



ELSEVIER

Genetic and genomic analysis of legume flowers and seeds

Claire Domoney¹, Gérard Duc⁴, TH Noel Ellis¹, Cristina Ferrándiz²,
Christian Firnhaber³, Karine Gallardo⁴, Julie Hofer¹,
Joachim Kopka⁵, Helge Küster³, Francisco Madueño²,
Nathalie G Munier-Jolain⁴, Klaus Mayer⁶, Richard Thompson⁴,
Michael Udvardi⁵ and Christophe Salon⁴

New tools, such as ordered mutant libraries, microarrays and sequence based comparative maps, are available for genetic and genomic studies of legumes that are being used to shed light on seed production, the objective of most arable farming. The new information and understanding brought by these tools are revealing the biological processes that underpin and impact on seed production.

Addresses

¹ John Innes Centre, Colney Lane, Norwich NR4 7UH, UK

² Instituto de Biología Molecular y Celular de Plantas, CSIC-Universidad Politécnica de Valencia, Avenida de los Naranjos, Valencia 46022, Spain

³ Institute of Genome Research, Center for Biotechnology, Bielefeld University, D-33594 Bielefeld, Germany

⁴ INRA Unité de Recherches en Génétique et Ecophysiologie des Légumineuses à Graines (URLEG), BP 86510, F-21065 Dijon, France

⁵ Max Planck Institute of Molecular Plant Physiology, Am Mühlenberg 1, 14476 Golm, Germany

⁶ GSF Forschungszentrum für Umwelt und Gesundheit GmbH, Ingolstaedter Landstrasse 1, 85764 Neuherberg, Germany

Corresponding author: Ellis, TH Noel (noel.ellis@bbsrc.ac.uk)

Current Opinion in Plant Biology 2006, 9:133–141

This review comes from a themed issue on
Genome studies and molecular genetics
Edited by Nevin D Young and Randy C Shoemaker

Available online 15th February 2006

1369-5266/\$ – see front matter

© 2006 Elsevier Ltd. All rights reserved.

DOI 10.1016/j.pbi.2006.01.014

Introduction

Legumes are striking in the range of forms of their flowers, fruits and seeds [1••]. In addition to naming the family, ‘legume’ can also mean ‘fruit’ and in French it means ‘vegetable’; so clearly legumes are central to our ideas of a useful plant. Most edible crop legumes are in the large monophyletic group of the Papilionoideae, so we are used to the idea that legumes (with the notable exception of *Cordia*) have pea-like flowers, but the diversity of forms in the other legume groups is astonishing [1••] and the ancestral form of the papilionoid flower is not clear [2]. Many legume seeds are large and protein-rich, presumably a corollary to the attributes, notably nitrogen fixation, that permitted early legumes to compete well in

nutrient-poor tropical soils. Not all legumes are large seeded, however, and many Caesalpinoid legumes do not have a nitrogen-fixing symbiosis [2]. A major distinction within legume seeds is represented by those that do, versus those that do not, accumulate free amino acids in their seeds [3••].

The legume pod (the fruit, the legume) is the defining characteristic of this group and, together with the immature seed, is often eaten as a vegetable or by animals as a component of forage or silage. The earliest known fossil flower that has an emerging pod dates from 50 million years ago [4], but this is apparently relatively late in the diversification of legumes [3••].

In this review, we draw together strands of research on the genetic and genomic analysis of legume flowers and seeds, especially those that have been brought together in the EU ‘Grain Legumes’ sixth framework Integrated Project (GLIP: www.eugrainlegumes.org). The article therefore concentrates on the subset of legume crops in the Galegoid group called ‘cool season’ legumes together with the associated model systems.

Developmental physiology of seeds

Seed development depends on interactions between the genotype and the environment of the plant that bears the seed; interactions between the plant and the seed are also relevant. These general statements hold for all seed plants, but in legumes, the contribution of nitrogen fixation and its relation to N assimilation provides an added complexity. The existence of this complexity suggests that additional regulatory mechanisms may be found in growth-related processes in legumes, although elements of N regulation seem to be conserved over wide evolutionary distances [5].

Seed development has two largely distinct phases, first, cell division and, second, cell expansion. The first phase seems to depend on the embryo genotype, but cell division rate also varies according to the trophic and thermic environment of the embryo. Furthermore, cell division rate increases with the reproductive node. Hormones [6], environmental factors, or carbon (C) and N supply [7,8] have been proposed to control this mitotic activity of the seed. Phloem sucrose flow has been specifically implicated [9]

and the apoplastic hexose concentration, together with acid invertase activity, is positively correlated with final cell number. Conversely, the flow of N towards seeds does not determine seed growth potential [9].

During the seed filling or cell expansion phase, the main carbon contribution seems to derive from recent photosynthate rather than from remobilized carbon. Seed nitrogen (N) depends upon both symbiotic N₂ fixation and the retrieval of mineral N from the soil. Although there are temporary transfers to the vegetative organs, N that is accumulated during seed filling is devoted predominantly to seeds. Overall N economy influences seed N content and its composition in protein fractions, but for any given level of N nutrition, the mode of N acquisition (fixation versus assimilation) does not seem to influence seed N content or the seed protein profile [10].

Seed growth rate depends on the cotyledonary cell number that is established in the first developmental phase, and this rate is rather insensitive to assimilate availability. The duration of seed filling is variable, however, and when assimilates are not limiting, seed filling ceases only when seeds have reached their maximal size, which is a function of cell number. Under conditions of abiotic stress, symbiotic N₂ fixation and, to a lesser degree, mineral N absorption decline sharply and the seed's demand for N cannot be sustained. Under these conditions, N is remobilised from vegetative organs [11], provoking the destruction of the photosynthetic machinery and hence lowering CO₂ assimilation and leading to senescence. There seems to be a delay in the availability of remobilised N, possibly because it requires the activation of proteolysis in the source tissues or because of a progressive increase in seed sink strength [11]. Understanding the underlying physiology of N remobilisation requires additional studies on the proteome of leaves and stems [12], on the genetic variability that is associated with N remobilisation, and on the interaction of these with plant architecture.

