

Development of EST-SSR markers and construction of a linkage map in faba bean (*Vicia faba*)

Walid El-Rodeny¹), Mitsuhiro Kimura²), Hideki Hirakawa²), Attia Sabah¹), Kenta Shirasawa²), Shusei Sato²), Satoshi Tabata²), Shigemi Sasamoto²), Akiko Watanabe²), Kumiko Kawashima²), Midori Kato²), Tsuyuko Wada²), Hisano Tsuruoka²), Chika Takahashi²), Chiharu Minami²), Keiko Nanri²), Shinobu Nakayama²), Mitsuyo Kohara²), Manabu Yamada²), Yoshie Kishida²), Tsunakazu Fujishiro² and Sachiko Isobe^{*2})

¹) Sakha Agricultural Research Station, Field Crop Research Institute, Agricultural Research Center, P.O. Box 33717, Kafr SI-Sheikh, Egypt

²) Kazusa DNA Research Institute, 2-6-7 Kazusa-Kamatari, Kisarazu, Chiba 292-0818, Japan

To develop a high density linkage map in faba bean, a total of 1,363 FBES (Faba bean expressed sequence tag [EST]-derived simple sequence repeat [SSR]) markers were designed based on 5,090 non-redundant ESTs developed in this study. A total of 109 plants of a ‘Nubaria 2’ × ‘Misr 3’ F₂ mapping population were used for map construction. Because the parents were not pure homozygous lines, the 109 F₂ plants were divided into three subpopulations according to the original F₁ plants. Linkage groups (LGs) generated in each subpopulation were integrated by commonly mapped markers. The integrated ‘Nubaria 2’ × ‘Misr 3’ map consisted of six LGs, representing a total length of 684.7 cM, with 552 loci. Of the mapped loci, 47% were generated from multi-loci diagnostic (MLD) markers. Alignment of homologous sequence pairs along each linkage group revealed obvious syntenic relationships between LGs in faba bean and the genomes of two model legumes, *Lotus japonicus* and *Medicago truncatula*. In a polymorphic analysis with ten Egyptian faba bean varieties, 78.9% (384/487) of the FBES markers showed polymorphisms. Along with the EST-SSR markers, the dense map developed in this study is expected to accelerate marker assisted breeding in faba bean.

Key Words: faba bean, EST-SSR, linkage map, comparative mapping, genetic diversity.

Introduction

Fabaceae (*Leguminosae*) is the third largest angiosperm family, containing ca. 18,000 species attributed to 650 genera (Zhu *et al.* 2005). Legumes are important to world agriculture, as they provide biological fixed nitrogen, break the cereal disease cycles, and contribute to locally grown food and feed, including forage (Stoddard *et al.* 2009). Faba bean (*Vicia faba* L.) is an excellent candidate crop for providing protein and starch for human diets and animal feed in many countries, such as China, Ethiopia, Egypt, France, and Australia. Faba bean is a diploid species with 2n = 12 chromosomes, which possesses one of the largest genomes among crop legumes (~13,000 Mb; Johnston *et al.* 1999). Faba bean is a partially cross-pollinated crop, with an average degree of cross-fertilization of approximately 40–50% (Link *et al.* 1994a). Although the breeding of partially heterogeneous synthetic varieties has been repeatedly recommended, most faba bean lines have been bred based on inbred lines due to the difficulty of controlling hybridization

on a large scale (Duc *et al.* 1992, Link *et al.* 1994b). Marker assisted selection (MAS) is expected to change the strategies used for faba bean breeding, as it has in many other crop species. However, the large genome size and the lack of genomic tools have slowed the introduction of MAS to faba bean breeding programs (Duc *et al.* 2010, Terzopoulos *et al.* 2008, Zong *et al.* 2009). Therefore, more reliable and efficient molecular markers are needed for faba bean breeding.

During the past two decades, several different types of molecular markers have been successfully used to characterize genetic diversity in faba bean accessions, including Restriction Fragment Length Polymorphism (RFLP), Amplified Fragment Length Polymorphism (AFLP), Random Amplified Polymorphic DNA (RAPD), Inter Simple Sequence Repeats (ISSR), and Sequence-Specific Amplification Polymorphism (SSAP) markers (Abo El-kheir *et al.* 2010, Gresta *et al.* 2010, Link *et al.* 1995, Ouji *et al.* 2012, Terzopoulos and Bebeli 2008, van de Ven *et al.* 1990, Zeid *et al.* 2003, Zong *et al.* 2009). These studies have been instrumental in elucidating the genetic diversity and relationships among accessions in faba bean *ex situ* germplasm collections. Microsatellites, or simple sequence repeats (SSR), have a number of advantages over other markers, as they have co-dominant inheritance and are relatively abundant,

Communicated by T. Anai

Received April 2, 2014. Accepted June 24, 2014.

*Corresponding author (e-mail: sisobe@kazusa.or.jp)

multiallelic, and readily transferable (Rafalski *et al.* 1996). Recently, large numbers of primer pairs of SSR markers were designed based on Roche 454 transcript sequences (Kaur *et al.* 2012, Yang *et al.* 2012). To the best of our knowledge, approximately 30,000 primer pairs have been designed to date, and 550 SSR markers were shown to be capable of amplification and identification of polymorphisms in faba bean germplasm collections (Akash *et al.* 2012, Gong *et al.* 2010, 2011, Kaur *et al.* 2012, Ma *et al.* 2011, Požárková *et al.* 2002, Yang *et al.* 2012, Zeid *et al.* 2009). These SSR markers are potent tools for genetic analysis, and mapping of these markers onto linkage maps or the faba bean genome would increase their efficiency.

Several genetic linkage maps of faba bean have been published during the past decade. Most of the early published maps exploited a combination of different types of markers, including RAPD, SCAR (sequence characterized amplified regions), AFLP, ITAP (Intron-targeted amplified polymorphic), and SSR markers (Arbaoui *et al.* 2008, Avila *et al.* 2005, Díaz-Ruiz *et al.* 2009, 2010, Ellwood *et al.* 2008, Román *et al.* 2002, 2004, Surahman 2001). Despite the substantial efforts made in previous studies, no linkage map has been published in which the number of LGs converged with the chromosome number of faba bean, i.e., six. In other words, all previously published linkage maps show limited saturation.

The developed linkage maps have been used to identify quantitative trait loci (QTLs) conferring resistance against diseases such as *Orobanche crenata* Forsk. (Díaz-Ruiz *et al.* 2009, 2010, Román *et al.* 2002) and *Ascochyta fabae* Speg. (Avila *et al.* 2004, Roman *et al.* 2003), as well as molecular markers closely linked to a resistance gene against *Uromyces viciae-fabae* (Avila *et al.* 2003). In terms of abiotic stress, Arbaoui *et al.* (2008) studied the quantitative trait loci of frost tolerance and physiologically related traits in faba bean. These QTLs are already available, along with markers linked to a gene controlling growth habit (Avila *et al.* 2006, 2007) or to traits affecting the nutritional value of seeds (Gutierrez *et al.* 2006, 2007, 2008). However, few studies performed to date have focused on improving breeding for agronomic traits using molecular markers. To our knowledge, only Ramsey *et al.* (1995) reported marker assisted selection approach with identified QTLs related to agronomic traits affecting yield, although their results were limited due to the low density of the map used. In addition to the insufficient density of the developed maps, the small number of mapped transferable markers has inhibited the application of identified QTLs to MAS. Therefore, the construction of a linkage map with transferable markers, such as SSR, is needed for faba bean molecular breeding and genetics.

The present study was performed to develop and validate a large number of potential expressed sequence tag (EST)-derived SSR markers in faba bean. We mapped the developed markers onto six LGs of a genetic map. We demonstrated the applicability of the markers and the linkage map

by performing comparative analysis with the model legumes *Medicago truncatula* and *Lotus japonicus*. We also performed polymorphic analysis of Egyptian faba bean cultivars to investigate the transferability of the developed SSR markers. We expect the resulting EST-SSR markers and linkage map to serve as resources for accurate QTL mapping and molecular breeding in the future.

