Legume anchor markers link syntenic regions between

*Phaseolus vulgaris, Lotus japonicus, Medicago truncatula and Arachis*

Birgit Kristine Hougaard *, Lene Heegaard Madsen *, Niels Sandal *, Marcio de Carvalho Moretzsohn †, Jakob Fredslund ‡1, Leif Schauer ‡2, Anna Marie Nielsen *3, Trine Rohde *4, Shusei Sato §, Satoshi Tabata §, David John Bertioli †† and Jens Stougaard *

* Laboratory of Gene Expression, Department of Molecular Biology, University of Aarhus, DK-8000 Aarhus C, Denmark
† Embrapa Recursos Genéticos e Biotecnologia, CEP 70.770-900, Brasília, DF, Brazil
‡ Bioinformatics Research Center, University of Aarhus, DK-8000 Aarhus C, Denmark
§ Kazusa DNA Research Institute, 1532-3 Yana, Kisarazu, Chiba 292, Japan
†† Universidade Católica de Brasília, SGAN 916, CEP 70.790-160 Brasilia, DF, Brazil

1

Present address: ¹ The Alexandra Institute Ltd., University of Aarhus, DK-8200 Aarhus N, Denmark
² Interdisciplinary Nanoscience Center, University of Aarhus, DK-8000 Aarhus C, Denmark
³ Centre for Membrane Pumps in Cells and Disease-PUMPKIN, Department of Molecular Biology, University of Aarhus, DK-8000 Aarhus C, Denmark
⁴ Danish Biomembrane Research Centre, Institute of Physiology, The faculty of Health Sciences, DK-8000 Aarhus C, Denmark
Running title: Bean-Lotus-Medicago-Arachis synteny

Key words: Bean, peanut, model legumes, gene specific markers, synteny

Corresponding author: Jens Stougaard, Laboratory of Gene Expression, Department of Molecular Biology, University of Aarhus, Gustav Wieds Vej 10, DK-8000 Aarhus C, Denmark,
Telephone +45 8942 5011, E-mail: stougaard@mb.au.dk
We have previously described a bioinformatics pipeline identifying comparative anchor tagged sequence loci, combined with design of intron-spanning primers. The derived anchor markers defining the linkage position of homologous genes are essential for evaluating genome conservation among related species and facilitate transfer of genetic and genome information between species. Here we validate this global approach in common bean and in the AA genome complement of allotetraploid peanut. We present the successful conversion of app. 50% of the bioinformatics defined primers into legume anchor markers in bean and diploid Arachis species. One hundred and four new loci representing single copy genes were added to the existing bean map. These new legume anchor marker loci enabled the alignment of genetic linkage maps through corresponding genes and provided an estimate of the extent of synteny and collinearity. Extensive macro-synteny between Lotus and bean was uncovered on eight out of the 11 bean chromosomes and large blocks of macrosynteny were also found between bean and Medicago. This suggests that anchor markers can facilitate a better understanding of the genes and genetics of important traits in crops with largely uncharacterized genomes using genetic and genome information from related model plants.
INTRODUCTION

The legume family (Leguminosae) is the third largest family of higher plants and includes more than 19,000 species (LEWIS et al. 2005). Plants belonging to the family are diverse, and trees, shrubby perennials, annual herbs, ornamentals as well as agriculturally important crops are represented. Legumes play a critical role in natural ecosystems, agriculture, and agroforestry, where their ability to establish symbiosis with nitrogen fixing rhizobial bacteria makes them efficient colonizers of low-nitrogen environments and desirable protein crops. *Phaseolus vulgaris* L. (common bean) is a particularly important source of protein. It serves as a staple food known as the poor man’s meat, and contains dietary fiber, minerals, vitamins and various health promoting compounds (GUILLON and CHAMP 2002; LETERME 2002). *Arachis hypogaea* (cultivated peanut) is in addition to being consumed in many human foods, the fifth most important oilcrop and also a rich source of dietary protein for the chicken and pork industries (GRAHAM and VANCE 2003).

The legume family is divided into three subfamilies: Caesalpinioideae, Mimosoideae and Papilionoideae. Most of the economically important legumes are members of the monophyletic subfamily Papilionoideae, which can be divided into four major clades. Although the legumes included in our study, *Lotus japonicus*, *Medicago truncatula*, bean and Arachis, are all in the Papilionoideae, they belong to three different clades: halogalegina, phaseoloid/millettiooid and aeschynomenooid/dalbergioid. Lotus and Medicago are equally closely related to bean and equally distantly related to Arachis (DOYLE and LUCKOW 2003). Also the agronomical and genome characteristics of the legumes in our study differ. Both of the model legumes are herbaceous plants of limited
agricultural use with relatively small genomes of app. 470 Mb, while common bean and cultivated peanut are major grain legumes with larger genomes of 588 Mbp and 2,813 Mbp, respectively (http://www.rbgkew.org.uk/cval/homepage.html). In bean there are more than 29,000 domesticated and 1,300 wild accessions in germplasm banks (BROUGHTON et al. 2003). However, the genetic base of the commercial cultivars of specific market classes is narrow. Less than 5% of the genetic diversity available has been used globally despite nearly a century of organized bean improvement (BROUGHTON et al. 2003). Several genetic maps representing different populations have been established on a backbone of RFLP markers (FREYRE et al. 1998; NODARI et al. 1993; VALLEJOS et al. 1992).