Prolonging symbiotic fixation during seed growth, or optimising the interaction between N₂ fixation and the assimilation of mineral N throughout the growth cycle, reduces the demand for N remobilisation and therefore has a consequence for CO₂ assimilation during seed filling. This reduction in the demand for N remobilisation has been demonstrated by the progressive replacement of symbiotically fixed N by root-assimilated mineral N when N availability increases in the soil [13]. Potentially, the demand for remobilised N could be satisfied either by increasing the number of nodules per plant or by increasing the duration of symbiotic fixation. Both soybean [14] and pea mutants [15,16] that have increased nodule number are already available and have been studied. There is also potential for the genetic improvement of the shallow root system of pea [17] to provide larger,

deeper or more fibrous roots, as observed in spontaneous or induced mutants [18]. These roots could increase the retrieval of mineral N and/or provide an opportunity for an increased number of nodulation events.

Genetics

Metabolic genetics and physiology are key to understanding the process of seed filling. Nevertheless, we also need to understand the genetics of the seed itself, and of the earlier processes that regulate the timing and structure of flower and inflorescence development, in order to manipulate seed yield and quality.

Mutations affecting legume seed development

Many of the genes affecting seed morphology and/or seed composition that have been described in legumes have no apparent homologues in *Arabidopsis* and *vice versa*. For example, a wrinkled-seeded mutant of *Arabidopsis* is unrelated at the genomic level to those of pea, and is caused by a splicing mutation in a gene that encodes an APETALA2 (AP2)/ethylene-responsive element binding (EREB) transcription factor [19]. Seed developmental processes in pea are affected by the *single cotyledon (sic)* and *cytokinesis-defective (cyd)* mutations that produce enlarged cells, mainly in the cotyledons, and by mutations that do not allow the seed to complete a normal developmental programme [20].

With the exception of the *rugosus* (or wrinkled-seeded) class of mutants, mutations that influence the morphology of the mature seed in pea do not seem to determine specific changes in seed composition. The *rugosus* mutants, such as Mendel's wrinkled seed mutant *r* (which has a mutation in Starch-branching enzyme 1 [21,22]), affect starch biosynthesis. These defects affect sugar content and osmotic potential during seed development, the shape of the seed, and the proportion of protein, starch and lipid in the seed. As the fine-map positions of quantitative trait loci (QTL) that influence seed composition become available, their influence on seed development and morphology should become apparent.

Mutations in genes that affect the accumulation of individual seed proteins, starch polymers, oligosaccharides and other seed constituents have been described. These include mutations affecting the active site of trypsin inhibitor proteins [23], and the structural genes or QTL that affect the accumulation of Pea Albumin 2 (PA2) (C Domoney *et al.*, unpublished), Lipoyxygenase-2 (LOX-2) [24], amylose [21], verbascose [25] and iron [26]. Two classes of mutants that have an abnormal iron content have been described: in the *degenerating leaves (dgl)* mutant, excess iron is transported from vegetative tissues into seeds, whereas in the *bronze (brz)* mutant, this is not the case.

It is worth noting that many genes that are active during seed development might have additional functions in

other aspects of vegetative plant growth; for example, lectin genes are involved in root nodulation. In other cases, related but distinct genes are differentially regulated; for example, the various trypsin inhibitor genes are expressed in seeds, roots and flowers [27], and the closely related *PA2* and *ENDOTHECIUM 1 (END1)* genes are expressed in seeds and flowers, respectively [28].

Mutations that affect the development of legume flowers

Genes that are involved in flower development have been identified in different legume species [29,30], and these can often be related to genes described in *Arabidopsis* (Table 1). A major conclusion from these analyses is that many of the flowering genes have a function that is similar to that of their homologues in the model species *Arabidopsis thaliana*. This might be expected from the close taxonomic relationship between the Leguminosae and Brassicaceae, both of which are in the Rosid clade of the flowering plants [3]; however, there are significant dif-

ferences that possibly reflect unique aspects of the development of legume plants. For example, several legume genes, such as *Unifoliata (Uni)* or *Stamina pistilloida (Stp)*, have acquired additional functions [30–33]. A single *Arabidopsis* gene, *TERMINAL FLOWER1 (TFL1)*, seems to play the dual role of the pea homologues *Late flowering (Lf)* and *Determinate (Det)*. Likewise, the functions of the *Lotus japonicus PHANTASTICA (PHAN)* homologues, *LjPHANa* and *LjPHANb*, are performed by a single *Arabidopsis* gene [30,34]. MADS-box-family genes encode transcription factors that have key regulatory roles in processes such as flowering or flower and fruit development in *Arabidopsis* and other model species [35]. Several of the *Arabidopsis* flower MADS-box genes also have two homologues in legume species [30]. The expression of homologues of some *Arabidopsis* floral MADS-box genes has been detected in developing nodules of *Medicago sativa* [36,37]. The nature of the roles of these genes in nodulation is an intriguing open question.

Table 1

Genes and/or mutants related to legume flower development.

Mutant	Effect	Gene	<i>A. thaliana</i> gene	Effect in <i>A. thaliana</i>	Reference(s)
Pea (<i>Pisum sativum</i>)					
<i>late flowering</i>	Early flowering	<i>Lf</i>	<i>TFL1</i>	Early flowering, determination of shoot apex	[34]
<i>determinate</i>	Determination of shoot apex	<i>Det</i>	<i>TFL1</i>	Early flowering, determination of shoot apex	[34]
<i>vegetative1</i>	Inflorescences converted into shoots	n.c.			[38]
<i>vegetative2</i>	Inflorescences converted into shoots	n.c.			[39]
<i>unifoliata</i>	Flowers converted into inflorescences, abnormal leaves	<i>Uni</i>	<i>LFY</i>	Flowers converted into inflorescences	[31]
<i>proliferating inflorescence meristems</i>	Flowers converted into inflorescences	<i>Pim</i>	<i>AP1</i>	Flowers converted into inflorescences	[42,81]
<i>stamina pistilloida</i>	Homeotic floral organ conversions, abnormal leaves	<i>Stp</i>	<i>UFO</i>	Homeotic floral organ conversions	[32]
<i>cochleata</i>	Effect in flowers, leaves and nodules	n.c.			[43–45]
	n.m.	<i>PsPI</i>	<i>PI</i>	Homeotic floral organ conversions	[46]
<i>Medicago truncatula</i>					
<i>M. truncatula proliferating inflorescence meristem (mtpim)</i>	Flowers converted into inflorescences	<i>MtPIM</i>	<i>AP1</i>	Flowers converted into inflorescences	[82]
<i>M. truncatula apetala (mtap)</i>	Homeotic floral organ conversion	n.c.			[83]
<i>Lotus japonicus</i>					
<i>proliferating floral meristems</i>	Flowers converted into inflorescences, abnormal leaves	<i>PFM</i>	<i>LFY</i>	Flowers converted into inflorescences	[30]
<i>proliferating floral organs</i>	Homeotic floral organ conversions, abnormal leaves	<i>PFO</i>	<i>UFO</i>	Homeotic floral organ conversions	[33]

n.c., gene not cloned; n.m., no mutant available.

