



LEGATO

LEGumes for the Agriculture of TTomorrow

Collaborative project
Grant agreement no: 613551
SEVENTH FRAMEWORK PROGRAMME

THEME [KBBE.2013.1.2-02]
**[Legume breeding and management for sustainable
agriculture as well as protein supply for food and feed]**

Deliverable 1.7

***Main achievements of WP 1 “Genetics and new breeding
tools and material”***

Due date: M48

Actual submission date: M48

Project start date: 1st January 2014 **Duration:** 48 months

Workpackage concerned: WP1

Concerned workpackage leader: Paolo Annicchiarico (CREA)

Concerned task leaders:

- Francisco Madueño (CSIC)
- David Lloyd (Aberystwyth Univ.)
- Petr Smykal (UPOL)

Dissemination level: PU (public)

Introduction

This Work Package focused on genetic aspects and breeding tools that may contribute to substantial improvement of grain legume yield, quality or tolerance to abiotic stresses (essentially drought). In particular, it aimed at identifying and map genes regulating pod production in pea and autofertility in faba bean; identifying and/or validating markers, genes and marker-based approaches for reducing key anti-nutritional compounds (vicine and convicine in faba bean and ODAP in grasspea) or increasing pea drought tolerance; developing genomic resources such as high density genetic maps in pea and white lupin, and introgression lines between cultivated wild pea; investigating the global white lupin genetic variation; and producing new germplasm with improved traits. The next sections summarize the main results achieved by these activities.

Genes regulating the plant reproduction in pea and faba bean

This research work was carried out within Task 1 of this WP under the responsibility of CSIC for pea (*Pisum sativum*) and IFAPA for faba bean (*Vicia faba*). The work had distinct objectives for pea and faba bean. For pea, it aimed to identify new allelic variation that may contribute to higher grain yield by increasing one of its components, namely, the number of nodes bearing pods. Earlier work identified from a mutant population the two MORE FLOWERING NODES (MFN) genes *mfn1* and *mfn2* (encoding two homologous transcription factors), which caused a remarkable increase in the number of fertile nodes and increased of production of seeds in the single and the double mutant. While further characterizing these genes in its WP 3, LEGATO searched for new allelic variation (or the variation for *mfn1* and *mfn2* genes) in different pea germplasm within WP 1.

The work on pea implied the molecular characterization of two germplasm collections, one including 28 wild *Pisum* accessions, and the other comprising 10 modern pea cultivars. In both cases, the accessions were selected subset from a larger number with the objective to maximize the phenotypic variation in number of flowering nodes, exploiting phenotyping data available from other partners (UPOL and CREA). In both sets of material, new polymorphisms in the cDNAs that lead to amino acid changes in the encoded proteins were found for both *MFN1* and *MFN2* genes, but these alleles showed no correlation with the number of flowering nodes. Apparently, the effect of these alleles was insufficient to modify the function of the relevant proteins and to cause a trait change comparable with that of the original *mfn1* and *mfn2* mutant alleles.

In conclusion, this work indicated that *mfn1* and *mfn2* mutants are the main genes at hand for improving the node fertility, and found no evidence for their occurrence in modern varieties or wild material. As a part of Task 1, these mutants have been introgressed into four recent cultivars via repeated backcross.

Faba bean is characterized by intermediate levels of cross-fertilization, which complicate its breeding by preventing both the selection of cross-pollinating varieties capable of thorough

heterosis exploitation (because of the partial self-pollination) and the selection of fully self-fertile pure lines. However, variation occurs in the level of cross-fertilization, and knowledge on genes that control this variation would allow the marker-assisted selection (MAS) either of pure lines or of synthetic varieties with maximized heterosis because of obliged cross-pollination.

