Molecular Genetic Studies on Vicia Faba

By

Kareem Ahmed Abdel-Maksoud Yousef

B.Sc. Agric. Sci., Genetics , Ain Shams University (2010)

A Thesis
submitted in partial fulfillment of the requirements
for the Degree of

Master of scince

In

Genetics and Genetic Engineering

To

Department of Genetics and Genetic Engineering

Faculty of Agriculture, Moshtohor,

Benha University

(2022)

Abstract

The purpose of this study was to create a new faba bean genetic database simple sequence repetition (SSR) primers and categorise them based on the genes and biological processes that they target. To find probable SSRs, researchers analysed previously published 75,605 and 148,196 faba sequences in the genome and transcriptome, respectively. The total number of SSRs found in each dataset was 25,502 and 12,319, respectively. Trinucleotides have the most repeat counts, followed by dinucleotides, according to the distribution of different repeat classes. Only 238 (21.8%) of the 1091 PCR primers were successful designed from the recovered genic SSR sequences target genomic sequences, while the remaining 853 PCR primers target transcriptome sequences. Our SSR primers addressed roughly 897 genes, according to the annotation of gene-targeted SSRs. There are around 1890 gene ontology (GO) identification codes available. The GO keywords were dispersed across many molecular cell characteristics. There were 554 technical words, 196 domains, and 160 molecular feature phrases with the most redundancy. The molecular function, cellular component, and biological process were among the GO codes that pertained to the general level of GO (544, 670, and 676 GOs, respectively). To test 12 Egyptian faba bean genotypes, 27 SSR PCR primers were created. Approximately 11 SSRs produced one to two PCR bands, whereas some only produced one sharp band with a varied band size. A total of 13 polymorphic primers were utilized in this study. The information content of the polymorphism was 0.3, indicating that it was fairly informative. Five ISSR and six SCoT primers were used to investigate the similarity and relationship among the twelve faba bean genotypes, gave positive results. Using ISSR generated a total of (71 bands) with an average of 14.2 bands/primer,

while a total of 91 bands were amplified using SCoT with an average of 16.2 bands/primer. The results of SCoT and ISSR for genetic diversity among the twelve faba bean genotypes indicated that the genetic similarity ranged from 0.77 to 0.92 for ISSR, while the genetic similarity ranged from 0.73 to 0.91 for SCoT.
Keywords : SSR, Database, PCR, Egypt, Faba bean, Genetic diversity.

List of Content

List of Tables

List of Figures

SUMMARY

This investigation was carried out in the Molecular Genetics and Genome Mapping Lab., Agricultural Genetic Engineering Research Institute (AGERI), Agricultural Research Center (ARC), Giza, Egypt, and The International Center for Agricultural Research in the Dry Areas (ICARDA) in Cairo, Egypt, during the period from 2017 to 2020. The aim of this work was to evaluate the genetic diversity and phylogenetic relationship among twelve faba bean genotypes using molecular markers and SSR-PCR technique to identify polymorphism, to create a new faba genetic database simple sequence repetition (SSR) primers and categorise them based on the genes and biological processes that they target, and to create genomic markers using faba bean genomic and transcriptomic data that could be beneficial for identifying and improving faba genetic resources. The goal is to make a new genetic database of SSR PCR primers for target genes and cell processes that can be annotated and categorized. In addition, the genetic structure of several selected faba bean genotypes, as well as the diversity and efficacy of SSR markers used to faba bean.

To achieve this aim, we used 12 faba bean genotypes from Field Crops Research Institute, Agricultural Research Center (ARC), Egypt, Wady-1, Masr-1, Noubaria-3, Noubaria-2, Noubaria-1, Sakha-1, Sakha-3, Sakha-4, Giza-716, Giza-843, Giza-40, and Giza-3.

The obtained results could be summarized as follows:

1- Simple sequence repeats identification

In this investigation, there were 156.5 MB of transcriptomic data and 82 MB of genomic data downloaded in total. A total of 25,502 bp and 12,319 bp of SSR motifs were identified. The biggest repeat count was di-nucleotide, followed by tri-nucleotide, according to the

allocation of different types of repeats in genomic and transcriptome samples.

The frequency of discovered SSR patterns in genomic sequences revealed that the AG/CT di-nucleotide had the most common repeat (16.17%), followed by the tri-nucleotide sequence AAG / CTT (6.61%).

