Advances in faba bean genetics and genomics

Article

Accepted Version

O'Sullivan, D. M. ORCID: https://orcid.org/0000-0003-4889-056X and Angra, D. ORCID: https://orcid.org/0000-0002-6681-0597 (2016) Advances in faba bean genetics and genomics. Frontiers in Genetics, 7. 150. ISSN 1664-8021 doi: 10.3389/fgene.2016.00150 Available at https://centaur.reading.ac.uk/67667/

Advances in faba bean genetics and genomics

Running title: Faba bean genetics and genomics

Donal M. O'Sullivan, Deepti Angra.
School of Agriculture, Policy and Development,
University of Reading,
PO Box 237,
Whiteknights,
Reading,
RG6 6AR,
United Kingdom.
d.m.osullivan@reading.ac.uk

Abstract:

Vicia faba L, is a globally important grain legume whose main centres of diversity are the Fertile Crescent and Mediterranean basin. Because of its small number (six) of exceptionally large and easily observed chromosomes it became a model species for plant cytogenetics the 70s and 80s. It is somewhat ironic therefore, that the emergence of more genomically tractable model plant species such as *Arabidopsis* and *Medicago* coincided with a marked decline in genome research on the formerly favoured plant cytogenetic model. Thus, as ever higher density molecular marker coverage and dense genetic and even complete genome sequence maps of key crop and model species emerged through the 1990s and early 2000s, genetic and genome knowledge of *Vicia faba* lagged far behind other grain legumes such as soybean, common bean and pea.

However, cheap sequencing technologies have stimulated the production of deep transcriptome coverage from several tissue types and numerous distinct cultivars in recent years. This has permitted the reconstruction of the faba bean meta-transcriptome and has fuelled development of extensive sets of Simple Sequence Repeat and Single Nucleotide Polymorphism (SNP) markers.

Genetics of faba bean stretches back to the 1930s, but it was not until 1993 that DNA markers were used to construct genetic maps. A series of Random Amplified Polymorphic DNA-based genetic studies mainly targeted at quantitative loci underlying resistance to a series of biotic and abiotic stresses were conducted during the 1990's and early 2000s. More recently, SNP-based genetic maps have permitted chromosome intervals of interest to be aligned to collinear segments of sequenced legume genomes such as the model legume *Medicago truncatula*, which in turn opens up the possibility for hypotheses on gene content, order and function to be translated from model to crop. Some examples of where knowledge of gene content and function have already been productively exploited are discussed.

The bottleneck in associating genes and their functions has therefore moved from locating gene candidates to validating their function and the last part of this review covers mutagenesis and genetic transformation, two complementary routes to validating gene function and unlocking novel trait variation for the improvement of this important grain legume.

Keywords: *Vicia faba*; transcriptome; single nucleotide polymorphism; genetic linkage map; synteny.

Introduction

Origins and importance of Vicia faba in world agriculture

Vicia faba L.(*Vf*) or faba bean is a grain legume of great importance in world agriculture due to its high yield potential compared to alternative grain legumes, its ability to fix nitrogen through symbiosis with *Rhizobium leguminosarum* in its root nodules, but most crucially for its role as a staple dietary protein source in North African and Middle Eastern cultures. The species is thought to have been domesticated in the Eastern Mediterranean region, perhaps somewhere between Afghanistan and the Eastern Mediterranean (van de Wouw et al., 2001), but no extant wild relative has yet been found although new species closely related to *Vf* have been found in recent decades (Maxted, 1993). Nonetheless, there is great variability within the domesticated genepool, with the major centre of diversity centred around the Mediterranean basin and secondary centres of diversity in the Nile Valley, South America, and Central and Eastern Asia (Duc et al., 2010), providing much untapped potential to breeders. This diversity of forms is exemplified by the botanical classification of '*major*' (large-seeded or 'broad' bean), '*equina*' (mid-sized or 'horse' bean) and '*minor*' (small, rounded seed) types. *Vf* is one of the earliest domesticated Old World agricultural crops, with credible archaeobotanical evidence linking it to the pre-pottery Neolithic period around 10,000 BP in a site in North-West Syria (Tanno and Willcox, 2006). However, again in keeping with its great diversity and adaptability, it has a long history of utilisation outside the centre of origin in diverse agroecological settings from the Boreal and Atlantic Maritime climates to arid and sub-tropical regions. The most current available statistics for dry beans show it ranking 7th in terms of global production of grain legumes with production in 2013 of 1.17 million tonnes (FAOSTAT, 2013), though this probably grossly underestimates the nutritional and societal importance of the species as a majority of faba bean worldwide is cultivated and consumed locally by subsistence farmers. Furthermore, FAOSTAT figures relate to dry beans and there are no reliable global data for large-seeded broad bean types picked green and consumed fresh. China is the leading producer, followed currently by Ethiopia, Australia, France, Egypt, Morocco, Sudan and the UK, though rank order amongst the latter seven countries has historically been somewhat fluid. Despite the strong cultural attachment of Mediterranean basin and Middle Eastern populations to consumption of faba bean, for a variety of reasons addressed below, many of these countries have become net importers.

Challenges facing faba bean

Yield stability

Faba bean is generally speaking the most productive of grain legumes in environments where rainfall is not limiting or in irrigated conditions, and can indeed be a highly profitable crop, especially if the economic benefits of biologically fixed nitrogen and enhanced weed and disease control in subsequent crops are considered (Preissel et al., 2015). However, in common with many legumes, its yields are relatively unstable, which is thought to be one important reason underlying the low inclusion rate of leguminous crops in European agriculture in particular (Cernay et al., 2015). This inherent yield instability is thought to be due, in part, to its apparently profligate flowering habit whereby many flowers spread over many nodes are produced containing far more ovules than the capacity of the plant to fertilize and fill those potential seed sites, resulting in a high but variable rate of flower and fertilized ovule abortion. Optimal yield production in faba bean is also dependent on symbiosis with *Rhizobium leguminosarum* biovar *viciae* to produce nitrogen-fixing root nodules as well as on the pollination services of wild bee populations to ensure both optimal seed set and outcrossing rates. Pollinator insufficiency has been shown recently in the UK to explain up to 64% yield loss (Nayak et al., 2015) while a large N-fixation response to rhizobial inoculation in the field has been shown even

in the presence of high natural rhizobial populations (Denton et al., 2013). On top of these variable symbiotic or mutualistic interactions are layered a host of abiotic and biotic stresses, combinations of which will exact further environment-dependent losses. The many specific pests and pathogens of faba bean include fungi (*Ascochyta fabae* – ascochyta blight, *Botrytis fabae* – chocolate spot, *Uromyces fabae* – bean rust, *Peronospora viciae* f.sp. *faba* - downy mildew), insects (*Aphis fabae* – black bean aphid, *Bruchus rufimanus* – the bruchid seed beetle), nematodes (*Ditylenchus gigas* – stem nematode) and parasitic plants (*Orobanche crenata* and *Orobanche foetida* – broomrape) to name just a few of the most prominent. Detailing the fascinating coevolutionary struggles between these diverse organisms and their common host is beyond the scope of this review and interested readers can find further details in one of many recent reviews of biotic stresses on faba bean (Sillero et al., 2010 Stoddard et al., 2010; Pérez-de-Luque et al., 2010). Chief amongst abiotic stresses are heat and drought, which play all the more important a role in yield determination because of the sensitivity of flower fertility to even mild or transient levels of stress. Fertility of developed flowers has recently been shown to be highly sensitive to heat stress, though total yield losses are mitigated by the shift of yield production to higher nodes formed after a stress event has passed (Bishop et al., 2016b). And a further study showed that bee pollination can significantly mitigate damage to fertility caused by heat stress (Bishop et al., 2016a). Cold/frost tolerance is an important trait for winter beans (Sallam et al., 2015), and tolerance to salinity soils is pertinent in semi-arid regions (Tavakkoli et al., 2012). Again, abiotic stress resistance is a wide topic which has been adequately covered in recent reviews (Patrick and Stoddard, 2010 Link et al., 2010, Khan et al., 2010).

