Available online at www.sciencedirect.com

ScienceDirect



RNA-Seq analysis and development of SSR and KASP markers in lentil (*Lens culinaris* Medikus subsp. *culinaris*)[☆]

Dong Wang^{a,b,1}, Tao Yang^{b,1}, Rong Liu^{b,1}, Nana Li^a, Xiaomu Wang^a, Ashutosh Sarker^d, Xiaodong Zhang^a, Runfang Li^a, Yanyan Pu^a, Guan Li^b, Yuning Huang^b, Yishan Ji^b, Zhaojun Li^{a,c}, Qian Tian^a, Xuxiao Zong^{b,*}, Hanfeng Ding^{a,c,*}

^aShandong Center of Crop Germplasm Resources, Shandong Research Station of Crop Gene Resource & Germplasm Enhancement, Ministry of Agriculture, Shandong Provincial Key Laboratory of Crop Genetic Improvement, Ecology and Physiology, Jinan 250100, Shandong, China

^bNational Key Facility for Crop Gene Resources and Genetic Improvement, Institute of Crop Sciences, Chinese Academy of Agricultural Sciences, Beijing 100081, China

^cCollege of Life Science, Shandong Normal University, Jinan 250014, Shandong, China

^dInternational Center for Agricultural Research in the Dry Areas (ICARDA), New Delhi 110012, India

ARTICLE INFO

Article history:

Received 10 October 2019

Received in revised form 7 March 2020

Accepted 2 June 2020

Available online 19 June 2020

Keywords:

Lentil

RNA-Seq

EST-SSR

KASP

Genetic diversity

ABSTRACT

Lentil (*Lens culinaris* Medikus subsp. *culinaris*, $2n = 14$) is a cool-season legume with high production potential for multiple uses. However, limited molecular research has been conducted in this species owing to its large genome, which impedes the generation of genome sequences and the development of molecular markers. In this study, more than 1.37 billion filtered clean reads were collected by RNA-Seq of six diverse lentil accessions and 217,836 transcripts and 161,095 unigenes were *de novo* assembled, yielding respectively 257.1 and 240.6 million nucleotides. The mean transcript length was 1180 bp and the N50 and N90 lengths were respectively 2075 and 479 bp. The mean length of the unigenes was 1494 bp and their N50 and N90 values were respectively 2203 and 714 bp. The unigenes were annotated against seven databases. The *FLOWERING LOCUS T* (*FT*) gene homolog in lentil showed high protein sequence similarity to the *FT* gene homologs of pea and alfalfa. On the basis of the RNA-Seq analysis, 26,449 EST-SSR markers were designed *in silico*, and 276 preliminarily screened markers were selected to evaluate polymorphism in 94 diverse lentil accessions. In total, 125 (45.29%) of 276 EST-SSR markers were found to be polymorphic. A total of 130,073 SNP loci were detected and 78 (61.41%) of 127 SNPs were successfully converted to KASP markers. Population genetic analyses of the lentil accessions with EST-SSR and KASP markers revealed similar genetic structures, suggesting that the RNA-Seq-generated resources and the developed markers are reliable for use in molecular marker-assisted breeding of lentil.

© 2020 Crop Science Society of China and Institute of Crop Science, CAAS. Production and hosting 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/>).

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

* Corresponding authors.

E-mail addresses: zongxuxiao@caas.cn (X. Zong), dinghanfeng@shandong.cn (H. Ding).

¹These authors contributed equally to this work.

<https://doi.org/10.1016/j.cj.2020.04.007>

2214-5141 © 2020 Crop Science Society of China and Institute of Crop Science, CAAS. Production and hosting 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

Lentil (*Lens culinaris* Medikus subsp. *culinaris*) is an annual, self-pollinated, herbaceous cool-season legume crop. It is a diploid ($2n = 2x = 14$) species with a genome size of approximately 4 Gb [1,2]. Lentil is the fifth most important pulse crop in the world and the annual production of lentil worldwide was 6,375,732 t in 2018, ranking behind only common bean (*Phaseolus vulgaris* L.), chickpea (*Cicer arietinum* L.), pea (*Pisum sativum* L.), and cowpea (*Vigna unguiculata* L.) [3]. Lentil is a source of protein, carbohydrates, vitamins, and micronutrients in the human diet. It also serves as a high-value feed for livestock. Lentil is often grown on marginal land and affected by drought, high-temperature stress, and various diseases such as blight, anthracnose, rust, and root diseases [4]. The worldwide average yield of lentil is only 1042 kg ha⁻¹ [3]. Breeding strategies including molecular approaches offer potential for increasing the yield of lentil.

Quality traits have been introduced into the main cultivars of lentil by conventional breeding methods. However, these methods are inefficient and time-consuming for the improvement of quantitative traits, whose expression is usually strongly affected by environment. Marker-assisted selection (MAS) in lentil breeding has been limited. In comparison with other legumes, such as soybean, mung bean, adzuki bean and pigeon pea, lentil genome research has progressed slowly [5]. Although a preliminary draft lentil genome assembly has been released and is available at the KnowPulse website [6,7], efforts to improve the genome assembly and annotation are still in progress. Recently, next-generation sequencing technology (NGS) has been applied to non-model plants [8–12]. The lentil RNA-Seq assembly provides valuable genetic resources for lentil genetic research and molecular breeding [13–16]. Kaur et al. [13] performed lentil transcriptome sequencing using Roche 454 GS-FLX Titanium system (Roche Diagnostics Corporation) for large-scale *de novo* assembly of unigenes and identified 2392 simple sequence repeats (SSRs) containing expressed sequence tags (ESTs). Based on this study, Kaur et al. [14] subsequently identified numerous single-nucleotide polymorphisms (SNPs) and used subsets of 546 SSRs and 768 SNPs for genetic mapping in lentil. Verma et al. [15] performed ribonucleic acid sequencing (RNA-Seq) using a different type of Illumina (<http://www.illumina.com>) paired-end sequencing and developed an expressed gene catalog as well as 5673 SSR primer pairs. Sudheesh et al. [16] applied RNA-Seq to generate a comprehensive assembled and annotated reference transcriptome for lentil, advancing functional gene identification and genomic annotation in the species. The accuracy and sensitivity of fluorescence-labeled SSR markers using capillary electrophoresis is superior to that of PAGE silver staining, which is more suitable for genotyping large samples. Fluorescence labeling of SSR markers using capillary electrophoresis has been used in genetic studies of barley [17], sugarcane [18], cherry [19], and lentil [13,14].

Kompetitive Allele-Specific PCR (KASP, from KBioscience or LGC Genomics), which is based on specific matching of primer ends, can be employed for SNPs and InDels (insertions and deletions) genotyping and affords accurate biallelic

identification of SNPs and InDels at specific sites in a wide range of genomic sample types. KASP is a uniplex SNP genotyping platform. KASP genotyping technology is used as both a product and a genotyping service by LGC Genomics Service Laboratories in North America and Europe [20,21]. KASP technology has been used in crops such as wheat [22,23], maize [24,25], soybean [26,27], peanut [28,29], and lentil [4,30–32]. Sharpe et al. [4] designed KASP assays with 28 SNPs and a GoldenGate array (Illumina, Inc.) with 1536 SNPs for genetic mapping in lentil. Majeed et al. [30] described KASP requirements in lentil with dry DNA (48 ng) and wet DNA (6 µL at 5 ng µL⁻¹). Fedoruk et al. [31] applied 56 KASP assays for genetic mapping and quantitative trait locus (QTL) analysis of seed quality characteristics in lentil based on the study of Majeed et al. Rodda et al. [32] combined the GoldenGate assay with 768 newly developed SNPs and the KASP assay with 200 previously published SNPs along with SSR markers to genotype a recombinant inbred line (RIL) population for genetic linkage mapping and perform QTL analysis for boron tolerance. However, the number of KASP markers shown to be polymorphic is still limited.

Although previous studies have identified EST-SSR markers and SNP markers in lentil, the number of markers that can be directly used in genetic studies of lentil remains limited. In the present study, we used six lentil accessions from different origins as experimental materials and performed RNA-Seq on a mixture of roots, stems and leaves of lentil seedlings to obtain reference transcriptome information. We developed EST-SSR and KASP markers and validated them in several lentil accession to provide a basis for future lentil molecular breeding.

2. Materials and methods

2.1. Plant materials

Six lentil accessions: A008, A094, A170, A370, A669, and A677, were selected for RNA-Seq and marker development, and 88 accessions were used for marker validation. RNA-Seq was performed with three biological repeats (samples) of each accession. The 94 accessions originated in seven countries across eastern Asia, western Asia, North America, and north Africa and included 79 Chinese germplasm resources from China (Xinjiang, Inner Mongolia, Ningxia, Qinghai, Shanxi, Shaanxi, Gansu, Hubei, and Yunnan) (Table S1). Seeds of all accessions were provided by the National Crop Gene Bank located in the Institute of Crop Sciences, Chinese Academy of Agricultural Sciences and planted in a greenhouse at the institute in Beijing, China in November 2016.