The activity on faba bean implied the exploitation of earlier phenotyping data and the generation of new data relative to a recombinant inbred line (RIL) population of 124 lines issued by a *Vicia faba* x *V. paucijuga* cross), recording six traits that dissected the transformation of flowers into pods and of ovules into seeds in the absence of cross-pollination. Fine mapping analyses of traits exploiting the collinearity of faba bean with *Medicago truncatula* and other legumes revealed two quantitative trait loci (QTLs) for the number of flowers per node in chromosome I and a few QTLs for the transformation of ovules into seeds that collocated in chromosome VI, all of them stable across years and evaluation conditions. QTLs were also found for the numbers of pods per node and pods per flower in chromosome IV but, in this case, their confidence interval would be too wide for envisaging a meaningful candidate-gene approach. Additional work aimed to elucidate mechanisms for self-fertility in a subset of contrasting genotypes. Self-sterile material exhibited production of largely abnormal pollen and narrower style-ovary angle than self-fertile one. On-going work will verify the importance for self-fertility of pollen availability synchronized with stigma receptivity by an early, timely rupture of the stigmatic cuticle.

In conclusion, the work on faba bean opened the way to the identification of self-fertility genes by future candidate-gene approaches that will exploit the generated information on stable, fine-mapped QTLs, as well as allowed for a deeper understanding of physiological mechanisms contributing to self-fertility.

Details of the scientific work described in this section can be found in the project Deliverable 1.4, whereas a few publications for international refereed journals are going to be finalized.

Marker-assisted selection for reduced anti-nutritional content in faba bean and grass pea

This research work, which was part of Task 2 of this WP, aimed to generate marker tools to select for low content of vicine and convicine in faba bean, and low content of β -N-oxalyl-L- α,β -diaminopropionic acid (β -ODAP) in grass pea (*Lathyrus sativus*). The work on faba bean was performed by IFAPA in collaboration with INRA, while AMBIS exploited its results to design and apply MAS tools to a few breeding programs. Vicine and convicine (v-c) are faba bean compounds with anti-nutritional effects in monogastric animals and potential toxicity in humans. There is a single gene (*vc-*) responsible for a 10-20 fold reduction, but its identification by a candidate-gene approach is complicated by lack of knowledge on the v-c synthetic pathway, the large faba bean genome size (~13 Gbp) and the lack of a reference genome.

The activity on faba bean implied the trait characterization for v-c content (at INRA) and the genetic analysis (at IFAPA) of the F₂ progeny of a bi-parental cross between two parent genotypes with contrasting v-c content (Vf6 x1268). The moderate synteny between the *Medicago truncatula* chromosome 2 and the faba bean chromosome 1 helped restrict the genomic region hosting the *vc-* gene. A genetic map including more than 4000 markers was generated using KASPar and DarTSeq markers, whereas fine-mapping work identified markers that were progressively close to the *vc-* gene. Eventually, two markers closely linked to the gene could explain about 86 % of the trait variation and showed 100 % efficiency in selecting for the low v-c allele in the studied population. Concurrently, candidate genes within the genomic interval of these markers are being defined, to identify the *vc-* gene via expression studies. In close contact with these developments, AMBIS designed and tested primers linked to low v-c in faba bean for use on breeding material provided by Saatzucht Gleisdorf and CREA.

In conclusion, the work on faba bean achieved the objective of producing a reliable molecular marker tool to select for low vicine-convicine content. Its utility for a range of other breeding populations is pending verification. In addition, this work set the basis for future candidate-gene research with high chances to eventually identify and sequence the *vc-* gene (and, generate, thereby, the most reliable marker-based tool for selection). In close contact with its developments, AMBIS designed and tested primers linked to low v-c in faba bean (partly exploiting results by subtask T1.2.3) for future use on breeding material provided by various partners. Finally, the dense genetic map that was produced can be useful for future genetic studies.

ITQB was responsible for the work on grass pea. ODAP is the main anti-nutritional compound of grass pea, and induces the neurodegenerative syndrome known as lathyrism in humans and animals that eat its grains frequently. Plant material with reduced ODAP content has been identified and there is evidence for low-ODAP as a quantitatively inherited trait, but information is lacking on genes and/or linked markers that could aid in selecting for this trait.