Quality

Meeting the protein demand of a growing global population represents a challenge not only from the yield perspective but also from the quality perspective. As yield potential continues to increase through the efforts of breeders, it is important for the protein density of the crop to be maintained, if not improved, from typical values of around 30% dry weight. The well-established relative deficiency of faba bean in essential sulphur-containing amino acids, cysteine and methionine, can of course be balanced in food and feed processing, but amelioration of the balance of essential amino acids would be a valuable breeding target which has not yet been addressed (Multari et al., 2015). Finally, in order for the nutritive value of faba protein to be maximized and nitrogen- and phosphorous-rich wastes in the food chain minimized, serious research is needed to identify and remove anti-nutritional factors that inhibit the normal digestion of starch and protein in the gut (amylase and protease inhibitors), that cause oxidative stress (e.g. vicine) or that sequester key nutrients (e.g. phytic acid). Much research on faba bean quality has focussed on two key anti-nutritional factors: tannins and vicine/convicine. Vicine and convicine are particularly intriguing as the metabolized derivatives of these pyrimidine glucosides cause a serious and potentially fatal condition known as favism in genetically predisposed humans and are found only in a handful of *Vicia* spp, the most notable of which is faba bean (Ray and Georges, 2010). Seed coat tannins and vicine-convicine are naturally high in a large majority of current varieties (e.g. Khamassi et al., 2013) and have been shown to lower protein digestibility and energy content in a variety of animal feeding studies. The nutritional quality of faba bean and status of research on anti-nutritional factors have been thoroughly reviewed by Crepon et al., (2010).

Against this backdrop of high yield potential and high utility in increasing the sustainability of nitrogen cycling in agro-ecosystems and security of protein supply in the food chain, it is both timely and appropriate to review how genetic and genomics are already contributing to the accelerated breeding of high yielding, climate-resilient and nutritious faba beans and how we can expect these technologies to be exploited in the future.

Genomics

Transcriptomes

The first major contribution to systematic faba bean transcriptome knowledge was the release of approximately 5,000 Expressed Sequence Tags (EST) from developing embryos of a broad bean variety 'Windsor' described by (Ray and Georges, 2010). This study provided a useful snapshot of the functional classification and relative expression level of the more abundant transcripts from the embryo transcriptome in the early to middle stages of its development, even though restricted to one genotype and one tissue. Kaur et al., (2012) and Yang et al., (2012) undertook 454 sequencing of the transcriptome specifically to underpin SSR discovery and this significantly increased the volume of transcriptome data available although in both cases from mixed genotypes. Subsequently, studies began to encompass multiple, separate inbred genotypes (Webb et al., 2016; Ocaña et al., 2015) as well as multiple tissues and genotypes (Ray et al., 2015). The deepest transcriptome coverage yet produced has come from Illumina sequencing of a library of mixed tissues enriched with embryo transfer cells from variety 'Fiord' (Arun-Chinnappa and McCurdy, 2015; Zhang et al., 2015). At the time of writing, transcriptomes from nine specifically identified single genotypes and a selection of tissues including whole seedling, root, shoot, leaf, seed coat and embryo were available (summarised in **Table 1**), which offers for the first time to the faba bean research community the possibility to conduct crude electronic Northern analyses and to mine genotypic variants from multiple genetic backgrounds.

SNPs

The extreme paucity of *Vf* sequence of any description in public databases, lamented by previous reviewers e.g. Gnanasambandam et al., (2012), meant that very small numbers of SNPs had been discovered prior to 2014. The first genetic linkage map of faba bean to explicitly target gene-based polymorphisms was reported by Ellwood et al., (2008), who adopted a strategy of cross-species amplification of conserved orthologues in order to identify polymorphic intron-targeted markers, which were implemented at first either as Cleaved Amplified Polymorphic Sequence (CAPS) or Single Nucleotide Primer Extension (SNuPE) assays. Later, many of these polymorphic intron sequences were converted from CAPS/SNuPE to Kompetitive Allele Specific PCR (KASP) format (Cottage et al., 2012), which made the exploitation of this first suite of SNPs more accessible. The advent of RNA-Seq datasets fuelled the next wave of SNP development. Kaur et al., (2014) designed an iSelect assay based on 768 SNPs, of which 551 were placed on a genetic map generated from a RIL population of the cross Icarus x Ascot. More recently, Webb et al., (2016) reported design of individual KASP assays for 845 SNPs mined from alignment of assembled transcriptomes of 'Albus' and 'BPL10' inbred lines; of these, 653 were successfully mapped. The burgeoning transcriptome datasets described in the previous section permit ever greater numbers of SNPs to be called, for example Ray et al., (2015) reported 5,300 unique variants where alternate alleles could be found in one more of the discovery genotypes SSNS-1, A01155, or CDC Fatima and Ocaña et al., (2015) reported 39,060 SNP and 3,669 InDel polymorphisms in their analysis of transcriptomes of INRA-29H and *Vf*136, though these larger SNP sets remain to be validated as working SNP assays.

We have taken this SNP mining exercise further by recently conducting a meta-transcriptome assembly using all *Vf* transcriptome datasets available on 1st October 2015 and scoring mapped reads from individual genotypes for SNP variants. This SNP mining exercise, detailed results of which will be presented elsewhere, resulted in identification so far of more than 320,000 SNPs, of which 69,407 have a 'Freebayes' SNP quality score >200. Our analysis of variant content of the *VC* interval provides an illustration of how SNP mining can be useful in practice. The *VC* gene was recently

mapped to a 4.6 cM interval on the tip of *Vf* chromosome I between SNP markers orthologous to Medtr2g007220 and Medtr2g010740 (Khazaei et al., 2015). As this interval lies within a larger region of largely uninterrupted collinearity with Mt2, it is reasonable to hypothesize that gene order and content within this interval is largely conserved in *Vf*. According to our preliminary analysis, of the 308 predicted Medicago (Mt4.0) genes within the interval Medtr2g007220-Medtr2g010740, there is at least partial transcript coverage in the *Vf* meta-transcriptome for 179, and of these, 146 contain a total of 840 high confidence SNPs. **Figure 1** shows a typical informative alignment for a gene towards the centre of this interval. Over the 1534 bp contig, which covers the full coding region predicted in Medicago together with some 5' and 3'-UTR sequence, there are 16 varietal *Vf* SNPs. We are currently in the process of validating a number of SNPs mined from genes predicted by synteny to belong to the *VC* region, ordering them genetically and using them to further delineate the *VC* interval (Angra & O'Sullivan, in preparation).

EST-SSRs and genomic SSRs

The other category of molecular marker which has received much attention from the faba bean research community was Expressed Sequence Tag - Simple Sequence Repeats (EST-SSRs). As inherently highly discriminatory, co-dominant markers embedded in genic locations which can be readily associated with orthologous positions in related species, EST-SSRs are an attractive category of marker and offer the advantage that repeats can be mined from a single genetic background with relatively low sequence coverage, so were in the first wave of marker sets to be developed from transcriptome data. Initially modest numbers of EST-SSRs reported by Gong et al., (2011); Ma et al., (2011) were followed by greater numbers of candidate and validated EST-SSRs (El-Rodeny et al., 2014; Kaur et al., 2012). Although highly portable and informative, EST-SSRs, as with all types of SSR marker, can be difficult to score in a fully automated fashion, and are inherently less amenable than SNPs and INDELs to conversion to high throughput parallel assay formats.

In contrast to EST-SSRs, genomic SSRs anchored in non-coding genomic sequence may be difficult to validate as locus-specific assays due to the prevalence of complex, repetitive DNA sequences outside coding regions, which constrain the design of locus-specific primers. Furthermore, most genomic SSRs which fall outside conserved coding regions cannot be used to make syntenic bridges to better characterised relatives. Nonetheless, genomic SSRs, mined from SSR-enriched genomic DNA libraries, have been successfully developed by Zeid et al., (2009) and Yang et al., (2012). **Table 2** summarises key molecular marker sets which have been developed to date for faba bean.

Genetics

Much progress has been made in the power and possibilities of genetic analysis in faba bean since the pioneering work of Adela Erith, working at the University of Reading in the 1920s, on inheritance of colour, size and form of seeds and other traits (Erith, 1930). Erith's seminal paper showed Mendelian inheritance of genes controlling readily observable (but nonetheless important) characters including hilum colour, seed and flower colour and height. However, many years were to pass before the genes underlying these or any other traits have been identified. Between then and now, *Vf* genetics has passed through distinct phases which will be described below in chronological order.