Cultivated peanut is an allotetraploid with an AABB genome complement. The polyploidization event in its origin left the tetraploid cultivated peanut reproductively isolated from its wild diploid relatives (HALWARD et al. 1993). This fact, combined with self-pollination, has entailed a narrow genetic base and a limited diversity for some traits of agricultural interest (MORETZSOHN et al. 2005). No linkage map is available for cultivated peanut. The maps that have been published in Arachis were based on interspecific crosses of diploid Arachis species or a cross of a synthetic tetraploid with cultivated peanut (BUROW et al. 2001; HALWARD et al. 1993; MORETZSOHN et al. 2005).

Characterizing syntenic relationships between model and crop legumes seems therefore to have wide applications. If collinearity (conservation of gene order) or synteny (conservation of linkage) is found between species, information from large-scale genome sequencing of the Lotus and Medicago gene-rich regions can be used to optimize the only limited genetic and genomic information available from the genomes of bean and Arachis. Following this rationale our objective was to investigate the level and structure of synteny
between the legume models (primarily Lotus) and bean and Arachis. The aim was to design robust PCR based markers requiring relatively simple laboratory equipment and for this purpose we focused on size markers based on insertions/deletions, CAPS (cleaved amplified polymorphic sequence) and dCAPS markers (derived cleaved amplified polymorphic sequence) (KONIECZNY and AUSUBEL 1993; NEFF et al. 1998). The bioinformatics pipeline which suggests comparative anchor tagged sequence loci (CATS) for use within a set of related organisms (FREDSLUND et al. 2006a) and the automated primer design program PriFi (FREDSLUND et al. 2005) were used to identify and rank intron-spanning gene based sets of degenerate primers. In order to identify conserved primer sites for use throughout the legume family, EST collections originating from Lotus, Medicago, soybean and, to a limited extent, bean and Arachis were used as input. The pipeline then applied a number of filters optimizing the selection of single copy orthologous genes and the likelihood of identifying polymorphisms between the mapping parents. Hence primers were designed for conserved exon regions allowing PCR amplification of an intron-spanning fragment, detection of polymorphism and gene verification by sequencing of the intron and flanking exon regions, respectively (FREDSLUND et al. 2006a; FREDSLUND et al. 2006b; FREDSLUND et al. 2005). Applying these criteria a list of 459 universal Leg-primer sets were generated (http://cgi-www.daimi.au.dk/cgi-chili/GeneticMarkers/table).

We have tested 159 pipeline generated primer pairs in bean and 124 in Arachis and converted approximately 50% into legume anchor markers. Subsequent comparison of the corresponding positions of these anchor markers on the linkage maps of Lotus, Medicago, bean and Arachis showed different levels of genome conservation. A high level of
macrosynteny was discovered between bean and Lotus where syntenic regions often span large regions, even entire linkage groups. Substantial regional synteny was also found between bean and Medicago and these regions occasionally cover segments with no apparent bean-Lotus synteny.

MATERIALS AND METHODS

Design of gene based PCR markers: Homologous ESTs from mainly *Lotus japonicus*, *Medicago truncatula* and soybean were aligned together with genomic sequence from Lotus or Medicago and primer sites in conserved exon regions were identified. The bioinformatics pipeline were set to optimize for the following criteria: 1) the targeted gene was either single copy or represented by two copies in Arabidopsis. 2) Primer design were based on ESTs where the topology of the gene tree were found to match the species tree indicating orthology of targeted genes. 3) Primer pairs anneal in conserved exon regions allowing PCR amplification of an intron-spanning fragment (Fredslund *et al.* 2006a; Fredslund *et al.* 2006b; Fredslund *et al.* 2005). The resulting primers are called Leg-primers. At first the work was done manually, later (starting with marker Leg060) the process was automated. In addition, the ESTs were compared to the full set of inferred protein-coding sequences of the reference species Arabidopsis in order to estimate the number of paralogous sequences. Only sequences expected to be single copy in the legumes were used.

Starting with Leg175, a fully automated bioinformatics approach was used for both grouping closest homologues, sequence alignment and primer design (Fredslund *et al.*
A number of primer sets (http://bioweb.abc.hu/cgi-mt/pisprim/pisprim.pl) developed by György Kiss’s group in Gödöllő, Hungary for the Grain Legume Integrated Project (GLIP) were also tested in bean. These primer sets were developed from essentially the same criteria. Three of them were mapped and given names of the type “mtmt_Gen_” followed by three numbers with spacing in between.

The criteria for selecting primers, from the bioinformatics pipeline, for experimental use/ marker development were that they should be single copy in Arabidopsis. Homologous single copy genes in any two genomes are likely to be orthologs, and only true orthologues are suitable for comparative studies. Moreover, the alignment underlying primer design should contain ESTs from at least three different species, or two species from different clades, to represent a broad span of sequence divergence and thereby make it likely that the PCR primers would perform well in distantly related legumes like Arachis. The predicted biological function of the gene, where the primers anneal, was never used as criteria in the selection process.