Finally, some pea mutants have no equivalent phenotypes described in *Arabidopsis*, suggesting the existence of genes that have unique functions in legumes. This is the case for the *Vegetative1* (*Veg1*) and *Veg2* genes, which are required for, or are correlated with, inflorescence architecture; noticeably, this trait is more complex in legumes than in *Arabidopsis* [38,39].

In pea, mutations that define a wide variety of genes controlling flowering time have also been identified and characterised thoroughly [39]. More recently, analysis of genomic data from model legume sequencing projects has led to the identification of homologues in pea, *Medicago truncatula*, *L. japonicus* and soybean for many of the key regulators of *Arabidopsis* flowering time [29*].

Flower and leaf development

Many legumes produce inflorescences and leaves during the reproductive phase of development. This is in contrast to *Arabidopsis*, which ceases leaf production at the transition to flowering, although tell-tale markers such as *AINTEGUMENTA* suggest that subtending leaf anlagen are present on the inflorescence [40]. In *Arabidopsis*, either leaf outgrowth is actively suppressed or the supply of cells from the apex is insufficient, such that flowers develop at the expense of leaves. In wildtype pea, neither of these potential obstacles occurs: lateral inflorescences and subtending compound leaves are formed at every post-floral transition node until the plant senesces and dies. In the prematurely terminating apical meristem of the *det* mutant, however, the incomplete development of subtending leaves is often seen and the second and final flowering node is usually subtended by an atypical, less-complex leaf [41]. The *Det* gene, which encodes a *TFL1* homologue [34], acts to maintain the indeterminacy of the apical meristem, and in so doing, maintains the balance between leaf and inflorescence production in the caulescent shoot.

This balancing act is also revealed by the *cochleata* (*coch*) mutant [42], which appears to have an increased meristematic potential during vegetative and reproductive phases. A surprising recent discovery is that this increased potential also extends to the roots, where bifurcating nodules carrying callus or root-like outgrowths have been observed [43–45].

The compound leaf of pea thus acts as a sensitive monitor of imbalances between leaf and inflorescence production, but also reveals common links in the regulation of leaf and flower development that are not apparent in *Arabidopsis*. Two classic examples of this are the mutants *uni* and *stp*, the pea orthologues of *leafy* (*lfy*) and *unusual floral organs* (*ufo*) [31,32]. Both mutants produce simplified compound leaves throughout development [30–32], but upon flowering, more familiar *Arabidopsis*-like roles of *Uni* and *Stp* in regulating floral meristem determinacy are revealed. A

more recently discovered example of a floral developmental gene that has a role in compound leaf patterning is the pea homologue of the *Arabidopsis* petal and stamen identity gene *PISTILLATA* (*PI*). This gene is capable of complementing the *Arabidopsis pi-1* floral mutant but was also found to be expressed in developing leaf primordia [46], thus hinting again at a second, as yet undiscovered, function in leaves.

Although differences between *Arabidopsis* and pea leaf and flower development have been uncovered, many aspects are conserved. The relationship between *Crispa*, the pea orthologue of *ASYMMETRIC LEAVES1* (*AS1*), and the *KNOX* homeobox genes is more like that in *Arabidopsis* than in tomato [47], even though the latter is a dicot species with compound leaves. A note of caution is required, however, as features uncovered in pea might not apply to all legumes. *Lotus* is known to carry a duplication of the *AS1* orthologous locus. The expression patterns of the two *Lotus* genes, called *LjPHANa* and *LjPHANb*, differ, leading Luo *et al.* [48] to postulate that they have divergent roles in leaves and flowers.

Tendrils are a specialised component of the pea leaf that are not available for study in other model plants, and they are important for the standing ability of this crop. Mutant rescue and phenocopy experiments using hormones and inhibitors [49], as well as differences in the expression levels of the auxin efflux carrier *Pisum sativum PINOID-LIKE2* (*PsPK2*) between tendrilled and non-tendrilled genotypes, suggest roles for auxin and gibberellin in regulating pea leaf form [50]. The relationship between hormones and *KNOX*-regulated leaf patterning events at the shoot apex is becoming clearer [51], but the nature of classical pea leaf-patterning genes, such as *afila*, *coch*, *tendrill-less* (*tl*) and *stipules reduced*, remains elusive. Berdnikov and Gorel [52] discuss the possibility that a newly identified gene, *Tl2*, which regulates both tendril morphology and flower colour, might have originally participated in the regulation of anthocyanin gene expression, before acquiring an additional role in pea leaf development.

Genomics

Both species-specific and comparative genomic tools are available in legumes, and for crops this is especially so for soybean. For genetic (including QTL) analysis, these tools provide a route to unravelling the molecular basis of given traits. For example, in comparisons between cowpea, mungbean and soybean, related genomic regions appeared to be associated with seed weight QTL [53,54]. In pea, QTL that affect total seed protein accumulation are being mapped in a number of laboratories, and lines that have elevated protein content are being subjected to metabolomic and genomic analyses. These studies should assist in the identification of genes in model systems that are candidates for the determinants of crop traits.

Between three [55] and seven [56] loci have been reported to contribute to variation in seed N content in crosses between round-seeded pea lines; in the latter work, *le* (internode length) was found to have a large effect. Further work is required to assign four of the linkage groups described by Tar'an *et al.* [55] to the consensus pea genetic map (e.g. by populating these linkage groups with reference markers). In *Arabidopsis*, transcription factors (TFs) that are preferentially expressed during seed development rather than in leaves are being defined by real-time reverse transcription (RT)-PCR [57], and the functions of these genes are being unravelled by RNA interference (RNAi). Ongoing examination of the relative expression of TFs during embryogenesis in *Medicago*, together with mutational analysis, will identify TFs that are 'master controllers' of seed composition; genomic sequencing projects will assist with identifying candidate TFs that are associated with the protein or nitrogen QTL.

