Short Communication

# Heterozygosity and diversity analysis using mapped SNPs in a faba bean inbreeding programme

Cottage A[1], Gostkiewicz K[1], Thomas JE[1], Borrows R[1], Torres A-M[2], and O'Sullivan D.M[1,✉]

1. The National Institute of Agricultural Botany, Huntingdon Road, Cambridge, CB2 0LE, UK.
2. IFAPA, Centro Alameda del Obispo, Area de Mejora y Biotecnología Apdo. 3092, E-14080 Córdoba, Spain.

✉ Corresponding author: Donal O'Sullivan,
John Bingham Laboratory,
NIAB,
Cambridge,
CB3 0LE
United Kingdom

Tel:+44 1223 342463
Fax:+44 1223 277602
E-mail: donal.osullivan@niab.com

**Abstract**

The aim of this study was to convert existing faba bean Single Nucleotide Polymorphisms (SNP) markers from Cleaved Amplification Polymorphic Sequences (CAPS) and SNAPshot formats which are expensive and time-consuming to the more convenient KBiosciences Competitive Allele-Specific PCR (KASP) assay format. Out of 80 assays designed, 75 were validated, though a core set of 67 of the most robust markers are recommended for further use. The 67 best KASP SNP assays were used across two generations of single seed descent to detect unintended outcrossing and to track and quantify loss of heterozygosity, a capability that will significantly increase the efficiency and performance of pure line production and maintenance. This same set of assays was also used to examine genetic relationships between the 67 members of the partly inbred panel, and should prove useful for line identification and diversity studies in the future.

**Keywords:** *Vicia faba* L.; KBiosciences competitive Allele-Specific PCR (KASP); single nucleotide polymorphism (SNP); haplotype

**Abbreviations:**

CAPS: Cleaved Amplification Polymorphic Sequences

KASP: KBiosciences Competitive Allele-Specific PCR

AFLP : Amplified Fragment Length Polymorphism

SSD: Single Seed Descent

Although faba bean is widely consumed in China, North African and some Sub-Saharan countries, forming an important source of dietary protein in these regions, as well as providing a nitrogen-fixing break crop in arable rotations wherever it is grown, research underpinning genetic improvement of the crop has attracted little support in recent decades. Two major reasons underlie the slow progress. Firstly, the partially allogamous habit of *Vicia faba* means that the development of pure-breeding lines from manually crossed $F_1$-hybrids or from open-pollinated populations takes many generations and faithful propagation of these lines (necessary before repeated field-scale measurements can be made) is vulnerable at all times to accidental cross-pollination. At 13 Gb, its large genome size - more than 20 times as large as the rice genome and almost as large as the hexaploid wheat genome - that made *Vicia faba* one of the cytogenetic models of choice in the 60s and 70s, has hampered progress in development of molecular markers with the result that today, it lags behind far less important crops in terms of genomic tools available (reviewed by Torres et al. 2010). One particular problem that researchers hoping to collate a reasonable density of markers to conduct genome-wide QTL or diversity studies is that those limited numbers of markers available such as EST-SSRs (Gong et al. 2011; Ma et al. 2011), genomic SSRs (Zeid et al. 2009), STS (Wang et al. 2011), CAPS (Gutierrez et al. 2006) and SCARs (Gutierrez et al. 2007) require a diversity of amplification, labelling and separation techniques and have not yet, to our knowledge, been systematically assessed across a unified set of germplasm.

The construction of comparative genetic maps relating gene content and order from sequenced model species, to those of unsequenced crop species has been explored as a means to partially unlock the sequence content of genetic intervals containing traits and QTL of interest. Ellwood et al. (2008) conducted just such a study relating gene order in the sequenced genome of *Medicago truncatula* to that inferred from genetic maps of grain legumes *Vicia faba* and *Lens culinaris* by linkage mapping of conserved orthologous

sequences containing mappable intronic polymorphisms in the target crop legume species. Using this strategy, a total of 127 conserved intron-targeted polymorphisms were mapped to 12 linkage groups in faba bean, constituting the first (and so far, the only) sequence-based linkage map for faba bean. Twenty-two of these polymorphisms had been developed as SNaPshot® primer extension assays suitable for capillary electrophoresis-based genotyping instruments and the remainder in the form of CAPS assays suitable for deployment on agarose gels (Ellwood et al. 2008).

Our aim in this work was to facilitate the systematic exploitation of mapped sequence-based markers in faba bean by converting a subset of the assays to a single low-cost, high throughput genotyping platform and to demonstrate their use in practice for both analysis of genetic diversity and the molecular quality control of an inbreeding programme.