**DNA extraction and marker development:** Genomic DNA from the mapping parents and for each of the recombinant inbred lines (RILs) were isolated from frozen, young leaf tissue using a modified cetyltrimethylammonium bromide (CTAB) method. Leaf tissue from a single leaf was crushed in liquid nitrogen using mortar and pestle. The tissue was then transferred to a 13 ml Sarstedt tube and CTAB extraction was carried out as described by Sandal et al. (2006) except that the chloroform/isoamylalcohol step was repeated and 10
µl (instead of 5 µl) of RNAse (10 mg/ml stock) was added. The DNA pellet was redissolved in about 100 µl of H₂O. Standard PCR conditions were used for the parental survey, where markers were amplified with 94°C for 5 min, then 40 cycles of 94°C denaturing for 30 sec, 48°C (or 45°C) annealing for 30 sec and 72°C extension for 2 min, followed by a 5 min final extension at 72°C. The PCR reaction was carried out in a final volume of 20 µl containing estimated 0.1 µg of genomic DNA. Single PCR products were purified in the tube using a combination of ExonucleaseI (New England Biolabs, M0293L) and Shrimp Alkaline Phosphatase (USB, 70092Y) in a buffer containing 20 mM Tris-HCl pH 8.0 and 10 mM MgCl₂. Multiple PCR products were gel extracted using the QIAquick Gel Extraction Kit (Qiagen, 28706), and except for eluting in 30 µl, the kit protocol was followed. Sequencing was done using ABI BigDye version 3.1. The identity of the amplification products was confirmed by Blastx searches against the Arabidopsis protein database at NCBI or comparison of the translated sequence from the amplification product to the translated sequence of the EST contig underlying primer design. Only amplification products reliably identified were considered further. If there was a significant size difference in the amplification product from the mapping parents, a size marker was made. Otherwise the web-based dCAPS Finder Ver. 2.0 software (NEFF et al. 2002) was used for suggestion of enzymes for CAPS markers and design of mismatched primers for development of dCAPS markers based on reliable SNPs.

**Map calculation and map drawing in bean:** Linkage analysis was performed using JoinMap Version 2.0 (STAM 1993). The linkage groups were numbered according to
Freyre et al. (1998). MapChart 2.2 was used to draw the map (Voerrips 2002), and map modifications were done in Microsoft PowerPoint. To test the quality of the data and the mapping, the colormapping procedure was used (Kiss et al. 1998).

The map position of the bean sequences representing 18 different standard bean markers with a single best match (E-value < $10^{-7}$) to a Lotus sequence with a known map position (Table S3) were placed on the bean-Lotus/Medicago/Arachis syntenic map (Figure 3) along with the Leg-markers. Some of these markers (D1032, Bng12, Bng228 and Bng200) were already positioned along with the Leg-primers using the BxJ RILs. Others had been mapped in either the BxJ F$_2$ population or in other bean populations, and positioning of these on the BxJ RIL map were made by manual evaluation according to shared markers and cM distances on the bean integrated linkage maps (http://www.ncbi.nlm.nih.gov/mapview/map_search.cgi?taxid=3885) (this applies to D1051, Bng72, Bng108, Bng148/R, Mng415, Bng57, Bng3, Hyp4-1, Bng179, Bng87, Bng95, Bng26, Bng206 and Bng1).

**Linkage analysis in Lotus:** For the anchor markers that could not be mapped in the Gifu x MG20 population, primers were tested for polymorphism between *L. japonicus* Gifu, and the related species *Lotus filicaulis* (Sandal et al. 2006). Lotus BAC/TAC sequences and the web-based computer program Tandem Repeat Occurrence Locator (TROLL) (http://wsmartins.net/webtroll/troll.html) was used for identifying microsatellites. Due to suppression of recombination and resulting poor resolution in parts of the Lotus genetic map (Sandal et al. 2006), some of the Leg-markers were not mapped to a specific location on the Lotus map, but rather to a region. If this region was smaller than 10 cM the marker
was artificially put on the map on a location equivalent to the midpoint of the interval where it was mapped (this applies to: 12M-Gm, 33M-Gm, Leg036, Leg055, Leg081, Leg183, Leg187, Leg191, Leg203, Leg213, Leg228 and Leg231). Leg-markers mapping to a region larger than 10 cM are not shown on the Lotus map (this applies to: Leg111, Leg177, Leg202, Leg206 and Leg762).

RESULTS

**Gene based PCR markers:** To validate the application of the Leg-primers for comparative mapping a subset corresponding to single copy genes in Arabidopsis were selected for generating legume anchor markers in bean and Arachis. The chosen Leg-primers were used to amplify PCR fragments from the BAT93 and Jalo EEP558 parents of the bean recombinant inbred lines (RILs) (Freyre et al. 1998), as well as the *Arachis duranensis* K7988 and *Arachis stenosperma* V10309 parents (Bertioli et al. 2008) of the Arachis population, subsequently used for mapping. Following the flow scheme outlined in Figure 1 an overlapping set of 159 and 124 primer pairs were tested in bean and Arachis, respectively. In order to verify the identity of amplified fragments, the terminal exon sequences of the PCR fragments were compared to a translated version of aligned ESTs used for primer design or alternatively a Blastx search was performed (Figure 1). For bean approximately 70% of the PriFi primer pairs based on ESTs from four, three or two different legume species amplified a product that was verified (Table S1). In Arachis approximately 55% of the primer pairs based on four or three ESTs produced a PCR fragment that could be verified (Table S1). Following a positive confirmation of gene
identity, sequence polymorphism between the mapping parents was used to develop the CAPS, dCAPS or size markers applied for mapping the Leg-marker loci. In bean 50% of the original primer sets tested were developed into markers and in Arachis the yield was 45% (Table 1). Prior to the development of the bioinformatics pipeline a number of markers were developed manually using the same principles and adding these markers a total of 111 bean markers were made.