Comparative genetics is an important approach in revealing legume gene function. Gene identification in complex genomes can be aided by anchoring genetic maps to reference genome sequence [58–60]. This has been used as a way of overcoming the problems of scale in the genomes of the Viciae or of segmental duplication in soybean. As in the examples discussed above, many legume genes have a novel function when compared to that described for *Arabidopsis* orthologues. The study of these genes will not only improve our understanding of legume genetics but will also increase our depth of knowledge of gene function. Comparative genetic tools therefore need to be augmented by legume-specific descriptions of the intermediate steps between genotype and phenotype. Post-genomic tools that are based on the emerging genomic sequence from the model legumes *L. japonicus* and *M. truncatula* are currently being constructed to assist in this task.

Transcriptome

A comprehensive set of expression profiling tools is becoming available for both model and crop legumes, in part developed within the GLIP. For *M. truncatula*, cDNA-based arrays have been used to profile gene expression in root endosymbioses [61–63]. The scope of these tools has been augmented by the addition of expressed sequence tag (EST) clusters from developing flowers and pods [64]. These experiments profiled transcriptional changes during flower and pod development, identifying more than 700 genes that are specifically regulated during the development of *M. truncatula* reproductive tissues.

To reduce cross-hybridization and to increase the number of probes, cDNA arrays are increasingly being replaced by 70mer oligonucleotide microarrays such as Mt16kOLI1-Plus, an upgrade of the tool described by [65], and Ps6kOLI1 for pea. In contrast to the *M. truncatula* micro-

array, the pea array focuses on genes that are relevant to legume seed formation. The Mt16kOLI1Plus arrays comprise 16 470 probes representing all tentative consensus sequences (TCs) of The Institute for Genomic Research *M. truncatula* Gene Index 5 plus 384 probes for transcription factors and other regulators. The Ps6kOLI1 arrays are based on 5177 probes that are targeted against EST clusters primarily derived from cotyledon and seed coat ESTs of pea [66].

Proteome

Proteomic analysis in *Medicago* using MALDI-TOF (matrix-assisted laser desorption/ionization-time of flight) mass spectrometry [67] identified 84 proteins whose abundance is regulated during seed filling. In congruence with the two phases of seed development mentioned above, proteins that are involved in cell division (e.g. β -tubulin and annexin) are abundant in early stages of seed development. However, these proteins decrease in abundance before the accumulation of the major storage proteins, which appear in the order: vicilins (14 days after pollination [DAP]), legumins (16 DAP) and convicilins (18 DAP) (see [68]). Other proteins that are putatively involved in protein deposition (protein disulfide isomerase and a heavy chain binding protein [BiP] homologue) and proteins that are implicated in cell expansion (such as a reversibly glycosylated polypeptide) are expressed concomitantly. Lipoxygenases and potential anti-nutritional factors, such as components of the precursor-accumulating vesicles (PV100), whose maturation releases a trypsin inhibitor, also accumulate during this phase.

At around 16 DAP, a temporary accumulation of proteins that are involved in carbon metabolism (e.g. sucrose synthase and starch synthase) and photosynthetic enzymes (i.e. oxygen-evolving enhancer protein and chlorophyll a/b binding protein) is observed, which might relate to the transient accumulation of starch in mid-maturation grains of *Medicago* [69]. The differential patterns of the appearance of enzymes that are involved in methionine metabolism provide new information about the regulation of metabolic activities during seed development. Methionine synthase, which provides methionine for protein synthesis, accumulates throughout seed filling, whereas S-adenosylmethionine synthetase is expressed early in seed filling and its accumulation might correlate with high metabolic activity at this stage. Enzymes that are involved in the recycling of methionine, including S-adenosylhomocysteine hydrolase and adenosine kinase, are expressed later. Finally, 1-aminocyclopropane-1-carboxylate (ACC) synthase, which is involved in the metabolic route from methionine to ethylene, is detected late in seed development, consistent with its role in the ripening process.

The seed proteome reference map has also been used to analyze the protein composition of seed tissues, yielding

specific markers for integuments, endosperm and embryo. Related proteomic analyses include studies on proteins that are expressed during somatic embryogenesis [70,71], and a survey of proteins isolated from the entire seed pod [72]. More recently, the analysis of the seed proteome has been extended and related to transcriptome data (K Gallardo *et al.*, pers. comm.).

Metabolome

Although not dependent on genome-sequence information, metabolomics (the non-biased analysis of all of the metabolites of an organism) has been facilitated by advances in technology in the post-genomic era. Common hardware configurations for metabolomics include gas chromatography coupled to mass spectrometry (GC-MS) or liquid chromatography (LC-MS) [73,74]. The full power of metabolomics is yet to be applied to legume flower and seed development. A non-biased GC-MS study of the different organs of *L. japonicus* showed that the metabolome of flowers was quite distinct from that of other organs [75]. Interestingly, the concentrations of a number of amino acids, including Gln, Pro, Trp, and Val, were significantly higher in flowers than in other organs, although the biological significance of this remains unknown (seeds were not included in this study). Several 'metabolite-profiling' studies have focused on specific types of compounds in legume seeds, for example saponins in *Medicago* seeds, pods, and other organs [76]. Open methods of metabolic profiling, such as NMR, also offer the potential for the genetic analysis of metabolic responses to environmental stresses [77]. Clearly, legume seed metabolomics represents an important research opportunity.

Bioinformatics

The application of high-throughput analytical platforms, and the associated volume of analytical data, require powerful strategies and technologies to manage, appropriately analyse and communicate the huge amount of genomic and functional genomic information produced by genome sequencing and genome-scale functional genomics projects.

There is a need to build the foundation for a legume genome research environment, which provides information and comprehensive analysis of genomic data, bioinformatics analysis, structuring of functional genomic data, and options to transfer the knowledge gained within the model genomes to legume crop species. The development of database tools for the storage, presentation and analysis of genetic diversity data has recently been presented [78] or can be found in the Legume Information System ([79]; www.comparative-legumes.org).