Classical genetic analysis

The period 1930-1993 was the era of molecular marker-free genetic analysis. During this period and beyond, useful foundations were laid showing simple inheritance patterns for a variety of heritable traits. As mentioned already, Erith, (1930) found single dominant or semi-dominant genes explaining

flower, seed and hilum colour as well as plant height, though there was no well-developed locus nomenclature in this early work. Faba bean received much attention in the 1960s to the early 1990s as a model for cytogenetics due to its uncommonly large haploid genome size of 13 Gb (Soltis et al., 2003) and modest haploid chromosomes number (n=6), which made for large, readily observable chromosomes. For a time, the study of chromosome breakage response to a range of physical and chemical agents became quite fashionable (e.g.Menke et al., 2000; Milan and Upadhyay, 2007; Rybaczek et al., 2008; Sobita and Bhagirath, 2005) and *Vf* was one of the early crop species whose chromosomes were found to be amenable to flow sorting (Kovarova et al., 2007). The more direct relevance of cytogenetics to modern molecular genetics was the identification of asynaptic mutants (Sjödin, 1970), on which a series of trisomic stocks were founded. Genetic analyses of crosses involving trisomic parents allowed genetic markers to be assigned to physical chromosomes (Patto et al., 1999). The definitive work on pre-molecular faba bean genetics was the monumental work of Sjödin, 1971 to collect, induce, classify and cross mutants and spontaneous variants in a wide variety of characters of interest. Although such comprehensive surveys of genetic variation have not since been carried out, induced and spontaneous mutants affecting a variety of traits have been described in this period. For instance, a gene controlling symbiosis, *sym-1*, was described by Duc and Picard, (1986). Duc et al., (1999) reported a recessive allele of the *i1* locus conferring a green cotyledon; this has a similar phenotype as the green cotyledon allele of the *Sc* locus, mapped by Khazaei et al., (2014) to chromosome IV.

Work on genetics was not confined to visible morphological phenotypes either. Ramsay, (1997) showed Mendelian inheritance of a major effect seed dormany gene, denoted *doz*, detected by scoring the timing of seed germination semi-quantitatively in a set of segregating Recombinatn Inbred Lines (RILs). Unfortunately, few of the populations generated and characterized in these early genetic studies have been maintained to the present day, though the trait sources (or at least sources carrying analogous phenotypic states) have for the most part. However, quantitative traits, especially those under oligogenic control, were always going to require a molecular mapping approach so that medium or even small size effects could be placed within a sufficiently dense phenotype-independent marker framework.

Molecular genetics

A combination of isozymes and RAPD markers formed the basis of the first *Vf* molecular marker map (Torres et al., 1993), consisting of 66 markers arranged in eleven linkage groups, in which, interestingly, the first indications of synteny between faba bean and other *Fabaceae* were noted.

From this point on, *Vf* genetic studies followed a standard pattern. Typically biparental RIL or F_2 populations generally consisting of no more than 200 progeny lines were genotyped with non-sequence based markers, chiefly RAPD. The highlights of this phase of faba bean genetics were a series of RAPD-based QTL studies targeting resistance to pathogens and parasites – for instance, to ascochyta (Avila et al., 2004; Diaz-Ruiz et al., 2009), rust (Avila et al., 2003), and *Orobanche* (Diaz-Ruiz et al., 2010; Gutierrez et al., 2013). This early period of integration of molecular marker technology in genetic mapping studies has been reviewed by Torres et al., (2010) and Gnanasambandam et al., (2012).

In parallel with the acceleration in growth of faba bean sequence and marker datasets, there has been a correspondingly encouraging increase in the density and utility of gene-based genetic maps. The first gene-based genetic map of *Vicia faba*, composed in the main of 127 co-dominant, portable, Intron-Targeted Amplified Polymorphism (ITAP) markers, was that of Ellwood et al., (2008). A linkage map comprising 128 EST-SSR markers was produced by Ma et al., (2013) and a 552- locus map

comprising loci generated from 235 faba bean-derived EST-SSRs and 162 markers derived by cross-amplification from red and white clover was reported recently by El-Rodeny et al., (2014). The consensus map of Satovic et al., (2013) was the first where a majority of markers mapped to just six chromosomally assigned linkage groups, though a minority of these were gene-based, transportable, co-dominant markers. More recently, 551 SNPs and 71 SSRs were combined in Kaur et al., (2014), though not all of the 12 linkage groups reported could be definitively assigned to one of the six physical chromosomes. The densest SNP coverage available in a fully physically anchored consensus linkage map to date is that reported by Webb et al., (2016). This combined 34 SNP markers discovered in *Vf*6/*Vf*27 backgrounds by Ellwood et al., (2008) and converted to KASP format by Cottage et al., (2012) with 653 new 'Albus' x 'BPL10' SNPs into a single 687-locus consensus map with all markers mapping to just six linkage groups each of which could be assigned to a physical chromosome.

Description of Synteny

The great benefit of the progress in mapping ever greater numbers of sequence-based markers is that the by now well-established conservation of gene order amongst related legume genomes could be used to anchor genetic maps from unsequenced legumes (in this instance faba bean) to the *Medicago truncatula* (*Mt*) genome. The first sequence-based genetic map of faba bean which allowed the global pattern of *Vf-Mt* synteny to be observed was that of Ellwood et al., (2008). Although the number of markers (127) was modest, a clear picture of extensive macrosynteny emerged. El-Rodeny et al., (2014) and Kaur et al., (2014) elaborated this picture with a greater marker density, but the clearest and most complete picture of the extent of macrosynteny between *Vicia faba* and *Medicago truncatula* comes from the Webb et al.,(2016) study. All 653 newly discovered and mapped SNP markers in this latter study were selected following highly conservative filtering for single copy *Vf* sequences with a clear best reciprocal BLAST hit relationship with a single copy gene in *M. truncatula*. Reflecting these stringent marker design criteria, the name of the each new Webb et al., (2016) *Vf* marker carries explicit reference to its presumed *Mt* orthologue. Combined with the fact that each linkage group in the Webb et al., (2016) consensus map corresponds to a full physical *Vf* chromosomes meant that for the first time all six *Vf* chromosomes could be aligned to the Mt sequence without spurious interruptions to macrosynteny pattern caused by lack of marker coverage.

Exploitation of synteny

The hallmark of 'pre-synteny' genetic studies in all species, *Vf* being no exception, was that the reporting of a given gene or QTL in proximity of a certain molecular marker was quite frequently the end of the story as there was no method to target markers to a region of interest without screening vast numbers of anonymous markers such as AFLPs or RAPDs.

The year 2012 marked a new departure for faba bean with sequence-based and synteny-anchored marker maps being applied to flowering time QTL (Cruz-Izquierdo et al., 2012). Likewise, detection of ascochyta resistance QTL by Kaur et al., (2014) permitted localization of QTL in regions with clear colinearity to fully sequenced model genomes. Khazaei et al., (2014) studying stomatal traits and Khazaei et al., (2015) mapping vicine-convicine content were able to do likewise. In theory now, a given QTL would be associated with an interval whose gene content and even order could be predicted by exploiting synteny with Medicago and marker development efficiently targeted to this predicted gene content.

As outlined in the previous sections, much of this progress has stemmed from the exploitation of sequencing and cost-effective genotyping technologies to achieve reasonably dense coverage of all six

Vf chromosomes with sequence-based molecular markers that allow the *Vf* gene-based genetic map to be confidently aligned in large part with fully sequenced reference genomes such as Medicago (Young et al., 2011) or soybean (Schmutz et al., 2010).

