THE CROP JOURNAL XX (2017) XXX–XXX

Available online at www.sciencedirect.com

ScienceDirect

Identification, development, and application of cross-species intron-spanning markers in lentil (*Lens culinaris* Medik.)☆

Debjyoti Sen Gupta[a,*,1], *Jitendra Kumar*[a,*,1], *Sunanda Gupta*[a], *Sonali Dubey*[a], *Priyanka Gupta*[a], *Narendra Pratap Singh*[b], *Gaurav Sablok*[c,1]

[a]*Division of Crop Improvement, ICAR-Institute of Pulses Research, Kanpur 208024, UP, India*
[b]*Division of Plant Biotechnology, ICAR-Institute of Pulses Research, Kanpur 208024, UP, India*
[c]*Department of Biodiversity and Molecular Ecology, Research and Innovation Centre, Fondazione Edmund Mach, Via E. Mach 1, 38010 San Michele all'Adige, Trento, Italy*

ARTICLE INFO

Article history:
Received 6 July 2017
Received in revised form 21 August 2017
Accepted 25 September 2017
Available online xxxx

Keywords:
Allele diversity
Cross-species
Genotypes
Introns
Lentil
Molecular markers

ABSTRACT

Lentil (*Lens culinaris* Medik) is one of the most important food legumes in the world. The use in lentil of molecular marker-assisted breeding is limited, owing to the low availability of polymorphic markers. In the present study, we developed a set of polymorphic intron-spanning markers (ISMs) using a cross-species mapping approach. In this approach, putative unique transcripts (PUTs) of *L. culinaris* were mapped onto the *Medicago truncatula* genome, exploiting its closeness with the lentil genome. Spliced alignment of the PUTs resulted in a total of 25,717 alignments, allowing the development of 1703 ISMs. From these, a subset of 105 ISMs were synthesized and validated with a 51% amplification success rate in 32 lentil genotypes. Of these ISMs, 40 (74%) were polymorphic and generated 2–11 alleles per locus in a genetically diverse panel of 32 lentil genotypes including wild species. This set of polymorphic ISMs along with their functional annotation data will be useful in lentil breeding.

1. Introduction

Lentil (*Lens culinaris* Medik.) is one of the most important food legumes and is grown in many parts of the world including Canada, Australia, northwestern USA, Turkey, Syria, Nepal, India, and Bangladesh [1]. The world annual production is nearly 5 Mt [1]. Lentil originated in the Fertile Crescent and is reported to be one of the earliest domesticated food crops [2]. Cultivation of lentil is affected by various biotic and abiotic stresses including foliar and root diseases, high temperature, drought, soil pH (<5.4), and waterlogging. Most lentil-producing countries use conventional breeding approaches in their active breeding programs for developing high-yielding lentil cultivars with better grain quality. However, in recent years, molecular markers are being widely used for accelerating precise breeding in several crops including major pulse crops such as chickpea and pigeonpea [3]. Further, the advent of next-generation sequencing technologies has allowed the rapid genome sequencing of pulses such as chickpea (*Cicer arietinum*) [4] and pigeonpea (*Cajanus cajan*) [5,6]. Availability of the genome sequences in these species

☆ Peer review under responsibility of Crop Science Society of China and Institute of Crop Science, CAAS.

* *Corresponding authors.*
E-mail addresses: debjyoti.gupta@icar.gov.in (D.S. Gupta), jitendra.kumar@icar.gov.in (J. Kumar).

[1] Debjyoti Sen Gupta, Jitendra Kumar and Gaurav Sablok contributed equally to this work.

https://doi.org/10.1016/j.cj.2017.09.004

Please cite this article as: D.S. Gupta, et al., Identification, development, and application of cross-species intron-spanning markers in lentil (*Lens culinaris* Medik...., The Crop Journal (2017), https://doi.org/10.1016/j.cj.2017.09.004

has permitted the development of genome-anchored maps, visualization of single nucleotide polymorphism SNPs, and identifying species-specific SNPs. In the past, limited efforts have been made toward the development of molecular marker systems such as simple sequence repeats (SSRs) and SNPs for enriching genomic resources in lentil [7–12]. The plant genomes have genes with larger intronsand spliced alignment of transcripts to the genome has revealed a large diversity in intron size. Despite being of diverse lengths, introns have been a major resource for molecular-marker development in several crop species [13–16] and have been recently leveraged to develop marker resources for legumes through the development of intron-spanning markers (ISMs), which are codominant or dominant, reproducible markers that show multi-allelic patterns [13–16]. Development of these markers is gaining importance in species for which a reference genome sequence is available. In lack of the reference genome cross-species reference genome mapping can be used to identify the intron-spanning markers. Briefly, development of intron-spanning markers can be done by performing a spliced alignment of the gene transcripts to the reference genome and developing polymerase chain reaction (PCR) primers that are anchored in conserved exons that span target introns. Previously, intron-spanning markers were developed in legumes using CSGM Designer [17], which provides algorithm-or alignment-based identification of intron spanning markers and these can be validated on a set of diverse genotypes. Similarly, in lentil, transcript sequence databases available in the public domain can be used for *de novo* assembly and identification and design of primers for the amplification of ISM regions. The objectives of this study were to (1) develop polymorphic ISMs in lentil using expressed sequence tag (EST) sequences, and (2) validate polymorphic ISM markers in a diverse panel of *Lens* genotypes including wild lentil species.

