

**MOLECULAR MARKERS AND BIO-DATA
MINING STUDIES IN *Vicia faba***

By

MORAD MOKHTAR MOKHTAR

B.Sc. Agric. Sci. (Biotechnology), Fac. Agric., Al-Azhar Univ., 2011

M.Sc. Agric. Sci. (Genetics), Fac. Agric., Cairo Univ., 2017

THESIS

**Submitted in Partial Fulfillment of the
Requirements for the Degree of**

DOCTOR OF PHILOSOPHY

In

**Agricultural Sciences
(Genetics)**

**Department of Genetics
Faculty of Agriculture
Cairo University
EGYPT**

2021

Format Reviewer

Vice Dean of Graduate Studies

Name of Candidate: Morad Mokhtar Mokhtar

Degree: Ph.D.

Title of Thesis: Molecular Markers and Bio-data Mining Studies in *Vicia faba*.

Supervisors: Dr. Ebtissam Hussein Aly Hussein

Dr. Salah El-Din Sayed Mohamed El-Assal

Dr. Mohamed Atia Mohamed Atia Omar

Department: Genetics

Branch:

Date: 11 / 4 /2021

ABSTRACT

Faba bean (*Vicia faba* L.) is an essential food and fodder legume crop worldwide due to its high content of proteins and fibers. Molecular markers tools represent an essential key for faba bean breeders toward rapid crop improvement. Despite the lack in the *V. faba* genome resources availability, few transcriptomes and mitochondrial genome sequence data have been released. These data in addition to previously developed genetic linkage maps represent a great raw material for developing functional markers and maps that can accelerate the faba bean breeding programs. In the present study, we present the *Vicia faba* Omics database (*VfODB*) as a comprehensive database integrating germplasm information, expressed sequence tags (ESTs), expressed sequence tags-simple sequence repeats (EST-SSRs), mitochondrial-simple sequence repeats (mtSSRs), microRNA-target markers and genetic maps in faba bean. Besides, Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway-based markers and maps are integrated as a novel class of annotation-based markers/maps. Collectively, we developed 31536 EST markers, 9071 EST-SSR markers and 3023 microRNA-targets markers were developed based on the *V. faba* RefTrans V2 mining. By mapping 7940 EST and 2282 EST-SSR markers against the KEGG pathways database 100 functional maps were successfully developed. Also, 40 mtSSR markers were developed based on mitochondrial genome mining. On the data curation level, 3461 markers representing twelve types of markers (CAPS, EST, EST-SSR, Gene marker, INDEL, Isozyme, ISSR, RAPD, SCAR, RGA, SNP and SSR), which mapped across 18 *V. faba* genetic linkage maps were retrieved. Also, the *VfODB* (<http://vfodb.easyomics.org/>) was designed to provide two user-friendly tools to identify, classify SSR motifs and *in silico* amplify their targets. This *VfODB* is expected to serve as a powerful database and helpful platform for faba bean research community as well as breeders interested in Genomics-Assisted Breeding.

Key words: Faba bean, ESTs, EST-SSRs, KEGG, SNPs, pathways, database.

CONTENTS

	Page
INTRODUCTION	1
REVIEW OF LITERATURE	5
1. Faba bean overview	5
2. Origin and distribution	6
3. Genetic diversity	8
4. Molecular markers, genetic maps and QTL mapping	11
MATERIALS AND METHODS	16
1. Source of mitochondrial and transcriptome sequences	16
2. Source of previously mapped DNA markers	16
3. Bio-data mining, curation and database construction ...	17
a. Mitochondrial-based markers development.....	17
b. Transcriptome-based markers.....	18
c. KEGG pathway-based markers.....	32
d. Genetic maps-based markers.....	33
4. Validation of the designed EST and EST-SSR primers ..	36
RESULTS AND DISCUSSION	42
1. <i>V. faba</i> mitochondrial genome mining	42
2. <i>V. faba</i> Ref-transcriptome mining	47
a. Distribution of the different SSR motifs in <i>V. faba</i> Ref-transcriptome.....	47
b. EST-SSR primers design.....	53
c. Development of EST markers.....	58

	Page
d. Devolvement of microRNA-target markers.....	59
3. <i>In vitro</i> validation of the designed primers.....	61
4. KEGG pathway-based markers.....	74
5. Genetic maps data curation.....	109
6. Constriction of the <i>VfODB</i> web-based database.....	110
CONCLUSIONS.....	172
SUMMARY.....	174
REFERENCES.....	184
ARABIC SUMMARY	

LIST OF TABLES

No.	Title	Page
1.	microRNA ID, target sequence and the target protein used to predict microRNA-target markers within the <i>V. faba</i> RefTrans. V2.....	22
2.	The previously developed and published genetic maps (GM) of <i>Vicia faba</i> and included in the VfODB	35
3.	The names, sequences, melting temperatures, sizes and GC content of the EST primer pairs used in the in vitro validation of the designed EST primers in the present investigation	37
4.	The names, repeat type, repeat sequence, forward (F), reverse (R), melting temperatures, sizes and GC content of the EST-SSR primer pairs used in the in vitro validation of the designed primers in the present investigation.....	38
5.	The names, repeat type, repeat sequence, forward (F), reverse (R), temperatures (T _m °C), sizes and GC content of the mitochondrial SSR primer pairs used in in vitro validation of the designed primers in the present investigation.....	39
6.	The details of the identified SSRs in <i>Vicia faba</i> mitochondrial genome	43
7.	The primers ID, repeat sequence, primer sequence forward (F), primer sequence reverse (R), and annealing temperatures (T _m °C) of the in silico designed mitochondrial SSR primers	45
8.	Summary of the identified SSR motifs in <i>V. faba</i> Ref-transcriptome	48