Materials and Methods

Plant materials

A linkage map was constructed using an F₂ mapping population derived from crosses between the Egyptian faba bean varieties ‘Nubaria 2’ and ‘Misr 3’. The female parental variety, ‘Nubaria 2’, is a faba bean cultivar adapted to the Nubaria region in Egypt (northern latitude of 30.1), with characteristics including late flowering, large seeds, and drought tolerance. The male parental variety, ‘Misr 3’, was produced by a breeding program for *Orobanche* tolerance at the Field Crop Research Institute (FCRI), Agricultural Research Center (ARC), in Egypt (northern latitude of 31.1). This variety exhibits early flowering, small seeds, and *Orobanche* tolerance. In each variety, plural plants were used for the parental cross. A total of 109 F₂ plants were obtained from three F₁ plants and used for linkage analysis. The two parental varieties are not pure homozygous lines, as heterozygosity is observed within each variety. Therefore, the amplified fragment sizes of SSR markers were sometimes different among the 109 F₂ plants, depending on the original F₁ plants. Thus, the 109 F₂ plants were divided into three subpopulations (Subpop1, Subpop2, and Subpop3) according to their segregation patterns. Subpop1 consisted of 65 F₂ plants, while Subpop2 and Subpop3 comprised 23 and 21 F₂ plants, respectively.

The allele frequency of the SSR markers developed in this study was investigated in ten Egyptian faba bean varieties as follows: ‘Sakha 1’, ‘Sakha 2’, ‘Sakha 3’, ‘Misr 1’, ‘Misr 3’, ‘Giza 3’, ‘Giza 843’, ‘Nubaria 1’, ‘Nubaria 2’, and ‘Cairo 1’. All of the varieties were bred in FCRI, ARC, in Egypt, and comprised different botanical types. One plant of each variety was used for polymorphic analysis.

Development of EST-SSR markers

Total RNA was extracted from seedlings, leaves, young pods, and flowers of the Japanese faba bean variety ‘Komasakae’ using Plant RNA Purification Reagent (Invitrogen, CA, USA). Purification of polyadenylated RNA and conversion to cDNA were performed as described previously (Asamizu *et al.* 1999). Synthesized cDNA was resolved by 1% agarose gel electrophoresis, and fragments ranging from 1 to 3 kb were recovered. The recovered fragments were cloned into the *Eco* RI-*Xho* I site of the pBluescript II SK- plasmid vector (Stratagene, CA, USA) and introduced into the *E. coli* ElectroTen-Blue strain (Stratagene) by electroporation. For generation of ESTs, plasmid DNAs were amplified from the colonies using

TempliPhi (GE Healthcare UK Ltd, Buckinghamshire, England) and subjected to sequencing using a BigDye Terminator Cycle Sequencing Ready Reaction Kit (Applied Biosystems, CA, USA). The reaction mixtures were run on an automated ABI PRISM 3730 DNA sequencer (Applied Biosystems).

Sequencing chromatograms were converted into nucleotide bases with PHRED (Ewing *et al.* 1998, Ewing and Green 1998), and the sequences derived from the vector and linkers were removed with CROSSMATCH (Ewing and Green 1998). The EST reads were quality-trimmed with TRIM2 (Huang *et al.* 2003) using the Phred quality score ≥ 20 , and ambiguous regions, including more than ten X or N bases, were trimmed. Contiguous, high-quality reads ≥ 100 bp were submitted to the DDBJ/EMBL/GenBank databases under the accession numbers HX900245 to HX915284. The PHRAP program with default parameters was used to cluster faba bean ESTs to identify non-redundant faba bean ESTs (Ewing and Green 1998). A similarity search was performed for the non-redundant faba bean ESTs using the BLASTX program against protein-encoding genes deduced in the genomes of *Arabidopsis thaliana* (TAIR10, Arabidopsis Genome Initiative 2000), *L. japonicus* (release 2.5, Sato *et al.* 2008), *M. truncatula* (release 3.5 version 4: <http://www.medicago.org/genome/>), and *Glycine max* (Glyma1; Schmutz *et al.* 2010). The EST contigs and singlets were classified into KOG categories according to the results of BLASTX searches with an E-value cutoff of $1e-10$ against amino acid sequences in the KOG database (<http://www.ncbi.nlm.nih.gov/COG/>) (Tatusov *et al.* 2003).

Simple sequence repeats (SSRs) ≥ 15 nucleotides in length, which contained all possible combinations of dinucleotide (NN), tri-nucleotide (NNN), and tetra-nucleotide (NNNN) repeats, were identified from the non-redundant faba bean ESTs using the fuzznuc program from EMBOSS (Rice *et al.* 2000) for SSRs within two mismatches. Primer pairs for amplification of SSR-containing regions were designed based on the flanking sequences of each SSR with the aid of the Primer3 program (Rozen and Skaletsky 2000) so that the amplified fragment sizes were between 90 bp and 300 bp in length. The newly developed markers were designated as FBES (Faba Bean EST-derived SSR) markers.

Amplification of SSR-containing regions and detection of polymorphisms

The DNA was extracted from young leaves using a DNeasy Plant Mini Kit (Qiagen Inc, CA, USA). DNA quantification and quality checks were performed using a Nanodrop ND1000 spectrophotometer (Nanodrop Technologies, DE, USA) and 0.8% agarose gel electrophoresis, respectively.

In addition to the developed FBES markers, 7,244 previously published red clover SSR (RCS) markers (Sato *et al.* 2005) and 1,900 white clover SSR (WCS) markers (Isobe *et al.* 2012) were used for verification of the DNA of 'Nubaria 2' and 'Misr 3' by PCR. PCR was performed in a 5 μ l reaction volume containing 0.6 ng of genomic DNA in 1X PCR

buffer (Bioline, London, UK), 3 mM MgCl₂, 0.08U of BIOTAQ DNA polymerase (Bioline), 0.8 mM dNTPs, and 0.4 μ M of each primer. A modified touchdown PCR protocol was followed as described by Sato *et al.* (2005). The PCR products were separated by 10% polyacrylamide gel electrophoresis in TBE buffer.

SSR markers showing solid amplification with the DNAs of parental varieties were subsequently used for polymorphic analysis of six plants each from Subpop1 and Subpop2 to screen polymorphic markers in each subpopulation. The selected polymorphic markers were subjected to subsequent segregation analysis in each subpopulation. The same set of screened polymorphic markers in Subpop2 was used for segregation analysis in Subpop3, as differences in the segregation patterns between Subpop2 and Subpop3 were less than those between Subpop1 and the other subpopulations. For segregation analysis of each subpopulation, the PCR products were separated by 10% polyacrylamide gel electrophoresis in TBE buffer or with an ABI 3730xl fluorescent fragment analyzer (Applied Biosystems) according to the polymorphic fragment sizes of the PCR amplicons. In the former case, the data were analyzed using Polyans software (<http://www.polyans.kazusa.or.jp>). In the latter case, each 5 μ l PCR amplicon was labeled with 10 μ M R6G-ddCTP or R110-ddUTP along with 0.16 unit Klenow fragment, 0.5 μ l 10X Klenow buffer (TAKARA BIO Inc., Shiga, Japan) and 0.16 unit Thermo Sequenase DNA Polymerase (GE Healthcare UK Ltd) before electrophoresis. Polymorphisms were investigated using GeneMapper software version 4.0 (Applied Biosystems).

Linkage analysis

The segregation data were scored using 'F2' population type codes employed in JoinMap[®] analysis (Van Ooijen 2006). Due to the existence of genetic diversity within the parental varieties, it was often difficult to estimate the sources of parental varieties for scored polymorphic alleles. In these cases, the allelic codes were temporarily set to 'a' or 'b' for polymorphic homozygous alleles. The segregation data sets of each subpopulation were individually classified into multiple LGs using the color map method (Kiss *et al.* 1998), which employs a comparison of graphical genotypes of the segregation data. Then, the robustness of the data sets for each linkage group was confirmed by the Grouping Module of JoinMap[®] version 4 (Van Ooijen 2006) using a logarithm of odds (LOD) threshold of 2.0. During the process of color mapping, temporary coded reciprocal genotypes were converted to coupling genotypes. Locus orders of each LG were temporally determined using the Regression Mapping Module of JoinMap[®] with the following parameters: Kosambi's mapping function, LOD ≥ 1.0 , REC frequency ≤ 0.4 , goodness of fit jump threshold for removal of loci = 5.0, number of added loci after which a ripple is performed = 1, and third round = yes.

Subsequently, multiple LGs in each subpopulation were combined to construct an integrated linkage map. Prior to

integration, the genotypes of dominant loci on temporally constructed LGs in each subpopulation were imputed to co-dominant genotypes according to the flanking genotypes of co-dominant loci. The combinations of integrated LGs were determined using commonly mapped markers across the LGs. The locus genotype data in each subpopulation were then combined into one dataset in each integrated LG using the Combine Groups for Mapping Integration Module, followed by locus ordering by the Regression Mapping Module of JoinMap. The parameters used for the mapping module of an integrated map were $\text{LOD} \geq 1.0$, $\text{REC frequency} \leq 0.4$, goodness of fit jump threshold for removal of loci = 5.0, number of added loci after which a ripple is performed = 1, and third round = yes, except for LG3. For the construction of LG3, $\text{LOD} \geq 1.2$ was employed. The adequateness of locus order was confirmed by drawing graphical genotypes in each LG. The subpopulation-specific maps were reconstructed according to the combinations of integrated LGs in the integrated map (details are described in the Results). Different thresholds were employed for LOD and REC values in each LG.