The exploitation of macrosynteny between crop and model genomes has a number of benefits. Firstly, once a trait has been mapped to a genetic interval which aligns well to a segment of a model sequenced genome, knowledge of gene function in the model species can be translated back to target crop species. In an instance illustrated in **Figure 2**, *ZT1* (controlling flower pigmentation as well as seed coat tannins) was mapped to the *Vf*_Mt3g092810_001 – *Vf*_Mt3g094760_001 interval, with clear synteny to a portion of Mt3. Perusal of the annotation information for the syntenic interval in Medicago revealed a logical biological candidate in the form of the *Transparent Testa Glabra 1* (*TTG1*) WD40 transcription factor (Medtr3g092840), which had previously been shown in *Mt* to determine flower colour, and in the follow-up of this hypothesis, a deletion in the recessive allele sequence was found to plausibly explain the *zt* (unpigmented) phenotype (Webb et al., 2016). This co-called translational genomics approach leverages prior investment in model species biology and fast-tracks causative allele identification. Secondly, quite independently of translation of biological information, molecular marker targeting to specific regions of interest is possible using the syntenic framework. This is relevant in situations where the target interval is large or initial candidate functional information is absent and is illustrated in **Figure 2** by the case of the *VC* locus. The published interval in which this gene maps is collinear with part of Mt2 (Khazaei et al., 2015) and as discussed earlier, SNPs discovered in sequences orthologous to Medicago genes in the syntenic interval can be used selectively and cost-effectively in further high-resolution mapping of the locus. Thirdly, this syntenic framework can in principle be used to reverse map genes in *Vf*. Here, we use the example of the *Vf TERMINAL FLOWER 1* (*VfTFL1*) gene. Avila et al., (2007) examined the translational hypothesis that an orthologue of the *TFL1* gene controls determinacy of flowering in *Vf* as it does in Arabidopsis, soybean and numerous other legume and non-legume species by showing correlation of the determinate type with a diagnostic non-synonymous substitution in a conserved residue of the coding region across a diverse panel of determinate and indeterminate types. They did not, however, genetically map *VfTFL1*. Once again, clear macro-colinearity between the *Vf* region corresponding to the *Mt* region harbouring *MtTFL1* suggests a working hypothesis (which remains to be proven) that the published *VfTFL1* sequence should map to the long arm of *Vf* chromosome 1.

Genomic Diversity

Another arena in which genomics is playing a crucial role is in the description and rational exploitation of diverse, but often poorly characterised, genetic resources. Despite the lack of a known wild progenitor species, the domesticated *Vf* genepool, as alluded to in the introduction, is extremely diverse. This diversity is manifest in the plethora of morphological forms and adaptations to diverse agroecological settings. However, full exploitation of this diversity for faba bean improvement necessitates efficient methods for quantifying and mapping genomic diversity. In parallel with the development of SSAP, AFLP, SSR and SNP marker types, a succession of studies have used increasingly powerful molecular marker sets to quantify and map genomic diversity.

One of the first surveys of genomic diversity in diverse *Vf* germplasm was carried out by Sanz et al., (2007) using retrotransposon-based Sequence-Specific Amplified Polymorphism (SSAP) markers. This study found relatively little fine genetic structure and the first molecular evidence that '*major*', '*minor*' and '*equina*' lines do not form distinct clades, but rather are completely dispersed across the *Vf* phylogeny. Due to the low numbers of *Vf* lines (n=20), it was not possible to draw any conclusions

on geographic partitioning of diversity, though it was noted that samples from the better-represented countries in the study were well dispersed across the phylogenetic tree. These themes were recapitulated by Zeid et al., (2003) using AFLP markers to genotype 79 inbred lines from Europe, North Africa and Asia. Again, '*major*' and '*minor*' botanical types were clearly shown not to be genetically distinct, though here for the first time was a suggestion that the (n=8) lines of Asian origin formed a genetically differentiated group. The AFLP-based study of Zong et al., (2009) took this further with a study of large numbers of Chinese landraces (n=204) from the Chinese Academy of Agricultural Sciences (CAAS) germplasm collection, mainly winter types. They found the mainland Chinese winter germplasm to be completely distinct from other Asian, African and European diverse lines included for comparison, in line with the long history of cultivation of winter-type faba bean in relatively isolated mountainous regions of China. Kwon et al. (2010), using Targeted Region Amplified Polymorphism (TRAP) markers, also observed Chinese landraces drawn from the United States Department of Agriculture (USDA) germplasm collection (n=107) to cluster completely separately from 30 comparator lines from Asia and Europe. However, the largest study of faba bean genomic diversity to date has been reported by Wang et al., (2012a), whose large sample (n=802) confirmed not only the genetic distinctness of Chinese germplasm from African, European and other Asian germplasm, but for the first time, convincingly showed differentiation amongst Chinese provinces and between winter and spring ecotypes. SNP genotyping platforms have been more recently used on small samples of diversity (Cottage et al., 2012; Kaur et al., 2014) and the prospect of wider deployment of SNP markers on larger panels of diverse material promises the possible foundations for future *Vf* genome-wide association studies.

Functional genomics

Mutagenesis

The potential of mutagenesis as a tool for breeding was amply demonstrated by Sjödin, (1971), reviewing not only observations made in the extensive mutagenesis programme run by Swedish breeders Svalov-Weibull in the 'Primus' genetic background, but also a series of spontaneous mutants reported by a host of previous researchers. The only other mutageneis programme reported in the literature was 23 years later, when the isolation of five nodulation mutants in a screen of 20,000 M_2 EMS-mutagenised lines in the cv. 'Ascott' background was reported by Duc, (1995).

Interest in mutagenesis has undergone something of a resurgence in recent years. The summary of ongoing public sector programmes provided in **Table 4**, collated from a series of personal communications, shows a number of target traits that correspond to specific mutations not likely to be found in nature. For example, the control of the devastating parasitic weed, *Orobanche crenata*, referred to in an earlier section, could be effectively controlled, together with a range of other troublesome broad-leaved weeds, if faba bean varieties with target-site mutations that render them insensitive to particular actives were developed Gressel, (2009). Examples are the Ser653 and Ala205 mutations in the AcetoLactate Synthase target of imidazolinone and amidosulfuron families of herbicide actives, which do not occur in nature, but have been documented to occur at low frequency under strong selection pressure in the field, and there are examples in numerous crop species of targeted isolation of induced mutants (reviewed in Tan et al., 2005). At the time of writing, a number of imazapyr resistance mutations have been identified by Mao et al., (2014) and are undergoing further characterization. As a further example of the ongoing mutagenesis programmes listed in Table 4, **Figure 3** shows a selection of phenotypes observed in an M_2 X-ray and EMS populations grown during 2015 at the University of Reading.

Whilst the primary driver for contemporary mutagenesis programmes remains the identification of novel induced variants for direct incorporation in active breeding programmes, a broader motivation can be ascribed to some of the recent activity. Until efficient genetic transformation systems have been adopted and made available as a service, reverse genetics in the form of TILLing (Till et al., 2003) may be one of the most amenable techniques for validation of candidate genes identified via synteny-based approaches, especially if combined with powerful re-sequencing techniques to identify the causative mutations in a single step (Wang et al., 2012b).

Genetic transformation/genome editing

Genetic modification represents both a research tool, permitting testing of hypotheses on gene function by over-, mis-expression or knockdown/knockout studies and an outlet for genetic research in generation of targeted phenotypic modifications based on knowledge of gene function. Stable germline transformation of *Vf* using *in vitro* regeneration of *Agrobacterium*-infiltrated (non-meristematic) internode stem segments was first reported by Böttinger et al., (2001). Adopting a somewhat different strategy, Hanafy et al., (2005) infiltrated excised (meristematic) embryo axes with *Agrobacterium* and successfully recovered stable transgenic lines. Both methods, however, reported low primary transformation efficiencies and relied on micro-grafting of putative transgenic shoot material onto non-transgenic roots, a slow and highly manual process. Hanafy et al., (2013) later reported abiotic stress resistance phenotypes of *Vf* transgenic lines overexpressing potato PR10a using their previous methods. This remains to our knowledge the sole successful demonstration to date of the feasibility of a biotechnological approach to *Vf* improvement. In the absence of a robust and efficient transformation method, some attention has been devoted to the task of decreasing generation time using tissue-culture based embryo rescue, with some success (Mobini et al., 2015). However, the search for an efficient transformation method has been recently taken up by other groups (e.g. Abdelwahd et al., (2014).

The prospects afforded by new insights into the phenotypic effects of allelic variation and the more refined biotechnological possibilities afforded by rapidly maturing genome editing technologies (Gaj et al., 2013) could potentially stimulate renewed interest in genetic transformation. An example of a game-changing product which could readily be generated using even a medium efficiency transformation system would be herbicide resistance obtained by directed mutagenesis of endogenous herbicide target genes e.g. introduction of heterologous glyphosate resistance of bacterial origin.