Annotation of the sequenced gene space of *M. truncatula* is an important component of post-genomic analysis and comparative genomics in legumes, so web-based search and browse interfaces are being developed as convenient

user interfaces. Within GLIP, a public *M. truncatula* genome database (UrMeLDB; European *Medicago* and Legume Database; <http://www.urmeldb.net>) has been implemented. This database houses all available *Medicago* genomic sequences as well as annotation and analysis data that are associated with genomic backbone templates. To ensure consistent structural annotation and to minimize the duplication of work, an International *Medicago* Genome Analysis Group (IMGAG) has agreed analytical standards and has distributed and shared the analytical workload between the contributing partners from Europe and the US.

It has become clear that most plant genomes have undergone successive rounds of polyploidisation and the restructuring of their genomes by successive deletion, inversion and intragenomic duplication events. With the emergence of extensive legume genome sequence, the history of genome duplication and restructuring events can be characterised in detail. For *Medicago*, meta-scale strategies have been planned to compare, relate and anchor the *Medicago* genomic template against other reference plant genomes. The close taxonomic proximity of *L. japonicus* and *M. truncatula* is especially fortunate for this analysis. This approach will provide a versatile means to reconstruct the evolutionary history of legume genomes, and will be instrumental in transferring functional information between organisms and in studying the fate and evolutionary dynamics of plant genomes. Already it seems that the large difference in the sizes of the genomes of *L. japonicus*, *M. truncatula* and the legumes in the Viciae is not a consequence of recent whole-genome or segmental duplication [58–60].

Within the GLIP, the bioinformatic groups are working towards the inter-operability of genome databases in the system called BioMOBY; this technology will, for example, be adapted for use within GLIP. The system provides application-level access to data and tools, while access to data and services will be integrated into a central web portal. Hence the system will provide a one-stop-shop for resources through the web portal and offer simple navigation combined with up-to-date data, simplifying data synchronization by removing the need to download and store data that is required for local analyses.

Conclusions

It is clear that a wide range of mutants and genetic variants are available in a diversity of legume species. The analysis of these resources with high-throughput methods is currently being undertaken in the context of a good description of the physiology of crops and other plants, and is underpinned by emerging genome sequence information. These analyses open up opportunities for the description and prediction of legume crop performance and allow the analysis of these processes in the context of the distinct physiology of legumes.

Although genome sequence analysis provides a firm anchor and description of legume gene content, it is clear that a remarkable degree of polymorphism exists within [80] and between legume genomes. The remodelling of metabolic and developmental processes in legumes is as diverse as their appearance, and will contribute to a deeper understanding of basic processes in plant biology.

Acknowledgements

The authors acknowledge the support of the European Commission under the Frameworks VI project FP6-2002-FOOD-1-506223 (GRAIN LEGUMES).

References and recommended reading

Papers of particular interest, published within the annual period of review, have been highlighted as:

- of special interest
- of outstanding interest

1. Lewis G, Schirer B, Mackinder B, Lock M (Eds): *Legumes of the World*. Kew: Royal Botanic Gardens; 2005.
This stunning book is an aesthetic delight. The book includes a current description of the phylogenetics and evolutionary history of legumes. Each genus is described and illustrated by wonderful photographs, line drawings and colour plates. The choice of illustrations is made with exceptional care. For example, the beautiful final plate that illustrates pea includes an illustration of each of the phenotypes described in Mendel's paper, including some mentioned only in passing.
2. Doyle JJ, Luckow MA: **The rest of the iceberg. Legume diversity and evolution in a phylogenetic context.** *Plant Physiol* 2003, **131**:900-910.
3. Wojciechowski M: **Reconstructing the phylogeny of legumes (Leguminosae): an early 21st century perspective.** *Adv Leg Syst* 2003, **10**:5-35.
This paper provides a comprehensive and informative description of the current synthesis of legume phylogenetics.
4. Lavin M, Herendeen PS, Wojciechowski MF: **Evolutionary rates analysis of Leguminosae implicates a rapid diversification of lineages during the tertiary.** *Syst Biol* 2005, **54**:575-594.
5. Schauser L, Roussis A, Stiller J, Stougaard J: **A plant regulator controlling development of symbiotic root nodules.** *Nature* 1999, **402**:191-195.
6. Ozga JA, van Huizen R, Reinecke DM: **Hormones and seed-specific regulation of pea fruit growth.** *Plant Physiol* 2002, **128**:1379-1389.
7. Egli DB, Ramseur EL, Zhen-Wen Y, Sullivan CH: **Source-sink alterations affect the number of cells in soybean cotyledons.** *Crop Sci* 1989, **29**:732-735.
8. Munier-Jolain NG, Ney B: **Seed growth rate in grain legumes. II. Seed growth rate depends on cotyledon cell number.** *J Exp Bot* 1998, **49**:1971-1976.
9. Munier-Jolain NG, Salon C: **Can sucrose content in phloem sap reaching field pea (*Pisum sativum* L.) seeds be a useful indicator of seed growth potential?** *J Exp Bot* 2003, **54**:2457-2465.
10. Munier-Jolain N, Boutin JP, Planchot V, Colonna P, Salon C, Duc G, Buleon A, Lhuillier-Soundele A, Page D, Quillien L *et al.*: **Elaboration de la qualité des graines de pois: influences environnementales sur leur qualité nutritionnelle et leur valeur technologique.** In *AIP-AGRAF Pour l'Elaboration de la Composition et de l'Aptitude à l'Utilisation des Graines et des Fruits 1996-1999*. Edited by Albagnac G, Colonna P, Doussinault G, Habib R. INRA, 2000. 43-66. [Translation of chapter title: Analysis of pea seed quality: environmental influences on nutritional quality and technological value.
11. Schiltz S, Munier-Jolain N, Jeudy C, Burstin J, Salon C: **Dynamics of exogenous nitrogen partitioning and nitrogen remobilization from vegetative organs in pea revealed by ¹⁵N in vivo labeling throughout seed filling.** *Plant Physiol* 2005, **137**:1463-1473.
12. Schiltz S, Gallardo K, Huart M, Negroni L, Sommerer N, Burstin J: **Proteome reference maps of vegetative tissues in pea. An investigation of nitrogen mobilization from leaves during seed filling.** *Plant Physiol* 2004, **135**:2241-2260.
13. Voisin AS, Salon C, Munier-Jolain NG, Ney B: **Effect of mineral nitrogen on nitrogen nutrition and biomass partitioning between the shoot and roots of pea (*Pisum sativum* L.).** *Plant Soil* 2002, **242**:251-262.
14. Searle IR, Men AE, Laniya TS, Buzas DM, Iturbe-Ormaetxe I, Carroll BJ, Gresshoff PM: **Long-distance signaling in nodulation directed by a CLAVATA1-like receptor kinase.** *Science* 2003, **299**:109-112.
15. Duc G, Messenger A: **Mutagenesis of pea (*Pisum sativum* L.) and the isolation of mutants for nodulation and nitrogen fixation.** *Plant Sci* 1989, **60**:207-213.
16. Sagan M, Ney B, Duc G: **Plant symbiotic mutants as a tool to analyse nitrogen nutrition and yield relationship in field-grown peas (*Pisum sativum* L.).** *Plant Soil* 1993, **153**:33-45.
17. Thorup-Kristensen K: **Root growth of green pea (*Pisum sativum* L.) genotypes.** *Crop Sci* 1998, **38**:1445-1451.
18. Sidorova KK, Shummy VK, Vlasova EY, Glianenko LN, Mishchenko TM: **The *Brt* (branched roots) and *Lrt* (long roots) genes control the development of roots in peas (*Pisum sativum* L.).** *Pisum Genet* 2002, **34**:23-24.
19. Cernac A, Benning C: **Wrinkled1 encodes an AP2/EREB domain protein involved in the control of storage compound biosynthesis in *Arabidopsis*.** *Plant J* 2004, **40**:575-585.
20. Johnson S, Liu CM, Wang TL: **An analysis of seed development in *Pisum sativum*. XVIII. The isolation of mutants defective in embryo development.** *J Exp Bot* 1994, **45**:1503-1511.
21. Wang TL, Domoney C, Hedley CL, Casey R, Grusak MA: **Can we improve the nutritional quality of legume seeds?** *Plant Physiol* 2003, **131**:886-891.
22. Bhattacharyya MK, Smith AM, Ellis THN, Hedley C, Martin C: **The wrinkled-seed character of pea described by Mendel is caused by a transposon-like insertion in a gene encoding starch-branching enzyme.** *Cell* 1990, **60**:115-122.
23. Page D, Aubert G, Duc G, Welham T, Domoney C: **Combinatorial variation in coding and promoter sequences of genes at the *Tri* locus in *Pisum sativum* accounts for variation in trypsin inhibitor activity in seeds.** *Mol Genet Genomics* 2002, **267**:359-369.
24. Forster C, Domoney C, Casey R: **Analysis of a lipooxygenase pseudogene in *Pisum*.** *Theor Appl Genet* 1999, **98**:835-839.
25. Peterbauer T, Lahuta LB, Blochl A, Mucha J, Jones DA, Hedley CL: **Analysis of the raffinose family of oligosaccharide pathway in pea seeds with contrasting carbohydrate composition.** *Plant Physiol* 2001, **127**:1764-1772.
26. Grusak MA: **Strategies for improving the iron nutritional quality of seed crops: lessons learned from the study of unique iron-hyperaccumulating pea mutants.** *Pisum Genet* 2000, **32**:1-5.
27. Domoney C, Welham T, Ellis N, Mozzanega P, Turner L: **Three classes of proteinase inhibitor gene have distinct but overlapping patterns of expression in *Pisum sativum* plants.** *Plant Mol Biol* 2002, **48**:319-329.
28. Gomez MD, Beltrán J-P, Cañas LA: **The pea *END1* promoter drives anther-specific gene expression in different plant species.** *Planta* 2004, **219**:967-981.
29. Hecht V, Foucher F, Ferrándiz C, Macknight R, Navarro C, Morin J, Vardy ME, Ellis N, Beltrán JP, Rameau C, Weller JL: **Conservation of *Arabidopsis* flowering genes in model legumes.** *Plant Physiol* 2005, **137**:1420-1434.
The authors of this paper use legume sequence information to locate homologues of *Arabidopsis* flowering genes in pea and *Medicago*. This revealed that some gene families have expanded differentially in legumes and that some important *Arabidopsis* flowering-time regulators are absent.