Comparative mapping

Syntenic regions between the genomes of faba bean and two model legumes, *M. truncatula* and *L. japonicus*, were detected by identifying the conservation of the relative locations of genes and genomic regions. The sources of the genome sequences of the two model legumes are described above (see *Development of EST-SSR markers*). The cDNA sequences adjacent to the mapped EST-SSR markers on the faba bean map were compared with the genic sequences in the reference genomes using the BLASTX program with E-value cutoff of $1e-20$. A synteny block between faba bean and *M. truncatula* was defined as the region where five or more conserved homologs were located within a 10-cM region in the faba bean linkage map and a 3-Mb DNA stretch in the *M. truncatula* genome. Between faba bean and *L. japonicus*, a synteny block was defined as the region where three or more conserved homologs were located within a 15-cM region in the faba bean linkage map and a 5 Mb DNA stretch in the *L. japonicus* genome. The identified syntenic regions were plotted using the Cicros program (<http://cicross.ca>).

Polymorphic analysis of FBES markers with ten Egyptian faba bean varieties

Polymorphisms in the ten Egyptian faba bean varieties were investigated with FBES markers that were confirmed to have solid amplification patterns in the parental varieties of the mapping population. The PCR protocol was the same as that used for segregation analysis in the F_2 mapping population. The amplified DNA fragments were separated by an ABI 3730xl fluorescent fragment analyzer (Applied Biosystems), and data analysis was performed with GeneMapper software version 4.0 (Applied Biosystems). The allelic data were converted into a binary matrix using

the scores 1/0 for presence or absence of the peak. The PIC (polymorphic information content) value of each identified peak was calculated using PowerMarker version 3.25 (Liu and Muse 2005). When polymorphic analysis was performed, the FBES markers often identified multiple peaks on each variety, which made it difficult to identify the exact number of loci. Therefore, the mean PIC value of the identified peaks generated by a marker was substituted for the PIC of each marker. The allelic binary data were also analyzed with genetic similarity coefficients between the varieties using the distance analysis module in PowerMarker employing the Nei's gene diversity method (Nei 1973). A UPGMA (unweighted pair group method with arithmetic average) dendrogram was constructed using MEGA version 5.05 (Tamura *et al.* 2011).

Results

Features of faba bean ESTs

A total of 15,202 cDNA clones were sequenced, including 3,936 clones from a flower library, 3,936 clones from a young pod library, 3,604 clones from a root library, 96 clones from a leaf library, 96 clones from a light-grown seedling library, and 3,534 clones from a dark-grown seedling library. Clustering of the EST sequences was performed using the PHRAP program (Ewing and Green 1998). As a result, 5,090 potential non-redundant EST sequences, including 2,325 contigs and 2,765 singletons, were generated (Supplemental Table 1). A total of 4,407,230 qualified bases were obtained; the average GC content was 40.7%.

The 5,090 faba bean non-redundant EST sequences were searched for similarity against proteome databases of two legume genomes (*L. japonicus* and *M. truncatula*) and the *Arabidopsis* genome. The numbers of ESTs showing significant similarity to *L. japonicus*, *M. truncatula*, and *Arabidopsis* were 4,108 (80.7%), 4,536 (89.1%), and 4,104 (80.6%), respectively. To investigate the functional classification of faba bean ESTs, non-redundant EST sequences were compared with the eukaryotic clusters of orthologous groups (KOGs) by BLASTX and classified into KOG categories (Tatusov *et al.* 2003). Among the 5,090 non-redundant faba bean EST sequences, 2,536 showed similarity to genes with KOG categories. The distribution of non-redundant faba bean EST sequences assigned to KOG functional categories is shown in Supplemental Fig. 1.

Microsatellite features and marker development

A total of 128 di-, tri-, and tetra-nucleotide SSRs that were equal to or longer than 15 bp were identified in the non-redundant EST sequences. Assuming the total size of the non-redundant faba bean EST sequences is 4.41 Mbp, the frequency of occurrence of the SSRs in transcribed regions of the faba bean genome was estimated to be one SSR in every 34.4 kb. Di-, tri-, and tetra-nucleotide SSRs accounted for 21.9%, 72.7%, and 5.5% of the identified SSRs, respectively (Table 1). Among the assigned 128 SSR

Table 1. SSRs in the non-redundant faba bean ESTs and designed EST-SSR primers

SSR pattern	SSR numbers in non-redundant ESTs	Frequency (%)	Designed EST-SSR primers			Total
			Mismatch 0 ^b	Mismatch 1 ^b	Mismatch 2 ^b	
AG	24	18.8	13	11	30	54
AT	3	2.3	2	6	20	28
AC	1	0.8	1	1	10	12
AAG	24	18.8	19	56	219	294
GGT	15	11.7	9	21	78	108
AAT	15	11.7	13	17	59	89
AAC	14	10.9	10	40	112	162
GGA	9	7.0	6	20	76	102
ATC	8	6.3	3	31	127	161
AGC	6	4.7	3	21	71	95
ACG	2	1.6	1	4	18	23
GGC	0	0.0	0	11	24	35
ACT	0	0.0	0	6	13	19
AAAT	3	2.3	3	7	38	48
AATC	2	1.6	2	0	8	10
AAAG	1	0.8	0	4	47	51
AATG	1	0.8	1	4	11	16
AAAC	0	0.0	0	1	33	34
AAGC	0	0.0	0	0	7	7
AATT	0	0.0	0	2	5	7
Other tetra-nucleotide repeats ^a	0	0.0	0	1	7	8
Total	128	100	86	264	1013	1363

^a Other tetra-nucleotide repeats include GGAT, GGGA, AACG, GGAC and AGGT.

^b Mismatch 0 represents perfect SSR. Mismatch 1 and Mismatch 2 indicates SSR which include one and two irregular bases against the nucleotide repeat, respectively.

regions, qualified primer pairs were designed on 86 SSR regions. In addition, SSRs that include one or two mismatches, i.e., irregular sequences against the nucleotide repeats, were identified by the fuzznuc program in order to increase the numbers of candidate EST-SSR markers. Then, an additional 265 and 1,012 primer pairs were designed on the flanking SSR regions allowing one base and two base mismatches, respectively (Table 1). As a result, a total of 1,363 EST-SSR markers were designed. Among the 1,363 generated markers, 94 (6.9%) were di-nucleotide repeats, whereas 1,088 (79.8%) and 181 (13.3%) were tri- and tetra-nucleotide repeats, respectively. Between the di-nucleotide repeats, poly (AG)n (n = 54, 57.4%) were most frequently observed, followed by poly (AT)n (28, 29.8%) and poly (AC)n (12, 12.8%). No poly (GC)n were identified. Ten types of tri-nucleotide repeats were observed, among which poly (AAG)n (294, 27.0%) were the most abundant, followed by poly (AAC)n (162, 14.9%) and poly (ATC)n (161, 14.8%). Among the tetra-nucleotide repeats, poly (AAAG)n (51, 28.2%), poly (AAAT)n (48, 26.5%), and poly (AAAC)n (34, 18.8%) were more commonly observed compared to the other motifs, which constitute 73.5% of the tetra-nucleotide repeats. The details of the designed faba bean EST-SSR primers, along with the corresponding SSR motif, product size, and primer sequence, are available at http://marker.kazusa.or.jp/Faba_bean.

Linkage map construction

A total of 10,507 SSR markers, including 1,363 FBES, 7,244 RCS (Red Clover SSR), and 1,900 WCS (White

Clover SSR) markers, were subjected PCR with the parental DNAs. As a result, 1,089 FBES, 1,743 RCS, and 261 WCS markers showed solid amplification. These markers were subsequently employed in polymorphic analysis of six plants each from Subpop1 and Subpop2 of the F₂ mapping population. In Subpop1, 365 markers showed polymorphisms, including 228 FBES, 121 RCS, and 16 WCS markers, while 282 markers, including 134 FBES, 126 RCS, and 22 WCS markers, showed polymorphisms in Subpop2. The 365 and 282 polymorphic markers were tested by subsequent segregation analysis in Subpop1 and Subpop2. The marker set showing polymorphisms in Subpop2 was also used for segregation analysis of Subpop3.