2. Materials and methods

2.1. Development of intron-spanning markers, primer design, and functional annotation

A cross-species mapping-based approach was used for developing intron-spanning markers. In this approach, a well annotated and curated reference genome of *Medicago truncatula* was used because of its close phylogenetic positioning with *L. culinaris*. The *L. culinaris* putative unique transcripts (PUTs) from PlantGDB version 187 [18] were mapped onto the *M. truncatula* genome. Prior to mapping, repeat masking of the genome was performed using RepeatMasker, available from http://www.repeatmasker.org/. The RepBase libraries available from http://www.girinst.org/repbase/ and the *L. culinaris* ESTs downloaded from Plant GDB version 187 were aligned to the genome using GeneSeqer [19], a spliced alignment tool available from http://brendelgroup.org/bioinformatics2go/GeneSeqer.php. Following alignment, intron-spanning coordinates were extracted and primers were designed for the respective coordinates using Primer3 version 1.1.4, available from http://primer3.sourceforge.net/releases.php [20].

Following the identification of intron-spanning regions, primer pairs were designed using Primer3 with parameters defined as minimum amplicon size 100 bp and maximum amplicon size 300 bp, primer size 18–27 bp, primer T_m 57–63 °C, primer GC content 30%–70%, CG clamp 0, maximum end stability 250, maximum T_m difference 2, maximum self-complementarity 6, maximum 3′ end self-complementarity 3, maximum Ns accepted 0, and maximum poly-X5. The aligned PUTs to the *Medicago* genome have been annotated and then after annotation gene ontology has been defined based on the annotation. Functional annotation and gene ontology of the intron-spanning markers were performed using BLASTx searches (E-value, 1×10^{-5}) against GenBank (http://www.ncbi.nlm.nih.gov/), UniProt (http://www.uniprot.org/), and TAIR (https://www.arabidopsis.org/) databases.

2.2. Plant materials and DNA extraction

Thirty-two *Lens* genotypes were used for genotyping with 105 primers. A diverse panel of thirty-two *Lens* genotypes consisting of *L. culinaris* released cultivars, advanced breeding lines, parents of mapping populations, and genotypes of *L. ervoides* and *L. culinaris* subsp. *orientalis* was tested to identify polymorphic markers (Table 1). DNA samples were extracted from individual plant leaf tissue when seedlings were two weeks old using the cetyltrimethylammonium bromide (CTAB) procedure [21]. The DNA concentrations of the extracted samples were recorded and were compared with after corresponding concentration with λDNA. The extracted DNA samples were diluted to a uniform concentration of 20 μg μL^{-1} for PCR amplification.

2.3. PCR amplification

One hundred and five primer pairs (Table S1) were synthesized from Imperial Life Sciences (P) Limited, Gurugram, India and used in this study. PCR reactions (in 25-μL volumes) were conducted in a G-Strom (model number GT-40319, UK) thermocycler. Each reaction contained 2.5 μL *Taq* buffer (Merck, Bangalore, India), 1.5 μL $MgCl_2$ (25 mmol L^{-1}) (Merck, Bangalore, India), 0.20 mmol L^{-1} of each dNTP (Merck, Bangalore India), 0.50 mmol L^{-1} of each primer [Imperial Life Sciences (P) Limited, Gurugram India], 0.5 U of *Taq* polymerase (Merck, Bangalore, India), and 20 ng of template DNA. Primers amplifying *Lens* DNA were validated in a set of 32 diverse *Lens* genotypes using the following PCR conditions: 94 °C for 4 min, followed by 35 cycles of 94 °C for 1 min, 58 °C for 1 min, and 72 °C for 1 min followed by a final elongation step of 72 °C for 15 min. PCR products were resolved by 10% polyacrylamide vertical gel electrophoresis (Sigma-Aldrich, New Delhi, India) and visualized by silver staining. Fragments were scored visually after staining.

2.4. Molecular data scoring and statistical analysis

Polymorphism information content (PIC) values were calculated following Botstein et al. [22]. The presence and absence of the band were scored as 1 and 0 and the binary data so obtained for all *Lens* genotypes for polymorphic markers were used to calculate a correlation matrix using Jaccard's similarity coefficient analysis [23]. The similarity coefficient was used to construct a dendrogram based on the unweighted pair group method with arithmetic average (UPGMA) using NTSYS pc-2.21q [24] software. The data were also subjected to principal coordinate analysis (PCA) using NTSYS.

Please cite this article as: D.S. Gupta, et al., Identification, development, and application of cross-species intron-spanning markers in lentil (*Lens culinaris* Medik...., The Crop Journal (2017), https://doi.org/10.1016/j.cj.2017.09.004