Conclusions

Vicia faba genetics and genomics is now in a much healthier state than it was just a few years ago. We can take heart from the accelerating progress in gene identification and in production of outputs relevant to contemporary methods of breeding. We have seen the transformative effects of transcriptome re-sequencing in opening the doors to high density gene-based marker discovery and mapping. In the near future, we expect that a high throughput SNP chip incorporating many 10's of thousands of genome-wide SNPs discovered from a wide variety of genetic backgrounds will be produced and made available for community use, as well as giving rise to an ultra-high density SNP map in which perhaps as much as half of all *Vf* genes are genetically mapped. In parallel, we should see progress in the application of genotype-by-sequencing methods, both for the unbiased assessment of genetic relationships, breeding applications such as genomic selection and in trait mapping approaches based on bulked segregant analysis. Faba bean's modest but well-linked community of researchers would do well however, to set its sights on a series of even more ambitious targets to enable the community as a whole to elevate its work onto a higher plane of achievement and impact.

For example, given the many highly cost effective sequencing and assembly technologies now available, a comprehensive genomic scaffold and haplotype map is surely now within reach. The history of modern agricultural genomics shows the transformative effects of a well-annotated reference genome. Communities of crop researchers who have organised themselves and published strategic roadmaps requiring centralised investment in a professional genome assembly and annotation have captured significant R&D investment and transformed the profile and fate of their communities (rice, wheat, soybean, *Phaseolus*, cowpea). Similarly, on functional genomics platforms: whilst there are no doubt training benefits to having numerous small mutagenesis programmes dotted around the globe, it could be argued that the global community needs one well-funded programme scoped to guarantee saturation mutagenesis, to catalogue the mutations obtained in a public sequence database and to distribute mutant seed to research groups freely on request, an unlikely outcome from nationally funded programmes. Likewise, transformation and genome editing technologies could advantageously be developed and provided as an efficient service from a centralized laboratory.

Tables:

Table 1: Key *Vf* transcriptome datasets

Bioproject	Genbank reference	Cultivar	Tissue	Reference
PRJNA225873	SRP033593	BPL10	10-d seedling	Webb et al., 2016
PRJNA225881	SRP033121	Albus	10-d seedling	Webb et al., 2016
PRJNA238140	SRX476199	CDC Fatima	6-d root	Ray et al., 2015
	SRX476200	CDC Fatima	6-d shoot	Ray et al., 2015
	SRX476493	CDC Fatima	seed coat	Ray et al., 2015
	SRX476217	SSNS-1	6-d root	Ray et al., 2015
	SRX476220	SSNS-1	6-d shoot	Ray et al., 2015
	SRX475907	A01155	6-d root	Ray et al., 2015
	SRX475873	A01155	6-d shoot	Ray et al., 2015
	SRX476566	A01155	seed coat	Ray et al., 2015
PRJNA277609	SRP055969	Fiord	Mixed tissues	Arun-Chinnappa and McCurdy, 2015
PRJEB8906	ERP009949	Fiord	Cotyledon epidermis and parenchyma	Zhang et al., 2015
NA	JR964201- JR970413*	Icarus, Ascot	Mixed tissues	Kaur et al., 2012
PRJNA253768	SRP043650	NS	leaves	Suresh et al., 2015
NA	GI:219212932 - GI:219282595	Windsor	2 week old embryo	Ray and Georges, 2010
NA	SRP045955	INRA-29H	leaf	Ocaña et al., 2015
NA	SRP045955	*Vf*136	leaf	Ocaña et al., 2015

NA: Not applicable; NS: Not stated

*Only assembled contigs available as TSA

460 **Table 2**: Key molecular marker sets

461

Marker Type	**Discovery genotypes**	**Number of validated polymorphic markers**	**Number mapped**	**Reference**
EST-SSR	Komasake	647	552	El-Rodeny et al., 2014
genomic SSR	Mixed	90	NA	Yang et al., 2012
EST-SSR	Icarus, Ascot	71	57	Kaur et al., 2014
SNP (KASP)	Albus, BPL10	824	687	Cottage et al., 2012,Webb et al., 2016
SNP (Illumina)	Icarus, Ascot	480	465	Kaur et al., 2014
ITAP	*Vf*6, *Vf*27	151	127	Ellwood et al., 2008

462

463

Table 3: Genetic maps

Cross	Number of loci	Number of linkage groups	Average marker interval	Map length (cM)	Reference
Consensus of six F_2 populations	687	6	2.04 cM	1403.8	Webb et al., 2016
'Nubaria 2' x 'Misr 3' F_2	552 EST-SSR	6	1.25 cM	687.7	El-Rodeny et al., 2014
Consensus of three RIL populations	729 RAPD, ITAP, SSR, morphological	6	6.31 cM	4602	Satovic et al., 2013
91825 x K1563 F_2	128 SSR	15	12.4 cM	1587	Ma et al, 2013 ref
Icarus x Ascot F5:6 RILs	522 (57 EST-SSR, 465 SNP)	12	2.33 cM	1216.8	Kaur et al., 2014
*Vf*6 x *Vf*27 RILs	127	12	13.27	1685.8	Cottage et al., 2012; Ellwood et al., 2008

468 **Table 4**: Current/ongoing mutagenesis programmes

Genetic background	**Institution/PI**	**Mutagen**	**Primary specific targets**	**No. independent lines in most advanced generation**
Hedin/2	Uni Reading/DM O'Sullivan	EMS	Imidizolanone resistance	1,097 M_2
Hedin/2	Uni Reading/DM O'Sullivan	X-ray	Imidizolanone resistance	985 M_2
SSNS-1	Uni Saskatchewan/B Vandenberg	NaN_3	Low phytate	1,500 M_4
SSNS-1	Uni Saskatchewan/B Vandenberg	EMS	Low phytate	1,500 M_4
BPL710	ICARDA/F Maalouf	EMS	Glyphosate tolerance	400 M_6
NA112	ICARDA/F Maalouf	EMS	Herbicide resistance	5,000 M_2
Nura	Uni Adelaide/J Paull	EMS	Imazapyr resistance	Four imazapyr resistant lines under investigation

Figure Legends:

Figure 1: Integrative Genomics Viewer (IGV) snapshot of SNPs in the *Vf* orthologue of Medtr2g009080 mined from public trancriptome datasets.The eleven horizontal tracks shown in this view from top to bottom are: the meta-transcriptome contig TR55082¦c0_g4_i2 is shown as a schematic, all intravarietal SNP positions; positions where the minor (non-consensus) allele is found in each of the following genotypes: 'Albus', 'Fiord', 'SRX641218', 'SSN1', 'Fatima', A01155' and 'BPL10', alignment of the reference *Vf* contig TR55082¦c0_g4_i2 to the Mt4.0 genome (coding region of Medtr2g009080 in this instance), and in the bottom panel is shown a 3-frame translation of the *Vf* sequence. In this example, there are a total of 16 intravarietal SNPs distributed across the gene - eleven unique to 'Fiord', two unique to 'A001155' and three minor alleles in common to 'Fiord' and 'A01155', with the remaining genotypes carrying the unchanged common allele identical to the consensus contig sequence.

Figure 2: Wheel representation of synteny between pseudomolecule sequence of Medicago truncatula chromosomes Mt1-Mt8 and the consensus *Vf* linkage map (*Vf*1-*Vf*6). Syntenic relationships were as described in Webb et al., 2016 and re-drawn for this figure using 'Circos' software (Krzywinski et al., 2009). Link lines to Medicago orthologues for markers on each *Vf* chromosome are shown in a different colour. The virtually uninterrupted collinearity between the whole of *Vf*3 and the whole of Mt1 is highlighted by representing these orthologous chromosomes in green. To illustrate how useful this syntenic framework is for translational genomics as well as for drawing together knowledge of the genome, three exemplar loci where there is a proven or presumed orthologous relationship are shown as red ticks on the orthologous chromosomes and linked by black lines. These are: the mapped *ZT1* locus and its Medicago orthologue *MtWD40-1*, the mapped genetic interval corresponding to the mapped *VC* (vicine-convicine) locus and its corresponding *Medicago* syntenic interval on *Mt2*, and the actual position of the *Medicago TERMINAL FLOWER LOCUS 1* (*TFL1*) with a link drawn to the region of *Vf*1 where it is presumed the *Vf* orthologue (*VfTFL1*) is likely to be located.