30. Dong ZC, Zhao Z, Liu CW, Luo JH, Yang J, Huang WH, Hu XH, Wang TL, Luo D: **Floral patterning in *Lotus japonicus***. *Plant Physiol* 2005, **137**:1272-1282.
31. Hofer J, Turner L, Hellens R, Ambrose M, Matthews P, Michael A, Ellis N: **UNIFOLIATA regulates leaf and flower morphogenesis in pea**. *Curr Biol* 1997, **7**:581-587.
32. Taylor S, Hofer J, Murfet I: **Stamina pistilloida, the pea ortholog of *Fim* and *UFO*, is required for normal development of flowers, inflorescences, and leaves**. *Plant Cell* 2001, **13**:31-46.
33. Zhang S, Sandal N, Polowick PL, Stiller J, Stougaard J, Fobert PR: **Proliferating Floral Organs (*Pfo*), a *Lotus japonicus* gene required for specifying floral meristem determinacy and organ identity, encodes an F-box protein**. *Plant J* 2003, **33**:607-619.
34. Foucher F, Morin J, Courtiade J, Cadioux S, Ellis N, Banfield MJ, Rameau C: **DETERMINATE and LATE FLOWERING are two TERMINAL FLOWER1/CENTRORADIALIS homologs that control two distinct phases of flowering initiation and development in pea**. *Plant Cell* 2003, **15**:2742-2754.
35. Jack T: **Plant development going MADS**. *Plant Mol Biol* 2001, **46**:515-520.
36. Heard J, Dunn K: **Symbiotic induction of a MADS-box gene during development of alfalfa root nodules**. *Proc Natl Acad Sci USA* 1995, **92**:5273-5277.
37. Zuccherro JC, Caspi M, Dunn K: **ngf9: a third MADS-box gene expressed in alfalfa root nodules**. *Mol Plant Microbe Interact* 2001, **14**:1463-1467.
38. Reid JB, Murfet IC: **Flowering in *Pisum*: a fifth locus**. *Veg. Ann Bot* 1984, **53**:369-382.
39. Reid JB, Murfet IC, Singer SR, Weller JL, Taylor SA: **Physiological-genetics of flowering in *Pisum***. *Sem Cell Dev Biol* 1996, **7**:455-463.
40. Long J, Barton MK: **Initiation of axillary and floral meristems in *Arabidopsis***. *Develop Biol* 2000, **218**:341-353.
41. Singer S, Hsiung LP, Huber SC: **Determinate (*det*) mutant of *Pisum sativum* (Leguminosae: Papilionoideae) exhibits an indeterminate growth pattern**. *Amer J Bot* 1990, **77**:1330-1335.
42. Taylor SA, Hofer JM, Murfet IC, Sollinger JD, Singer SR, Knox MR, Ellis TH: **PROLIFERATING INFLORESCENCE MERISTEM, a MADS-box gene that regulates floral meristem identity in pea**. *Plant Physiol* 2002, **129**:1150-1159.
43. Gourlay CW, Hofer JM, Ellis TH: **Pea compound leaf architecture is regulated by interactions among the genes UNIFOLIATA, COCHLEATA, AFILA, and TENDRIL-LESS**. *Plant Cell* 2000, **12**:1279-1294.
44. Yaxley JL, Jablonsky W, Reid JB: **Leaf and flower development in pea (*Pisum sativum* L.) mutants COCHLEATA and UNIFOLIATA**. *Ann Bot* 2001, **88**:225-234.
45. Ferguson BJ, Reid JB: **Cochleata: getting to the root of legume nodules**. *Plant Cell Physiol* 2005, **46**:1583-1589.
46. Berbel A, Navarro C, Ferrándiz C, Cañas LA, Beltrán JP, Madueño F: **Functional conservation of PISTILLATA activity in a pea homolog lacking the PI motif**. *Plant Physiol* 2005, **139**:174-185.
47. Tattersall AD, Turner L, Knox MR, Ambrose MJ, Ellis THN, Hofer JM: **The mutant *crispa* reveals multiple roles for PHANTASTICA in pea compound leaf development**. *Plant Cell* 2005, **17**:1046-1060.
48. Luo JH, Yan J, Weng L, Yang J, Zhao Z, Chen JH, Hu XH, Luo D: **Different expression patterns of duplicated PHANTASTICA-like genes in *Lotus japonicus* suggest their divergent functions during compound leaf development**. *Cell Res* 2005, **15**:665-677.
49. DeMason DA: **Auxin-cytokinin and auxin-gibberellin interactions during morphogenesis of the compound leaves of pea (*Pisum sativum*)**. *Planta* 2005, **222**:151-166.
50. Bai F, Watson JC, Walling J, Weeden N, Santner AA, DeMason D: **Molecular characterization and expression of *PsPK2*, a PINOID-like gene from pea (*Pisum sativum*)**. *Plant Sci* 2005, **168**:1281-1291.
51. Hay A, Craft J, Tsiantis M: **Plant hormones and homeoboxes: bridging the gap?** *Bioessays* 2004, **26**:395-404.
52. Berdnikov VA, Gorel FL: **A mutation, *tl2*, in pea (*Pisum sativum* L.) affects leaf development only in the heterozygous state**. *Theor Appl Genet* 2005, **110**:1086-1091.
53. Fatokun CA, Menancio-Hautea DI, Danesh D, Young ND: **Evidence for orthologous seed weight genes in cowpea and mung bean based on RFLP mapping**. *Genetics* 1992, **132**:841-846.
54. Maughan PJ, Saghai Maroof MA, Buss GR: **Molecular-marker analysis of seed-weight: genomic locations, gene action, and evidence for orthologous evolution among three legume species**. *Theor Appl Genet* 1996, **93**:574-579.
55. Tar'an B, Warkentin T, Somers DJ, Miranda D, Vandenberg A, Blade S, Bing D: **Identification of quantitative trait loci for grain yield, seed protein concentration and maturity in field pea (*Pisum sativum* L.)**. *Euphytica* 2004, **136**:297-306.
56. Burstin J, Marget P, Huart M, Munier-Jolain N, Loridon K, Aubert G, Rameau C, Duchene C, Desprez B, Duc G: **QTLs of seed nitrogen content in pea**. In *Proceedings of the Fifth European Conference on Grain Legumes*; Dijon, France. AEP, 2004:151-152.
57. Czechowski T, Bari RP, Stitt M, Scheible WR, Udvardi MK: **Real-time RT-PCR profiling of over 1400 *Arabidopsis* transcription factors: unprecedented sensitivity reveals novel root- and shoot-specific genes**. *Plant J* 2004, **38**:366-379.
58. Kaló P, Seres A, Taylor SA, Jakab J, Kevei Z, Kereszt A, Endre G, Ellis THN, Kiss GB: **Comparative mapping between *Medicago sativa* and *Pisum sativum***. *Mol Genet Genom* 2004, **272**:235-246.
59. Choi H-K, Kim D-J, Zhu H, Mun J-H, Baek J-M, Roe B, Ellis N, Young ND, Doyle J, Kiss G, Cook DR: **Conserved gene order between crop and model legume species**. *Proc Natl Acad Sci USA* 2004, **101**:15289-15294.
60. Zhu H, Choi H-K, Cook DR, Shoemaker RC: **Bridging model and crop legumes through comparative genomics**. *Plant Physiol* 2005, **137**:1189-1196.
61. Küster H, Hohnjec N, Krajinski F, El Yahyaoui F, Manthey K, Gouzy J, Dondrup M, Meyer F, Kalinowski J, Brechenmacher L, van Tuinen D *et al.*: **Construction and validation of cDNA-based Mt6k-RIT macro- and microarrays to explore root endosymbioses in the model legume *Medicago truncatula***. *J Biotechnol* 2004, **108**:95-113.
62. El Yahyaoui F, Küster H, Ben Amor B, Hohnjec N, Pühler A, Becker A, Gouzy J, Vernié T, Gough C, Niebel A *et al.*: **Expression profiling in *Medicago truncatula* identifies more than 750 genes differentially expressed during nodulation, including many potential regulators of the symbiotic program**. *Plant Physiol* 2004, **136**:3159-3176.
63. Manthey K, Krajinski F, Hohnjec N, Firnhaber C, Pühler A, Perlück AM, Küster H: **Transcriptome profiling in root nodules and arbuscular mycorrhiza identifies a collection of novel genes induced during *Medicago truncatula* root endosymbioses**. *Mol Plant Microbe Interact* 2004, **17**:1063-1077.
64. Firnhaber C, Pühler A, Küster H: **EST sequencing and time course microarray hybridizations identify more than 700 *Medicago truncatula* genes with developmental expression regulation in flowers and pods**. *Planta* 2005, **222**:269-283.
- This paper reports the first microarray-based approach to characterize gene expression in *M. truncatula* reproductive tissues. Marker genes for specific stages of flower and pod development are proposed on the basis of unsupervised clustering techniques.
65. Hohnjec N, Vieweg MF, Pühler A, Becker A, Küster H: **Overlaps in the transcriptional profiles of *Medicago truncatula* roots**