*KASP assay design and validation*

Ellwood et al. (2008) used primers designed to anneal to intron-flanking sequences conserved between *Medicago truncatula* and one or more other crop legumes such as *Lupinus* spp., *Glycine max*, or *Pisum sativum* coding sequences to amplify and sequence orthologous gene fragments from faba bean genomic DNA of contrasting genotypes Vf6 (‘equina’ type) and Vf27 (‘paucijuga’ type). For KASP assay design, these amplified Vf6 and Vf27 sequences (Genbank accession codes FH893713 - FH937528) were aligned, sequence mismatches were highlighted and the particular SNPs conferring the restriction enzyme recognition sequences used for the CAPS marker development identified. The latter SNPs, in preference to any others present, were targeted for KASP assay development on the grounds that they were already experimentally verified as true SNPs rather than sequencing errors. Typically, 100-200bp of Vf6/Vf27 consensus sequence, with target SNP indicated, was used as a template for KASP assay design; with other sequence mismatches and repetitive regions counter-indicated. For marker GLIP307, an alternative SNP to the one on which the published CAPS assays was based was selected (indicated by prefix ‘new_’),

as the sequence surrounding the CAPS cleavage site was unsuitable for placement of allele-specific primers required for implementation as a KASP assay. Assays were designed for a total of 79 polymorphic sites reported by Ellwood et al. (2008) as well as for a single non-synonymous SNP (designated Vf_TFL) thought to be the causative polymorphism in the *Vicia faba* orthologue of the Terminal Flowering Locus (*TFL*) gene controlling the determinate versus indeterminate flowering habit character (Avila et al. 2007).

For the purposes of SNP validation, the parental genotypes and sources of the polymorphic sequences on which assays were designed – Vf6 and Vf27 – as well as a panel of 65 additional genotypes at various stages of inbreeding were assembled. The 65 additional genotypes consisted of part-inbred lines derived from diverse landrace materials selected from the ICARDA genebank originating from Iran, Afghanistan, Egypt, Sudan, Ethiopia, Pakistan, Syria, Palestine, Tajikistan and Azerbaijan, as well as a selection of ten ICARDA bean pure lines (BPL). Other lines included were either derived by single seed descent from commercial variety seed stocks, or were sources of traits of interest ranging from white flower (e.g. Kasztelan), determinate growth habit (e.g. Granit) and closed flower, to resistance to the stem nematode *Ditylenchus gigas* (e.g. INRA-29H) or tolerance to the parasitic broomrape *Orobanche crenata* (e.g. Vf136). A summary of the lines, their origins and identifiers are presented in **Table 1**. Since the purpose of collecting this panel was to derive highly inbred lines, a single seed from each line was sown to establish a first single seed descent generation (SSD1) and DNA extracted from this individual plant. Therefore, although names and sources of lines are given, these are for general information and our inbred derivatives may or may not carry certain characteristics associated with the historic line name/id. It should also be noted that many of the lines were inbred lines before this work began, so our denomination of the first SSD generation (SSD1) is only intended to refer to selfing generations in our hands. For some lines (noted in Table 1), we sourced DNA only. For example, from Vf27 to have a known positive control for the non-Vf6 allele, from Verde Bonita as a positive control for the determinate growth habit allele of the *VfTFL1* locus, and from a group of 4 University of Göttingen lines related to each other by pedigree. In addition

to the trait sources mentioned above, some members of the panel were included to extend the geographic and phenotypic diversity represented, whilst others were included by virtue of their status as parental lines for mapping populations owned by or currently under development by ourselves and various members of the faba bean community. DNA was prepared from young leaf material in accordance with the method of Fulton et al. (1995), and diluted to a concentration of 7ng/µl and the full panel genotyped using the 79+1=80 newly designed KASP assays (Assay1/SSD1). From a total of 80 target polymorphisms submitted, 75 were validated in this set as working assays, as judged by good allelic discrimination between Vf6 and Vf27 control DNAs and a clear clustering of allele calls into homozygous Vf6, homozygous Vf27 and heterozygous clusters. AnMtL6SNP, LG025, LG093SNP, PNDKN1SNP and PUT3SNP failed validation. However, eight further markers were eliminated from subsequent analysis as they contained >10% missing data leaving a core set of 67 robust, high quality, mostly mapped SNP markers. Detailed information permitting reproduction of any of the 75 validated SNPs including the eight markers excluded due to missing data is given in **Table 2** together with the location of the KASP markers on the linkage map of Ellwood et al. (2008).

*Towards a simple molecular hybridity test for faba bean*

We took single seeds from the genotyped founding plants of 32 lines from the panel described above into a second generation of SSD and extracted DNA from these 32 lines. The overall proportion of heterozygous genotype calls in the SSD1 generation was 3.8% but with heterozygosity in individual lines ranging from 0.0 to 0.24. In fact, 14 of the 32 lines were fully homozygous at all the loci studied including most of those indicated by their donors as ‘inbred’ lines. If seed from partially heterozygous lines was brought through further generations of SSD in pollinator-free conditions, the independent assortment of sister chromatids at meiosis and subsequent random self-fertilisation would lead to resolution of half of all heterozygous tracts into homozygous form and a halving of the overall heterozygosity rate after each round of SSD on average. In order to empirically verify that