2.2. cDNA library preparation, sequencing, and transcriptome assembly

Four-week-old seedlings (with root, stem and leaves) of accessions A008, A094 A170, A370, A669, and A677 were sampled. Total RNA was extracted from each plant using the TRIzol-based RNA extraction kit (Novogene Co. Ltd., Beijing, China). The sequencing library was prepared using the

NEBNext Ultra RNA Library Prep Kit for Illumina (NEB, Ipswich, MA, USA) according to the manufacturer's instruction. Clustering of the index-encoded samples was performed on a cBot Cluster Generation System using the TruSeq PE Cluster Kit v3-cBot-HS (Illumina, San Diego, California, USA) according to the manufacturer's instructions. After cluster generation, the library preparations were sequenced on the Illumina HiSeq platform, and paired-end reads were produced. Raw data were stored in the Sequence Read Archive (SRA) of NCBI (<https://submit.ncbi.nlm.nih.gov/subs/sra/SUB5351043/overview>) with the ID SUB5351043.

The FASTQ-format raw data were first filtered to generate clean data by deleting reads containing adapter or poly-N sequences (with an unknown-base proportion greater than 10%) as well as low-quality reads (for which bases with Q value ≤ 20 accounting for more than 50% of the entire read) using an in-house Perl script. The Q20, Q30, GC content and sequence repeat levels of the clean sequences were calculated. All downstream analyses were based on the high-quality clean sequence. Transcriptome assembly was performed using Trinity [33], with `min_kmer_cov` set to 2 and all other parameters set to their default values. Data nonredundancy was achieved with Corset (<https://code.google.com/p/corset-project/>), which used the number of reads and expression patterns of the aligned transcripts to hierarchically cluster transcripts [34]. Finally, the longest transcript in each cluster was retained as a unigene for subsequent annotations.

2.3. Unigene functional annotation

To obtain comprehensive gene function information, unigenes were annotated based on seven databases: 1) NCBI non-redundant protein sequences (NR, using diamond 0.8.22 (<https://github.com/bbuchfink/diamond>) with an e-value = $1e^{-5}$) [35]; 2) NCBI nucleotide sequences (NT, using NCBI BLAST 2.2.28+ (<ftp://ftp.ncbi.nlm.nih.gov/blast/executables/LATEST/>) with an e-value = $1e^{-5}$) [36]; 3) Protein family (PFAM, using HMMER 3.0 package and hmmscan (<https://www.ebi.ac.uk/Tools/hmmer/>) with an e-value = 0.01) [37]; 4) euKaryotic Ortholog Groups (KOG, using diamond 0.8.22 with an e-value = $1e^{-3}$) [35]; 5) SwissProt (using diamond 0.8.22 with an e-value = $1e^{-5}$) [35]; 6) Kyoto Encyclopedia of Genes and Genomes (KO, using the KEGG Automatic Annotation Server (KAAS, <https://www.genome.jp/tools/kaas/>) with an e-value = $1e^{-10}$) [38]; and 7) Gene Ontology (GO, using Blast2GO 2.5 (<https://www.blast2go.com>) and a custom script with an e-value = $1e^{-6}$) [39].

2.4. Alignment of an annotated gene in lentil with homologous genes in alfalfa and pea

Protein sequences related to those encoded by the FT gene in the model organism alfalfa (*Medicago truncatula*) were searched against the NCBI database. It was found that the sequence of AET05551.2 was FT and the TFL1-like protein in alfalfa. The protein sequence was subjected to a BLAST search against the lentil unigene database using BioEdit 7.0.9.0 (<https://bioedit.software.informer.com/>) with default parameters. The lentil sequence with ID Cluster-18821.11917,

for which the alignment score was $3e^{-41}$, was most similar. The gene was functionally annotated as the flowering gene Ta1 in pea (*Pisum sativum* L.). Cluster-18821.11917 was a homolog of the FT gene in alfalfa and pea. The FT-related protein sequences in alfalfa and pea were searched against the NCBI database. FT protein sequences including AEI99551.1, AEI99552.1, AEI99553.1, AEI99554.1, AEI99555.1 in and ADZ05704.1, ADZ05700.1, ADZ05701.1, ADZ05702.1, and ADZ05703.1 were found. Finally, the ClustalW Multiple function of BioEdit was used to perform multiple alignment of these protein sequences.

2.5. Development and validation of EST-SSR markers

Genomic DNA was extracted from the fresh leaves of 4-week-old seedlings (mixtures of three plants) using a new Plant Genomic DNA Isolation kit (DP320-03, Tiangen Biotechnology, Beijing, China) and stored at -20°C for later PCR.

High-throughput SSR detection was performed with MISA (MicroSatellite identification tool, <http://pgrc.ipk-gatersleben.de/mwasa/mwasa.html>) (Table S2). The equipment parameters were set to minimum SSR repeat motif lengths of mono-10, di-6, tri-5, tetra-5, penta-5, and hexa-5. For complex situations, the maximum allowed interval between two different SSRs was 100 bp. Microsatellite loci were selected using Primer 3.0 (<https://SourceMeme.net/Projects/primer3/>), and SSR primers were designed based on flanking conserved sequences [40].

Of all EST-SSRs detected, 480 were selected randomly for preliminary screening with accessions A036, A158, A376, and A667 from diverse origins. The primers (Beijing Zixi Biotechnology Co. Ltd., Beijing, China) and the DL500 DNA marker (TaKaRa Biotechnology Co. Ltd., Dalian, China) was used for PCR amplification and electrophoresis. The volume of PCR system was 10 μL , containing 5.0 μL of $2\times$ Premix Taq buffer (TaKaRa), 2.0 μL DNA ($10\text{ ng }\mu\text{L}^{-1}$), 1.0 μL of the primer pair ($2.0\text{ pmol }\mu\text{L}^{-1}$), and 2.0 μL ddH₂O. The amplification reaction was performed in a Biometra TAdvanced PCR Thermal Cycler (Biometra, Jena, Germany) with the program of 94°C for 5 min; 35 cycles of 94°C for 30 s, 52°C for 45 s, and 72°C for 1 min; 72°C for 10 min. The amplification product was separated on an 8% nondenaturing polyacrylamide gel.

Based on the results of the preliminary screening, the selected markers were fluorescently labeled for the validation of polymorphism in 94 lentil accessions and the amplified products were separated by capillary electrophoresis. The fluorescent label FAM (6-carboxy-fluorescein) or HEX (5-hexachloro-fluorescein) was added to the 5' end of each pair of primers for amplification of EST-SSR markers. The fluorescently labeled primers used in the experiment were synthesized by Beijing Zixi Biotechnology Co. Ltd. (Beijing, China). The primers were dissolved in sterile ddH₂O and diluted to $10\text{ pmol }\mu\text{L}^{-1}$ for use. Other reagents were purchased from TaKaRa Biotechnology (Dalian, China) Co. Ltd. Capillary electrophoresis PCR amplification was performed as described above and the PCR product was diluted 50-fold. Autofluorescence detection was performed on an ABI 3730XL DNA Analyzer (Applied Biosystems, Foster City, USA). Finally, the size standard and parameters were set up for data collection and image analysis using Data Collection and

GeneMapper software (Thermo Fisher Scientific, Inc., Waltham, Mass, USA). To further verify the quality of the designed primers, we performed BLAST searches of the sequences of the polymorphic EST-SSR markers (Table S3) against the pre-released lentil genome sequence with BLASTn at the website KnowPulse (<https://knowpulse.usask.ca/blast/nucleotide/nucleotide>) [6,7] with a typical cutoff E-value of $1e^{-5}$ and a similarity level $\geq 90\%$ to search for homologs [41,42].

2.6. Development and validation of KASP markers

Picard-Tools V1.41 (<https://broadinstitute.github.io/picard/>) and SAMtools v0.1.18 [43] were used for sorting, removal of duplicate reads, and merging alignment results for each example. GATK 3 software was used to call SNPs [44]. Filtering of the original VCF file with the GATK standard method and other parameters was performed as follows: cluster, 3; WindowSize, 35; QD <2.0 or FS >60.0; MQ <40.0 or SOR >4.0; MQRanksum <-12.5 or ReadPosRanksum < -8.0 or DP <10.

KASP primer design followed the standards recommended by the KASP developer. The 3' ends of the two specific forward primers were located at the SNP locus, and fluorescently labeled FAM (GAAGGTGACCAAGTTCATGCT) and HEX (GAAGGTGGGAGTCAACGGATT) were added at the 5' end. Three primers were included in the primer mix. The basic KASP reaction system for the 384-well plates consisted of 4 μ L DNA template, 2.0 μ L of 2 $\times$ master mix, 0.064 μ L primer mix, and 1.936 μ L ddH₂O. The PCR procedure for KASP was as follows: 94 °C for 5 min; 94 °C for 20 s, 61 °C for 1 min, 10 cycles; 94 °C for 20 s, 55 °C for 1 min, 26 cycles; and 72 °C for 1 min. Fluorescence-labeled FAM and HEX were used for genotyping. The main instruments used in the experiment were an ABI 7900HT Real-Time PCR instrument and a GeneAmp PCR System 9700 (Thermo Fisher Scientific, Inc., Waltham, Mass, USA). Fluorescently labeled primers were synthesized by Beijing Zixi Biotechnology Co. Ltd. (Beijing, China).