A total of 428 segregation locus genotypes were generated from 365 markers in Subpop1, whereas 292 and 316 segregation locus genotypes were identified from 249 and 282 markers in Subpop2 and Subpop3, respectively. Using a color mapping approach and JoinMap analysis, temporary maps of Subpop1, Subpop2 and Subpop3 were generated with 18, 21, and 16 LGs, respectively (Table 2). The genotypes of dominant loci on each LG were imputed to co-dominant genotypes according to the flanking genotypes of co-dominant loci prior to integration of the three subpopulation maps. The number of imputed dominant loci was 124, 115 and 75, respectively. Of the generated LGs, 14, 18, and 11 LGs in Subpop1, Subpop2, and Subpop3, respectively, were combined with the six LGs of the integrated 'Nubaria 2' × 'Misr 3' map. Then, the 14, 18, and 11 LGs in Subpop1, Subpop2, and Subpop3, respectively, were re-integrated into six LGs in each subpopulation-specific

Table 2. Number of mapped loci, length, locus density in the ‘Nubaria 2’ × ‘Misr 3’ linkage map, and numbers of integrated LGs and segregation distortion of sub-populations

Linkage group	Nubaria 2 × Misr 3 map			Subpop1			Subpop2			Subpop3		
	Number of mapped loci	Length (cM)	Locus Density (cM)	Number of integrated LGs	Number of integrated loci	Distortion (%) ^a	Number of integrated LGs	Number of integrated loci	Distortion (%) ^a	Number of integrated LGs	Number of integrated loci	Distortion (%) ^a
LG1	182	161.5	0.89	3	122	76.2	4	42	21.4	2	49	8.2
LG2	130	136.2	1.05	5	78	19.2	4	46	6.5	2	57	12.3
LG3	64	129.8	2.03	1	20	15.0	3	42	21.4	3	24	4.2
LG4	67	104.2	1.56	1	33	21.2	3	37	2.7	1	25	12.0
LG5	84	88.9	1.06	3	56	14.3	3	30	6.7	2	25	12.0
LG6	25	64.1	2.56	1	9	0.0	1	18	16.7	1	8	0.0
Total	552	684.7	1.24	14	318	40.7	18	215	10.5	11	188	24.5
Unintegrated ^b	–	–	–	4	29	13.8	3	29	31.0	5	31	7.9

^a A significant level at $P < 0.05$.

^b The numerics indicate numbers of un-integrated LGs, total numbers of mapped loci on the un-integrated LGs, and percentages of distorted loci.

map using JoinMap version 4.0 with optimal thresholds of LOD and REC values (Supplemental Table 2). The adequacy of locus order of the six re-integrated LGs in each subpopulation-specific map was confirmed by drawing graphical genotypes. The results indicate that the linkage maps of Subpop1, Subpop2, and Subpop3 consist of 10, 9, and 11 LGs, respectively. The number of mapped loci, length, and locus density in the three subpopulation-specific linkage maps ranged from 224 to 367, from 617.3 cM to 839.1 cM, and from 2.58 cM·locus⁻¹ to 3.61 cM·locus⁻¹, respectively (Supplemental Table 2). The highest segregation distortions ($p < 0.05$) were observed in Subpop1, which ranged from 0.0% (LG6) to 76.2% (LG1), with a mean value from the six integrated LGs of 40.7% (Table 2). The average segregation distortions of the six LGs in the other two subpopulation-specific maps were 10.5% (Subpop2) and 24.5% (Subpop3), respectively (Table 2).

The integrated ‘Nubaria 2’ × ‘Misr 3’ map consisted of six LGs, representing a total length of 684.7 cM, with 552 loci (Table 2, Fig. 1). The LG numbers were named in order of the length (from the longest to the shortest). Of the 552 loci, 21 were mapped on the three subpopulation-specific maps, while 128 were located on the two of three subpopulation-specific maps. The adequacy of locus order of each LG was confirmed by drawing graphical genotypes (Supplemental Fig. 2). The locus order of each LG (except LG3) was not substantially altered when the LOD threshold was increased by more than 1.0 in the Regression Mapping Module of JoinMap. The locus order on LG3 was improved when the LOD threshold was changed from 1.0 to 1.2. Therefore, $\text{LOD} \geq 1.2$ was employed for LG3. The number of mapped loci and the length of each LG ranged from 25 (LG6) to 182 (LG1) and from 64.1 cM (LG6) to 161.5 cM (LG1), respectively. The mean locus density was 1.24 cM·locus⁻¹ and ranged from 0.89 cM·locus⁻¹ (LG1) to 2.56 cM·locus⁻¹ (LG6). The largest gap was 24.7 cM, which occurred between RCS3037_LG3 and FBES0241 on LG2, followed by a 23.3 cM gap between RCS2383_LG2a and RCS0349_LG2 on LG2 (Supplemental Table 3). The ratios of mapped FBES, RCS, and WCS loci were 54.5%, 38.6%,

and 6.8%, respectively. The ratio of mapped FBES loci in each LG ranged from 37.5% (LG3) to 64.8% (LG1).

The number of markers that generated mapped loci on the integrated map was 397, including 235 FBES, 136 RCS, and 26 WCS markers. Of the 397 markers, 283 (71.2%) were mapped onto single positions of the integrated map, while the other 114 (28.7%) were mapped onto multiple positions. The former 283 and later 114 markers were regarded as single-loci diagnostic (SLD) and multi-loci diagnostic (MLD) markers, respectively. The ratios of MLD markers to mapped FBES, RCS, and WCS markers were 25.4% (44/217), 66.3% (53/133), and 30% (6/26), respectively. The loci generated from MLD markers were mapped across the linkage map (Supplemental Table 3). The ratio of mapped loci generated from MLD markers was 49% (269/552) in total and ranged from 31% (LG5) to 60% (LG1) in each LG.

Comparison with the genome of a model legume, *M. truncatula*

The 552 loci mapped on the integrated ‘Nubaria 2’ × ‘Misr 3’ map were generated from 235 FBES, 136 RCS, and 26 WCS markers. Among the mapped markers, significant similarities to the *M. truncatula* gene sequences were observed in 178 faba bean ESTs, 95 red clover genome or EST sequences, and 23 white clover ESTs from which FBES, RCS, and WCS markers were designed, respectively. As a result, 412 mapped loci (74.6% of the mapped loci) showed significant similarities to the genes of *M. truncatula*. In addition, significant similarities between corresponding sequences of mapped markers and the *L. japonicus* gene sequences were observed in 82 faba bean ESTs, 58 red clover genome or EST sequences, and nine white clover ESTs. As a result, 188 mapped loci (34.1% of mapped loci) showed significant similarities to the genes of *L. japonicus*.

Alignment of homologous sequence pairs along each linkage group revealed obvious syntenic relationships between the LGs in faba bean (fb-LG) and the chromosomes in *M. truncatula* (Mt-Chr) and between fb-LG and chromosomes in *L. japonicus* (Lj-Chr; Fig. 2). In a comparison with the *M. truncatula* genome, single syntenic blocks in fb-LG2 and fb-LG6 were identified in Mt-Chr5 and Mt-Chr3,