Figure 3: Mutant phenotypes observed in X-ray and EMS M_2 lines grown at the University of Reading in 2015. A. disease lesion mimic B. determinate flowering B. curled wing petal C. disease lesion mimic D. reduced flower pigmentation.

References

Abdelwahd, R., Udupa, S.M., Gaboun, F., Diria, G., Mentag, R., Ibriz, Mohamed and Iraqi, D. (2014) Agrobacterium-mediated transformation of cotyledonary node of Vicia faba L. *Rom Agric Res* **31**, online first.

Arun-Chinnappa, K.S. and McCurdy, D.W. (2015) De novo assembly of a genome-wide transcriptome map of Vicia faba (L.) for transfer cell research. *Front Plant Sci* **6**.

Avila, C.M., Atienza, S.G., Moreno, M.T. and Torres, A.M. (2007) Development of a new diagnostic marker for growth habit selection in faba bean (Vicia faba L.) breeding. *Theor Appl Genet* **115**, 1075-1082.

Avila, C.M., Satovic, Z., Sillero, J.C., Rubiales, D., Moreno, M.T. and Torres, A.M. (2004) Isolate and organ-specific QTLs for ascochyta blight resistance in faba bean (Vicia faba L). *Theor Appl Genet* **108**, 1071-1078.

Avila, C.M., Sillero, J.C., Rubiales, D., Moreno, M.T. and Torres, A.M. (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.

Bishop, J., Jones, H.E., Lukac, M. and Potts, S.G. (2016a) Insect pollination reduces yield loss following heat stress in faba bean (Vicia faba L.). *Agriculture, Ecosystems & Environment* **220**, 89-96.

Bishop, J., Potts, S.G. and Jones, H.E. (2016b) Susceptibility of Faba Bean (Vicia faba L.) to Heat Stress During Floral Development and Anthesis. *J Agron Crop Sci*, n/a-n/a.

Böttinger, P., Steinmetz, A., Schieder, O. and Pickardt, T. (2001) Agrobacterium-mediated transformation of Vicia faba. *Mol Breeding* **8**, 243-254.

Cernay, C., Ben-Ari, T., Pelzer, E., Meynard, J.-M. and Makowski, D. (2015) Estimating variability in grain legume yields across Europe and the Americas. *Sci Rep-Uk* **5**, 11171.

Cottage, A., Gostkiewicz, K., Thomas, J.E., Borrows, R., Torres, A.M. and O'Sullivan, D.M. (2012) Heterozygosity and diversity analysis using mapped single nucelotide polymorphisms in a faba bean inbreeding programme. *Mol Breeding* **30**, 1799-1809.

Crepon, K., Marget, P., Peyronnet, C., Carrouee, B., Arese, P. and Duc, G. (2010) Nutritional value of faba bean (Vicia faba L.) seeds for feed and food. *Field Crop Res* **115**, 329-339.

Cruz-Izquierdo, S., Avila, C.M., Satovic, Z., Palomino, C., Gutierrez, N., Ellwood, S.R., Phan, H.T.T., Cubero, J.I. and Torres, A.M. (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.

Denton, M.D., Pearce, D.J. and Peoples, M.B. (2013) Nitrogen contributions from faba bean (Vicia faba L.) reliant on soil rhizobia or inoculation. *Plant Soil* **365**, 363-374.

Diaz-Ruiz, R., Satovic, Z., Avila, C.M., Alfaro, C.M., Gutierrez, M.V., Torres, A.M. and Roman, B. (2009) Confirmation of QTLs controlling Ascochyta fabae resistance in different generations of faba bean (Vicia faba L.). *Crop Pasture Sci* **60**, 353-361.

Diaz-Ruiz, R., Torres, A.M., Satovic, Z., Gutierrez, M.V., Cubero, J.I. and Roman, B. (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. (1995) Mutagenesis of Faba Bean (Vicia-Faba L) and the Identification of 5 Different Genes-Controlling No Nodulation, Ineffective Nodulation or Supernodulation. *Euphytica* **83**, 147-152.

Duc, G., Bao, S.Y., Baum, M., Redden, B., Sadiki, M., Suso, M.J., Vishniakova, M. and Zong, X.X. (2010) Diversity maintenance and use of Vicia faba L. genetic resources. *Field Crop Res* **115**, 270-278.

Duc, G., Moussy, F., Zong, X. and Ding, G. (1999) Single gene mutation for green cotyledons as a marker for the embryonic genotype in faba bean, Vicia faba. *Plant Breeding* **118**, 577-578.

Duc, G. and Picard, J. (1986) Note on the Presence of the Sym-1 Gene in Vicia-Faba Hampering Its Symbiosis with Rhizobium-Leguminosarum. *Euphytica* **35**, 61-64.

El-Rodeny, W., Kimura, M., Hirakawa, H., Sabah, A., Shirasawa, K., Sato, S., Tabata, S., Sasamoto, S., Watanabe, A., Kawashima, K., Kato, M., Wada, T., Tsuruoka, H., Takahashi, C., Minami, C., Nanri, K., Nakayama, S., Kohara, M., Yamada, M., Kishida, Y., Fujishiro, T. and Isobe, S. (2014) Development of EST-SSR markers and construction of a linkage map in faba bean (Vicia faba). *Breeding Sci* **64**, 252-263.

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

Erith, A.G. (1930) The inheritance of colour, size and form of seeds, and of flower colour in Vicia Faba L. *Genetica* **12**, 477-510.

Gaj, T., Gersbach, C.A. and Barbas Iii, C.F. (2013) ZFN, TALEN, and CRISPR/Cas-based methods for genome engineering. *Trends in Biotechnology* **31**, 397-405.

Gnanasambandam, A., Paull, J., Torres, A., Kaur, S., Leonforte, T., Li, H., Zong, X., Yang, T. and Materne, M. (2012) Impact of Molecular Technologies on Faba Bean (Vicia faba L.) Breeding Strategies. *Agronomy* **2**, 132-166.

Gong, Y.M., Xu, S.C., Mao, W.H., Li, Z.Y., Hu, Q.Z., Zhang, G.W. and Ding, J. (2011) Genetic Diversity Analysis of Faba Bean (Vicia faba L.) Based on EST-SSR Markers. *Agr Sci China* **10**, 838-844.

Gressel, J. (2009) Crops with target-site herbicide resistance for Orobanche and Striga control. *Pest Manag Sci* **65**, 560-565.

Gutierrez, N., Palomino, C., Satovic, Z., Ruiz-Rodriguez, M.D., Vitale, S., Gutierrez, M.V., Rubiales, D., Kharrat, M., Amri, M., Emeran, A.A., Cubero, J.I., Atienza, S.G., Torres, A.M. and Avila, C.M. (2013) QTLs for Orobanche spp. resistance in faba bean: identification and validation across different environments. *Mol Breeding* **32**, 909-922.

Hanafy, M., Pickardt, T., Kiesecker, H. and H.-J., J. (2005) Agrobacterium-mediated transformation of faba bean (Vicia faba L.) using embryo axes. *Euphytica* **142**, 227-236.

Hanafy, M.S., El-Banna, A., Schumacher, H.M., Jacobsen, H.J. and Hassan, F.S. (2013) Enhanced tolerance to drought and salt stresses in transgenic faba bean (Vicia faba L.) plants by heterologous expression of the PR10a gene from potato. *Plant Cell Rep* **32**, 663-674.

Kaur , S., Cogan , N.O.I., Forster , J.W. and Paull, J.G. (2014) Assessment of Genetic Diversity in Faba Bean Based on Single Nucleotide Polymorphism. *Diversity* **6**, 88-101.

Kaur, S., Kimber, R.B.E., Cogan, N.O.I., Materne, M., Forster, J.W. and Paull, J.G. (2014) SNP discovery and high-density genetic mapping in faba bean (Vicia faba L.) permits identification of QTLs for ascochyta blight resistance. *Plant Sci* **217**, 47-55.

Kaur, S., Pembleton, L.W., Cogan, N.O.I., Savin, K.W., Leonforte, T., Paull, J., Materne, M. and Forster, J.W. (2012) Transcriptome sequencing of field pea and faba bean for discovery and validation of SSR genetic markers. *Bmc Genomics* **13**.