2.7. Data analysis

GeneMapper 4.0 software provided the value of each allele according to each fluorescent label. The size of each allele was recorded and values below 500 were considered invalid data. The input files for different software platforms were prepared according to the formatting requirements of each type of software used subsequently. The genetic diversity parameters of each EST-SSR were calculated using PowerMarker 3.25 [45], including major allele frequency (MAF), genotype number (NG), allele number (NA), gene diversity (GD), heterozygosity (He), and polymorphic information content (PIC). Principal component analysis (PCA) was performed using NTSYSpC 2.10e (<https://downloadbull.com/portable-ntsyspc-v2-10e-free-download/>). PowerMarker 3.25 was used to calculate the genetic distance [46] and MEGA 6.0 [47] was used to construct a phylogenetic tree using the UPGMA method. The population genetic structure of the test materials was analyzed using STRUCTURE 2.3.4 [48,49]. The parameters were set as follows: length of burn-in period = 10,000, number of MCMC reps after burn-in = 10,000, number of populations (K) = 1–20, number of iterations = 20. The optimal population structure and subgroup number were determined according to the delta K

(ΔK) value [50] with the website program STRUCTURE HARVESTER (http://taylor0.biology.ucla.edu/struct_harvest/) [51].

3. Results

3.1. Illumina sequencing and transcriptome assembly of lentil

Using the Illumina HiSeq system, a total of 1415.2 million reads were obtained from the 18 lentil samples (six accessions with three replicates) with a mean of 78.6 million, of which 1377.4 million were clean reads. These reads were assembled into 217,836 transcripts and 161,095 unigenes. The mean lengths of transcripts and unigenes were 1180 and 1494 bp, respectively (Table 1). The N50 values of transcripts and unigenes were 2075 and 2203 bp, respectively.

3.2. Gene function annotation

Gene function annotation revealed 161,095 unigenes. Among the seven databases, the largest number of genes were annotated by NT, 117,830 (73.14%), and the fewest were annotated by KOG, 32,244 (20.01%) (Fig. 1a). Venn plots of annotation results for five databases including NR, NT, PFAM, GO, and KOG showed that 25,974 genes could be annotated with the five databases (Fig. 1b).

After KO annotation, 48,384 unigenes of lentil were classified according to their participation in KEGG metabolic pathways, including cellular processes (2434), environmental information processing (1759), genetic information processing (9789), metabolism (21,743), and organismal systems (1448) (Fig. 2a). The largest number of genes were involved in the metabolism branch, under which the most common pathway was the carbohydrate metabolism, with the most genes (4435).

After GO annotation, the successfully annotated 85,585 unigenes were classified according to the three levels of gene function: biological process, cellular component, and molecular function (Fig. 2b). The greatest numbers of unigenes (51,049) were associated with cellular processes. Among the cellular components, the numbers of unigenes associated with cells were greatest and reached 30,325. For molecular function, the greatest numbers of unigenes (49,117) were associated with binding.

A total of 32,244 KOG-annotated unigenes were classified according to KOG groups and divided into 26 groups (Fig. 2c).

Table 1 – Characteristics of the *de novo* assembly of lentil using the Trinity software.

Item	Transcript	Unigene
Total nucleotides (bp)	257,080,794	240,633,780
Sequence number	217,836	161,095
Max length (bp)	33,174	33,174
Mean length (bp)	1180	1494
Median length (bp)	698	1113
Min length (bp)	201	201
N50 (bp)	2075	2203
N90 (bp)	479	714

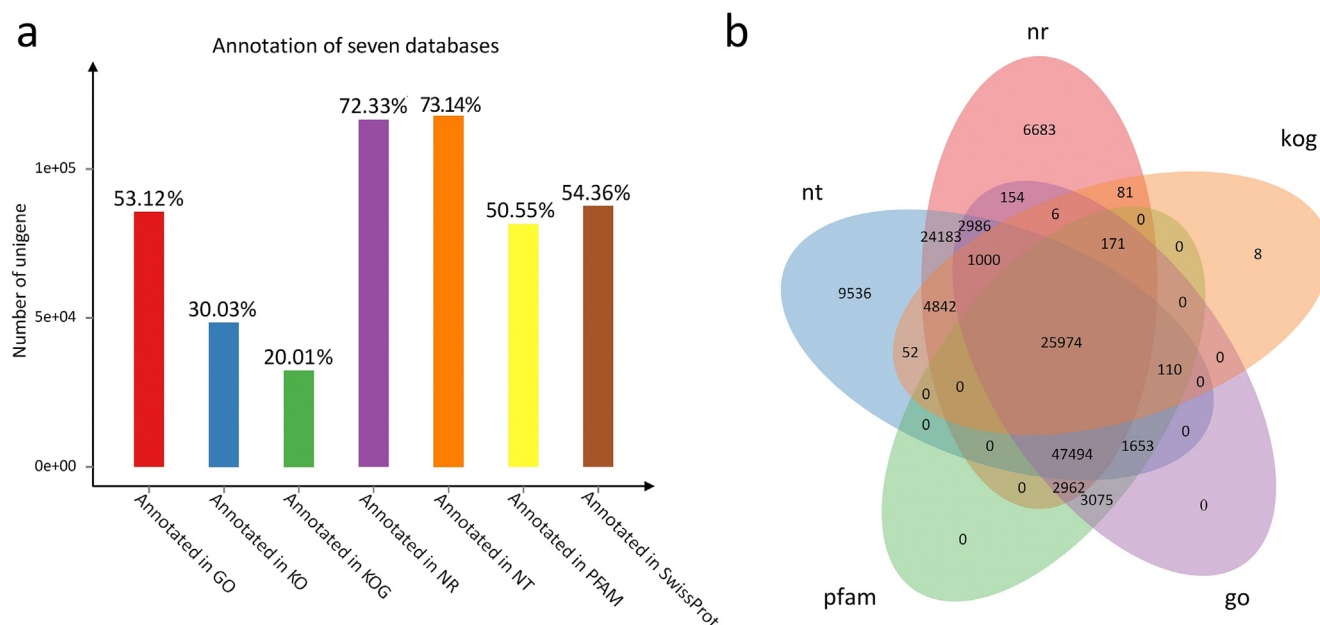


Fig. 1 – Unigene annotation results. (a) Frequency distribution of the annotation results of unigenes in GO, KO, KOG, NR, NT, PFAM, and SwissProt databases. (b) Venn diagram of gene annotation for the NR, NT, PFAM, GO, and KOG databases.

Among these genes, the number associated with posttranslational modification, protein turnover, and chaperones was highest, at 4393; the number in the general function prediction category ranked second, at 4131; and the number associated with translation, ribosomal structure, and biogenesis ranked third, at 2895.

To verify the accuracy of assembly and annotation, we selected an unigene annotated as a FT homologous gene in lentil to compare with those of pea and alfalfa. The lentil RNA-Seq analysis and gene function annotation revealed that the Cluster-18,821.11917 was the FT gene homolog of lentil. Protein sequence alignment showed that the FT gene homolog of lentil displayed high similarity with the FT gene families of pea and alfalfa (Fig. 3).

3.3. Development and validation of EST-SSR and KASP markers

A total of 26,449 pairs of EST-SSR primers were designed based on the RNA-Seq results (Table S2). A total of 480 primer pairs were randomly selected for verification (Tables S3 and S4). EST-SSR primers were identified in accessions A036 (Shanxi, China), A158 (Yunnan, China), A376 (Canada), and A667 (Turkey). The results showed that 276 pairs of primers (Tables S3 and S4) amplified clear products. The remaining 204 primers (Table S4) showed no amplification products. The 276 primer pairs were further verified with fluorescently labeled EST-SSR markers using capillary electrophoresis for 94 lentil accessions. The results showed that among the 276 primer pairs, 125 were polymorphic (Table S3) and 43 monomorphic (Table S4), accounting for respectively 45.29% and 15.58% of the 276 pairs. The remaining 108 primer products were poorly separated by capillary electrophoresis

and were classified as primers that amplified complex products (Table S4). Fig. S1 shows the results of electrophoresis of some test samples with the fluorescently labeled marker P233. To further verify the quality of the screened markers, sequences of the selected 125 polymorphic markers were aligned with the pre-released lentil genome sequence, showing that 123 of the 125 markers could be mapped to the reference genome, with 110 (88%) mapping to unique positions (Table S3).

Based on RNA-Seq analysis, a total of 130,073 SNPs were identified (Table S5). Of these, 127 were selected for KASP marker development with 127 pairs of designed KASP primers (Table S6). Of the 127, 78 (61.41%) could be successfully transformed into KASP markers. Seventy-eight KASP markers were verified in 94 lentil accessions, among which 76 markers were polymorphic and only two monomorphic, accounting for respectively 59.84% and 1.57% of markers. Fig. S2 shows the results of the detection of the WD_SNP1 primer based on Sequence Detection Systems.