The research activity on grass pea encompassed the phenotyping for β -ODAP content in two locations (Alvaiázere and Córdoba) and the molecular characterization by about 3500 polymorphic DarTseq markers of 115 accessions representative of the worldwide species distribution. Marker-trait analysis revealed six markers with statistically significant linkage to the trait, which could be useful for selection work. Sequence information on these markers was used for BLAST searches, to unveil possible alleles associated with the metabolism of ODAP compounds. Also, a RIL population was developed from a cross between genotypes with contrasting trait levels, and its F₅ generation is undergoing field evaluation with the aim of detecting QTLs for β -ODAP content within a future genetic linkage map.

In conclusion, the work on grass pea identified the first reported molecular markers that could assist in the selection for low β -ODAP. The definition of these markers from a broad

genetic base should ensure applicability for a wide range of genetic material, while awaiting for the development of gene-based markers (or other markers closer to the gene).

The scientific work described in this section has been reported in greater detail within the project Deliverable 1.3, and will be the object of two articles for internationally refereed journals.

Genome-based approaches to select pea for drought tolerance

Drought is expected to increase in timing and intensity in Europe and the Mediterranean basin as a consequence of climate change. Despite its considerable importance, pea drought tolerance through genome- or marker-enabled approaches has hardly been investigated. LEGATO explored two approaches for this trait. The former, investigated by INRA, implied the search for allelic variation in candidate genes. The latter, studied by CREA, implied the validation of a genomic selection procedure for pea tolerance to severe terminal drought that was recently developed within the Era-Net project REFORMA. While the former approach can be useful to exploit the effect of a few major genes with high positive effect on the tolerance trait, the latter can capitalize on both major and minor genes by estimating their effect (as accounted for by linked markers) in a statistical model used for marker-based selection. The latter approach is particularly promising for complex quantitative traits (such as yield and drought tolerance), but its application has been hindered until very recently by its need for large marker numbers to reasonably saturate the plant genome.

Possible candidate genes were preliminarily searched for by assessing transcriptome changes in pea plants subjected to water stress during the seed filling period, which identified 233 genes with distinct up- or down-regulation under stress, from which were selected five genes on the basis of further experiments and literature information on *Medicago* and *Arabidopsis* orthologs. These genes encode (1) a glutamine amidotransferase, (2) a sucrose transporter from the SWEET family, (3) a phosphoenolpyruvate carboxykinase, (4) H⁺-coupled K⁺ transporter, and (5) a carotenoid cleavage dioxygenase named *RMS1* involved in strigolactone biosynthesis. Putative loss-of-function alleles for these genes were identified within a pea TILLING (Targeting Induced Local Lesions in Genomes) population. The effect of mutant alleles for the *RMS1* and the glutamine amidotransferase genes was assessed in TILLING lines of each trait that were previously backcrossed twice with the wild-type (cv. Cameor), which were grown along with the wild-type under stress and non-stress conditions in Dijon's phenotyping platform. Plants with the *RMS1* mutant gene proved more susceptible to drought in two experiments, suggesting that the biosynthesis of strigolactone could play a key regulatory role in the plant's response to drought. Work in progress is assessing the importance of the glutamine amidotransferase gene, as well as defining a regulatory network connecting sulfur nutrition and drought to locate relevant candidate genes.

The candidate-gene research work provided evidence for the important role of strigolactone biosynthesis for pea drought tolerance, although the relevant allelic variation available in modern pea germplasm requires exploration before finalizing the definition of marker-assisted selection tools. Work in progress may produce useful MAS tools for other key traits associated with drought tolerance.