Khamassi, K., Ben Jeddi, F., Hobbs, D., Irigoyen, J., Stoddard, F., O'Sullivan, D.M. and Jones, H. (2013) A baseline study of vicine-convicine levels in faba bean (Vicia faba L.) germplasm. *Plant Genet Resour-C* **11**, 250-257.

Khan, H.R., Paull, J.G., Siddique, K.H.M. and Stoddard, F.L. (2010) Faba bean breeding for drought-affected environments: A physiological and agronomic perspective. *Field Crop Res* **115**, 279-286.

Khazaei, H., O'Sullivan, D.M., Jones, H., Pitts, N., Sillanpaa, M.J., Parssinen, P., Manninen, O. and Stoddard, F.L. (2015) Flanking SNP markers for vicine-convicine concentration in faba bean (Vicia faba L.). *Mol Breeding* **35**.

Khazaei, H., O'Sullivan, D.M., Sillanpaa, M.J. and Stoddard, F.L. (2014) Use of synteny to identify candidate genes underlying QTL controlling stomatal traits in faba bean (Vicia faba L.). *Theor Appl Genet* **127**, 2371-2385.

Kovarova, P., Navratilova, A., Macas, J. and Dolezel, J. (2007) Chromosome analysis and sorting in Vicia sativa using flow cytometry. *Biol Plantarum* **51**, 43-48.

Krzywinski, M., Schein, J., Birol, İ., Connors, J., Gascoyne, R., Horsman, D., Jones, S.J. and Marra, M.A. (2009) Circos: An information aesthetic for comparative genomics. *Genome research* **19**, 1639-1645.

Link, W., Balko, C. and Stoddard, F.L. (2010) Winter hardiness in faba bean: Physiology and breeding. *Field Crop Res* **115**, 287-296.

Ma, Y., Bao, S.Y., Yang, T., Hu, J.G., Guan, J.P., He, Y.H., Wang, X.J., Wan, Y.L., Sun, X.L., Jiang, J.Y., Gong, C.X. and Zong, X.X. (2013) Genetic linkage map of Chinese native variety faba bean (Vicia faba L.) based on simple sequence repeat markers. *Plant Breeding* **132**, 397-400.

Ma, Y., Yang, T., Guan, J.P., Wang, S.M., Wang, H.F., Sun, X.L. and Zong, X.X. (2011) Development and Characterization of 21 Est-Derived Microsatellite Markers in Vicia Faba (Fava Bean). *Am J Bot* **98**, E22-E24.

Mao, D., Paull, J., Oldach, K.H., Preston, C., Yang, S.Y., Davies, P. and McMurray, L. (2014) The development of multiple herbicide tolerances in faba bean (Vicia faba L.) through induced mutation. In: *15th Autralasian Plant Breeding Conference*. Melbourne, Australia.

Maxted, N. (1993) Towards a faba bean progenitor! *FABIS Newsletter* **31**.

Menke, M., Meister, A. and Schubert, I. (2000) N-methyl-N-nitrosourea-induced DNA damage detected by the comet assay in Vicia faba nuclei during all interphase stages is not restricted to chromatid aberration hot spots. *Mutagenesis* **15**, 503-506.

Milan, P.R. and Upadhyay, S. (2007) Impact of food additives on mitotic chromosomes of Vicia faba L. *Caryologia* **60**, 309-314.

Mobini, S.H., Lulsdorf, M., Warkentin, T.D. and Vandenberg, A. (2015) Plant growth regulators improve in vitro flowering and rapid generation advancement in lentil and faba bean. *In Vitro CellDevBiol-Plant* **51**, 71-79.

Multari, S., Stewart, D. and Russell, W.R. (2015) Potential of Fava Bean as Future Protein Supply to Partially Replace Meat Intake in the Human Diet. *Comprehensive Reviews in Food Science and Food Safety*, n/a-n/a.

Nayak, G.K., Roberts, S.P.M., Garratt, M., Breeze, T.D., Tscheulin, T., Harrison-Cripps, J., Vogiatzakis, I.N., Stirpe, M.T. and Potts, S.G. (2015) Interactive effect of floral abundance and semi-natural habitats on pollinators in field beans (Vicia faba). *Agr Ecosyst Environ* **199**, 58-66.

Ocaña, S., Seoane, P., Bautista, R., Palomino, C., Claros, G.M., Torres, A.M. and Madrid, E. (2015) Large-Scale Transcriptome Analysis in Faba Bean (*Vicia faba* L.) under *Ascochyta fabae* Infection. *Plos One* **10**, e0135143.

Patrick, J.W. and Stoddard, F.L. (2010) Physiology of flowering and grain filling in faba bean. *Field Crop Res* **115**, 234-242.

Patto, M.C.V., Torres, A.M., Koblizkova, A., Macas, J. and Cubero, J.I. (1999) Development of a genetic composite map of Vicia faba using F-2 populations derived from trisomic plants. *Theor Appl Genet* **98**, 736-743.

Pérez-de-Luque, A., Eizenberg, H., Grenz, J.H., Sillero, J.C., Ávila, C., Sauerborn, J. and Rubiales, D. (2010) Broomrape management in faba bean. *Field Crop Res* **115**, 319-328.

Preissel, S., Reckling, M., Schläfke, N. and Zander, P. (2015) Magnitude and farm-economic value of grain legume pre-crop benefits in Europe: A review. *Field Crop Res* **175**, 64-79.

Ramsay, G. (1997) Inheritance and linkage of a gene for testa-imposed seed dormancy in faba bean (Vicia faba L.). *Plant Breeding* **116**, 287-289.

Ray, H., Bock, C. and Georges, F. (2015) Faba Bean: Transcriptome Analysis from Etiolated Seedling and Developing Seed Coat of Key Cultivars for Synthesis of Proanthocyanidins, Phytate, Raffinose Family Oligosaccharides, Vicine, and Convicine. *Plant Genome-Us* **8**.

Ray, H. and Georges, F. (2010) A genomic approach to nutritional, pharmacological and genetic issues of faba bean (Vicia faba): prospects for genetic modifications. *GM crops* **1**, 99-106.

Rybaczek, D., Zabka, A., Pastucha, A. and Maszewski, J. (2008) Various chemical agents can induce premature chromosome condensation in Vicia faba. *Acta Physiol Plant* **30**, 663-672.

Sallam, A., Martsch, R. and Moursi, Y. (2015) Genetic variation in morpho-physiological traits associated with frost tolerance in faba bean (Vicia faba L.). *Euphytica* **205**, 395-408.

Sanz, A.M., Gonzalez, S.G., Syed, N.H., Suso, M.J., Saldana, C.C. and Flavell, A.J. (2007) Genetic diversity analysis in Vicia species using retrotransposon-based SSAP markers. *Mol Genet Genomics* **278**, 433-441.

Satovic, Z., Avila, C.M., Cruz-Izquierdo, S., Diaz-Ruiz, R., Garcia-Ruiz, G.M., Palomino, C., Gutierrez, N., Vitale, S., Ocana-Moral, S., Gutierrez, M.V., Cubero, J.I. and Torres, A.M.