3.4. Genetic diversity evaluation

For EST-SSR markers, MAF ranged from 0.1882 to 0.9946, with a mean of 0.7143; GD ranged from 0.0107 to 0.8836, with a mean of 0.3800; He ranged from 0 to 1, with a mean of 0.2541; and PIC ranged from 0.0106 to 0.8727 with a mean of 0.3407 (Table S7). For KASP markers, MAF ranged from 0.5172 to 1.0000, with a mean of 0.8198; GD ranged from 0 to 0.4994, with a mean of 0.2665; He ranged from 0 to 0.9524, with a mean of 0.0846; and PIC values ranged from 0 to 0.3747, with a mean of 0.2229 (Table S8). A total of 727 genotypes and 608 alleles were obtained for EST-SSR markers, while 215 genotypes and 154 alleles were generated with KASP markers

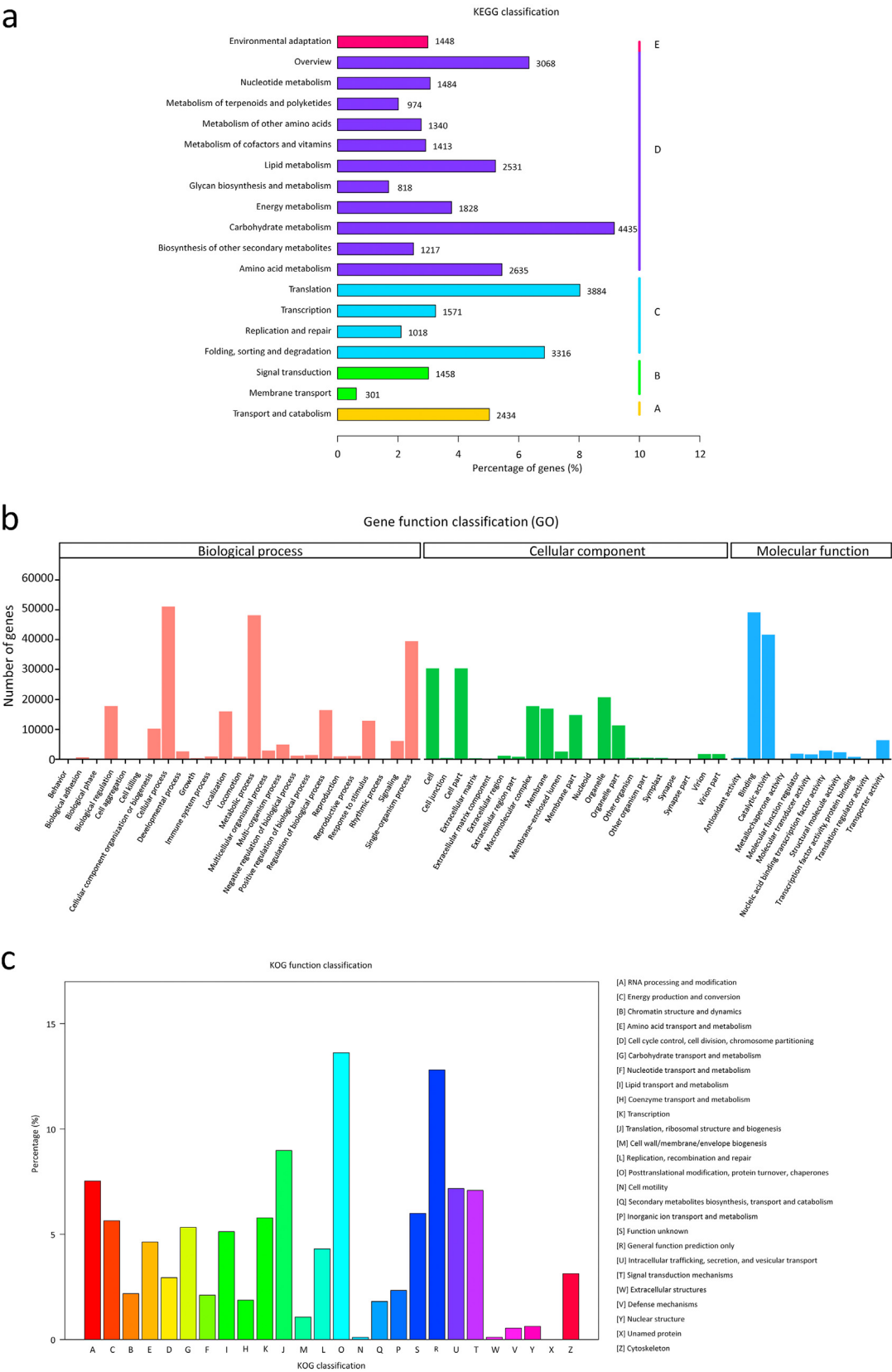


Fig. 2 – Classification of unigenes in KEGG, GO, and KOG. (a) KEGG classification of unigenes associated with cellular processes, environmental information processing, genetic information processing, metabolism, and organismal systems. (b) GO function classification of unigenes associated with biological processes, cellular components, and molecular functions. (c) KOG function classification of unigenes.

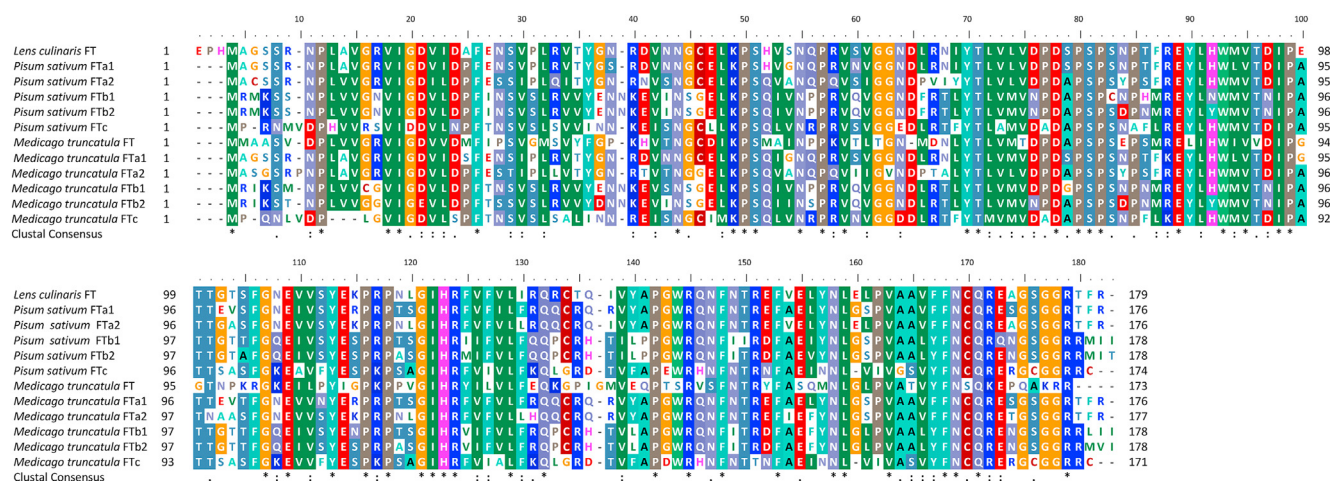


Fig. 3 – Alignment between the protein sequences of FT homolog unigene Cluster-18821.11917 in lentil and FT gene homologs in pea and alfalfa.

(Table 2). The information contents of EST-SSR and KASP markers were classified into slight ($PIC < 0.25$), moderate ($0.25 \leq PIC < 0.5$), and high ($PIC \geq 0.5$). According to these standards, there were 30 highly and 49 moderately informative EST-SSR markers, and 27 KASP markers were classed as moderately informative (Table 2). In general, the genetic diversity of SNP markers was lower than that of EST-SSR markers.

3.5. Population genetic structure analyses

The cumulative contribution rates of the first three principal components were 98.45% and 79.35% for the EST-SSR and KASP markers, respectively. PCA based on EST-SSR markers (Fig. 4a) showed that the spatial distributions of the Chinese accessions (red ellipse) clustered together and were clearly separated from the foreign accessions (blue ellipse) except for several accessions, indicating that the genetic basis of Chinese accessions was significantly different from that of exotic accessions. Only one sample from Japan of the exotic accessions was clustered into the Chinese group, indicating that this sample was closely related to the Chinese accessions. PCA analysis with KASP markers displayed a similar genetic variation pattern with those of EST-SSR markers despite the different marker system (Fig. 4b). The Chinese resources were more prominently clustered with the exception of two accessions, while the distribution of the

exotic materials was broader. These results indicated that the two types of markers showed essentially identical effectiveness for classifying genetic materials.

The UPGMA dendrogram based on EST-SSR markers divided 94 lentil accessions into two major groups (Fig. 4c). The first group was composed mainly of Chinese lentil accessions, whereas most members of the second group were from outside China. The first group was further divided into two subgroups, of which one small subgroup included three genotypes from western Asia and the second subgroup consisted Chinese accessions with one Japanese accession. The UPGMA dendrogram based on KASP markers also identified two major groups in the 94 lentil accessions mainly corresponding to Chinese and foreign accessions (Fig. 4d), which were highly consistent with the classification using EST-SSR markers. The first group was also split into two subgroups. In slight contrast with EST-SSR, the first subgroup consisted of all Chinese samples and the second subgroup of only two accessions, one from Japan and the other from Syria.

Structure analysis based on EST-SSR markers divided all the samples into two groups (Fig. 4e), when the ΔK value was highest at $K = 2$ (Fig. S3a). Red represents the first group, consisting mostly of samples from China and one Japanese sample (A677). Green indicates the second subgroup, consisting mostly of exotic germplasm, and two Chinese accessions (A170 and A491). The structure analysis of the KASP markers was consistent with the results obtained from

Table 2 – Summary of alleles and diversity indices of the EST-SSR and KASP markers tested in 94 lentil accessions.