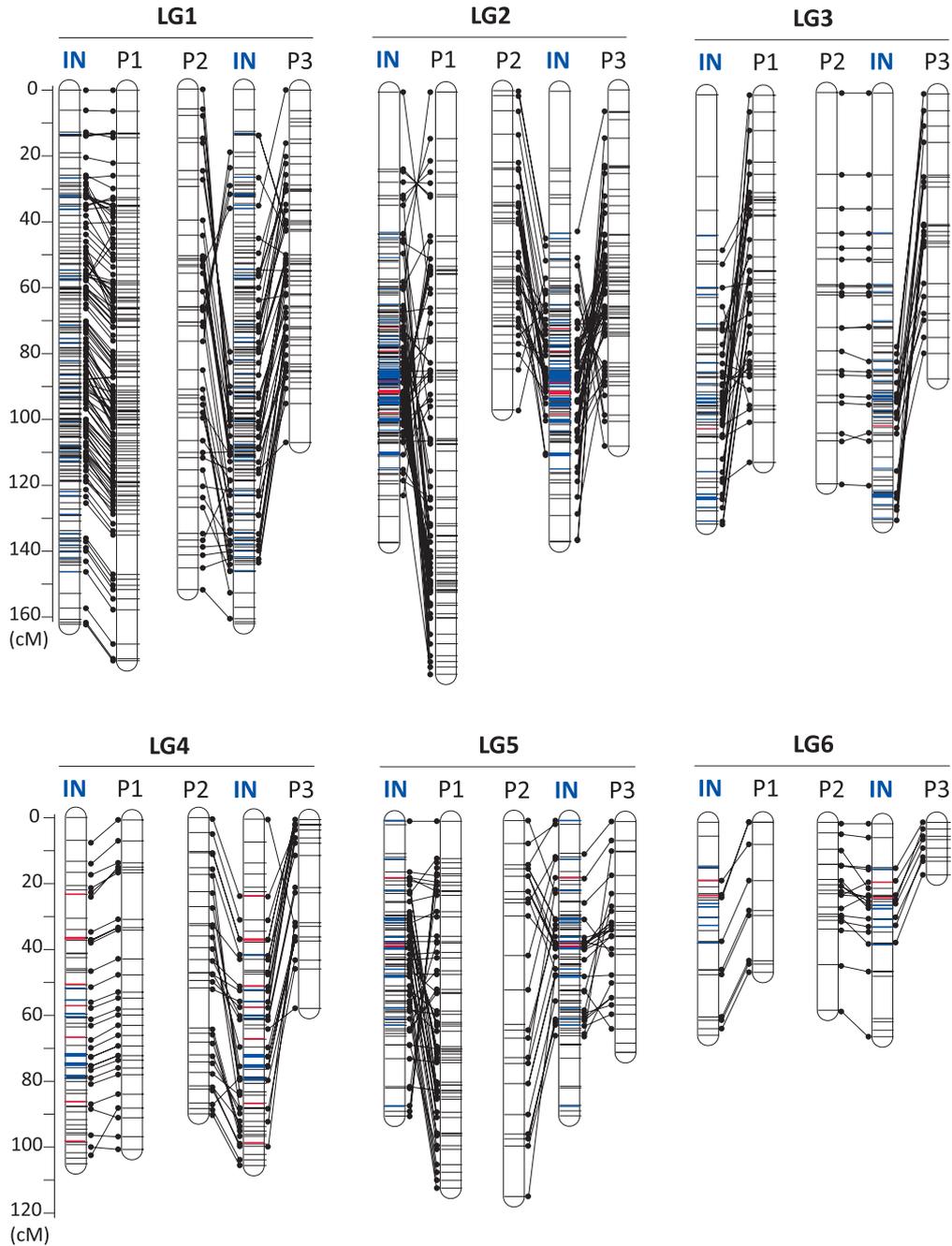


Fig. 1. An integrated linkage map of ‘Nubaria 2’ × ‘Misr 3’ and subpopulation-specific maps. IN, P1, P2 and P3 represents the integrated map, Subpop1, Subpop2 and Subpop3 maps, respectively. The corresponding loci between the integrated map and subpopulation specific maps are connected by lines with dots. Red and blue bars indicate commonly mapped loci on the integrated map across three or two of the tree maps, respectively.

respectively, while fb-LG4 shared two synteny blocks with Mt-Chr2 and Mt-Chr7. The other fb-LGs (fb-LG1, fb-LG3, and fb-LG5) shared multiple synteny blocks with between five and all Mt-Chrs. In a comparison with the *L. japonicus* genome, fb-LG4 shared a single synteny block with Lj-Chr6. On the same Lj-Chr region (16.7–27.9 Mb on Lj-Chr6), synteny blocks were identified between all fb-LGs except fb-LG5. In addition, fb-LG5 shared two synteny blocks with Lj-Chr3 and Chr4, while fb-LG2 and fb-LG6

each shared three synteny blocks on Lj-Chr2, Chr5, and Chr6, and Lj-Chr1, Chr3, and Chr6, respectively. Multiple synteny blocks from fb-LG1 and LG3 were identified in all Lj-Chrs.

Polymorphic analysis of FBES markers with ten Egyptian faba bean varieties

A total of 521 FBES markers less than 500 bp in fragment size were selected from the 1,089 FBES markers that

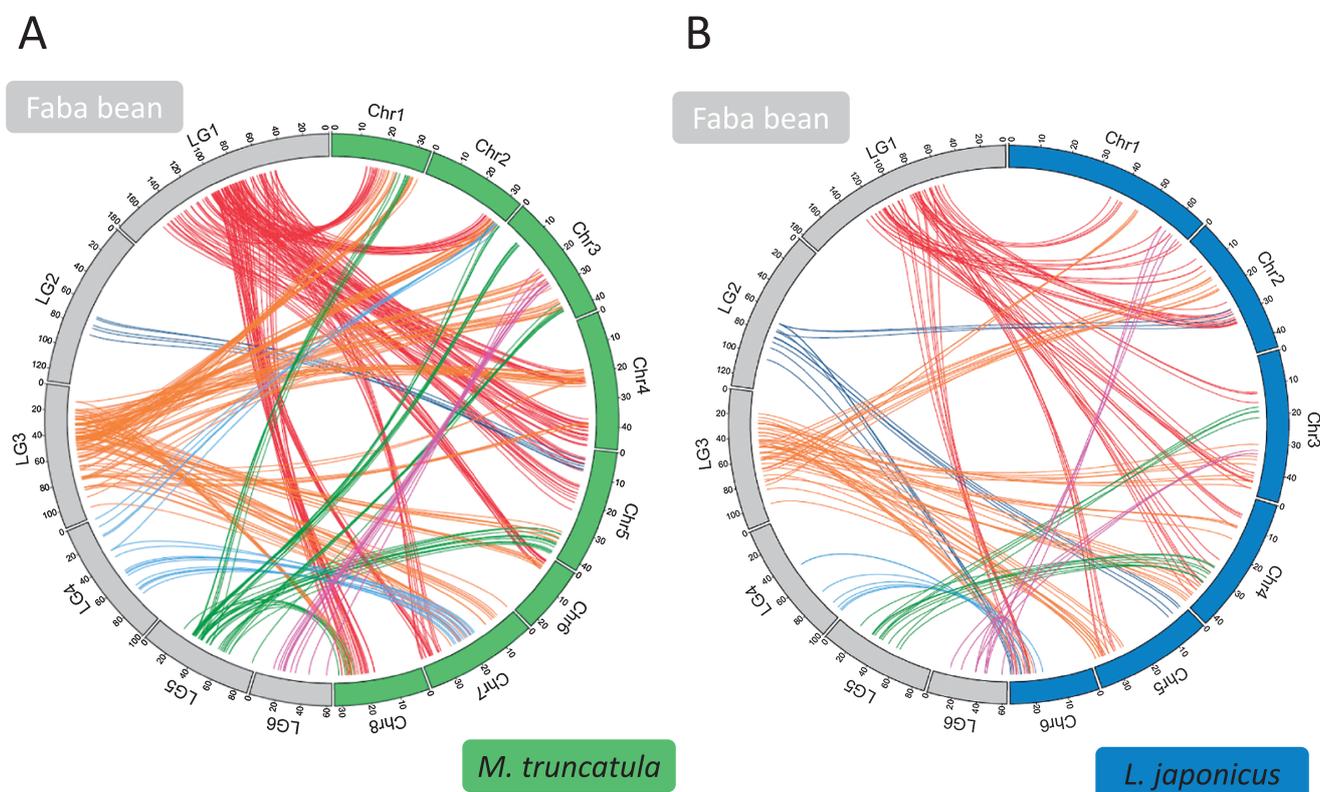


Fig. 2. Graphical view of syntenic relationship between A) faba bean and *M. truncatula*, and B) faba bean and *L. japonicus*. Syntenic regions between the two species are connected by colored lines. The colors of the lines represent LGs on the ‘Nubaria 2’ × ‘Misr 3’ map. Scales represent genetic position on LGs (cM, faba bean) and physical position on chromosomes (Mbp, *M. truncatula* and *L. japonicus*).

Table 3. Similarity coefficients among the ten Egyptian varieties examined

	Cairo1	Giza3	Giza843	Misr1	Misr3	Nubaria1	Nubaria2	Sakha1	Sakha2
Cairo1									
Giza3	0.379								
Giza843	0.346	0.319							
Misr1	0.336	0.335	0.286						
Misr3	0.371	0.403	0.407	0.394					
Nubaria1	0.380	0.326	0.362	0.372	0.406				
Nubaria2	0.370	0.362	0.338	0.344	0.403	0.394			
Sakha1	0.376	0.369	0.341	0.338	0.404	0.383	0.373		
Sakha2	0.384	0.324	0.365	0.334	0.385	0.344	0.372	0.349	
Sakha3	0.390	0.353	0.342	0.325	0.390	0.343	0.399	0.311	0.347

showed solid amplification with the parental DNAs of the map. The amplification status of the 521 FBES markers was assessed with ten *Vicia faba* accessions using a fluorescent fragment analyzer. Amplification of DNA fragments was observed for 487 markers, while 34 markers did not generate amplicons. Using the fluorescent fragment analyzer, 78.9% (384/487) of the amplified markers were found to be polymorphic among the ten faba bean varieties, whereas 21.1% (103/487) was monomorphic (Supplemental Table 4). A total of 1,217 polymorphic alleles (peaks) were identified from the 384 markers. The number of observed peaks in each marker ranged from 1 to 11; the mean value was 2.82. Seventy-eight markers identified null alleles among the ten varieties. The maximum number of observed peaks on a marker in each variety ranged from 1 to 5 for the 487 FBES

markers, with an average of 1.95. Twenty percent (95/487) of the markers produced more than two peaks in each variety, while 31.2% (152/487) and 49.3% (240/487) of the markers produced a maximum of one and two peaks in each variety, respectively. The PIC ranged from 0.0 to 0.38, with an average of 0.19.