660 (2013) A reference consensus genetic map for molecular markers and economically important
661 traits in faba bean (Vicia faba L.). *Bmc Genomics* **14**.
662 Schmutz, J., Cannon, S.B., Schlueter, J., Ma, J., Mitros, T., Nelson, W., Hyten, D.L., Song, Q.,
663 Thelen, J.J., Cheng, J., Xu, D., Hellsten, U., May, G.D., Yu, Y., Sakurai, T., Umezawa, T.,
664 Bhattacharyya, M.K., Sandhu, D., Valliyodan, B., Lindquist, E., Peto, M., Grant, D., Shu, S.,
665 Goodstein, D., Barry, K., Futrell-Griggs, M., Abernathy, B., Du, J., Tian, Z., Zhu, L., Gill, N.,
666 Joshi, T., Libault, M., Sethuraman, A., Zhang, X.C., Shinozaki, K., Nguyen, H.T., Wing,
667 R.A., Cregan, P., Specht, J., Grimwood, J., Rokhsar, D., Stacey, G., Shoemaker, R.C. and
668 Jackson, S.A. (2010) Genome sequence of the palaeopolyploid soybean. *Nature* **463**, 178-
669 183.
670 Sillero, J.C., Villegas-Fernández, A.M., Thomas, J., Rojas-Molina, M.M., Emeran, A.A., Fernández-
671 Aparicio, M. and Rubiales, D. (2010) Faba bean breeding for disease resistance. *Field Crop*
672 *Res* **115**, 297-307.
673 Sjödin, J.A.N. (1970) Induced asynaptic mutants in Vicia faba L. *Hereditas* **66**, 215-232.
674 Sjödin, J.A.N. (1971) Induced morphological variation in Vicia faba L. *Hereditas* **67**, 155-179.
675 Sobita, K. and Bhagirath, T. (2005) Effects of some medicinal plant extracts on Vicia faba root tip
676 chromosomes. *Caryologia* **58**, 255-261.
677 Soltis, D.E., Soltis, P.S., Bennett, M.D. and Leitch, I.J. (2003) Evolution of genome size in the
678 angiosperms. *Am J Bot* **90**, 1596-1603.
679 Stoddard, F.L., Nicholas, A.H., Rubiales, D., Thomas, J. and Villegas-Fernández, A.M. (2010)
680 Integrated pest management in faba bean. *Field Crop Res* **115**, 308-318.
681 Suresh, S., Kim, T.-S., Raveendar, S., Cho, J.-H., Yi, J.-Y., Lee, M.C., Lee, S.-Y., Baek, H.-J., Cho,
682 G.-T. and Chung, J.-W. (2015) Transcriptome characterization and large-scale identification
683 of SSR/SNP markers in symbiotic nitrogen fixation crop faba bean (Vicia faba L.). *Turk J*
684 *Agric For* **39**, 459-469.
685 Tan, S., Evans, R.R., Dahmer, M.L., Singh, B.K. and Shaner, D.L. (2005) Imidazolinone-tolerant
686 crops: history, current status and future. *Pest Manag Sci* **61**, 246-257.
687 Tanno, K. and Willcox, G. (2006) The origins of cultivation of Cicer arietinum L. and Vicia faba L.:
688 early finds from Tell el-Kerkh, north-west Syria, late 10th millennium BP. *Veg Hist*
689 *Archaeobot* **15**, 197-204.
690 Tavakkoli, E., Paull, J., Rengasamy, P. and McDonald, G.K. (2012) Comparing genotypic variation in
691 faba bean (Vicia faba L.) in response to salinity in hydroponic and field experiments. *Field*
692 *Crop Res* **127**, 99-108.
693 Till, B.J., Reynolds, S.H., Greene, E.A., Codomo, C.A., Enns, L.C., Johnson, J.E., Burtner, C.,
694 Odden, A.R., Young, K., Taylor, N.E., Henikoff, J.G., Comai, L. and Henikoff, S. (2003)
695 Large-scale discovery of induced point mutations with high-throughput TILLING. *Genome*
696 *research* **13**, 524-530.
697 Torres, A.M., Avila, C.M., Gutierrez, N., Palomino, C., Moreno, M.T. and Cubero, J.I. (2010)
698 Marker-assisted selection in faba bean (Vicia faba L.). *Field Crop Res* **115**, 243-252.
699 Torres, A.M., Weeden, N.F. and Martin, A. (1993) Linkage among isozyme, RFLP and RAPD
700 markers in Vicia faba. *Theor Appl Genet* **85**, 937-945.
701 van de Wouw, M., Enneking, D., Robertson, L.D. and Maxted, N. (2001) Vetches. In: *Plant Genetics*
702 *Resources of Legumes in the Mediterranean* pp. 134-158.
703 Wang, H.F., Zong, X.X., Guan, J.P., Yang, T., Sun, X.L., Ma, Y. and Redden, R. (2012a) Genetic
704 diversity and relationship of global faba bean (Vicia faba L.) germplasm revealed by ISSR
705 markers. *Theor Appl Genet* **124**, 789-797.
706 Wang, T.L., Uauy, C., Robson, F. and Till, B. (2012b) TILLING in extremis. *Plant Biotechnology*
707 *Journal* **10**, 761-772.
708 Webb, A., Cottage, A., Wood, T., Khamassi, K., Hobbs, D., Gostkiewicz, K., White, M., Khazaei, H.,
709 Ali, M., Street, D., Duc, G., Stoddard, F.L., Maalouf, F., Ogbonnaya, F.C., Link, W., Thomas,
710 J. and O'Sullivan, D.M. (2016) A SNP-based consensus genetic map for synteny-based trait
711 targeting in faba bean (Vicia faba L.). *Plant Biotechnol J* **14**, 177-185.
712 Yang, T., Bao, S.Y., Ford, R., Jia, T.J., Guan, J.P., He, Y.H., Sun, X.L., Jiang, J.Y., Hao, J.J., Zhang,
713 X.Y. and Zong, X.X. (2012) High-throughput novel microsatellite marker of faba bean via
714 next generation sequencing. *Bmc Genomics* **13**.

Young, N.D., Debelle, F., Oldroyd, G.E., Geurts, R., Cannon, S.B., Udvardi, M.K., Benedito, V.A., Mayer, K.F., Gouzy, J., Schoof, H., Van de Peer, Y., Proost, S., Cook, D.R., Meyers, B.C., Spannagl, M., Cheung, F., De Mita, S., Krishnakumar, V., Gundlach, H., Zhou, S., Mudge, J., Bharti, A.K., Murray, J.D., Naoumkina, M.A., Rosen, B., Silverstein, K.A., Tang, H., Rombauts, S., Zhao, P.X., Zhou, P., Barbe, V., Bardou, P., Bechner, M., Bellec, A., Berger, A., Berges, H., Bidwell, S., Bisseling, T., Choisne, N., Couloux, A., Denny, R., Deshpande, S., Dai, X., Doyle, J.J., Dudez, A.M., Farmer, A.D., Fouteau, S., Franken, C., Gibelin, C., Gish, J., Goldstein, S., Gonzalez, A.J., Green, P.J., Hallab, A., Hartog, M., Hua, A., Humphray, S.J., Jeong, D.H., Jing, Y., Jocker, A., Kenton, S.M., Kim, D.J., Klee, K., Lai, H., Lang, C., Lin, S., Macmil, S.L., Magdelenat, G., Matthews, L., McCorrison, J., Monaghan, E.L., Mun, J.H., Najar, F.Z., Nicholson, C., Noirot, C., O'Bleness, M., Paule, C.R., Poulain, J., Prion, F., Qin, B., Qu, C., Retzel, E.F., Riddle, C., Sallet, E., Samain, S., Samson, N., Sanders, I., Saurat, O., Scarpelli, C., Schiex, T., Segurens, B., Severin, A.J., Sherrier, D.J., Shi, R., Sims, S., Singer, S.R., Sinharoy, S., Sterck, L., Viollet, A., Wang, B.B., Wang, K., Wang, M., Wang, X., Warfsmann, J., Weissenbach, J., White, D.D., White, J.D., Wiley, G.B., Wincker, P., Xing, Y., Yang, L., Yao, Z., Ying, F., Zhai, J., Zhou, L., Zuber, A., Denarie, J., Dixon, R.A., May, G.D., Schwartz, D.C., Rogers, J., Quetier, F., Town, C.D. and Roe, B.A. (2011) The Medicago genome provides insight into the evolution of rhizobial symbioses. *Nature* **480**, 520-524.

Zeid, M., Mitchell, S., Link, W., Carter, M., Nawar, A., Fulton, T. and 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., Schon, C.C. and Link, W. (2003) Genetic diversity in recent elite faba bean lines using AFLP markers. *Theor Appl Genet* **107**, 1304-1314.

Zhang, H.M., Wheeler, S., Xia, X., Radchuk, R., Weber, H., Offler, C.E. and Patrick, J.W. (2015) Differential transcriptional networks associated with key phases of ingrowth wall construction in trans-differentiating epidermal transfer cells of Vicia faba cotyledons. *BMC plant biology* **15**, 103.

Zong, X.X., Liu, X.J., Guan, J.P., Wang, S.M., Liu, Q.C., Paull, J.G. and Redden, R. (2009) Molecular variation among Chinese and global winter faba bean germplasm. *Theor Appl Genet* **118**, 971-978.