Marker type	Marker number	Total genotype number	Total allele number	Mean MAF	Mean GD	Mean He	Mean PIC	Informative type		
								Slight	Moderate	High
EST-SSR	125	727	608	0.7143	0.3800	0.2541	0.3407	46	49	30
KASP	78	215	154	0.8198	0.2665	0.0846	0.2229	51	27	0

Markers were classified as highly informative ($PIC \geq 0.5$), moderately informative ($0.25 \leq PIC < 0.5$), and slightly informative ($PIC < 0.25$). MAF, major allele frequency; He, heterozygosity; GD, gene diversity; PIC, polymorphic information content.

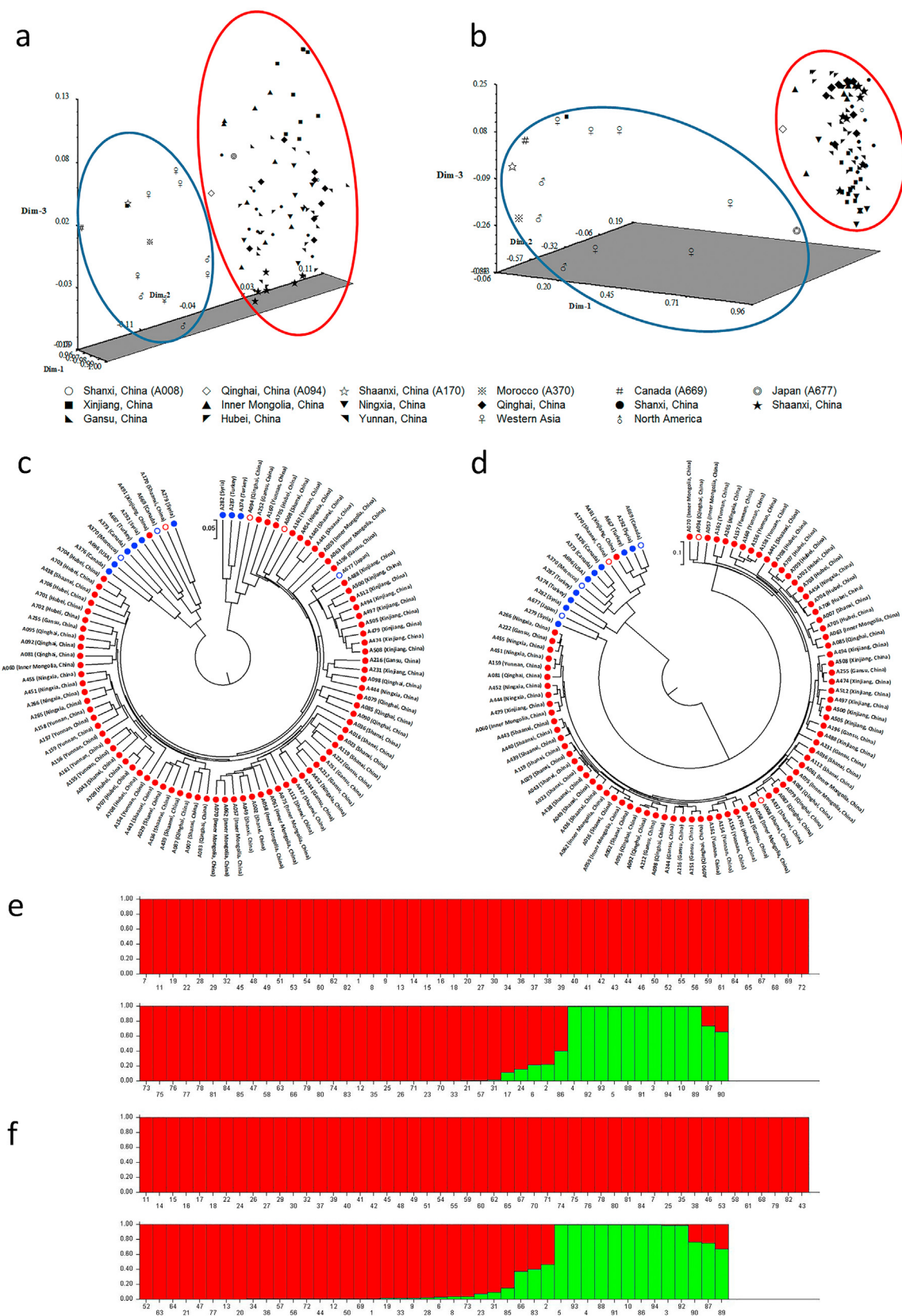


Fig. 4 – Results of population genetic structure analyses of 94 lentil accessions using EST-SSR and KASP markers. Three-dimensional PCA of 94 lentil accessions based on 125 EST-SSR markers (a) and 78 KASP markers (b). UPGMA dendrograms showing the genetic relationships among 94 lentil accessions based on 125 EST-SSR markers (c) and 78 KASP markers (d). STRUCTURE analyses of 94 lentil accessions based on 125 EST-SSR markers (e) and 78 KASP markers (f).

EST-SSR markers in general. When the ΔK value was highest at $K = 2$ (Fig. S3b), two genetic groups were identified (Fig. 4f). The first group in red contained all the Chinese samples and two foreign accessions (A677 from Japan, A279 from Syria). The second group in green consisted mostly of exotic germplasm along with the same two Chinese accessions (A170 and A491) with the results of EST-SSR markers. In summary, the results of population genetic structure analyses were consistent among different approaches and markers.

4. Discussion

4.1. A high-quality transcriptome assembly and annotation for lentil

As a global food legume crop, the harvest area and consumption demand of lentils are expanding year by year according to FAO statistics [3]. With the increasing in global agronomic importance, more and more attentions were paid to the development of genomic and genetic resources for crop improvement of lentils. Except that a preliminary draft genome assembly of lentil has been released [6,7], several studies using transcriptome analysis have also provided valuable genetic resources for lentils [13,15,16,52]. An earlier study of transcriptome sequencing generated a total of 15,354 contigs ranging between 114 bp and 6479 bp with a mean of 717 bp, as well as 68,715 singletons ranging between 101 bp and 560 bp, with an average of 286 bp. A total of 20,419 unique matches were identified when the lentil unigene set was compared against the EST sequence databases of *Glycine max* [13]. Verma et al. performed RNA-Seq in lentil (cv. Precoz) and obtained 42,196 nonredundant high-quality transcripts with an average length of 810 bp and an N50 of 1432 bp [15]. Sudheesh et al. [16] conducted RNA-Seq in the lentil variety Cassab and obtained a high-quality reference unigene set containing 58,986 contigs and scaffolds with N50 length of 1719 bp. A total of 46,372 transcripts were annotated in at least one of five databases, while 22,621 transcripts were annotated in all five databases. When a drought-sensitive lentil cultivar (cv. Sultan) was subjected to *de novo* RNA-Seq analysis after the establishment of drought conditions, *de novo* assembly generated 207,076 transcripts with a mean length of 950 nt [52]. In this study, a new reference transcriptome assembly including 217,836 nonredundant high-quality transcripts was obtained with N50 value of 2075 bp (Table 1), which is better in total numbers and transcript average length than that of previous studies [13,15,16,52]. In addition, more unigenes (25,974) were annotated in all five databases (Fig. 1) than previous studies [13,16].

To verify the accuracy of assembly and annotation, we selected an unigene annotated as a FT homologous gene in lentil and performed sequence alignment with those FT homologous genes of pea and alfalfa. Flowering time is an agronomic trait closely related to adaptation and productivity [53]. Phenological adaptation is considered to be the major evolutionary force underlying the ecological differentiation of lentil [54,55]. The FT gene encodes a small protein that binds to the protein domain (PEBP) of mammalian phosphatidyl ethanolamine analogues,

promoting flowering and initiating grain production [56,57]. Hecht et al. showed that FT gene homologs and the flowering induction function are highly conserved among model legumes including soybean (*Glycine max*), crowfoot (*Lotus japonicus*), alfalfa (*Medicago truncatula*) and pea (*Pisum sativum*) [58]. Tadege et al. suggested that the five FT genes of the leguminous plant alfalfa exhibit different expression patterns and responses to environmental cues [59]. However, little is known about the FT gene homologs of lentil, and direct comparisons among lentil, pea and alfalfa are lacking. In the present study, the Cluster-18,821.11917 annotated as an FT gene homolog in lentil, showed high sequence similarity to those FT homologous genes of pea and alfalfa. To sum up, a high-quality lentil reference transcript was obtained and comprehensively annotated in this study, which can provide a powerful tool for lentil gene mining.