our overwinter glasshouse conditions were sufficiently favourable to guarantee a very high rate of self-pollination, we genotyped DNA extracted from individual plants of the second SSD generation (SSD2) using the 67 most robust KASP markers (Assay2/SSD2). Of the 32 lines assayed after a further generation of single seed descent, 31 had the same or reduced levels of heterozygosity. We detected one individual (ig11290-2/SSD2) which had evidently undergone outcrossing in the previous generation as there were 18 additional heterozygous loci in SSD2 relative to the SSD1 individual. The combination of these 18 loci therefore allowed the unambiguous identification of the pollen donor in this case as belonging to ig132660-2 as illustrated in Table 3. Excluding this outcrossed individual, for the 31 second SSD generation lines judged by genotype to have been successful selfs, a total of 42 of the 78 previously heterozygous positions were homozygous in Assay2 – a loss of heterozygosity of 53.4%, a little more than the predicted 50%. The line with the highest heterozygosity was NV604 (Borington bulk) accounting for 16 heterozygous calls in SSD1 of which 9 loci resolved to homozygosity in SSD2. As illustrated in Figure 1, the overall regression of heterozygosity in SSD2 on heterozygosity in SSD1 was close to b=0.5 as expected. This demonstrates that a rapid molecular hybridity test, which can be conducted at minimal cost per line per generation efficiently detects undesired outcrossing events, and allows an empirical assessment of when a line is sufficiently inbred to be considered ‘pure’. The investment in time to extract and assay DNA is more than repaid in the saving made through not growing and maintaining derivatives of lines which are far from ‘pure’ due to occasional accidental outcrossing and we are confident that it would be possible to further streamline a molecular hybridity quality control test by reducing the number of markers and by not testing every generation. Although defining a minimal molecular hybridity assay is beyond the scope of the present work, in general two possible strategies could be envisaged. The first involves finding private alleles that distinguish each line to be inbred from all other lines likely to be in flower within the same environment as the targets. In this case, a single assay per line should be sufficient to tell whether the line has been successfully selfed or not, though the disadvantage of this approach is that each line needs to be tested with a different marker

assay. Also, where the number of available markers does not exceed the number of lines, private alleles will only be available for a small proportion of lines. An alternative approach would employ the smallest number of high information content markers that could uniquely identify each non-redundant founding genotype and keep these as a standard set. In theory, *n* assays could distinguish between $2^n$ unique genotypes, although in practice (as in the present example) it can be seen by inspection that as many as 10 markers rather than the theoretical minimum of 5 markers ($2^5$=32) may be required to uniquely identify at least one of the haplotypes present in each of the 31 lines brought to SSD2. The minimum number of markers for a given application depends on the number and average PIC of available markers as well as the size and diversity of the particular set of lines being multiplied in the same environment.

An additional way of retrospectively assessing whether a line has been propagated without any accidental outcrossing is to genotype many separate individuals from a seed stock supposed to have derived from single seed descent. To gain a provisional impression of how uniform randomly-selected seed from a single stock of an inbred line might be, three sub-lines of BPL10, two each of BPL11 and 16 individuals from a single plant progeny of Hedin/2, representing lines which were explicitly inbred for several generations prior to coming into our hands, were genotyped in Assay1/SSD1. The BPL10 and 11 sub-lines were near identical and, apart from a single off-type homozygous call in one individual, no heterozygosity or difference was seen in or between any of the 16 Hedin/2 single plant progeny siblings assayed (data not shown), allowing us to create bulks of seed from multiple individual sub-lines with a minimized risk of perpetuating any potential heterogeneity present.

*Validation of a diagnostic KASP assay for determinate growth habit*

Given that the KASPar genotyping platform is ideally suited to molecular breeding, we felt it would be beneficial to faba bean breeders who wished to rapidly introgress the determinate growth habit – to be able to reproduce the CAPS marker described by Avila et al. (2007) in

the form of a simple single-tube KASP assay. To validate the Vf_TFL KASP assay, we required that it would a) reproduce the allele calls made using the CAPS protocol for reference indeterminate (Vf6) and determinate (Verde Bonita) genotypes, b) that the 64 indeterminate lines other than Vf6 in our study would be called as possessing the indeterminate allele, and c) that the previously untested determinate Polish variety, 'Granit' would be called as having the determinate 'Verde Bonita' allele. Since the only two 'G:G' calls at the non-synonymous Leu-9 to Arg mutation in our panel were 'Granit' and 'Verde Bonita', and all the indeterminate lines including Vf6 were called as 'T:T', we considered that this KASP assay to be validated as giving equivalent information to the CAPS assay previously described.