**Gene based genetic map of bean:** To develop the bean genetic linkage map, 77 lines from the recombinant inbred (RI) population BAT93 x JaloEEP558 (BxJ) (FREYRE et al. 1998) were used. The segregation information collected from the 111 Leg-markers in this study was analyzed, along with data from 150 framework markers from the core linkage map (FREYRE et al. 1998). Using JoinMap version 2.0 (STAM 1993) and a 3.5 LOD threshold, we were able to assign 104 Leg-markers to 99 loci in 11 linkage groups, corresponding to the haploid chromosome number of bean (Figure 2). The linkage groups range from 40.8 cM ($Pv$LG10) to 99.8 cM ($Pv$LG2), and the total map length is 758 cM (Table 2). The number of Leg-markers per chromosome range from three to 19 with an average of 9.5 and the average interval between Leg-markers is 7.3 cM. The generated Leg-markers include 15 size polymorphic markers, 48 CAPS and 48 dCAPS markers. Marker information can be seen in the supplementary Table S2.

Besides the described linkage groups, one Leg-marker (7-Gm) showed linkage to one of the framework markers (DROD3a) and four Leg-markers (Leg191, Leg194, Leg223 and Leg232) came out as unlinked loci.
Mapping in both the BxJ and the DOR364 x G19833 (DxG) population, had previously proved useful for positioning microsatellites that were monomorphic in one or the other of the populations (Blair et al. 2003). In an attempt to improve the marker yield seventeen different Leg-primer pairs, which amplified a correct but monomorphic product in BxJ, were tested in the DxG mapping parents. However, a SNP polymorphism was detected in only two of the PCR fragments (data not shown) implying that a mapping population originating from more distantly related parents are optimal for this approach.

**Allele frequencies in the BxJ RI population:** The allele frequency was calculated for each marker locus to evaluate the degree of deviation from the expected 0.5 transmission frequency for each parental allele at each locus of the RIL lines. All markers with segregation data were subjected to this analysis, including the framework markers and interestingly there were large regional differences in allele frequencies. At the bottom part of **PvLG7** there is a major segregation distortion. The maternal BAT93 alleles are overrepresented compared to the paternal Jalo EEP558 alleles (Figure S1). The segregation distortion has a maximum at the bottom of **PvLG7** with 85% of the alleles originating from BAT93 but is normalized to the expected value of 50% at the middle of **PvLG7** after about 35 cM. This linkage group was previously reported to have a preference for the BAT93 allele (Freyre et al. 1998; Nodari et al. 1993). On **PvLG10**, a modest but consistent segregation distortion was observed. **PvLG10** carries only 11 markers, but at all loci, the transmission frequency was higher for the BAT93 alleles. Moreover, on **PvLG1**, **PvLG4** and **PvLG6** large regions of segregation distortion were found. On **PvLG1** and **PvLG6** the Jalo EEP558 alleles were favoured in regions of about 30 cM, whereas on **PvLG4** the
BAT93 alleles were more prominent, again in a region of about 30 cM. The preference for the Andean Jalo EEP558 alleles on linkage group \( Pv\)LG1 is consistent with the observations of Freyre et al. (1998) and Nodari et al. (1993). Preferential transmission of parental alleles could be caused by an allele specific advantage in viability or fertility and Leg-markers may represent or be linked to alleles selected during the eight generations used for development of the RILs (Freyre et al. 1998).
Bean-Lotus synteny: In order to evaluate the extent of macrosynteny between Lotus and bean the individual positions of Leg-markers on the two genetic linkage maps were compared. Most of the Leg-marker positions in Lotus were known from their presence on BAC or TAC clones, which were mapped in the intraspecific cross between *L. japonicus* ecotypes Gifu and *L. japonicus* Miyakojima MG-20 (HAYASHI et al. 2001; KATO et al. 2003) and the collinear *Lotus filicaulis x L. japonicus* Gifu map (SANDAL et al. 2006). Where map positions were absent, we took advantage of the Lotus genome-sequencing program. A corresponding BAC or TAC clone was included in the genome sequencing and genetically mapped using the standard microsatellite marker procedure. A total of eighty-one of the legume anchor markers were positioned on the genetic map of both bean and Lotus (Figure 3). To add more loci, ninety-nine of the previously developed standard bean markers from the Build 1.1 linkage map (http://www.ncbi.nlm.nih.gov/mapview/map_search.cgi?taxid=3885) that had sequence information attached was searched (Blastn) against the Lotus genome sequence. This approach found bean sequences representing 18 different bean markers with a single best match (E-value < 10\(^{-7}\)) to a Lotus sequence with a known map position on the Lotus genetic map (Table S3).

A total of 99 unique positions were shared between bean and Lotus, ranging from two (*Pv*LG5) to 20 (*Pv*LG2) shared markers per linkage group and an average of nine common markers per linkage group. Extensive synteny between Lotus and bean was found for about half of the bean linkage groups. In *Pv*LG4, *Pv*LG7, *Pv*LG9 and *Pv*LG11 the syntenic regions span entire linkage groups (with a maximum of one non-syntenic marker), and in *Pv*LG2 and *Pv*LG8 half a linkage group. *Pv*LG5 and *Pv*LG10 have only two and three markers, respectively, shared with Lotus. Only four of the 11 bean linkage groups
have markers that map to more than two different Lotus linkage groups, and even among these, three clearly show synteny towards one particular Lotus chromosome. For example, six out of 11 markers on PvLG3 belong to LjLG4, seven out of 12 markers on PvLG6 belong to LjLG6 and five out of eight markers on PvLG8 belong to LjLG2. PvLG1 contains nine shared markers mapping to three different Lotus linkage groups.