No.	Title	Page
9.	The identified SSRs with compound repeats in the <i>V. faba</i> ref-transcriptome	48
10.	The details of the identified SSRs in <i>V. faba</i> Ref-transcriptome	54
11.	Primer ID, biological process, protein ID, protein name, primer type, amplified bands range, polymorphic bands length, and number of polymorphic loci for in vitro validated EST and EST-SSR primers	71
12.	Primer ID, genomic region, gene name, gene product, amplified bands length, polymorphic bands length, and number of polymorphic bands for in vitro validated mitochondrial SSR primers	73
13.	Statistics of the <i>V. faba</i> KEGG pathway-based maps	75
14.	The mapped enzymes and their assigned KEGG pathways..	84
15.	Summary of the genetic maps data curation in the <i>Vf</i> ODB database	110
16.	Population type and size, markers type, linkage group (LG), map length, mapped traits and reference of the genetic maps used in the present study	111
17.	Marker name, marker type and reference of 1322 unique markers representing twelve types of markers (CAPS, EST, EST-SSR, Gene marker, INDEL, Isozyme, ISSR, RAPD, SCAR, RGA, SNP and SSR) mapped across the 18 genetic linkage maps curated in the present study.....	115

LIST OF FIGURES

No.	Title	Page
1.	The workflow of mitochondrial-based markers development.....	19
2.	The workflow of transcriptome-based markers	20
3.	The workflow of transcriptome-based markers	33
4.	The workflow of genetic maps-based markers	34
5.	Histogram revealing the distribution of the various SSR classes in the genic and intergenic regions of the <i>V. faba</i> mitochondrial genome	44
6.	Histogram illustrating the distribution of the various SSR classes in <i>V. faba</i> Ref-transcriptome	51
7.	Histogram revealing the total number of EST sequences, number of SSR containing sequences, number of designed EST-SSR primers, number of EST-SSR with gene-ontology and number of EST-SSR assigned to enzymes.....	57
8.	Histogram illustrating the total number of EST sequences, number of designed EST primers, number of EST with gene-ontology and number of EST assigned to enzymes	58
9.	Histogram revealing the total number of EST sequences, number of microRNA-targets and number of microRNA-target markers	60

No.	Title	Page
10.	Statistics layout of the <i>V. faba</i> Ref-transcriptome mining-based markers. The circles from the inner to outer represent: 1- Total No. of EST sequence used in the analysis (37,378), 2- The distribute of designed EST primers (31,536), 3- Distribute of EST primers with GO ontology (17,081), 4- Distribute of EST primers assigned with enzymes (7,940), 5- Distribute of predicted micro-RNA (21,236), 6- Distribute of micro-RNA primers (3,023), 7- Distribute of designed EST-SSR primers (9,071), 8- Distribute of EST-SSR primers with GO ontology (5,217) and 9- Distribute of EST-SSR primers assigned with enzymes (2,282)	62
11.	Agarose gels showing the polymorphic patterns of the 9 EST, EST-SSR, and Mito-SSR markers in the genomic DNA of the 20 faba bean cultivars. Lane (M) Thermo Scientific GeneRuler 100bp DNA Ladder, lanes 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19 and 20 represent cvs. Misr1, Misr3, Wadi1, Giza 843, Nobaria 1, Nobaria 2, Nobaria 3, Sakhi 1, Sakhi 3, FAB 6719, FAB 5911, FAB 495, FAB 391, FAB 6901, FAB 6004, FAB 7391, FAB 6391, FAB 222, FAB 6899 and FAB 297, respectively.....	64
12.	Agarose gels showing the monomorphic patterns of the 14 EST, EST-SSR, and Mito-SSR markers in the genomic DNA of the 20 faba bean cultivars. Lane (M) Thermo Scientific GeneRuler 100bp DNA Ladder, lanes 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19 and 20 represent cvs. Misr1, Misr3, Wadi1, Giza 843, Nobaria 1, Nobaria 2, Nobaria 3, Sakhi 1, Sakhi 3, FAB 6719, FAB 5911, FAB 495, FAB 391, FAB 6901, FAB 6004, FAB 7391, FAB 6391, FAB 222, FAB 6899 and FAB 297, respectively.....	67

No.	Title	Page
13.	The annotation of the 26 enzymes assigned using 129 EST and EST-SSR primers in the plant pathogen interaction pathway. Each enzyme code is highlighted with red color	81
14.	The annotation of the 34 enzymes assigned using 223 EST and EST-SSR primers in the plant hormone signal transduction pathway. Each enzyme code is highlighted with red color	83
15.	Screenshot of the VfODB homepage	154
16.	Screenshot for one of the VfODB EST markers search page.....	156
17.	Screenshot for one of the VfODB EST-SSR markers search page	157
18.	Screenshot for one of the VfODB in vitro validated EST primers page	158
19.	Screenshot of the VfODB in vitro validated EST-SSR primers pages	159
20.	Screenshot for one of the VfODB mitochondrial SSR markers page	160
21.	Screenshot of the VfODB <i>in vitro</i> validated mitochondrial SSR primers pages	161
22.	Screenshot for one of the VfODB predicted microRNA-targets page	162
23.	Screenshot for one of the VfODB genetic map page	164
24.	Screenshot for one of the VfODB KEGG pathway-based maps page	165

No.	Title	Page
25.	Screenshot for one of the VfODB MicroSatellite identification tool page	168
26.	Screenshot for one of the VfODB <i>in silico</i> PCR tool page	169