*Marker information content and genotypic diversity in the inbred panel*

For all further analyses, relating to the use of the panel of KASP SNP assays in diversity studies/line identification, the most advanced (and therefore homozygous) generation of the line in question was taken (i.e. from Assay 2/SSD2 where available) as this represented the most inbred state of the line in question. Alleles at three loci - SAT, LG033 and GLIP71 - were private to the 'discovery' genotype Vf6, but all other markers had minimum allele frequencies ranging from 0.03 (2/67) to 0.50. Gene diversity, often referred to as expected heterozygosity, was calculated (for the biallelic loci we are concerned with here) as 2pq where *p* is the frequency of the first allele and *q* (the frequency of the alternate allele). Values ranged from 0.026 (for Vf6-private markers) to 0.5, with a mean of 0.285 across the whole set. Within this distribution of PIC values, marker subsets with varying degrees of information content can be picked out depending on the requirements of the user. Genetic relationships between the 67 faba bean lines were examined utilising the UPGMA algorithm implemented in TASSEL (Bradbury et al. 2007) and results were plotted using the interactive Tree-of-Life website (Letunic and Bork 2007). The resulting dendrogram is displayed in **Figure 2** with large (thousand seed weight >1,100g), medium (thousand seed weight 300-

1,100g) and small-seeded (thousand seed weight <300g) lines identified by red, green and blue leaf colours respectively (missing data indicated by blank space).

Vf6 and Vf27 are predictably most distant from each other due to the fact that 66/67 markers used were, by definition, polymorphic between these two lines and therefore we must underline that the relative distances of all lines in this comparison will be more a reflection of their degree of similarity or dissimilarity compared to the 'discovery' lines than an unbiased representation of true genetic distance. A much larger number of SNPs developed from a far broader discovery base than just two contrasting genotypes or an unbiased technique such as genotype-by-sequencing will be required to remove any significant ascertainment bias. Nonetheless, some of the closest genetic relationships detected – e.g. between ig12747-1 and ig12747-2 or between Hedin/2 and GOET405, 407 and 485 – are readily explained as sub-lines of the same accession or by explicit pedigree relationships. Vf27 and Vf172 are derived from the same original stock and their apparent redundancy is therefore perhaps not too surprising. Lastly, it should be noted that since the KASP assays presented here are co-dominant and mostly mapped, the data can be visualised in graphical genotype form to display chromosomal haplotypes, and used to calculate linkage disequilibrium decay rates and/or to focus on relatedness beween lines across a particular chromosomal segment of interest and in this respect, the marker system presented offers advantages over dominant unmapped anonymous Amplified Fragment Length Polymorphism (AFLP) markers previously used to assess genetic diversity amongst inbred faba bean lines (Zeid et al, 2003).

In summary, the set of KASP assays presented here provides a new and flexible tool to faba bean geneticists and breeders. The assays can be run in-house using purchased reagents or cost-effectively outsourced. The SNPs are numerous and informative enough to be used as an identification tool and have been shown to be potentially useful in carrying out molecular quality control over several generations of inbreeding. They can also be used to determine relationships between lines as long as due caution is exercised in the interpretation of these relationships. Although there may be scope to convert additional markers

described by Ellwood et al (2008) to the KASP platform to boost the number of KASP SNPs available, our efforts are currently directed towards the validation and mapping of polymorphisms discovered *in silico* from new EST sequence of contrasting lines.

**Acknowledgements**

This work has been supported by Defra as part of the Pulse Crop Genetic Improvement Network II (PCGINII) project. We thank Ken Street at ICARDA for identifying and donating seed of diverse genotypes from the ICARDA genebank, as well as Wolfgang Link (University of Goettingen), Mike Ambrose (JIC) and Gerard Duc (INRA-Dijon) for kind donations of seed from their respective germplasm collections. The full genotype dataset described here is available at http://www.viciatoolbox.org.

**Figure legends:**

**Figure 1.** Predicted versus empirical loss of heterozygosity across 38 lines genotyped over two generations of single seed descent.

**Figure 2.** UPGMA Dendrogram showing genetic relationships between accessions used in the study.

**Table 1**: Names and origins of the 67 lines used in this study.

| Source ID | NIAB acc. | Country of Origin* | Source | Noteworthy feature |
|---|---|---|---|---|
| BPL1 | NV646-3 | | ICARDA | Bean Pure Line |
| BPL10 | NV648-1 | | ICARDA | Bean Pure Line |
| BPL11 | NV649-1 | | ICARDA | Bean Pure Line |
| BPL12 | NV650-4 | | ICARDA | Bean Pure Line |
| BPL183 | NV656-3 | | ICARDA | Bean Pure Line |
| BPL21 | NV651-3 | | ICARDA | Bean Pure Line |
| BPL23 | NV652-3 | | ICARDA | Bean Pure Line |
| BPL27 | NV653-2 | | ICARDA | Bean Pure Line |
| BPL40 | NV654-1 | | ICARDA | Bean Pure Line |
| BPL63 | NV655-2 | | ICARDA | Bean Pure Line |
| ig11276 | NV013-1 | IRN | ICARDA | landrace |
| ig11290 | NV020-1 | IRN | ICARDA | landrace |
| ig11531 | NV053-5 | IRN | ICARDA | landrace |
| ig11656 | NV073-6 | AFG | ICARDA | landrace |
| ig11687 | NV079-7 | AFG | ICARDA | landrace |
| ig11695 | NV082-5 | AFG | ICARDA | landrace |
| ig11749 | NV100-5 | AFG | ICARDA | landrace |
| ig11903 | NV103-1 | AFG | ICARDA | landrace |
| ig12137 | NV129-2 | AFG | ICARDA | landrace |
| ig12159 | NV134-1 | EGY | ICARDA | landrace |
| ig12263 | NV138-1 | SDN | ICARDA | landrace |
| ig124126 | NV474-1 | PAL | ICARDA | landrace |
| ig124126 | NV474-2 | PAL | ICARDA | landrace |
| ig124213 | NV490-3 | TJK | ICARDA | landrace |
| ig124300 | NV511-1 | TJK | ICARDA | landrace |
| ig124301 | NV512-1 | TJK | ICARDA | landrace |
| ig12613 | NV150-2 | ETH | ICARDA | landrace |
| ig12658 | NV153-1 | ETH | ICARDA | landrace |
| ig12747 | NV163-2 | ETH | ICARDA | landrace |
| ig12747 | NV163-3 | ETH | ICARDA | landrace |
| ig13004 | NV175-1 | SYR | ICARDA | landrace |
| ig130596 | NV565-2 | IRN | ICARDA | landrace |
| ig130638 | NV574-1 | GEO | ICARDA | landrace |
| ig130674 | NV576-1 | IND | ICARDA | landrace |
| ig130734 | NV589-2 | AZE | ICARDA | landrace |
| ig131693 | NV590-1 | TJK | ICARDA | landrace |