2- SSR primer design

Our primer design method resulted in the creation of 1091 PCR primers, 238 of them target genomic sequences and 853 of them target transcriptome sequences. The average PCR product size was 200 bp, with a range of 150 to 368 bp. With 129 and 50 PCR primers, for designed primers, the most popular SSR types were TC/AG and TTC/AAG. SSR patterns with 5 unit patterns are targeted by 512 PCR primers, while 318 target patterns with 6 components are targeted by 318 PCR primers

3- Gene annotation

The annotation of gene targeted SSRs has shown that approximately 897 genes are specific through our engineered SSR primers. Genes belonging to the kinase, zinc finger, leucine rich, and phosphatase families are the destination of the largest amount of PCR primers, with 46, 24, 17, and 11 PCR primers targeting these genes. The disease resistance gene analog (RGA) gene category is targeted with 10 SSR primers. One thousand eight hundred ninety gene ontology (GO) identification codes were obtained using gene ontology and biological pathway analysis. These GOs feature prevalent keywords divided between distinct cell molecular functions, where the highest number of redundant sentences is the technical word, domain, and molecular function phrases with 554, 196, and 160 GOs, respectively. The high redundancy of the technical term of genetic ontology is a semantic unit

so that it can clearly be regarded as a GO word. Over redundancy of the terms of the GO domains indicates that the majority of SSR primers target protein-like domain sequences that could check their specificity for functional genes.

According to the Unipart database, the aspects of molecular activity identify operations that occur at the molecular stage, such as “catalysis” or “transport” This could show that the elevated volume of the intended primers targets sequences with the feasible functional operation. These GOs belong to the overall level of gene ontology, such as molecular structure, cellular component, and biological process with 544, 670, and 676 GOs, respectively. In the case of molecular function, the maximum number of GOs pertained to binding and catalytic activity functions at 257 and 209, respectively. About 33 primers are specific for binding protein activities, including calmodulin, RNA, nucleotide, and ATP binding proteins. The minimum amount of GOs corresponds to the antioxidant activity and the GOs 3 and 7 molecular function regulators. At the level of the cellular component, the maximum number of GOs belongs to the cell part with 238 GOs, while the minimum number of GOs belongs to the mitochondrial matrix and the microtubule with 2 GOs for each. In the case of biological processes, the cellular process and the metabolic process have the maximum number of GOs with208 and 168 GOs, respectively. For instance, about 37 SSR primers are associated with faba cell chloroplast and mitochondria. We have designed 164 SSR primers that are specific for different faba enzymes. The enzyme commission number (EC number) has been obtained for 73 genes; among them, 34 IDs were transferases and 24 hydrolases enzymes.

4- **Validation of PCR primers and molecular analysis**

PCR primers validation and molecular analysis Twenty seven SSR PCR primers were synthesized for in vitro validation. These primers are linked to resistance and plant development linked genes such as NBS-LRR and F box gene families.

All SSR primers have been effectively implemented and generated scorable PCR bands. Among 27 SSR primers, 13 were polymorphic (48%). Eleven SSR primers offered one to two PCR bands, while others offered only one strong band with a polymorphic band width. The information content of polymorphism (PIC) measures the capacity of the marker to identify genetic variation among the diversity group studied. In addition, marker information levels range from 0 to 1 where markers with a PIC greater than 0.5 are extremely informative, while the PIC value between 0.5 and 0.25 indicates mild informativeness. The numbers of the SSR PIC varied from 0 to 0. 57. If we could ignore the monomorphic indicators (with a value of 0 PIC), the average of our PIC scores would be 0.3, which would imply mild in formativeness. The phylogenetic tree built using SSR markers consists of three distinct clusters in which Sakha1 and Noubaria 1 are divided into two clusters. The other clusters were split into three distinct groups, with Noubaria3, Giza3 and Sakha 3 in one sub cluster, Giza 716 in one sub cluster, and the remaining genotypes in one sub cluster.

5- <u>ISSR and SCoT markers</u>

Five ISSR and six SCoT primers were used to investigate the similarity and relationship among the twelve faba bean genotypes. A total of (71for ISSR) band was amplified with average of 14.2 bands/primer, while total of (91 for SCoT) bands were amplified with average of 16.2 bands/primer. The results data of SCoT and ISSR among the twelve faba bean genotypes the indicating the genetic similarity ranged from 0.77 to 0.92 to ISSR, while the genetic similarity ranged from 0.73 to 0.91 to SCoT.