Similarity coefficients were investigated to examine the genetic relationships between the ten Egyptian varieties based on 1,217 polymorphic alleles. The mean similarity coefficient in all possible genotypes was 0.362, ranging from 0.286 to 0.407 (Table 3). The minimum and maximum similarity coefficients were observed between ‘Misr 1’ and ‘Gisa 843’ and between ‘Misr 3’ and ‘Gosa 843’, respectively (Supplemental Fig. 3).

Discussion

In this study, we identified 5,090 potential non-redundant ESTs, including 2,325 contigs and 2,765 singletons. Large-scale EST and transcript sequences were previously obtained by Akash *et al.* (2012), Kaur *et al.* (2012), and Yang *et al.* (2012). Our 5,090 ESTs showed significant matches to 2,289, 600, and 63 of the EST or transcript sequences developed by Akash *et al.* (2012), Kaur *et al.* (2012), and Yang *et al.* (2012), respectively, as revealed by BLASTN searches with an E-value cutoff of $1e-20$. Of the 5,090 ESTs, 80.7%, 89.1%, and 80.6% showed significant similarity to protein-encoding genes of *L. japonicus*, *M. truncatula*, and *Arabidopsis*, respectively. These higher similarities confirmed the results of a previous phylogenetic analysis (Young *et al.* 2003).

Di-, tri-, and tetra- nucleotide repeats were identified in the perfect match SSRs at a ratio of 7:80:13. This ratio nearly matches that identified by Kaur *et al.* (2012, 8:73:19) and differs from that identified by Akash *et al.* (2012, 34:59:6). A total of 1,363 primers were designed as FBES markers on flanking regions of perfect SSRs and those with one or two base mismatches. By performing BLAST searches, we confirmed that the sequences of all primer pairs were not published by Akash *et al.* (2012), Kaur *et al.* (2012), or Yang *et al.* (2012). Of the 1,363 FBES markers that we designed, 1,089 (80%) showed solid amplification with the DNAs of the mapping parents. This amplification success ratio was slightly lower than that of Kaur *et al.* (2012, 84%) but higher than that of Yang *et al.* (2012, 68%).

We constructed a 'Nubaria 2' × 'Misr 3' map by integrating LGs grouped in each subpopulation map. The number of LGs, which were grouped using the color map method, ranged from 16 to 21 in the three subpopulations. The color map method employs a comparison of graphical genotypes of the segregation data, and the result is generally stricter than that obtained by other mapping programs, such as JoinMap. When the population size is small, the number of generated LGs tends to be high due to the low frequency of recombination. Therefore, the combination of the small sizes of the subpopulations (21–65 plants) and the strict grouping method led to a large number of generated LGs.

The grouped LGs in each subpopulation were integrated using the Combine Groups for Mapping Integration Module in JoinMap. In each subpopulation, three to five LGs were not integrated due to the smaller number of commonly mapped markers. The 'Nubaria 2' × 'Misr 3' map consists of six LGs. This is the densest and the first faba bean map in which the number of LGs is equal to the chromosome number. The map contains 552 loci covering 684.7 cM. In the previously published faba bean maps, the number of mapped loci and the total length ranged from 77 to 277 and 984.5 cM to 2,856.7 cM, respectively (Arbaoui *et al.* 2008, Avila *et al.* 2005, Cruz-Izquierdo *et al.* 2012, Díaz-Ruiz *et al.* 2009, Ellwood *et al.* 2008, Román *et al.* 2002, 2004, Surahman 2001). Although the number of mapped loci in

the present map is the highest among the maps developed to date, the total length of the map is the shortest. It is possible that integration of segregation data derived from the subpopulations led to an underestimation of the map length during the process of 'join' the maps. In addition, the use of different software in linkage analysis may have led to the large differences in map distance among the maps. The genetic distances in most of the previously published maps, except for the map developed by Ellwood *et al.* (2008), were estimated using MAPMAKER/EXP version 2.0 or 3.0 (Lander *et al.* 1987, Lincoln *et al.* 1993). In our experience, JoinMap tends to generate shorter genetic distances than MAPMAKER/EXP. Due to the absence of commonly mapped markers, the present map is not compatible with the previously published maps. Location of our mapping markers on the previously published maps would more clearly reveal the possible causes for the shorter length of the present map.

The large ratio (49%) of mapped loci derived from MLD markers is another distinctive feature of the present map. The high ratio of MLD markers was also revealed in polymorphic analysis of the ten varieties. When a marker identifies more than two peaks in a single variety with a fluorescent fragment analyzer, the marker can be designated as an MLD marker, since faba bean is a diploid species. When two peaks are identified, the peaks are considered to be 'heterozygous alleles on a locus' or 'two homozygous alleles on two loci'. In this study, 20% of the tested FBES markers showed more than two peaks in a faba bean variety, and 31% identified two peaks. This result suggests that the possible ratio of MLD markers among the tested FBES markers ranged from 20% to 51%. Of the 235 mapped FBES markers, 187 were used in the polymorphic analysis of the ten varieties. We investigated the relationship between the maximum number of identified peaks in a variety (Supplemental Table 4) and MLD marker diagnosis (Supplemental Table 3) for the 187 FBES markers (Supplemental Table 5). We consider the four of 19 markers to be mislabeled as MLD markers, while 39 of the 54 markers were mislabeled as SLD markers. The result suggests that the possibility of over estimation of number of MLD markers is higher than that of SLD markers, and the presumed ratio of mapped loci generated from MLD markers (49%) in this study might be underestimated. Since the FBES markers were derived from ESTs, the large ratio of loci generated from MLD markers suggests that a large number of paralogous genes exist in the faba bean genome.

Previously reported comparative analyses suggest that a simple macrosyntenic relationship exists between faba bean and *M. truncatula* (Cruz-Izquierdo *et al.* 2012, Ellwood *et al.* 2008). In the current study, a more complex macrosyntenic relationship was observed between these two model legumes. Perhaps the large number of mapped loci in our map led to the identification of a more complex macrosyntenic relationship between these species than was previously reported. Despite the relatively high similarity observed

in faba bean ESTs against the *M. truncatula* genome compared to that of the *L. japonicus* genome, the complexity of the macrosyntentic relationships was not highly different between the two model legumes. For example, fb-LG2 and fb-LG6 each identified a single synteny block on the *M. truncatula* genome, while a single synteny block was also identified on the *L. japonicus* genome. On the other hand, fb-LG4 identified a single synteny block on the *L. japonicus* genome, and two synteny blocks were identified on the *M. truncatula* genome. Both fb-LG1 and fb-LG3 identified more synteny blocks than the other fb-LGs on the two model legumes. The result suggests that more genomic rearrangements occurred in chromosomes corresponding to the two LGs.

Understanding the genetic diversity in breeding materials is critical for crop improvement. In the past decade, genetic diversity among collected germplasm accessions of faba bean has been delineated using various types of genetic marker systems, including isozymes (Käser and Steiner 1983, Mancini *et al.* 1989), phenotypic markers (such as morphological markers) (Liu and Zong 2008, Wang *et al.* 2009, Zhang and Liu 2009), and molecular markers, which have been increasingly used to reveal the diversity of faba bean germplasm. The average similarity coefficients among the ten faba bean varieties examined, 0.362, was higher than that observed among the thirteen Chinese varieties (Zong *et al.* 2009) or the nine Tunisian varieties (Ouji *et al.* 2012). Egypt has the longest faba bean cultivation history in the world; this crop has been cultivated since the ancient Egyptian era. This long history may have affected the large genetic diversity observed in Egyptian faba bean varieties.

Due to its large genome size, genetic and genomic studies in faba bean have lagged behind those of other legume crops. In particular, the absence of a dense linkage map has inhibited the progress of MAS for traits controlled by QTLs. Along with the EST-SSR markers, the dense map developed in this study is expected to accelerate MAS in faba bean breeding. Moreover, the high ratio of MLD markers suggests that a large number of paralogous genes exist in the faba bean genome. Therefore, care should be taken when performing QTL mapping, especially due to the additive effects caused by paralogous genes.