| Source ID | NIAB acc. | Country of Origin* | Source | Noteworthy feature |
|---|---|---|---|---|
| ig132660 | NV594-2 | AZE | ICARDA | landrace |
| ig132813 | NV596-2 | GEO | ICARDA | landrace |
| ig14096 | NV266bw-5 | IRN | ICARDA | landrace |
| ig14189 | NV271-1 | PAK | ICARDA | landrace |
| ig14197 | NV275-2 | SDN | ICARDA | landrace |
| ig70718 | NV284-5 | EGY | ICARDA | landrace |
| ig72355 | NV318-5 | SYR | ICARDA | landrace |
| ig72423 | NV336-1 | SYR | ICARDA | landrace |
| ig72495 | NV357-1 | SYR | ICARDA | landrace |
| CGN07715 cf-3 | NV658-2 | | University of Göttingen | EU breeding line |
| FAB4000104-13 | NV660-1 | | University of Göttingen | EU breeding line |
| GOET405 | (DNA only) | | University of Göttingen | EU breeding line |
| GOET407 | (DNA only) | | University of Göttingen | EU breeding line |
| GOET485 | (DNA only) | | University of Göttingen | EU breeding line |
| Hedin/2 | NV639-1 | | University of Göttingen | derived from EU varietal material |
| Pietranera | (DNA only) | | University of Göttingen | EU breeding line |
| Albus | NV643-3 | | NIAB | derived from EU varietal material |
| Fuego | NV641-4 | | NIAB | derived from EU varietal material |
| Granit | NV642-1 | | NIAB | derived from EU varietal material |
| Kasztelan | NV644-1 | | NIAB | derived from EU varietal material |
| Maris Bead | NV640-3 | | NIAB | derived from EU varietal material |
| Verde_Bonita | (DNA only) | | IFAPA, Cordoba | EU breeding line |
| Vf136 | NV662-1 | | IFAPA, Cordoba | EU breeding line |
| Vf27 | (DNA only) | | IFAPA, Cordoba | Reference for Parental allele B |
| Vf6 | NV661-1 | | IFAPA, Cordoba | Reference for Parental allele A |
| Vf172 | NV620-1 | | IFAPA, Cordoba | paucijuga type |
| Borington Bulk | NV604-1 | | JIC | derived from EU varietal material |
| Puma | NV605-4 | | JIC | derived from EU varietal material |
| Robin | NV606-3 | | JIC | derived from EU varietal material |
| INRA-29H | NV657-1 | | INRA-Dijon | EU breeding line |
| GLV45 | NV638-2 | | INRA-Dijon | French breeding line |

*Country of origin is only given for geo-referenced landrace accessions

**Table 2:** gives KASP marker names adapted from Ellwood et al. 2008. The suffix "SNP" is used to denote the change of marker type from CAPS to SNP, the KBiosciences SNP assay identification is also given. Competitive allele-specific primers were chosen using the KBiosciences proprietary assay design process.

