HOSTED BY

ELSEVIER

Available online at www.sciencedirect.com

ScienceDirect

Identification of SNPs in barley (*Hordeum vulgare* L.) by deep sequencing of six reduced representation libraries

Ganggang Guo[a], *Dawa Dondup*[a,b], *Lisha Zhang*[a], *Sha Hu*[a], *Xingmiao Yuan*[a], *Jing Zhang*[a,*]

[a]*Key Laboratory of Crop Germplasm Resources and Utilization (MOA), The National Key Facility for Crop Gene Resources and Genetic Improvement, Institute of Crop Science, Chinese Academy of Agricultural Sciences, Beijing 100081, China*
[b]*Tibet Academy of Agricultural and Animal Husbandry Sciences, Lhasa 850032, China*

ARTICLE INFO

Article history:
Received 21 February 2014
Received in revised form
11 June 2014
Accepted 3 July 2014
Available online 14 July 2014

Keywords:
Barley
SNP discovery
Reduced representation libraries
Allele-specific PCR

ABSTRACT

High-density genetic markers are required for genotyping and linkage mapping in identifying genes from crops with complex genomes, such as barley. As the most common variation, single nucleotide polymorphisms (SNPs) are suitable for accurate genotyping by using the next-generation sequencing (NGS) technology. Reduced representation libraries (RRLs) of five barley accessions and one mutant were sequenced using NGS technology for SNP discovery. Twenty million short reads were generated and the proportion of repetitive sequences was reduced by more than 56%. A total of 6061 SNPs were identified, and 451 were mapped to the draft sequence of the barley genome with pairing reads. Eleven SNPs were validated using length polymorphic allele-specific PCR markers.

1. Introduction

Newly developed high-throughput SNP genotyping platforms have revolutionized genetic mapping and genome-wide association studies (GWAS) in plants [1,2] and animals [3]. Biparental and association mapping populations are powerful genetic materials to study induced mutation and natural variation [4]. SNPs are highly abundant genetic markers and are ideal for GWAS and genetic fine mapping [5]. Genome re-sequencing-based SNP discovery relies on low-coverage sequencing of individual samples as well as the presence of a high-quality draft genome sequence [5]. However, the cost of complete genome re-sequencing for SNP discovery is prohibitive, especially for species with large genomes. In order to meet this challenge genotyping methods based on next generation sequencing (NGS) have been developed and widely used, such as Complexity Reduction of Polymorphic Sequences (CRoPS) [6], Restriction site Associated DNA (RAD) [7], genotyping by sequencing (GBS)-narrow sense [8], and Multiplex Shotgun Genotyping (MSG) [9]. It is particularly noteworthy that GBS has become a powerful tool for association studies and genomics-assisted breeding in a range of species including

* *Corresponding author*. Tel.: +86 10 62189624.
E-mail address: zhangjing03@caas.cn (J. Zhang).
Peer review under responsibility of Crop Science Society of China and Institute of Crop Science, CAAS.

http://dx.doi.org/10.1016/j.cj.2014.06.008

those with complex genomes. As a popular GBS tool, the strategy of a restriction enzyme-based reduced representation library (RRL) is feasible and flexible for SNP identification because it reduces the complexity of the genome by orders of magnitude [10]. RRLs were used for SNP discovery first in human genomics by Sanger sequencing [11]. Later, as an efficient and cost-effective method, RRL was used in maize [12] and cattle [13] for SNP discovery by deep sequencing.