Acknowledgments

This work was supported by the Kazusa DNA Research Institute Foundation, the Field Crop Research Institute ARC Foundation, Invitation Fellowship Programs for Research in Japan funded by the JSPS (Japanese Society for the Promotion of Science), and a ParOwn Fellowship for Postgraduate Research funded by the Egyptian Government.

Literature Cited

Abo El-kheir, Z.A., M.S. Abdel-Hady, H.M.H. El-Naggar and A.R. Abd El-Hamed (2010) Molecular and biochemical markers of some

- Vicia faba* L. Cultivars in response to broomrape infestation. *Nature and Science* 8: 252–260.
- Akash, M.W and G.O. Myers (2012) The development of faba bean expressed sequence tag-simple sequence repeats (EST-SSRs) and their validity in diversity analysis. *Plant Breed.* 131: 522–530.
- Arabidopsis Genome Initiative (2000) Analysis of the genome sequence of the flowering plant *Arabidopsis thaliana*. *Nature* 408: 796–815.
- Arbaoui, M., W. Link, Z. Satovic and A.M. Torres (2008) Quantitative trait loci of frost tolerance and physiologically related trait in faba bean (*Vicia faba* L.). *Euphytica* 164: 93–104.
- Asamizu, E., Y. Nakamura, S. Sato, H. Fukuzawa and S. Tabata (1999) A large scale structural analysis of cDNA in a unicellular green alga *Chlamydomonas reinhardtii*. I. Generation of 3433 non-redundant expressed sequence tags. *DNA Res.* 31: 369–373.
- Avila, C.M., J.C. Sillero, D. Rubiales, M.T. Moreno and A.M. Torres (2003) Identification of RAPD markers linked to the Uvf-1 gene conferring hypersensitive resistance against rust (*Uromyces viciae-fabae*) in *Vicia faba* L. *Theor. Appl. Genet.* 107: 353–358.
- Avila, C.M., Z. Satovic, J.C. Sillero, D. Rubiales, M.T. Moreno and A.M. Torres (2004) Isolate and organ-specific QTLs for ascochyta blight resistance in faba bean. *Theor. Appl. Genet.* 108: 1071–1078.
- Avila, C.M., Z. Šaypvia, J.C. Sillero, S. Nadal, D. Rubiales, M.T. Moreno and A.M. Torres (2005) QTL Detection for Agronomic Traits in Faba Bean (*Vicia faba* L.). *Agriculturae Conspectus Scientificus.* 70: 65–73.
- Avila, C.M., S. Nadal, M.T. Moreno and A.M. Torres (2006) Development of a simple PCR-based marker for the determination of growth habit in *Vicia faba* L. using a candidate gene approach. *Mol. Breed.* 17: 185–190.
- Avila, C.M., S. Atienza, M. Moreno and A.M. Torres (2007) Development of a new diagnostic marker for growth habit selection in faba bean (*Vicia faba* L.) breeding. *Theor. Appl. Genet.* 115: 1075–1082.
- Cruz-Izquierdo, S., C.M. Avila, Z. Satovic, C. Palomino, N. Gutierrez, S.R. Ellwood, H.T.T. Phan, J.I. Cubero and A.M. Torres (2012) Comparative genomics to bridge *Vicia faba* with model and closely-related legume species: stability of QTLs for flowering and yield-related traits. *Theor. Appl. Genet.* 125: 1767–1782.
- Díaz-Ruiz, R., A. Torres, M.V. Gutierrez, D. Rubiales, J.I. Cubero, M. Kharrat, Z. Satovic and B. Román (2009) Mapping of quantitative trait loci controlling *Orobanche foetida* Poir. resistance in faba bean (*Vicia faba* L.). *Afr. J. Biotechnol.* 8: 2718–2724.
- Díaz-Ruiz, R., A.M. Torres, Z. Satovic, M.V. Gutierrez, J.I. Cubero and B. Román (2010) Validation of QTLs for *Orobanche crenata* resistance in faba bean (*Vicia faba* L.) across environments and generations. *Theor. Appl. Genet.* 120: 909–919.
- Duc, G., N. Berthaut, J. Le Guen and J. Picard (1992) Hybrid breeding in faba bean (*Vicia faba* L.). *In: Proc. 1st Conf. on Grain Legumes, Angers, France*, pp. 45–46.
- Duc, G., S. Bao, M. Baumc, B. Redden and M. Sadiki (2010) Diversity maintenance and use of *Vicia faba* L. genetic resources. *Field Crops Res.* 115: 270–278.
- Ellwood, S.R., H.T.T. Phan, M. Jordan, J. Hane, A.M. Torres, C.M. Avila, S. Cruz-Izquierdo and R.P. Oliver (2008) Construction of a comparative genetic map in faba bean (*Vicia faba* L.); conservation of genome structure with *Lens culinaris*. *BMC Genomics* 9: 380.
- Ewing, B. and P. Green (1998) Base-calling of automated sequencer traces using phred. II. Error probabilities. *Genome Res.* 8: 186–194.