| SNP Name | SNP ID | LG | ~ CM | Sequence | CAPS | SNaPshot |
|---|---|---|---|---|---|---|
| 13n10_1 | 9720001 | 1 | 150 | CTTTCCCTTTAATAA[C/T]AACCAACCAAACATT | | • |
| 13n10_4SNP | 9720002 | 1 | 82 | CCTAGAGAATTTCGA[A/C]AGAGGTTTATTTTCC | • | |
| 19a15_1SNP | 9720003 | 1 | 75 | CCGAGTTTTAATCTT[C/T]CGACGTCTAGAGTCA | • | |
| 19a15_2SNP | 9720004 | 1 | 79 | TTTTCTATTCAAACW[A/C]ATGTTCAATTAATTT | • | |
| 19a15_3 | 9720005 | 4 | 83 | TGATCTTTCCTCCTA[G/T]CCCTCTTGTTTCTTC | | • |
| 28d22_4SNP | 9720006 | 1 | 82 | GCTGAAATAAACTCT[C/T]GAGTCTTGATAATAG | • | |
| 6DCSSNP | 9720007 | 5 | 0 | ATTAATCTTTTTTRA[C/T]GTAAGTTAATCATTG | • | |
| AIGPa | 9720008 | 9 | 12 | CATGCGAGAAGAAAA[C/G]GAAGAAAATGGTTCT | | • |
| AnMtL8 | 9720010 | 5 | 108 | TTTCCAATCCGGTAG[A/G]TTTTCCTTGCCATTG | | • |
| AnMtS37SNP | 9720011 | 7 | 20 | TTCAGCGGAAAAGGT[C/T]CGGTTAGCTGGAAAT | • | |
| BGALSNP | 9720012 | 6 | 0 | CGGTCGTCATTGGCC[C/T]GGATATRTTGCTCAT | • | |
| CALTL | 9720013 | 2 | 194 | AATTGTGGCAGGTAA[A/G]NTGTTAATTATGTTT | | • |
| cgP137FSNP | 9720014 | 2 | 281 | CTTCAACAGACTCCG[A/G]CGACTCTTTACGATT | • | |
| CNGC4 | 9720015 | 1 | 102 | CTCTGTTTTTTCTTT[A/C]TTCACTTAACTATAC | | • |
| CTPSNP | 9720016 | 4 | 158 | GTCGGGGAAATGTCC[A/G]CTTCCCATTTGTGTT | • | |
| CULLSNP | 9720017 | 0 | 0 | TTTGTCTTTTTATTG[A/T]TAAAGGGCAAGGATG | • | |
| FENRSNP | 9720018 | 4 | 102 | AGTTAATATCTGAGC[C/T]CTTTCTCTAGTCATG | • | |
| GLIP063SNP | 9720019 | 4 | 138 | TATATTGAATCTTTG[C/T]TATAGTGAGTTTCAT | • | |
| GLIP065SNP | 9720020 | 6 | 39 | TTTCAATTGTAAGTG[C/G]CGGTGTTACGGAGTG | • | |
| GLIP071 | 9720021 | 6 | 22 | TCTCATAGTAACTGT[C/T]TTTTAAGGTAATGTG | | • |
| GLIP081SNP | 9720022 | 3 | 165 | CATCYGAATGCAAAA[C/G]CTTCCATGAACAAAT | • | |
| GLIP089 | 9720023 | 1 | 1 | ATGCCCAATAGTTCT[C/T]ACTCATGAACATTTG | | • |
| GLIP099SNP | 9720024 | 7 | 0 | AAGAATAATTGGTGA[C/T]TTAGTAGGAGTGGTA | • | |
| GLIP107SNP | 9720025 | 0 | 0 | GTTAGTTGTATAATG[C/T]GGTACTCTGATTTGA | • | |
| GLIP133SNP | 9720026 | 2 | 190 | TATGACTTTATGTTT[A/C]GAGTAAYGAATACAA | • | |
| GLIP135 | 9720027 | 0 | 0 | AATTCTTGTTAATGC[C/T]TGCGCCCGCAAAAAA | | • |
| GLIP137 | 9720028 | 1 | 274 | GGGCTCATGACTTTT[C/G]TTTTTTGTTGGGGTT | | • |
| GLIP139SNP | 9720029 | 1 | 178 | GGGGTTGATTTAGAT[A/T]AATTTTTATCACTTG | • | |