The syntenic marker positions on Lotus and bean chromosomes shown in Figure 3 indicate that only a limited number of large-scale rearrangements occurred during evolution of legume chromosomes. To evaluate this suggestion further marker collinearity was investigated. Three examples of this analysis are shown in Figure 4. It is clear that the blocks of syntenic loci are only partly collinear and that smaller internal chromosome rearrangements have occurred in addition to the large scale rearrangements. Figure 4c shows an example of one bean linkage group having blocks of anchor loci shared with two Lotus linkage groups and one Lotus linkage group having syntenic blocks shared with two bean linkage groups.

**Bean-Medicago synteny:** Positions of Leg-markers in Medicago were found by Blastn searches against the Medicago pseudomolecules (http://www.medicago.org/genome/cvit_blast.php), using a combination of bean sequences amplified with the Leg-primers and the Medicago EST sequence included in the CATS alignment for Leg-primer design. Search information is presented in Table S4. For the previously developed bean standard markers with sequence information known, the bean sequence was used (Table S3). Among the Leg-markers, 65 had a match with E-value $< 10^{-7}$ in Medicago, and among the 18 marker sequences 10 had a match with E-value $< 10^{-7}$. All
in all this amounts to 75 shared positions between bean and Medicago, ranging from three (PvLG5 and PvLG7) to 15 (PvLG2) markers per bean linkage group and an average of 6.8 markers per linkage group (Figure 3). Substantial synteny is seen. Two of the bean linkage groups, having three (PvLG5) and five (PvLG9) shared positions to Medicago respectively, have only markers mapping to a single Medicago chromosome and at the bottom part of PvLG2 a large syntenic segment spanning 52 cM is seen. Six out of 11 bean linkage groups have markers that map to more than two different Medicago linkage groups, but five of these have one (PvLG3 and PvLG11) or two (PvLG2, PvLG6 and PvLG8) uninterrupted syntenic regions of at least three markers. A total of 11 uninterrupted regions with three or more shared markers are distributed over the bean linkage groups.

Figure 5 shows two detailed alignments between bean and Medicago. In both examples, blocks of loci show synteny, but only partial collinearity indicating small internal chromosome rearrangements.

**Bean-Arachis synteny:** Positions of Leg-markers in Arachis were found by mapping in a diploid F2 mapping population, originating from a cross between *Arachis duranensis* K7988 and *Arachis stenosperma* V10309 (BERTIOLI et al. 2008). Eighty-two Leg-markers were mapped in Arachis and 38 of them have a linkage map position in bean. The number of shared loci between bean and Arachis range from zero (PvLG1) to eight (PvLG3), with an average of 3.5 shared loci per bean linkage group. In three of the bean linkage groups (PvLG2, PvLG4 and PvLG8) non-disrupted syntenic segments of three or more loci mapped to the same Arachis linkage group. Of these, PvLG4 and PvLG8 only shared
markers to one Arachis linkage group. However, several of the bean linkage groups have a patchy pattern of loci shared with more Arachis linkage groups (Figure 3).

Figure 6 shows two detailed examples of macrosynteny between Medicago, bean, Lotus and Arachis. The syntenic blocks between Arachis and Lotus consist of three and four markers respectively.

Bean-Lotus-Medicago-Arachis- synteny: Besides giving a quick overview of the pairwise synteny observed between bean, Lotus, Medicago and Arachis, respectively, the colorscheme given in Figure 3 reflects the previously described synteny blocks between Lotus and Medicago (CANNON et al. 2006). Accordingly, it is possible to relate the synteny detected by Leg-markers to the chromosome correspondances between the two model legumes. In the large-scale sequence based comparison reported by CANNON et al. (2006) 10 large-scale synteny blocks were identified. Noteably most of the syntenic regions we have identified fall in these blocks. For example three large syntenic segments between bean, Lotus and Medicago were found on \( P_vLG2 \) and \( P_vLG9 \) (Figure 3 and Table 2). Our synteny analysis therefore supports and is supported by the large-scale sequence based comparison.

Defining syntenic segments as non-interrupted regions with at least three markers, 48% of the bean linkage map is syntenic to Lotus (367 cM out of 758 cM) and 35% is syntenic to Medicago (263 cM out of 758 cM). However, bean has more markers shared with Lotus than with Medicago and therefore the numbers are not directly comparable.
Bean-Arabidopsis synteny: In our set-up for primer design, we have used the GenBank accession number of the Arabidopsis orthologue as a reference for the alignment of legume EST sequences, and therefore we have shared positions for some 100 loci in bean and Arabidopsis. Five of the bean linkage groups (PvLG3, PvLG5, PvLG6, PvLG8 and PvLG11) have uninterrupted regions of three or more loci extending up to 33 cM (Leg133-Leg220 on PvLG11) which share linkage on an Arabidopsis chromosome (Figure S2). Given the roughly 112 million years thought to separate Arabidopsis from the legumes (BOWERS et al. 2003), the gene linkage is remarkable although macrosynteny has been observed before (GRANT ET AL. 2000). Moreover, our experimental set-up focuses on single copy and highly conserved genes, and this group of genes is optimal for identifying ancestral synteny.