4.2. EST-SSR marker development and validation in lentil

SSR markers are codominant markers with abundant polymorphisms in many eukaryotic genomes [60]. SSR markers are widely used in major crops for genetic and genomic analyses such as DNA fingerprinting, genetic linkage mapping, genetic diversity assessment, QTL mapping, and gene mining [61–65]. The reported research on lentil involving SSR markers has included the development of SSR markers, association analysis based on SSR markers, and genetic diversity analysis [66–68]. Kaur et al. [13] performed transcriptome sequencing in six distinct lentil varieties and designed 2393 EST-SSR primer pairs, 192 of which were screened for validation in 12 lentil cultivars and one wild relative. A total of 166 (86.4%) EST-SSR primer pairs showed successful amplification, among which 47.5% revealed genetic polymorphism. Verma et al. [15] designed 5673 pairs of EST-SSR primers using RNA-Seq, 96 of which were randomly selected to amplify the parental genotype, and 82 pairs (85.4%) were successfully amplified. Of these, 54 primer pairs were selected to genotype 24 lentil cultivars and plants of the related species *Medicago*, *Glycine*, and *Vigna* for genetic diversity analysis, and 23 pairs (43%) were found to be polymorphic. In the present study, the transcriptomes of six lentil accessions from different regions were sequenced from roots, stems and leaves, and 26,449 EST-SSR markers were generated based on the sequencing results. A total of 480 pairs were first selected for validation in four lentil genotypes using PAGE, among which fluorescent labels were added to 276 pairs with clear bands, and the marker pairs were then verified in 94 lentil accessions using capillary electrophoresis. The results showed that 125 (45.29%) markers were polymorphic, 43 (15.58%) markers were monomorphic, and 108 (39.13%) markers were complex and unrecognizable (Tables S3 and S4). BLAST alignment of the 125 polymorphic EST-SSR markers with the preliminary genome of lentil [6,7] showed that 123 could be mapped to the reference genome and that 110 exhibited unique mapping positions, resulting in a unique mapping rate of 88.00% (Table S3). These results were similar to those of a study in chickpea showing a unique mapping rate of 91% [41]. Such results further verified the reliability of the EST-SSR markers developed in the present study. These 125 polymorphic EST-SSR markers

will greatly benefit molecular marker-assisted breeding in lentil.

4.3. KASP marker analysis is cost-effective in lentil

KASP is a commonly used technique for SNP genotyping. It is used in several research areas, including plant breeding, disease identification, and species identification [69]. KASP genotyping offers several advantages over other genotyping methods. KASP detection is fast and cost-effective [69–71]. Multiple genotyping can be performed simultaneously in a 96-well plate using KASP, increasing the likelihood of detecting rare hybrids. KASP is very flexible and can genotype sequences containing multiple SNPs if degenerate or mixed bases are added to the primer sequences. KASP marker development and validation have been reported in some cool-season legumes such as grass pea and lentil. Hao et al. [72] identified 50 KASP markers and validated them in 43 grass pea accessions, with 42 (84%) KASP markers showing polymorphism and displaying PIC values from 0 to 0.375, with a mean of 0.246. For lentil, Sharpe et al. [4] designed KASP assays based on 28 SNPs within a unique contig using a targeted transcriptome profiling strategy, and validation in 11 genotypes showed four failed assays. Fedoruk et al. [31] applied 56 KASP assays for genetic mapping and QTL analysis of seed quality characteristics in lentil based on the above study. Rodda et al. [32] used 200 KASP markers to genotype 178 F₆ lines and the parents of a RIL population for genetic linkage mapping and found that 106 KASP markers were polymorphic. However, the number of KASP markers verified to be polymorphic was still limited. In the present study, 127 KASP markers were designed and validated in 94 lentil accessions. The results showed that 78 (61.42%) KASP markers were available, among which 76 (59.84%) markers were polymorphic, and 2 (1.57%) were monomorphic, and the PIC values ranged from 0 to 0.3747, with an average of 0.2229. These 76 polymorphic KASP markers are novel markers for molecular-assisted breeding in lentil.

4.4. Comparison of SSR and SNP markers

Both SSR and SNP markers are commonly used tools for studying crop genetic diversity and population structure. Li et al. [73] performed genetic diversity analysis of 303 cultivated and wild soybean accessions using 99 SSR markers and 554 SNP markers, and the genetic diversity of 0.77 for SSR markers was significantly higher than that of 0.35 for SNP markers. In the present study, 125 EST-SSR markers and 78 KASP markers were used to analyze the genetic diversity and population structure of 94 lentil accessions. The mean genetic diversity was 0.3800 and the mean PIC value was 0.3407 for the EST-SSR markers, greater than the values of 0.2665 and 0.2229 obtained for the KASP markers. These results were consistent with previous research reports. SNP markers carry only two alleles and SSR markers carry multiple alleles. Comparison of the EST-SSR markers used in this study revealed that their genetic diversity was lower than that of the SSR markers developed in previous studies. The reason for this disparity is that EST-SSRs are located in coding regions, which are more conserved than non-coding regions. Comparison of the EST-SSR and

KASP markers via PCA, cluster analysis, and population structure analysis of 94 lentil accessions showed that the two types of markers were fairly consistent with each other, further demonstrating that RNA-Seq analysis and the two types of markers developed were reliable and effective for application in molecular marker-assisted breeding in lentil. However, the lentil samples used for marker validation were diverse. We suggest that future studies should investigate a wider range of germplasm.

5. Conclusions

Six lentil accessions from different regions were used for RNA-Seq with three replicates for each accession. The total sequencing data of the 18 samples consisted of 206.64 Gb of clean bases, with a mean of 11.48 Gb per sample. The sequencing data were *de novo* assembled, yielding 217,836 transcripts and 161,095 unigenes. In addition, 26,499 EST-SSR markers and 130,073 SNP loci were identified. After validation in 94 lentil accessions from different areas of the world, 125 of 276 (45.29%) EST-SSR markers and 78 of 127 (61.42%) KASP markers were found to be polymorphic. The results of the two marker verification analyses of population structure were highly consistent. The EST-SSR and KASP markers obtained in this study will assist in the molecular marker-assisted breeding of lentil. Future studies should investigate a wider range of lentil germplasm.

Supplementary data to this article can be found online at <https://doi.org/10.1016/j.cj.2020.04.007>.

Declaration of competing interest

The authors declare no conflicts of interest.

Acknowledgments

This work was supported by the funding of Subject Team of Science and Technology Innovation Project of Shandong Academy of Agricultural Sciences (CXGC2018E15), National Key Research and Development Program of China (2017YFE0105100), Industry Team of Science and Technology Innovation Project of Shandong Academy of Agricultural Sciences (CXGC2016A02), Crop Germplasm Resources Protection (2130135), Coarse Cereals Innovation Team of Modern Agricultural Industry Technology System of Shandong Province (SDAIT-15-01), China Agriculture Research System (CARS-08), Youth Research Fund of Shandong Academy of Agricultural Sciences (2016YQN19), and Agricultural Science and Technology Innovation Program (ASTIP) of CAAS.

Author contributions

Dong Wang, Tao Yang, Hanfeng Ding, and Xuxiao Zong designed and conceived the experiment. Dong Wang, Tao

Yang, Xuxiao Zong, Rong Liu, Yanyan Pu, Guan Li, Yuning Huang, Yishan Ji and Zhaojun Li performed the experiment. Dong Wang, Tao Yang, Rong Liu, Nana Li, Xiaomu Wang, and Qian Tian analyzed the RNA-Seq data. Dong Wang, Rong Liu, Tao Yang, and Hanfeng Ding wrote the manuscript. Ashutosh Sarker, Hanfeng Ding, and Xuxiao Zong revised the manuscript.