- Ewing, B., L. Hillier, M.C. Wendl and P. Green (1998) Base-calling of automated sequencer traces using phred. I. Accuracy assessment. *Genome Res.* 8: 175–185.
- Gong, Y., S. Xu, W. Mao, Q. Hu, G. Zang, J. Ding and Z. Li (2010) Generation and characterization of 11 novel EST-derived microsatellites from *Vicia faba* (Fabaceae). *Am. J. Bot.* 97: e69–e71.
- Gong, Y., S. Xu, W. Mao, Z. Li, Q. Hu, G. Zhang and J. Ding (2011) Genetic diversity analysis of faba bean (*Vicia faba* L.) based on EST-SSR markers. *Agric. Sci. China* 10: 838–844.
- Gresta, F., G. Avola, E. Albertini, L. Raggi and V. Abbate (2010) A study of variability in the Sicilian faba bean landrace ‘Larga di Leonforte’. *Genet. Resour. Crop. Ev.* 57: 523–531.
- Gutierrez, N., C. Avila, G. Duc, P. Marget, M. Suso, M. Moreno and A.M. Torres (2006) CAPs markers to assist selection for low vicine and convicine contents in faba bean (*Vicia faba* L.). *Theor. Appl. Genet.* 114: 59–66.
- Gutierrez, N., C. Avila, C. Rodriguez-Suarez, M. Moreno and A. Torres (2007) Development of SCAR markers linked to a gene controlling absence of tannins in faba bean. *Mol. Breed.* 19: 305–314.
- Gutierrez, N., C. Avila, M. Moreno and A. Torres (2008) Development of SCAR markers linked to *zt-2*, one of the genes controlling absence of tannins in faba bean. *Aust. J. Agric. Res.* 59: 62–68.
- Huang, X., J. Wang, S. Aluru, S.P. Yang and L. Hillier (2003) PCAP: a whole-genome assembly program. *Genome Res.* 13: 2164–2170.
- Isobe, S.N., H. Hisano, S. Sato, H. Hirakawa, K. Okumura, K. Shirasawa, S. Sasamoto, A. Watanabe, T. Wada, Y. Kishida *et al.* (2012) Comparative genetic mapping and discovery of linkage disequilibrium across linkage groups in white clover (*Trifolium repens* L.). *G3 (Bethesda)* 2: 607–617.
- Johnston, J.S., M.D. Bennett, A.L. Rayburn, D.W. Galbraith and H.J. Price (1999) Reference standards for determination of DNA content of plant nuclei. *Am. J. Bot.* 86: 609–613.
- Kaur, S., L.W. Pembleton, N.O.I. Cogan, K.W. Savin, T. Leonforte, J. Paull, M. Materne and J.W. Forster (2012) Transcriptome sequencing of field pea and faba bean for discovery and validation of SSR genetic markers. *BMC Genomics* 13: 104.
- Käser, H.R. and A.M. Steiner (1983) Subspecific classification of *Vicia faba* L. by protein and isozyme patterns. *Fabis. Newsletter* 7: 19–20.
- Kiss, G.B., A. Kereszt, P. Kiss and G. Endre (1998) Colormapping; a non-mathematical procedure for genetic mapping. *Acta Biol. Hung.* 19: 125–142.
- Lander, E.S., P. Green, J. Abrahamson, A. Barlow, M.J. Daly, S.E. Lincoln and L. Newburg (1987) MAPMAKER: an interactive computer program for constructing genetic linkage maps of experimental and natural populations. *Genomics* 1: 174–181.
- Lincoln, S.E., M.J. Daly and E.S. Lander (1993) Constructing genetic linkage maps with MAPMAKER/EXP version 3.0: A Tutorial and reference manual. Whitehead Kubitute o]for Biomedical Resarch. USA.
- Link, W., W. Ederer, P. Metz, H. Buiel and A.E. Melchinger (1994a) Genotypic and environmental variation for degree of cross-fertilization in faba bean. *Crop Sci.* 34: 960–964.
- Link, W., D. Stelling and E. Ebmeyer (1994b) Factors determining the performance of synthetics in *Vicia faba* L. 1. Heterogeneity, heterozygosity, and degree of cross-fertilization. *Euphytica* 75: 77–84.
- Link, W., C. Dixkens, M. Singh and M. Schwall (1995) Genetic diversity in European and Mediterranean faba bean germplasm revealed by RAPD markers. *Theor. Appl. Genet.* 90: 27–32.
- Liu, K. and S.V. Muse (2005) PowerMarker: an integrated analysis environment for genetic marker analysis. *Bioinformatics* 21: 2128–2129.
- Liu, Y.J. and X.X. Zong (2008) Morphological diversity analysis of faba bean (*Vicia faba* L.), germplasm resources from Qinghai. *J. Plant Genet. Res.* 9: 79–83.
- Ma, Y., T. Yang, J. Guan, S. Wang, H. Wang, X. Sun and X. Zong (2011) Development and characterization of 21 EST-derived microsatellite markers in *Vicia faba* (fava bean). *Am. J. Bot.* 98: e22–e24.
- Mancini, R., C. Pace, G.T.S. Mugnozza, V. Delre and D. Vittori (1989) Isozyme gene markers in *Vicia faba* L. *Theor. Appl. Genet.* 77: 657–667.
- Nei, M. (1973) Analysis of gene diversity in subdivided populations. *Proc. Natl. Acad. Sci. USA* 70: 3321–3323.
- Ouji, A., S. El-Bok, N.H. Syed, R. Abdellaoui, M. Rouaissi, A.J. Flavell and M. El-Gazzah (2012) Genetic diversity of faba bean (*Vicia faba* L.) populations revealed by sequence specific amplified polymorphism (SSAP) markers. *Afr. J. Biotechnol.* 11: 2162–2168.
- Požárková, D., A. Koblížková, B. Román, A.M. Torres, S. Lucretti, M. Lysák, J. Doležel and J. Macas (2002) Development and characterization of microsatellite markers from chromosome 1-specific DNA libraries of *Vicia faba*. *Biologia Plantarum* 45: 337–345.
- Rafalski, J.A., I.M. Vogel, M. Morgant, W. Powell, C. Andre and S.V. Tingey (1996) Generating and using DNA markers in plants. *In: Non mammalian, genic analysis: a practical guide Academic, San Diego, ISBN: 0121012859*, pp. 75–135.
- Ramsey, G., W. Van de Ven, R. Waugh, D. Griffiths and W. Powell (1995) Mapping quantitative trait loci in faba beans. *In: AEP (eds.) Proceedings of the 2nd European conference on grain legumes, Copenhagen, Denmark*, pp. 444–445.
- Rice, P., I. Longden and A. Bleasby (2000) EMBOSS: the European Molecular Biology Open Software Suite. *Trends Genet.* 16: 276–277.
- Román, B., A.M. Torres, D. Rubiales, J.I. Cubero and Z. Satovic (2002) Mapping of quantitative trait loci controlling broomrape (*Orobanche crenata* Forsk.) resistance in faba bean (*Vicia faba* L.). *Genome* 45: 1057–1063.
- Román, B., Z. Satovic, C.M. Avila, D. Rubiales, M.T. Moreno and A.M. Torres (2003) Locating genes associated with *Ascochyta fabae* resistance in *Vicia faba*. *Aust. J. Agric. Res.* 54: 85–90.
- Román, B., Z. Satovic, D. Pozarkova, J. Macas, J. Dolezel, J.I. Cubero and A.M. Torres (2004) Development of a composite map in *Vicia faba*, breeding applications and future prospects. *Theor. Appl. Genet.* 108: 1079–1088.
- Rozen, S. and H. Skaletsky (2000) Primer3 on the WWW for general users and for biologist programmers. *In: Krawetz, S. and S. Misener (eds.) Bioinformatics Methods and Protocols: Methods in Molecular Biology. Humana Press. Totowa, NJ, USA*, pp. 365–386.
- Sato, S., S. Isobe, E. Asamizu, N. Ohmido, R. Kataoka, Y. Nakamura, T. Kaneko, N. Sakurai, K. Okumura, I. Klimenko *et al.* (2005) Comprehensive structural analysis of the genome of red clover (*Trifolium pretense* L.). *DNA Res.* 12: 301–364.
- Sato, S., Y. Nakamura, T. Kaneko, E. Asamizu, T. Kato, M. Nakao, S. Sasamoto, A. Watanabe, A. Ono, K. Kawashima *et al.* (2008) Genome structure of the legume, *Lotus japonicus*. *DNA Res.* 15: 227–239.
- Schmutz, J., S.B. Cannon, J. Schlueter, J. Ma, T. Mitros, W. Nelson, D.L. Hyten, Q. Song, J.J. Thelen, J. Cheng *et al.* (2010) Genome sequence of the paleopolyploid soybean. *Nature* 463: 178–183.
- Stoddard, F.L., S. Hovinen, M. Kontturi, K. Lindström and A. Nykänen (2009) Legumes in Finnish agriculture: history, present status and future prospects. *Agric. Food Sci.* 18: 191–205.
- Surahman, M. (2001) The construction of genetic map of faba bean

- (*Vicia faba* L.) using RAPD markers. *Bul. Agron.* 29: 31–33.
- Tamura, K., D. Peterson, N. Peterson, G. Stecher, M. Nei and S. Kumar (2011) MEGA5: Molecular evolutionary genetics analysis using likelihood, distance, and parsimony methods. *Mol. Biol. Evol.* 28: 2731–2739.
- Tatusov, R.L., N.D. Fedorova, J.D. Jackson, A.R. Jacobs, B. Kiryutin, E.V. Koonin, D.M. Krylov, R. Mazumder, S.L. Mekhedov, A.N. Nikolskaya *et al.* (2003) The COG database: an updated version includes eukaryotes. *BMC Bioinformatics* 4: 41–54.
- Terzopoulos, P.J. and P.J. Bebeli (2008) Genetic diversity analysis of Mediterranean faba bean (*Vicia faba* L.) with ISSR markers. *Field Crops Res.* 108: 39–44.
- Terzopoulos, P.J., P.J. Kaltsikes and P.J. Bebeli (2008) Determining the sources of heterogeneity in Greek faba bean local populations. *Field Crops Res.* 105: 124–130.
- van de Ven, M., W. Powell, G. Ramsay and R. Waugh (1990) Restriction fragment length polymorphisms as genetic markers in *Vicia*. *Heredity* 65: 329–342.
- Van Ooijen, J.W. (2006) JoinMAP® 4, Software for the calculation of genetic linkage maps in experimental populations, Kyazma, B.V., Wageningen, Netherlands.
- Wang, X.J., X.S. Qi and X.R. Wang (2009) Genetic diversity analysis of main agronomic characters in faba bean germplasm. *Res. Agric. Modernization* 30: 633–636.
- Yang, T., S.Y. Bao, R. Ford, T.J. Jia, J.P. Guan, Y.H. He, X.L. Sun, J.Y. Jiang, J.J. Hao, X.Y. Zhang *et al.* (2012) High-throughput novel microsatellite marker of faba bean via next generation sequencing. *BMC Genomics* 13: 602.
- Young, N.D., J. Mudge and T.H.N. Ellis (2003) Legume genomes: more than peas in a pod. *Curr. Opin. Plant Biol.* 6: 199–204.
- Zeid, M., C.C. Schon and W. Lin (2003) Genetic diversity in recent elite faba bean lines using AFLP markers. *Theor. Appl. Genet.* 107: 1304–1314.
- Zeid, M., S. Mitchell, W. Link, M. Carter, A. Nawar, T. Fulton and S. Kresovich (2009) Simple sequence repeats (SSRs) in faba bean: New loci from Orobanche-resistant cultivar ‘Giza 402’. *Plant Breed.* 128: 149–155.
- Zhang, X.T. and Y.J. Liu (2009) Morphological diversity analysis of faba bean (*Vicia faba* L.) germplasm resources. *Qinghai Sci. Tech.* 5: 17–19.
- Zhu, H., H.K. Choi, D.R. Cook and R.C. Shoemaker (2005) Bridging model and crop legumes through comparative genomics. *Plant Physiol.* 137: 1189–1196.
- Zong, X., X. Liu, J. Guan, S. Wang, Q. Liu, J.G. Paull and R. Redden (2009) Molecular variation among Chinese and global winter faba bean germplasm. *Theor. Appl. Genet.* 118: 971–978.