DISCUSSION

Here we present the application of an automated and systematic approach for developing anchor markers in two distantly related legumes, bean and Arachis, chosen for testing the feasibility of the approach. Among the 159 Leg-primer pairs created by the bioinformatics pipeline, which were exhaustively tested in bean and an overlapping set of 124 Leg-primer pairs tested in Arachis, 113 (71%) amplified the correct product in bean and 67 (54%) amplified the correct product in Arachis. PriFi suggests and ranks four different primer pairs per alignment. For each of these we have only used the highest ranking and tested rankings from 187 to 22 without finding a difference in efficiency between high ranking and low ranking pairs. Use of one or more of the three alternative PriFi primer sets would therefore most likely increase the marker yield. Similarly, there were no major differences
in efficiency between primer pairs based on ESTs from four, three or two different legume species. The PriFi settings used to generate the 459 primer sets appear therefore to be robust. Since the Arachis mapping parents were more polymorphic than the bean-mapping parents the marker yield was nearly the same, 50% versus 45%. The use of the Arachis markers will be presented by BERTIOLI et al. (2008). Here we present a bean linkage map with 104 new legume anchor marker positions. The linkage map adds additional genetic resolution to the existing bean maps and expands the map with 1.9 cM at the outermost top of \( P_vLG6 \) and 4.1 cM at the top of \( P_vLG10 \). Leg-markers distribute over all of the 11 bean linkage groups, and in only five cases do two Leg-markers share the same locus. This absence of clustering may indicate that suppression of recombination is rare in the BxJ RILs. The distribution of Leg-markers and thus the genetic resolution varies. For example, linkage group \( P_vLG2 \) has 19 Leg-markers, whereas \( P_vLG10 \) has only three Leg-markers. This difference might be a consequence of the bioinformatics set-up designing Leg-primers. The bioinformatics pipeline selects primers in regions of high sequence conservation and gives preference to highly expressed genes represented in ESTs collections. Neither conserved genes nor highly expressed genes are necessarily randomly distributed in genomes and BERTIOLI et al. (2008) has reported an inverse relationship between the frequency of single copy genes and clustered retrotransposons. Most importantly our results provide proof of concept for the bioinformatics driven identification of anchor tagged loci which, together with positional information from homologous loci in Lotus, Medicago and Arachis, provides a basis for comparative genomics and an overview of the synteny among these four legumes.
**Synteny:** One of the ideas behind the model legume concept was to exploit synteny between model and crops to accelerate isolation and comparative characterization of genes from the less characterized crop legumes. This approach requires that the target genome regions contain the same genes in approximately the same order in models and crops. A substantial effort has been invested in investigating this question. Regional synteny between Medicago and/or Lotus models and pea (Aubert et al. 2006), alfalfa (Choi et al. 2004a), red clover (Sato et al. 2005), lens (Phan et al. 2006; Phan et al. 2007) and lupin (Nelson et al. 2006; Phan et al. 2006) lend support to this approach. In the present analysis, we have anchor tagged loci covering 758 cM of the bean genetic map and a set of 99 shared loci made it possible to compare this map with the genetic map of Lotus. All of the 11 bean linkage groups had non-interrupted regions of at least two markers also showing linkage in the Lotus genome, and in several of the bean linkage groups synteny spanned entire linkage groups. All in all, the legume anchor markers revealed a broad conservation of gene linkage (macrosynteny) between bean and Lotus.

Seventy-five loci were shared between bean and Medicago and substantial conservation of gene linkage between bean and Medicago was found (Figure 3). However, even though Medicago and Lotus phylogenetically are equally distant from bean, the conserved bean-Medicago segments are generally shorter than observed for bean-Lotus. Looking at uninterrupted syntenic regions larger than 20 cM, five can be found between bean and Medicago (PvLG2 [two regions] PvLG6, PvLG9, and PvLG10) whereas seven are seen between bean and Lotus (PvLG2 [two regions], PvLG4, PvLG7, PvLG8, PvLG9 and PvLG11). In certain regions linkage is more conserved between bean and Medicago than between bean and Lotus. For example the region in PvLG3 flanked by Leg187 and
Leg164N. Here five Leg-markers all map to MtChr8, whereas the same region maps to four different Lotus linkage groups. Likewise, in PvLG6 the region flanked by Leg217 and Leg081 have four Leg-markers all mapping to MtChr7, whereas in Lotus, they map to three different linkage groups. However, in PvLG2, PvLG4, PvLG7 and PvLG11 macrosynteny between bean and Lotus is more pronounced than between bean and Medicago.

Despite the fact that bean and Arachis have similar numbers of chromosomes (11 and 10, respectively), the wild Arachis species have genome sizes that are considerably larger than the genomes of bean and Lotus. *Arachis duranensis* has a gametic DNA content (1C) of 1243 Mbp (http://www.kew.org/genomesize/homepage.html). Moreover, Arachis is a rather distant relative of bean and Lotus, so the evidence of macrosynteny documented between Arachis, bean, Lotus and Medicago is remarkable. In both of the detailed examples shown in Figure 6, the loci shared between Lotus and Arachis span a larger region in Lotus than in Arachis. Given the difference in DNA content between these two species this seems surprising. However, whereas the average physical equivalent of 1 cM corresponds to 1.97 Mbp in Arachis (MORETZSOHN et al. 2005), Lotus chromosome 1 has a mean ratio of 0.77 Mbp/cM (PEDROSA et al. 2002).

A comparative map spanning cold- and warm-season legumes was presented by CHOI et al. (2004b). In their study, marker segregation was analyzed in *M. truncatula*, *Medicago sativa*, *P. sativum*, *V. radiata* and bean. Using a limited number of markers, eight out of 11 bean linkage groups were included from their already described relation to *V. radiata* (mung bean) (BOUTIN et al. 1995). No markers were directly shared between bean and Medicago and bean-Medicago synteny was inferred using several bridging legume species. The bean-Medicago relations presented here are documented by directly
corresponding shared loci and include all of the 11 bean linkage groups. In addition, nine of the bean linkage groups show synteny to more than one Medicago linkage group, and in the presented study we have precise indication of which regions on the individual bean linkage groups share homology to which Medicago chromosomes.