REFERENCES

- [1] S. Kumar, K. Rajendran, J. Kumar, A. Hamwieh, M. Baum, Current knowledge in lentil genomics and its application for crop improvement, *Front. Plant Sci.* 6 (2015), 78.
- [2] M. Lombardi, M. Materne, N.O.I. Cogan, M. Rodda, H.D. Daetwyler, A.T. Slater, J.W. Forster, S. Kaur, Assessment of genetic variation within a global collection of lentil (*Lens culinaris* Medik.) cultivars and landraces using SNP markers, *BMC Genet.* 15 (2014), 150.
- [3] FAOSTAT, 2018, FAO, Rome, Italy, <http://www.fao.org/faostat/en/#home>.
- [4] A.G. Sharpe, L. Ramsay, L.A. Sanderson, M.J. Fedoruk, W.E. Clarke, R. Li, S. Kagale, P. Vijayan, A. Vandenberg, K.E. Bett, Ancient orphan crop joins modern era: gene-based SNP discovery and mapping in lentil, *BMC Genomics* 14 (2013), 192.
- [5] S. Kumar, A. Hamweih, A. Manickavelu, J. Kumar, T.R. Sharma, M. Baum, Advances in lentil genomics, in: S. Gupta, N. Nadarajan, D.S. Gupta (Eds.), *Legumes in Omics Era*, Springer Science + Business Media, New York, USA 2014, pp. 111–130.
- [6] K. Bett, L. Ramsay, C. Chan, A. Sharpe, D. Cook, R.V. Penmetsa, P. Chang, C.J. Coyne, R. McGee, D. Main, J. Dolezel, D. Edwards, S. Kaur, S. Kumar, S.M. Udupa, A. Vandenberg, The Lentil Genome-From the Sequencer to the Field. Paper Presented at the International Conference on Pulses Conference, April 18–20, 2016, Marrakesh, Morocco, 2016.
- [7] L.A. Sanderson, C.T. Caron, R. Tan, Y. Shen, R. Liu, K.E. Bett, KnowPulse: a web-resource focused on diversity data for pulse crop improvement, *Front. Plant Sci.* 10 (2019), 965.
- [8] R. Garg, M. Jain, RNA-Seq for transcriptome analysis in non-model plants, in: R.J. Rose (Ed.), *Legume Genomics: Methods and Protocols*, Humana Press, Totowa, USA 2013, pp. 43–58.
- [9] Z. Wang, M. Gerstein, M. Snyder, RNA-Seq: a revolutionary tool for transcriptomics, *Nat. Rev. Genet.* 10 (2009) 57–63.
- [10] J. Zhang, R. Chiodini, A. Badr, G. Zhang, The impact of next-generation sequencing on genomics, *J. Genet. Genomics* 38 (2011) 95–109.
- [11] M. Xiao, Y. Zhang, X. Chen, E.J. Lee, C.J.S. Barber, R. Chakrabarty, I. Desgagné-Penix, T.M. Haslam, Y.B. Kim, E. Liu, G. MacNevin, S. Masada-Atsumi, D.W. Reed, J.M. Stout, P. Zerbe, Y. Zhang, J. Bohlmann, P.S. Covelto, V.D. Luca, J.E. Page, D.-K. Ro, V.J.J. Martin, P.J. Facchini, C.W. Sensen, Transcriptome analysis based on next-generation sequencing of non-model plants producing specialized metabolites of biotechnological interest, *J. Biotechnol.* 166 (2013) 122–134.
- [12] S. Sudheesh, T.I. Sawbridge, N.O. Cogan, P. Kennedy, J.W. Forster, S. Kaur, De novo assembly and characterisation of the field pea transcriptome using RNA-Seq, *BMC Genomics* 16 (2015), 611.
- [13] S. Kaur, N.O. Cogan, L.W. Pemberton, M. Shinozuka, K.W. Savin, M. Materne, J.W. Forster, Transcriptome sequencing of lentil based on second-generation technology permits large-scale unigene assembly and SSR marker discovery, *BMC Genomics* 12 (2011), 265.
- [14] S. Kaur, N.O.I. Cogan, A. Stephens, D. Noy, M. Butsch, J.W. Forster, M. Materne, EST-SNP discovery and dense genetic mapping in lentil (*Lens culinaris* Medik.) enable candidate gene selection for boron tolerance, *Theor. Appl. Genet.* 127 (2014) 703–713.
- [15] P. Verma, N. Shah, S. Bhatia, Development of an expressed gene catalogue and molecular markers from the de novo assembly of short sequence reads of the lentil (*Lens culinaris* Medik.) transcriptome, *Plant Biotechnol. J.* 11 (2013) 894–905.
- [16] S. Sudheesh, P. Verma, J. Forster, N. Cogan, S. Kaur, Generation and characterisation of a reference transcriptome for lentil (*Lens culinaris* Medik.), *Int. J. Mol. Sci.* 17 (2016), 1877.
- [17] A. Castillo, H. Budak, R.K. Varshney, G. Dorado, A. Graner, P. Hernandez, Transferability and polymorphism of barley EST-SSR markers used for phylogenetic analysis in *Hordeum chilense*, *BMC Plant Biol.* 8 (2008), 97.
- [18] A. Chandra, M.P. Grisham, Y.B. Pan, Allelic divergence and cultivar-specific SSR alleles revealed by capillary electrophoresis using fluorescence-labeled SSR markers in sugarcane, *Genome* 57 (2014) 363–372.
- [19] C. Liang, T. Wan, S. Xu, B. Li, X. Li, Y. Feng, Y. Cai, Molecular identification and genetic analysis of cherry cultivars using capillary electrophoresis with fluorescence-labeled SSR markers, *3 Biotech* 8 (2017), 16.
- [20] K. Semagn, R. Babu, S. Hearne, M. Olsen, Single nucleotide polymorphism genotyping using Kompetitive Allele Specific PCR (KASP): overview of the technology and its application in crop improvement, *Mol. Breed.* 33 (2014) 1–14.
- [21] J.T. Michael, High-throughput SNP genotyping to accelerate crop improvement, *Plant Breed. Biotechnol.* 2 (2014) 195–212.
- [22] A. Rasheed, W. Wen, F. Gao, S. Zhai, H. Jin, J. Liu, Q. Guo, Y. Zhang, S. Dreisigacker, X. Xia, Z. He, Development and validation of KASP assays for genes underpinning key economic traits in bread wheat, *Theor. Appl. Genet.* 129 (2016) 1843–1860.
- [23] C.T. Tan, H. Yu, Y. Yang, X. Xu, M. Chen, J.C. Rudd, Q. Xue, A. M.H. Ibrahim, L. Garza, S. Wang, M.E. Sorrells, S. Liu, Development and validation of KASP markers for the greenbug resistance gene Gb7 and the hessian fly resistance gene H32 in wheat, *Theor. Appl. Genet.* 130 (2017) 1867–1884.
- [24] B.T. Ertiro, V. Ogugo, M. Worku, B. Das, M. Olsen, M. Labuschagne, K. Semagn, Comparison of Kompetitive Allele Specific PCR (KASP) and genotyping by sequencing (GBS) for quality control analysis in maize, *BMC Genomics* 16 (2015), 908.
- [25] J. Mammadov, W. Chen, J. Mingus, S. Thompson, S. Kumpatla, Development of versatile gene-based SNP assays in maize (*Zea mays* L.), *Mol. Breed.* 29 (2012) 779–790.
- [26] G. Patil, J. Chaudhary, T.D. Vuong, B. Jenkins, D. Qiu, S. Kadam, G.J. Shannon, H.T. Nguyen, Development of SNP genotyping assays for seed composition traits in soybean, *Int. J. Plant Genomics* 2017 (2017), 6572969.
- [27] Z. Shi, S. Liu, J. Noe, P. Arelli, K. Meksem, Z. Li, SNP identification and marker assay development for high-throughput selection of soybean cyst nematode resistance, *BMC Genomics* 16 (2015), 314.
- [28] S.C.M. Leal-Bertioli, U. Cavalcante, K. Shirasawa, E.G. Gouvea, P.M. Guimarães, C. Ballén-Taborda, S.A. Jackson, D.J. Bertioli, M.C. Moretzsohn, Identification of QTLs for rust resistance in the peanut wild species “*Arachis magna*” and the development of KASP markers for marker-assisted selection, *G3-Genes Genomes Genet.* 5 (2015) 1403–1413.
- [29] R. Chopra, C.E. Simpson, A. Hillhouse, P. Payton, J. Sharma, M. D. Burow, SNP genotyping reveals major QTLs for plant architectural traits between A-genome peanut wild species, *Mol. Gen. Genomics* 293 (2018) 1477–1491.
- [30] U. Majeed, E. Darwish, S.U. Rehman, X. Zhang, Kompetitive allele specific PCR (KASP): a singleplex genotyping platform and its application, *J. Agric. Sci.* 11 (2018) 11–20.