Genomic selection exploited thousands of Genotyping-by-Sequencing (GBS)-generated SNP markers and phenotyping data for grain yield under severe terminal drought in Lodi's phenotyping platform and under field conditions in inland Morocco of 315 RILs belonging to three connected crosses between elite modern varieties. Predictions for line mean yield across environments displayed high predictive ability ($r = 0.69$ as correlation between modelled and observed data). Results from a genome-wide association study (GWAS) were exploited to develop a MAS criterion based on allelic variation at six independent loci, which, while being potentially alternative to genomic selection, had the distinct aim of selecting for intrinsic drought tolerance (as estimated by readjusting yields to eliminate the effect of drought escape via an early flowering). Genomic selection and MAS approaches were assessed on the basis of yield responses in managed-stress experiments of different sets of independent lines that were predicted to be top-, mid- or bottom-performing according to these selection criteria, along with phenotypically-selected material and parent lines. Both genomic selection and MAS were able to produce remarkable genetic progress. Genomic selection proved at least as efficient as phenotypic selection in terms of yield gain per selection cycle. Complementary work verified the superiority for genomic predictions of markers issued by the adopted GBS technique (based on *ApeKI* restriction enzyme) relative to another based on two restriction enzymes.

For genomic selection, displaying yield gain per selection cycle similar to phenotypic selection is very promising, considering that it is at least 4-fold less expensive per evaluated line and has at least 2-fold shorter selection cycle duration than phenotypic selection. On the other hand, the applicability of its model to other pea germplasm is pending verification and may require - as frequently foreseen by this approach - the development of specific models for other major reference populations. The MAS tool is of particular interest for selecting drought-tolerant germplasm for autumn-sown, cold-prone environments, where drought escape via an early flowering may produce material with greater susceptibility to low winter temperatures.

The scientific activities described in this section have been reported in detail within the project Deliverable 1.3 for the candidate-gene research work and the Deliverable D 1.5 for genomic selection and MAS. Their results have been reported in two scientific articles listed below, while three future articles for international refereed journals have been planned.

Gallardo, K., P.E. Courty, C. Le Signor, D. Wipf, V. Vernoud (2014). Sulfate transporters in the plant's response to drought and salinity: regulation and possible functions. *Frontiers in Plant Science* 5, 580 (doi: 10.3389/fpls.2014.00580).

Annicchiarico, P., N. Nazzicari, Y. Wei, L. Pecetti, E.C. Brummer (2017). Genotyping-by-sequencing and its exploitation for forage and cool-season grain legume breeding. *Frontiers in Plant Science* 8, 679 (doi: 10.3389/fpls.2017.00679).

High-density marker collections of pea and white lupin, and trait-marker studies in lupin

This section summarizes various research activities that were carried out within the Tasks 3 and 4 of this WP. The activities on pea aimed to create new germplasm with extensive molecular characterization for use as a genomic resource in future studies, including (1) a collection of Chromosome Segment Substitution lines (CSSL) derived from crosses of cultivated pea (*Pisum sativum*) with two wild species (*P. fulvum* and *P. elatius*), to be developed by UPOL and characterized in collaboration with GenXPro; (2) a very large RIL population originated by contrasting parent cultivars, to be developed by Aberystwyth University, whose high-density sequence-based map would be integrated with other existing maps and support the on-going pea genome sequencing effort by INRA. Work on white lupin (*Lupinus albus*) performed by CREA and IPG Poznan aimed at producing a dense genetic map and mapping a few key traits using data for a RIL population, as well as characterizing with high marker numbers a comprehensive landrace world collection and assessing the species diversity and genetic structure.

Introgression via three backcross generations of *P. fulvum* into cultivated pea, monitored by 28 microsatellite and 44 gene-specific markers, led to select 145 CSSL. These lines were genotyped by DARTseq markers, obtaining about 49000 SNP and 56000 PAV markers of which 1880 could be placed into linkage groups using synteny with *Medicago*. Fifty additional lines were genotyped using a 13k pea Illumina assay. A subset of 50 lines was phenotyped for 14 morphological and agronomic traits and several grain quality traits in two field experiment. MACE technology was used for a transcriptomic study of seed coat on another subset of lines, as this structure differs greatly between *P. fulvum* and cultivated pea. Introgression of *P. elatius* into cultivated pea (also involving three backcross generations) produced over 100 CSSL that await molecular characterization.