Macrosynteny between bean and Lotus was described by Zhu et al. (2005). A circle diagram included eight out of 11 bean linkage groups and the relations to Lotus were inferred through positions shared between Lotus and Medicago. In contrast to the study of Zhu et al. (2005), we present markers directly bridging the genomes of bean and Lotus and thereby we define regions of shared macrosynteny directly on each of the 11 bean linkage groups.

Previous studies have demonstrated substantial macrosynteny between Lotus and Medicago (Choi et al. 2004b; Young et al. 2005; Cannon et al. 2006), and in a bioinformatics approach comparing 121 Mbp of non-redundant map-anchored genomic sequence from the Lotus build with 149 Mbp from the Medicago build, 10 large-scale synteny blocks were identified (Cannon et al. 2006). On the bean - Lotus/Medicago/Arachis synteny map (Figure 3), the color code reflects the described synteny between Lotus and Medicago. As seen, in large parts of bean genome, the synteny to Lotus and Medicago converges on the 10 large-scale synteny blocks defined by direct sequence comparison. We infer that large syntetic segments are conserved in the genomes of legumes and that anchor markers can link these segments between different species and directly to the genome sequence of the model legumes.

Given the extensive linkage observed between bean and the two model legumes in this study, we therefore conclude that Lotus and Medicago will have a significant potential
for future identification of tightly linked markers for marker-assisted selection, and map-
based isolation of candidate genes.
ACKNOWLEDGEMENTS

Paul Gepts and Valérie Geffroy kindly provided seeds for the BAT93 x Jalo EEP558 population. Joe Tohme and Matthew Blair provided seeds for the DOR364 x G19833 mapping parents and DNA from the corresponding mapping population.

The work of BKH was funded by the Council for Development Research (RUF) under the Danish Ministry of Foreign Affairs (Project number 91210). In addition, there has been funding by The European Union INCO-DEV Programme (ARAMAP reference: ICA4-2001-10072); the Generation Challenge Program (Project number 31); the EU FP6 Grain Legume Integrated Project (GLIP reference: FOOD-CT-2004-506223); and the Danish National Research Foundation (Centre for Carbohydrate Recognition and Signalling).


Burow, M. D., C. E. Simpson, J. L. Starr and A. H. Paterson, 2001 Transmission genetics of chromatin from a synthetic amphidiploid to cultivated peanut (Arachis
*hypogaea* L.), broadening the gene pool of a monophyletic polyploid species. Genetics **159**: 823-837.


VOORRIPS, R. E., 2002 MapChart: software for the graphical presentation of linkage maps and QTLs. J Hered. **93**: 77-78.

SUPPLEMENTARY MATERIAL

Table S1. Efficiency of primer pairs based on 2, 3 and 4 ESTs.

Table S2. Bean Leg-marker information.

Table S3. Bean framework markers and the position of their homologues in *Lotus japonicus* and *Medicago truncatula*.

Table S4. Position of Leg-markers in *Medicago truncatula*.

**Figure S1.** Segregation of codominant markers in the bean RIL population.

**Figure S2.** Genetic linkage map of common bean with synteny to Arabidopsis given by a color code.
Table 1. Summary of primer efficiency for exhaustively tested primer pairs.

<table>
<thead>
<tr>
<th>Primer pairs</th>
<th>Amplify correct</th>
<th>Markers developed</th>
</tr>
</thead>
<tbody>
<tr>
<td>tested</td>
<td>gene</td>
<td></td>
</tr>
<tr>
<td>Bean (BxJ)</td>
<td>159 113 (71.1%)</td>
<td>80 (50.3%)</td>
</tr>
<tr>
<td>Arachis (KxV)</td>
<td>124 67 (54.0%)</td>
<td>56 (45.2%)</td>
</tr>
</tbody>
</table>
Table 2 Synteny summary of the 11 *Phaseolus vulgaris* linkage groups. The main affinities observed towards, *Lotus japonicus* (*Lj*), *Medicago truncatula* (*Mt*) and *Arachis* (Ara) are given. The approx. sizes of the synteny blocks in bean are based on bean-Lotus synteny, allowing one non-syntenic marker.