Barley, one of the first crops to be cultivated by humans, is the world's fourth-largest widely grown cereal. Its genome was sequenced in 2012 [14]. Among the sequenced genomes of major crops, high-density SNPs were developed from rice and maize by the re-sequencing method [2,15]. However, SNP discovery in barley was limited to ESTs and unigene fragments in relevant germplasms or array-based transcriptome analysis [16–18]. Development of genotyping by sequencing (GBS) technology was gradually optimized and adopted in barley for SNP identification and QTL mapping. Recently, a novel two-enzyme GBS protocol was developed and bi-parental populations were genotyped with GBS to develop SNPs in barley and wheat [19]. To test new semiconductor sequencing platforms for GBS, Mascher et al. genotyped a recombinant inbred line (RIL) population of barley and concluded that GBS technology can easily be modified as an advanced sequencing technology and genomic analysis tool [20]. A procedure for constructing GBS libraries by reducing genome complexity using restriction enzymes (REs) was reported. This procedure is simple, quick, highly reproducible with high specificity, and may reach important regions of the genome that are inaccessible by sequence capture approaches [8]. In addition, a high-density consensus genetic map in barley was available and GWAS of morphological traits had been performed. A short awn gene, *Breviaristatum-e* (*ari-e*), was mapped to a small genetic interval on chromosome 5H [21], and a master switch gene for anthocyanin production, *ANTHOCYANINLESS 2* (*ANT2*), encoding a basic helix–loop–helix protein (HvbHLH1) was also fine mapped [22]. The gene *HvCEN*, a homolog of *Antirrhinum CENTRORADIALIS* contributing to spring growth habit and environmental adaptation was identified in cultivated barley by the use of the 9K iSelect platform and GWAS [23]. A highly specific in-solution hybridization-based whole exome capture platform was developed and it provides a powerful tool for re-sequencing the genomes of other accessions of barley and its relatives [24]. In this study, we used the restriction enzyme-based RRL method and a parallel sequencing platform to discover *de novo* SNPs in six barley accessions. Some of the SNPs were converted into allele-specific PCR (AS-PCR) markers for marker validation. These converted markers have advantages of low cost per sample and ease of use, thus making them suitable for genetic diversity analysis of barley germplasm resources and marker-assisted breeding. They can also be used in fine mapping of genes controlling important traits in barley.

2. Materials and methods

2.1. Plant growth and DNA preparation

Four barley germplasm accessions from China (ZDM01159, ZDM01467, ZDM00014, ZDM08324), one accession from Mexico (ZDM08233), and one mutant (93–597) with multi-node and stem branching, obtained by γ-irradiation of the accession ZDM08324, and selfing for fifteen generations, were used for RRL construction. Seeds of the six accessions were sterilized with 3% H_2O_2 for 5 min, and washed three times for 5 min with purified water. Subsequently, they were germinated and grown in darkness at 18 ± 2 °C for 14 days. Etiolated seedlings were individually harvested and frozen in liquid nitrogen and then stored at −80 °C for DNA extraction. DNA was extracted and purified with a DNeasy plant mini kit (Qiagen, Hilden, Germany).

2.2. RRL construction and deep sequencing

Ten μg of DNA from each sample was digested with 100 units *Mse* I (New England Biolabs, Beverly, MA, USA) in a 200 μL reaction system. In order to digest the sample completely, the reaction was carried out overnight at 37 °C. The digested DNA was fractionated on a 3.0% agarose gel. Digestion products between 350 and 450 bp were recovered with a MinElute Gel Extraction Kit (Qiagen) according to the manufacturer's protocol.

Sequences were generated from the six barley RRLs on the Illumina GA II DNA sequencing platform (Illumina, San Diego, CA, USA). Raw data were assigned to individual samples using the barcode sequence and trimmed to 40 bp at each end. For the sequencing of barcode ligation, ligation product amplification and sequencing were completed by the BIOMARKER Company (http://www.biomarker.com.cn/).

2.3. SNP discovery and phylogenetic analysis

The raw reads were firstly blasted against the Triticeae Repeat Sequence Database [25] (TREP, http://wheat.pw.usda.gov/ITMI/Repeats/). The matched sequences were filtered, and proportions of repetitive elements were evaluated. Non-repetitive reads were mapped against the whole genome shotgun assembly of barley cultivar Morex [14] with the CLC Genomics Workbench 6.02 (http://www.clcbio.com/), and reads in pairs, or in broken pairs, and the average length of pairing reads were counted. Finally, the reads with >3 × coverage were used for polymorphism analysis and SNP discovery. All of the identified SNPs and their reference sequences (20 bp flanking sequence of mapped pairing reads) are listed in Supplemental Table S1. SNPs identified in at least four to six RRLs were used for phylogenetic analysis. The phylogenetic tree was constructed by the Dnapars program using PHYLIP software [26].

2.4. SNP validation

Twenty one SNPs distributed evenly on all 7 barley chromosomes were randomly selected and converted to AS-PCR markers for SNP validation. Two pairs of primers were designed for identification and genotyping of each SNP as described previously [27]. The SNP is present at the 3′ end of the allele-specific PCR primer to ensure specificity of amplification. Primer design was performed using WASP software [28] (http://bioinfo.biotec.or.th/WASP) with default parameters. The two primer pairs were multiplexed in a single-tube PCR assay to assess the allelic status at each SNP locus. Two AS-PCR products of different lengths were generated. The PCR products were electrophoresed on 1.5% agarose gel and visualized under UV light.