| | | | | | | |
|---|---|---|---|---|---|---|
| GLIP245SNP | 9720030 | 2 | 240 | TATTGTTACACTGTT[C/T]AATGGTTTTGGTTGA | | |
| GLIP253SNP | 9720031 | 1 | 147 | AATAATGCTCAAGAG[C/T]CAAAACCTCTTATAT | • | |
| GLIP265SNP | 9720032 | 3 | 168 | CTACTCTAGGCATTA[A/T]NGTCACCTTCGATGA | • | |
| GLIP291SNP | 9720033 | 4 | 141 | CAATATTTATTGTTA[C/T]GTCTGCAGATACATT | • | |
| GLIP337SNP | 9720034 | 5 | 24 | TTAAGTTTTTTAATT[A/C]TGTTAGATTTTRCTC | • | |
| GLIP427SNP | 9720035 | 5 | 8 | CCGTCTGATTGATCT[G/T]GATGGCTAGCACTGC | • | |
| GLIP451SNP | 9720036 | 2 | 238 | AAAGCAAAAGCTTTC[A/G]CGATTTCAGTTAGTA | • | |
| GLIP621aSNP | 9720037 | 1 | 233 | TTTTATTATTGTTTG[C/T]AGAGCTCTGCACTTG | • | |
| GLIP651aSNP | 9720038 | 2 | 286 | GTATNNGCGCGCGTA[C/T]GTAGGTACTGAACTT | • | |
| GLPSNP | 9720039 | 2 | 226 | TAACACTCTTTTCAC[C/G]AAAGTATTGAACAAA | • | |
| HBP2SNP | 9720040 | 3 | 159 | ACACKCATTAAAAAT[A/T]TATATACCTCAAAAT | • | |
| HYPTE3SNP | 9720041 | 3 | 22 | TCCGAGTTGGAAAAT[A/T]TATCGATTGTTTTTA | • | |
| SNP Name | SNP ID | LG | ~ CM | Sequence | CAPS | SNaPshot |
| LG007SNP | 9720042 | 8 | 2 | TTATTTGCCAAAGCT[C/T]GAATGTATAGTTATC | • | |
| LG018SNP | 9720043 | 8 | 6 | ACAAAATAATTTAAC[C/T]GAGAATCTTACCTTC | • | |
| LG023SNP | 9720044 | 8 | 9 | TTCAACCTGTAATAT[G/T]CACAAAATCATGTAT | • | |
| LG031SNP | 9720046 | 2 | 179 | TCTTGGCCAAGTGTT[C/G]ACATATACCTTGCAC | • | |
| LG033SNP | 9720047 | 2 | 172 | TTGTGCAGTGCCGAC[A/G]CTTTTGTTTAGAAGT | • | |
| LG038SNP | 9720048 | 2 | 182 | GCCTGTAAGACCCTC[A/G]TGCTTACTGAATATA | • | |
| LG041SNP | 9720049 | 2 | 170 | TTCACTTTTCATTCA[C/T]CTTTCAGTCACTAAT | • | |
| LG054SNP | 9720050 | 2 | 153 | CCACAATTAAACAGG[A/G]TGATAGTTTTCAGCA | • | |
| LG068SNP | 9720051 | 3 | 17 | TACATTAACCTCTGG[C/T]GGTGATTAAACAGTA | • | |
| LG085SNP | 9720052 | 1 | 249 | CTTCTAGTTTTATGT[A/G]CACGGTTTTTTTTAT | • | |
| LG101 | 9720054 | 2 | 157 | GTGTTTCTCTTATTC[A/G]TTTGAGCTTGTGTAA | | • |
| LG102 | 9720055 | 9 | 33 | AATTTAATATAAAAA[A/C]NATAATATACTAAAG | | • |
| LG107 | 9720056 | 9 | 64 | CTGCATCCCTATAAA[A/T]TTTCATATGGTACTG | | • |
| LSSR9bSNP | 9720057 | 2 | 31 | TTCCNTTTTAAAGTT[G/T]TAAACTAATTGACTT | • | |
| LUP052SNP | 9720058 | 8 | 20 | AAGTGGGTGCAATCC[C/T]CACCTTACAAGCCGG | • | |
| LUP066 | 9720059 | 2 | 83 | CATAAAAACNACAAY[A/T]GACAATTGGAATTGA | • | |
| LUP091SNP | 9720060 | 4 | 31 | GACTCGTGGAAGTAA[A/T]TTTTTGTTTTTGTTT | • | |
| LUP108SNP | 9720061 | 10 | 0 | TTTTTCCTCTTTCTC[G/T]AGTGTTATGTGTTGC | • | |
| LUP266cSNP | 9720062 | 1 | 220 | TTTNNTNCTGTTCAW[A/T]TTTTATGGCTATATA | • | |
| MMK1A | 9720063 | 3 | 122 | CATGTTTTTGATGGG[A/G]AYGGAAAATGGATTT | | • |

| | | | | | | |
|---|---|---|---|---|---|---|
| New_GLIP307SNP | 9720064 | 12 | 24 | AAGTGGGAAGAAGTT[A/T]NTTTTTTTTTNNNGA | | ● |
| PEPCASESNP | 9720065 | 8 | 51 | TGGAACTGTTGGAAG[A/G]GGAGGTGGACCTACT | ● | |
| PGDH | 9720066 | 0 | 0 | GGGTGAACTCGCCCG[A/T]ATCTGGAAAGGAGGC | | ● |
| PPHSNP | 9720068 | 1 | 134 | TTTGCAAAGGAGACT[C/T]TGGAGATTTTTGCAC | ● | |
| PRATSNP | 9720069 | 1 | 34 | GACGGTGCTGGTGCA[C/T]ATTCTTTTTCTGATG | ● | |
| RBPC_0SNP | 9720071 | 5 | 40 | GGGTGGWACWACCCA[A/G]TACACTGTCAACAAC | ● | |
| REPSNP | 9720072 | 2 | 165 | AACTTTACGTTTAAT[C/T]CGATTGTAATTSTTT | ● | |
| RNARSNP | 9720073 | 3 | 218 | TTCATACTGGGTGTA[A/C]AGATTCTCATATTCT | ● | |
| SARBSNP | 9720074 | 4 | 212 | TCCATCTTTAATCCT[C/T]CATCATCTTATAAAA | ● | |
| SATSNP | 9720075 | 2 | 111 | GAGAAGGTATGTCCC[A/G]CTTTTCCGTTTTTGA | ● | |
| SUSYSNP | 9720076 | 1 | 282 | TTCAGGGTAGAACGA[C/T]GTCAACCTGCGGCTA | ● | |
| TBB2SNP | 9720077 | 10 | 42 | GACATTCATTGGAAA[C/T]TCAACATCGATTCAG | ● | |
| tRALSSNP | 9720078 | 3 | 31 | GTACAAATAAAACTC[A/G]AGCAGAAACAAAGAA | ● | |
| UNK28SNP | 9720079 | 7 | 78 | TCCYGTAGAATTGTC[C/T]GGAGAGGGAAGACCT | ● | |
| Vf_TFLSNP | 9720080 | 0 | 0 | TGGCTCAAGAACCAC[G/T]AATTGTTGGAAGAGT | | ● |