A RIL population of 760 F₄ lines was developed to produce a high-density map, but its genotyping by GBS produced data of insufficient quality. An alternative genotyping strategy based on an Affymetrix Axiom array was adopted, with application on F₆ lines that have become available in the meantime. The SNP choice for this 90k array has been performed by INRA, and the synthesis of the array is ongoing.

In conclusion, the two sets of pea CSSL, of which one thoroughly genotyped, can be precious as a genetic resource as well as a genomic resource for future studies, particularly for traits to which wild pea may provide useful variability (e.g., tolerance to biotic and abiotic stresses), or for several domestication traits that need be selected for by breeders whenever devising crosses between modern and wild pea – whose selection via efficient MAS tools

would be particularly attractive. The large pea RIL population represents as well a useful genetic resource and, once genotyped, will be an important genomic resource.

For white lupin, GBS produced about 4800 polymorphic SNP markers for a RIL population of 189 genotypes, and about 6200 SNPs for a world collection of 323 genotypes belonging to 86 landraces from 8 historical cropping regions and 15 recent varieties. RIL data were used to generate a dense consensus linkage map including about 3600 SNPs with no segregation distortion and 465 markers issued from previous maps. Mapping of sequence-based markers onto the newly-sequenced narrow leafed lupin (NLL, *Lupinus angustifolius*) genome revealed matching syntenic blocks for 79 % of the white lupin genome. Linkage analysis for key traits identified four QTLs for responsiveness to vernalization (the main trait affecting climatic adaptation), two major QTLs for anthracnose resistance (the main disease), and one QTL for alkaloid content. These QTLs did not correspond to previously mapped NLL loci for these traits, revealing different trait control in the two species. The analysis of landrace genetic structure, while detecting sizeable variation at all diversity levels (regions; landraces within region; genotypes within landrace), revealed large variation within landraces that was not expected in a predominantly self-pollinated crop. Landrace diversity reflected a longitudinal gradient, with outstanding diversity in the Near East that agrees with the hypothesis of crop origin in this region. Variety diversity represented a fairly limited and separate portion of the overall crop diversity. Multi-environment phenotyping data for 83 landraces and 8 varieties were used to verify the predictive ability of genomic selection models in this species, which proved high for grain yield ($0.80 < r < 0.50$ as correlation between predicted and observed data) and several other traits (onset of flowering, cold tolerance, seed weight, etc.).

The dense white lupin genetic map can be useful for future studies, whereas markers linked to three key traits could be exploited either via GBS genotyping or by implementation in a PCR-based genotyping tool. The different control of key traits in white lupin and NLL suggests a limit for our ability to exploit information from a close sequenced species such as NLL, reinforcing the need for genome sequencing of white lupin. The genetic structure of white lupin can drive the exploitation of landrace genetic resources (e.g., by exploiting also variation within landraces), whose interest is reinforced by the modest portion of genetic variation exploited by modern varieties. The high predictive ability of genomic selection for yield and other traits is quite promising for breeding, also considering that predictions for pure lines (as needed by breeders) may be simpler than the current ones for populations.

The scientific activities described here have been reported in detail within the project Deliverable 1.1 for white lupin and the Deliverable D 1.6 for pea CSSL material. Some results have been reported in the scientific article below, while at least three future articles for international refereed journals have been planned.

Książkiewicz, M., N. Nazzicari, H. Yang, M.N. Nelson, D. Renshaw, S. Rychel, B. Ferrari, M. Carelli, M. Tomaszewska, S. Stawiński, B. Naganowska, B. Wolko, P. Annicchiarico (2017). A high-density consensus linkage map of white lupin highlights synteny with narrow-leafed lupin and provides

markers tagging key agronomic traits. Scientific Reports 7, 15335 (doi: 10.1038/s41598-017-15626-w).