synonymous changes. The non-synonymous amino acid substitutions might be responsible for a change in the functional activity of the gene products.

To assess the accuracy of SNPs detection by high-throughput sequencing, we randomly selected five SNPs to conduct PCR amplification and Sanger sequencing. A total of 10 PCR products were successfully amplified and sequenced between DXWR and the cultivated rice Nipponbare (online Supplementary Table S3). The 10 SNPs detected by high-throughput sequencing were all confirmed by PCR-based Sanger sequencing analysis (Fig. 1 and online Supplementary Table S3), indicating a high accuracy of SNPs detection in this study.

In conclusion, genome-wide SNPs were identified between DXWR and cultivated rice Nipponbare, and their accuracy was confirmed by PCR-based Sanger sequencing. The SNP markers identified in this study will provide a valuable resource for future research on genomics, evolution, genetic linkage, resource assessment, and genome-wide association in DXWR.

Supplementary material

The supplementary material for this article can be found at https://doi.org/10.1017/S1479262117000211.

Acknowledgements

This research was partially supported by the National Natural Science Foundation of China (31660386 and 31360327), the Key Projects of Jiangxi Education Department (KJLD12059), the Major Projects in Jiangxi Province (20161ACF60022), and the Sponsored Program for Distinguished Young Scholars in Jiangxi Normal University.

References

Arai-Kichise Y, Shiwa Y, Nagasaki H, Ebana K, Yoshikawa H, Yano M and Wakasa K (2011) Discovery of genome-wide DNA polymorphisms in a landrace cultivar of *japonica* rice by whole-genome sequencing. *Plant and Cell Physiology* 52: 274–282.

Atwell BJ, Wang H and Scafaro AP (2014) Could abiotic stress tolerance in wild relatives of rice be used to improve *Oryza sativa*? *Plant Science* 215–216: 48–58.

Ayed RB, Kallel I, Hassen HB and Rebai A (2014) SNP marker analysis for validating the authenticity of Tunisian olive oil. *Journal of Genetics* 93: e148–e154.

Chagné D, Crowhurst RN, Troggio M, Davey MW, Gilmore B, Lawley C, Vanderzande S, Hellens RP, Kumar S, Cestaro A, Velasco R, Main D, Rees JD, Iezzoni A, Mockler T, Wilhelm L, Van de Weg E, Gardiner SE, Bassil N and Peace C (2012) Genome-wide SNP detection, validation, and development of an 8K SNP array for apple. *PLoS ONE* 7: e31745.

Feltus FA, Wan J, Schulze SR, Estill JC, Jiang N and Paterson AH (2004) An SNP resource for rice genetics and breeding based on subspecies *indica* and *japonica* genome alignments. *Genome Research* 14: 1812–1819.

Gurgul A, Semik E, Pawlina K, Szmatoła T, Jasielczuk I and Bugno-Poniewierska M (2014) The application of genome-

wide SNP genotyping methods in studies on livestock genomes. *Journal of Applied Genetics* 55: 197–208.

Hu BL, Wan Y, Li X, Cai YH, Luo XD and Xie JK (2012) Research progress on genetic diversity and its ecological conservation of endangered Dongxiang wild rice (*Oryza rufipogon* Griff.). *Agricultural Science and Technology* 13: 113–119.

Kawahara Y, de la Bastide M, Hamilton JP, Kanamori H, McCombie WR, Ouyang S, Schwartz DC, Tanaka T, Wu J, Zhou S, Childs KL, Davidson RM, Lin H, Quesada-Ocampo L, Vaillancourt B, Sakai H, Lee SS, Kim J, Numa H, Itoh T, Buell CR and Matsumoto T (2013) Improvement of the *Oryza sativa* Nipponbare reference genome using next generation sequence and optical map data. *Rice* 6: 4.

Li P, Guo MZ, Wang CY, Liu XY and Zou Q (2015) An overview of SNP interactions in genome-wide association studies. *Briefings in Functional Genomics* 14: 143–155.

Mao DH, Yu L, Chen DZ, Li LY, Zhu YX, Xiao YQ, Zhang DC and Chen CY (2015) Multiple cold resistance loci confer the high cold tolerance adaptation of Dongxiang wild rice (*Oryza rufipogon*) to its high-latitude habitat. *Theoretical and Applied Genetics* 128: 1359–1371.

Ohyanagi H, Tanaka T, Sakai H, Shigemoto Y, Yamaguchi K, Habara T, Fujii Y, Antonio BA, Nagamura Y, Imanishi T, Ikeo K, Itoh T, Gojobori T and Sasaki T (2006) The Rice Annotation Project Database (RAP-DB): hub for *Oryza sativa* ssp. *japonica* genome information. *Nucleic Acids Research* 34: D741–D744.

Porebski S, Bailey LG and Baum BR (1997) Modification of a CTAB DNA extraction protocol for plants containing high polysaccharide and polyphenol components. *Plant Molecular Biology Reporter* 15: 8–15.

Sang T and Ge S (2013) Understanding rice domestication and implications for cultivar improvement. *Current Opinion in Plant Biology* 16: 139–146.

Shavrukov Y (2016) Comparison of SNP and CAPS markers application in genetic research in wheat and barley. *BMC Plant Biology* 16(Suppl. 1): 11.

Stein LD, Mungall C, Shu S, Caudy M, Mangone M, Day A, Nickerson E, Stajich JE, Harris TW, Arva A and Lewis S (2002) The generic genome browser: a building block for a model organism system database. *Genome Research* 12: 1599–1610.

Wang HT, Li GS, Kwon WS and Yang DC (2016) Development of EST intron-targeting SNP markers for *Panax ginseng* and their application to cultivar authentication. *International Journal of Molecular Sciences* 17: 884.

Xie JK, Agrama HA, Kong DL, Zhuang JY, Hu BL, Wan Y and Yan WG (2010) Genetic diversity associated with conservation of endangered Dongxiang wild rice (*Oryza rufipogon*). *Genetic Resources and Crop Evolution* 57: 597–609.

Zhang FT, Zhang LX, Cui FL, Luo XD, Zhou Y and Xie JK (2015) Identification of novel insertion-deletion markers for Dongxiang wild rice (*Oryza rufipogon* Griff.) using high-throughput sequencing technology. *Journal of Genetics* 94: e51–e55.

https://doi.org/10.1017/S1479262117000211 Published online by Cambridge University Press