**Table 3.** Graphical genotypes of line ig11290-2 SSD1 (ig11290-2/S**1**) and SSD2 (ig11290-2/S**2**) generations. Only the 19 loci which showed non-maternal alleles are shown. The last SNP shown – PRATSNP – corresponds neither to the maternal nor the proposed paternal parent, and as such cannot be properly explained other than by invoking genotype error.

| | 19a15_1SNP | | 19a15_2SNP | | AnMtL8 | | CTPSNP | | FENRSNP | | GLIP099SNP | | GLIP135 | | GLIP265SNP | | GLIP291SNP | | GLIP651aSNP | | GLPSNP | | LG031SNP | | LG038SNP | | LG102 | | LG107 | | LUP108SNP | | New_GLIP307SN P | | PGDH |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| **ig11290-2/S1** | T | T | C | C | G | G | G | G | T | T | T | C | T | C | A | A | C | C | C | C | G | C | G | C | A | A | C | C | T | T | T | G | T | T | A |
| **ig11290-2/S2** | T | C | C | A | G | A | G | A | T | C | T | C | T | C | T | A | T | C | ? | ? | C | C | C | C | G | A | C | A | T | A | T | G | T | A | T |
| **ig132660-2** | C | C | A | A | A | A | A | A | C | C | T | T | T | T | T | T | T | T | T | T | C | C | C | C | G | G | A | A | A | A | G | G | A | A | T |

**References:**

Avila C, Atienza S, Moreno M, Torres A (2007) Development of a new diagnostic marker for growth habit selection in faba bean ( Vicia faba L.) breeding. TAG Theoretical and Applied Genetics 115:1075-1082

Bradbury PJ, Zhang Z, Kroon DE, Casstevens TM, Ramdoss Y, Buckler ES (2007) TASSEL: software for association mapping of complex traits in diverse samples. Bioinformatics 23:2633-2635

Ellwood S, Phan H, Jordan M, Hane J, Torres A, Avila C, Cruz-Izquierdo S, Oliver R (2008) Construction of a comparative genetic map in faba bean (Vicia faba L.); conservation of genome structure with Lens culinaris. BMC Genetics 9:380

Fulton TM, Chunwongse J, Tanksley SD (1995) Microprep protocol for extraction of DNA from tomato and other herbaceous plants. Plant Molecular Biology Reporter 13:207-209

Gong Y-M, Xu S-C, Mao W-H, Hu Q-Z, Zhang G-W, Ding J, Li Z-Y (2011) Generation and characterization of 11 novel EST derived microsatellites from Vicia faba (Fabaceae). American Journal of Botany 98:e22-e24

Gutierrez N, Avila C, Rodriguez-Suarez C, Moreno M, Torres A (2007) Development of SCAR markers linked to a gene controlling absence of tannins in faba bean. Molecular Breeding 19:305-314

Gutierrez N, Avila CM, Duc G, Marget P, Suso MJ, Moreno MT, Torres AM (2006) CAPs markers to assist selection for low vicine and convicine contents in faba bean (Vicia faba L.). Theoretical and Applied Genetics 114:59-66

Letunic I, Bork P (2007) Interactive Tree Of Life (iTOL): an online tool for phylogenetic tree display and annotation. Bioinformatics 23:127-128

Ma Y, Yang T, Guan J, Wang S, Wang H, Sun X, Zong X (2011) Development and characterization of 21 EST-derived microsatellite markers in Vicia faba (fava bean). American Journal of Botany 98:e22-e24

Torres AM, Avila CM, Gutierrez N, Palomino C, Moreno MT, Cubero JI (2010) Marker-assisted selection in faba bean (Vicia faba L.). Field Crops Research 115:243-252

Wang H, Yang T, Guan J, Ma Y, Sun X, Zong X (2011) Development and characterization of 20 novel polymorphic STS markers in Vicia faba (fava bean). American Journal of Botany 98:e189-e191

Zeid M, Mitchell S, Link W, Carter M, Nawar A, Fulton T, Kresovich S (2009) Simple sequence repeats (SSRs) in faba bean: new loci from *Orobanche*-resistant cultivar 'Giza 402'. Plant Breeding 128:149-155

Zeid M, Schön CC, Link W (2003) Genetic diversity in recent elite faba bean lines using AFLP markers. Theoretical and Applied Genetics 107:1304-1314