<table>
<thead>
<tr>
<th>Bean linkage group</th>
<th>Length of LG (cM)</th>
<th>No. of anchor markers</th>
<th>Synteny identified</th>
<th>Cannon syntenic block</th>
<th>Approx. size (cM)</th>
</tr>
</thead>
<tbody>
<tr>
<td><em>PvLG1</em></td>
<td>56.3</td>
<td>9</td>
<td>5 1 -</td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>1 7 -</td>
<td>5</td>
<td>4</td>
</tr>
<tr>
<td><em>PvLG2</em></td>
<td>99.8</td>
<td>23</td>
<td>4 4 3</td>
<td>7</td>
<td>34</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>2 5 -</td>
<td>10</td>
<td>58</td>
</tr>
<tr>
<td><em>PvLG3</em></td>
<td>95.0</td>
<td>15</td>
<td>4 - -</td>
<td>8</td>
<td>15</td>
</tr>
<tr>
<td><em>PvLG4</em></td>
<td>75.0</td>
<td>7</td>
<td>2 4 6 5</td>
<td>-</td>
<td>65</td>
</tr>
<tr>
<td><em>PvLG5</em></td>
<td>56.3</td>
<td>4</td>
<td>3 2 -</td>
<td>-</td>
<td>23</td>
</tr>
<tr>
<td><em>PvLG6</em></td>
<td>71.6</td>
<td>15</td>
<td>6 2 -</td>
<td>3</td>
<td>37</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>1 7 -</td>
<td>5</td>
<td>3</td>
</tr>
<tr>
<td><em>PvLG7</em></td>
<td>70.7</td>
<td>9</td>
<td>5 1 9</td>
<td>2</td>
<td>64</td>
</tr>
<tr>
<td><em>PvLG8</em></td>
<td>70.8</td>
<td>12</td>
<td>2 5 7</td>
<td>9</td>
<td>30</td>
</tr>
<tr>
<td><em>PvLG9</em></td>
<td>42.9</td>
<td>9</td>
<td>1 3 2,8</td>
<td>4</td>
<td>43</td>
</tr>
<tr>
<td><em>PvLG10</em></td>
<td>40.8</td>
<td>4</td>
<td>3 - -</td>
<td>-</td>
<td>8</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>-----</td>
<td>--------</td>
<td>------</td>
<td>----</td>
<td>----</td>
<td>----</td>
</tr>
<tr>
<td>PyLG11</td>
<td>78.8</td>
<td>15</td>
<td>3</td>
<td>4,8</td>
<td>3</td>
</tr>
<tr>
<td></td>
<td>758</td>
<td>122&lt;sup&gt;a&lt;/sup&gt;</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<sup>a</sup> Including 104 Leg-markers and 18 standard bean markers (http://www.ncbi.nlm.nih.gov/mapview/map_search.cgi?taxid=3885) with a single best sequence match (E-value <10<sup>-7</sup>) to a Lotus sequence with a known map position in Lotus.
FIGURE LEGENDS

Figure 1.—Flow scheme of marker production. The Leg-primers were tested on DNA from mapping parents. Single products were enzymatically purified before sequencing, verification of gene identity and detection of polymorphism. Multiple products were gel extracted individually before sequencing. Only sequence confirmed PCR products were converted into markers.

Figure 2.—Genetic linkage map of bean. Each vertical bar represents individual bean linkage groups (LG). Marker names are shown on the right side of each linkage group, connecting lines indicate the position of each locus on the linkage map and centiMorgan values are given on the left side. Each of the Leg-markers (red and bold) represents a single-copy gene in bean except for one marker (“Baja”) which is marked with #. Non-Leg-markers are shown in italics.

Figure 3.—Synteny between bean, Lotus, Medicago and Arachis defined by legume anchor markers. Synteny to Lotus japonicus (Lj) and Medicago truncatula (Mt) is given by the color code. The color scheme reflects the synteny observed between Mt and Lj. For example pronounced synteny has been found between LjLG2 and MtChr5 (CANNON et al. 2006) and accordingly they were given the same color. The gene content of LjLG1 is for the main part shared between MtChr3 and MtChr7 and accordingly they were given similar colors. Two markers sharing the same position in bean are indicated by a vertical bar, in the color window. The correspondence between Leg-marker positions in Lotus and the 10
large-scale synteny blocks described by Cannon et al. (2006) is indicated by dotted lines. The bean linkage groups (LG) are shown as previously presented in Figure 2. The locations of the 14 non-Leg-markers marked with * are only approximate and therefore no centiMorgan position is given.

Figure 4.—Detailed examples of macrosynteny between bean (Pv) and Lotus japonicus (Lj). Each vertical bar represents individual bean and Lotus linkage groups (LG). Marker names are shown on one side of each linkage group and centiMorgan values are given on the other side. Each of the Leg-markers (shown in bold) represents a single-copy gene in bean except for one marker which is marked with #. Non-Leg-markers are shown in italics. Lines connect anchor marker loci. LjLG5, LjLG3 and LjLG4 are shown in reverse orientation.

Figure 5.—Detailed examples of macrosynteny between bean (Pv) and Medicago truncatula (Mt). Vertical bars represent individual bean linkage groups (LG) or segments of Medicago pseudochromosomes (hatched). Marker positions are given in centiMorgan in case of bean and 1:100,000 bp in case of Medicago. Names of Leg-markers are shown in bold whereas non-Leg-markers are shown in italics. Lines connect anchor marker loci. MtChr4 and MtChr5 are shown in reverse orientation.

Figure 6.—Detailed examples of macrosynteny between bean (Pv), Lotus japonicus (Lj), Arachis (Ara) and Medicago truncatula (Mt). Vertical bars represent individual bean, Lotus and Arachis linkage groups (LG) or segments of Medicago pseudochromosomes (hatched).
Marker positions are given in centiMorgan in case of bean, Lotus and Arachis and 1:100,000 bp in case of Medicago. Names of Leg-markers are shown in bold whereas non-Leg-markers are shown in italics. Lines connect anchor marker loci. MtChr5, LjLG2, AraLG7 and AraLG2 are shown in reverse orientation. Only the upper part of AraLG2 is shown.
Leg primers from PriFi

- PCR
  - Check on gel
    - No product: Abandon
    - Single product: Sequence
      - Verify identity from exon sequences
        - Detect polymorphism in intron sequence
          - Develop marker
            - Genotype plants

- Optimize PCR

- Multiple products
  - Optimize PCR

Fig 1
Fig 3
Fig 4