- inoculated with two different *Glomus* fungi provide insights into the genetic program activated during arbuscular mycorrhiza. *Plant Physiol* 2005, **137**:1283-1301.
66. Radchuk R, Radchuk V, Weschke W, Borisjuk L, Weber H: **Repressing the expression of the SUCROSE NONFERMENTING-1-RELATED PROTEIN KINASE gene in pea embryo causes pleiotropic defects of maturation similar to an abscisic acid-insensitive phenotype.** *Plant Physiol* 2006, **140**:263-278.
67. Gallardo K, Le Signor C, Vanderkerckhove J, Thompson R, ●● Burstin J: **Proteomics of *Medicago truncatula* seed development establishes the timeframe of metabolic processes related to reserve accumulation.** *Plant Physiol* 2003, **133**:1-19.
- This paper reports the identity of 84 stage-specific seed proteins and provides evidence for a programme of metabolic activities that accompany seed development.
68. Casey R, Domoney C, Ellis THN: **Legume storage proteins and their genes.** *Oxford Surveys in Plant Molecular Cell Biology* 1986, **3**:1-96.
69. Djemel N, Guedon D, Lechevalier A, Salon C, Miquel M, Prosperi JM, Rochat C, Boutin JP: **Development and composition of the seeds of nine genotypes of the *Medicago truncatula* species complex.** *Plant Physiol Biochem* 2005, **43**:557-566.
70. Imin N, Nizamidin M, Daniher D, Nolan KE, Rose RJ, Rolfe BG: **Proteomic analysis of somatic embryogenesis in *Medicago truncatula*. Explant cultures grown under 6-benzylaminopurine and 1-naphthaleneacetic acid treatments.** *Plant Physiol* 2005, **137**:1250-1260.
71. Imin N, de Jong F, Mathesius U, van Noorden G, Saeed NA, Wang X, Rose RJ, Rolfe BG: **Proteome reference maps of *Medicago truncatula* embryogenic cell cultures generated from single protoplasts.** *Proteomics* 2004, **4**:1883-1896.
72. Watson BS, Asirvatham VS, Wang L, Sumner LW: **Mapping the proteome of barrel medic (*Medicago truncatula*).** *Plant Physiol* 2003, **131**:1104-1123.
73. Fiehn O, Kopka J, Dormann P, Altmann T, Trethewey RN, Willmitzer L: **Metabolite profiling for plant functional genomics.** *Nat Biotechnol* 2000, **18**:1157-1161.
74. Huhman DV, Sumner LW: **Metabolic profiling of saponins in *Medicago sativa* and *Medicago truncatula* using HPLC coupled to an electrospray ion-trap mass spectrometer.** *Phytochem* 2002, **59**:347-360.
75. Desbrosses GG, Kopka J, Udvardi MK: ***Lotus japonicus* metabolic profiling: development of GC-MS resources for the study of plant-microbe interactions.** *Plant Physiol* 2005, **137**:1302-1318.
76. Huhman DV, Berhow M, Sumner LW: **Quantification of saponins in aerial and subterranean tissues of *Medicago truncatula*.** *J Agric Food Chem* 2005, **53**:14-20.
77. Charlton A, Allnut T, Holmes S, Chisholm J, Bean S, Ellis N, Mullineaux P, Oehlschlager S: **NMR profiling of transgenic peas.** *Plant Biotech J* 2004, **2**:27-35.
78. Lee JM, Davenport GF, Marshall D, Ellis THN, Ambrose MJ, Dicks J, van Hintum TJJ, Flavell AJ: **GERMINATE. A generic database for integrating genotypic and phenotypic information for plant genetic resource collections.** *Plant Physiol* 2005, **139**:619-631.
79. Gonzales MD, Archuleta E, Farmer A, Gajendran K, Grant D, Shoemaker R, Beavis WD, Waugh ME: **The Legume Information System (LIS): an integrated information resource for comparative legume biology.** *Nucleic Acids Res* 2005, **33**(database issue):D660-D665.
80. Jing R, Knox MR, Lee JM, Vershinin AV, Ambrose M, Ellis THN, Flavell AJ: **Insertional polymorphism and antiquity of *PDR1* retrotransposon insertions in *Pisum* species.** *Genetics* 2005, **171**:741-752.
81. Berbel A, Navarro C, Ferrándiz C, Cañas LA, Madueno F, Beltrán JP: **Analysis of *PEAM4*, the pea *AP1* functional homologue, supports a model for *AP1*-like genes controlling both floral meristem and floral organ identity in different plant species.** *Plant J* 2001, **25**:441-451.
82. Benlloch R, d'Effurth I, Ferrándiz C, Cosson V, Caballero T, Beltrán JP, Cañas L, Ratet P, Madueño F: **Reverse genetic strategies in *Medicago truncatula*: use of *Tnt1* mutagenesis and RNAi in the analysis of floral initiation and development.** 5th European Congress on Grain Legumes. Abstract Book. Dijon, France; 2004:207.
83. Penmetsa RV, Cook DR: **Production and characterization of diverse developmental mutants of *Medicago truncatula*.** *Plant Physiol* 2000, **123**:1387-1398.