- [31] M.J. Fedoruk, A. Vandenberg, K.E. Bett, Quantitative trait loci analysis of seed quality characteristics in lentil using single nucleotide polymorphism markers, *Plant Genome* 6 (2013) 1–10.
- [32] M.S. Rodda, S. Sudheesh, M. Javid, D. Noy, A. Gnanasambandam, A.T. Slater, G.M. Rosewarne, S. Kaur, Breeding for boron tolerance in lentil (*Lens culinaris* Medik.) using a high-throughput phenotypic assay and molecular markers, *Plant Breed.* 137 (2018) 492–501.
- [33] M.G. Grabherr, B.J. Haas, M. Yassour, J.Z. Levin, D.A. Thompson, I. Amit, X. Adiconis, L. Fan, R. Raychowdhury, Q. Zeng, Z. Chen, E. Mauceli, N. Hacohen, A. Gnirke, N. Rhind, F. Palma, B.W. Birren, C. Nusbaum, K. Lindblad-Toh, N. Friedman, A. Regev, Full-length transcriptome assembly from RNA-Seq data without a reference genome, *Nat. Biotechnol.* 29 (2011) 644–652.
- [34] N.M. Davidson, A. Oshlack, Corset: enabling differential gene expression analysis for de novo assembled transcriptomes, *Genome Biol.* 15 (2014), 410.
- [35] B. Buchfink, C. Xie, D.H. Huson, Fast and sensitive protein alignment using DIAMOND, *Nat. Methods* 12 (2014) 59–60.
- [36] S.F. Altschul, T.L. Madden, A.A. Schäffer, J. Zhang, Z. Zhang, W. Miller, D.J. Lipman, Gapped BLAST and PSI-BLAST: a new generation of protein database search programs, *Nucleic Acids Res.* 25 (1997) 3389–3402.
- [37] R.D. Finn, J. Tate, J. Misty, P.C. Coghill, S.J. Sammut, H.R. Hotz, G. Ceric, K. Forslund, S.R. Eddy, E.L.L. Sonnhammer, A. Bateman, The Pfam protein families database, *Nucleic Acids Res.* 36 (2008) D281–D288.
- [38] Y. Moriya, M. Itoh, S. Okuda, A.C. Yoshizawa, M. Kanehisa, KAAS: an automatic genome annotation and pathway reconstruction server, *Nucleic Acids Res.* 35 (2007) W182–W185.
- [39] S. Götz, J.M. García-Gómez, J. Terol, T.D. Williams, S.H. Nagaraj, M.J. Nueda, M. Robles, M. Talón, J. Dopazo, A. Conesa, High-throughput functional annotation and data mining with the Blast2GO suite, *Nucleic Acids Res.* 36 (2008) 3420–3435.
- [40] S. Rozen, H. Skaletsky, Primer3 on the WWW for general users and for biologist programmers, *Methods Mol. Biol.* 132 (2000) 356–386.
- [41] A. Asadi, A. Ebrahimi, S. Rashidi-Monfared, M. Basiri, J. Akbari-Afjani, Comprehensive functional analysis and mapping of SSR markers in the chickpea genome (*Cicer arietinum* L.), *Comput. Biol. Chem.* 84 (2020), 107169.
- [42] D. Nigam, S. Saxena, G. Ramakrishna, A. Singh, N.K. Singh, K. Gaikwad, De novo assembly and characterization of *Cajanus scarabaeoides* (L.) thours transcriptome by paired-end sequencing, *Front. Mol. Biosci.* 4 (2017), 48.
- [43] H. Li, B. Handsaker, A. Wysoker, T. Fennell, J. Ruan, N. Homer, G. Marth, G. Abecasis, R. Durbin, 1000 Genome Project Data Processing Subgroup, The sequence alignment/map format and SAMtools, *Bioinformatics* 25 (2009) 2078–2079.
- [44] A. McKenna, M. Hanna, E. Banks, A. Sivachenko, K. Cibulskis, A. Kernysky, K. Garimella, D. Altshuler, S. Gabriel, M. Daly, M.A. DePristo, The genome analysis toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data, *Genome Res.* 20 (2010) 1297–1303.
- [45] K. Liu, S.V. Muse, PowerMarker: an integrated analysis environment for genetic marker analysis, *Bioinformatics* 21 (2005) 2128–2129.
- [46] M. Nei, A.K. Roychoudhury, Gene differences between Caucasian, Negro, and Japanese populations, *Science* 177 (1972) 434–436.
- [47] K. Tamura, G. Stecher, D. Peterson, A. Filipski, S. Kumar, MEGA6: molecular evolutionary genetics analysis version 6.0, *Mol. Biol. Evol.* 30 (2013) 2725–2729.
- [48] J.K. Pritchard, M. Stephens, P. Donnelly, Inference of population structure using multilocus genotype data, *Genetics* 155 (2000) 945–959.
- [49] D. Falush, M. Stephens, J.K. Pritchard, Inference of population structure using multilocus genotype data: linked loci and correlated allele frequencies, *Genetics* 164 (2003) 1567–1587.
- [50] G. Evanno, S. Regnaut, J. Goudet, Detecting the number of clusters of individuals using the software STRUCTURE: a simulation study, *Mol. Ecol.* 14 (2005) 2611–2620.
- [51] D.A. Earl, B.M. von Holdt, Structure Harvester: a website and program for visualizing STRUCTURE output and implementing the Evanno method, *Conserv. Genet. Resour.* 4 (2012) 359–361.
- [52] H. Morgil, M. Tardu, G. Cevahir, İ.H. Kavakli, Comparative RNA-Seq analysis of the drought-sensitive lentil (*Lens culinaris*) root and leaf under short- and long-term water deficits, *Funct. Integr. Genomics* 19 (2019) 715–727.
- [53] A. Kahriman, H. Yilmaz Temel, A. Aydoğan, M.B. Tanyolac, Major quantitative trait loci for flowering time in lentil, *Turk. J. Agric. For.* 39 (2015) 588–595.
- [54] W. Erskine, Y. Adham, L. Holly, Geographic distribution of variation in quantitative traits in a world lentil collection, *Euphytica* 43 (1989) 97–103.
- [55] A. Sarker, W. Erskine, B. Sharma, M. Tyagi, Inheritance and linkage relationship of days to flower and morphological loci in lentil (*Lens culinaris* Medikus subsp. *culinaris*), *J. Hered.* 90 (1999) 270–275.
- [56] Y. Kobayashi, H. Kaya, K. Goto, M. Lwabuchi, T. Araki, Activation tagging of the floral inducer FT, *Science* 286 (1999) (1962–1965).
- [57] Y. Kobayashi, H. Kaya, K. Goto, M. Lwabuchi, T. Araki, A pair of related genes with antagonistic roles in mediating flowering signals, *Science* 286 (1999) (1960–1962).
- [58] V. Hecht, R.E. Laurie, J.K. Vander Schoor, R. Stephen, C.L. Knowles, L.C. Liew, F.C. Sussmilch, I.C. Murfet, R.C. Macknight, J.L. Weller, The pea GIGAS gene is a FLOWERING LOCUS T homolog necessary for graft-transmissible specification of flowering but not for responsiveness to photoperiod, *Plant Cell* 23 (2011) 147–161.
- [59] M. Tadege, K.S. Mysore, J. Putterill, J.L. Weller, R.C. Macknight, R.E. Laurie, P. Diwadkar, M. Jaudal, L. Zhang, V. Hecht, J. Wen, The medicago FLOWERING LOCUS T homolog, MtFTa1, is a key regulator of flowering time, *Plant Physiol.* 156 (2011) 2207–2224.
- [60] E. Zietkiewicz, A. Rafalski, D. Labuda, Genome fingerprinting by simple sequence repeat (SSR)-anchored polymerase chain reaction amplification, *Genomics* 20 (1994) 176–183.
- [61] G.Y. Siew, W.L. Ng, S.W. Tan, N.B. Alitheen, S.G. Tan, S.K. Yeap, Genetic variation and DNA fingerprinting of durian types in Malaysia using simple sequence repeat (SSR) markers, *PeerJ* 6 (2018), 4266.
- [62] T. Yang, L. Fang, X. Zhang, J. Hu, S. Bao, J. Hao, L. Li, Y. He, J. Jiang, F. Wang, S. Tian, X. Zong, High-throughput development of SSR markers from pea (*Pisum sativum* L.) based on next generation sequencing of a purified Chinese commercial variety, *PLoS One* 10 (2015), e0139775.
- [63] Ş. Göl, M. Göktay, J. Allmer, S. Doğanlar, A. Frary, Newly developed SSR markers reveal genetic diversity and geographical clustering in spinach (*Spinacia oleracea*), *Mol. Genet. Genomics* 292 (2017) 847–855.
- [64] A. Ohshima, K. Shirasawa, H. Matsunaga, S. Negoro, K. Miyatake, H. Yamaguchi, T. Nunome, H. Iwata, H. Fukuoka, T. Hayashi, Bayesian QTL mapping using genome-wide SSR markers and segregating population derived from a cross of two commercial F1 hybrids of tomato, *Theor. Appl. Genet.* 130 (2017) 1601–1616.
- [65] S. Taheri, T. Lee Abdullah, M.Y. Rafii, J.A. Harikrishna, S.P.O. Werbrouck, C.H. Teo, M. Sahebi, P. Azizi, De novo assembly of transcriptomes, mining, and development of novel EST-SSR markers in *Curcuma alismatifolia* (Zingiberaceae family) through Illumina sequencing, *Sci. Rep.* 9 (2019), 3047.

- [66] M. Bakır, A. Kahraman, Development of new SSR (simple sequence repeat) markers for lentils (*Lens culinaris* Medik.) from genomic library enriched with AG and AC microsatellites, *Biochem. Genet.* 57 (2019) 338–353.
- [67] A. Singh, V. Sharma, H.K. Dikshit, M. Aski, H. Kumar, N. Thirunavukkarasu, B.S. Patil, S. Kumar, A. Sarker, Association mapping unveils favorable alleles for grain iron and zinc concentrations in lentil (*Lens culinaris subsp. culinaris*), *PLoS One* 12 (2017), e0188296.
- [68] P.M. Koul, S. Kumar, V. Sharma, T.R. Sharma, M. Rana, R.K. Chahota, Analysis of genetic structure and interrelationships in lentil species using morphological and SSR markers, *3 Biotech* 7 (2017) (83–83).
- [69] E.L. Patterson, M.B. Fleming, K.C. Kessler, S.J. Nissen, T.A. Gaines, A KASP genotyping method to identify northern watermilfoil, eurasian watermilfoil, and their interspecific hybrids, *Front. Plant Sci.* 8 (2017), 752.
- [70] P. Khera, H.D. Upadhyaya, M.K. Pandey, M. Roorkiwal, M. Sriswathi, P. Janila, Y. Guo, M.R. McKain, E.D. Nagy, S.J. Knapp, J. Leebens-Mack, J.A. Conner, P. Ozias-Akins, R.K. Varshney, Single nucleotide polymorphism-based genetic diversity in the reference set of peanut (*Arachis* spp.) by developing and applying cost-effective kompetitive allele specific polymerase chain reaction genotyping assays, *Plant Genome* 6 (2013) 1–11.
- [71] D.L. Lister, H. Jones, M.K. Jones, D.M. O'Sullivan, J. Cockram, Analysis of DNA polymorphism in ancient barley herbarium material: validation of the KASP SNP genotyping platform, *Taxon* 62 (2013) 779–789.
- [72] X. Hao, T. Yang, R. Liu, J. Hu, Y. Yao, M. Burlyaeva, Y. Wang, G. Ren, H. Zhang, D. Wang, J. Chang, X. Zong, An RNA sequencing transcriptome analysis of grasspea (*Lathyrus sativus* L.) and development of SSR and KASP markers, *Front. Plant Sci.* 8 (2017), 1873.
- [73] Y. Li, W. Li, C. Zhang, L. Yang, R. Chang, B.S. Gaut, L. Qiu, Genetic diversity in domesticated soybean (*Glycine max*) and its wild progenitor (*Glycine soja*) for simple sequence repeat and single-nucleotide polymorphism loci, *New Phytol.* 188 (2010) 242–253.