Theor Appl Genet (2006) 113:225–238
DOI 10.1007/s00122-006-0288-0

ORIGINAL PAPER

The first gene-based map of *Lupinus angustifolius* L.-location of domestication genes and conserved synteny with *Medicago truncatula*

Matthew N. Nelson · Huyen T. T. Phan · Simon R. Ellwood · Paula M. Moolhuijzen · James Hane · Angela Williams · Clare E. O'Lone · John Fosu-Nyarko · Marie Scobie · Mehmet Cakir · Michael G. K. Jones · Matthew Bellgard · Michał Książkiewicz · Bogdan Wolko · Susan J. Barker · Richard P. Oliver · Wallace A. Cowling

Received: 21 December 2005 / Accepted: 31 March 2006 / Published online: 11 May 2006

Abstract We report the first gene-based linkage map of *Lupinus angustifolius* (narrow-leafed lupin) and its comparison to the partially sequenced genome of *Medicago truncatula*. The map comprises 382 loci in 20 major linkage groups, two triplets, three pairs and 11 unlinked loci and is 1,846 cM in length. The map was generated from the segregation of 163 RFLP markers, 135 gene-based PCR markers, 75 AFLP and 4 AFLP-derived SCAR markers in a mapping population of 93 recombinant inbred lines, derived from a cross between domesticated and wild-type parents. This enabled the mapping of five major genes controlling key domestication traits in *L. angustifolius*. Using marker sequence data, the *L. angustifolius* genetic map was compared to the partially completed *M. truncatula* genome sequence. We found evidence of conserved synteny in some regions of the genome despite the wide evolutionary distance between these legume species. We also found new evidence of widespread duplication within the *L. angustifolius* genome.

Electronic Supplementary Material Supplementary material is available for this article at http://dx.doi.org/10.1007/s00122-006-0288-0 and is accessible for authorised users.

Communicated by J. S. Heslop-Harrison

M. N. Nelson · C. E. O'Lone · S. J. Barker · W. A. Cowling
School of Plant Biology, Faculty of Natural and Agricultural Sciences, The University of Western Australia, Perth, WA 6009, Australia

H. T. T. Phan · S. R. Ellwood (✉) · J. Hane · A. Williams · R. P. Oliver
Australian Centre for Necrotrophic Fungal Pathogens, WA State Agricultural Biotechnology Centre, Division of Health Sciences, Murdoch University, Perth, WA 6150, Australia
e-mail: sellwood@murdoch.edu.au

P. M. Moolhuijzen · M. Bellgard
Centre for Bioinformatics and Biological Computing, Murdoch University, Perth, WA 6150, Australia

J. Fosu-Nyarko · M. Scobie · M. Cakir · M. G. K. Jones
WA State Agricultural Biotechnology Centre, School of Biological Sciences and Biotechnology, Division of Science and Engineering, Murdoch University, Perth, WA 6150, Australia

M. Książkiewicz · B. Wolko
Institute of Plant Genetics, Polish Academy of Sciences, 60-479 Poznan, Poland

Introduction

Narrow-leafed lupin (*Lupinus angustifolius* L.) is the most important grain legume crop in southern Australia. Up to 1.4 M tonnes of high protein grain are harvested annually for animal feed (Edwards and van Barneveld 1998) and human nutrition (Petterson 1998), and for major benefits in crop rotation with cereals (Nelson 1994). Interest is growing in Europe and North America for this crop as a useful protein crop (Bhardwaj 2002; Carruthers et al. 2000; Reeves 1991; Wilkins and Jones 2000).

The domestication history of narrow-leafed lupin is relatively short (Cowling et al. 1998) and has involved

the introduction of key domestication traits controlled by mutations at five or six loci. Most domestication alleles are recessive mutants (Cowling 1999). *Iucundis* (*Iuc*), controls alkaloid production and bitterness in wild lupins, but the recessive mutant *iuc* is exploited to produce "sweet" low alkaloid forms (Gladstones 1977). *Mollis* (*Moll*) controls water permeability of seed, "hard" seeds being important for long term survival of the species in the wild, but the recessive mutant *moll* is necessary to allow immediate germination upon sowing (Mikolajczyk 1966). Two genes are known to be responsible for pod shattering, *Tardus* (*Ta*) and *Lentus* (*Le*), and the additive effect of the recessive mutants *ta* and *le* prevents pod shattering at harvest (Gladstones 1967). The *le* allele has a pleiotropic effect of deposition of a red pigment in the cortex at the base of the stem in green plants, and a red pigment inside dry mature pod walls and is therefore more easily identified than the *ta* gene (Gladstones 1967). Early flowering is promoted by the dominant mutant allele *Ku*, which is important for adaptation to short growing seasons in Australia (Gladstones 1977). *Leucospermus* (*Leuc*) controls pigment production in seeds, cotyledons, and flowers and the recessive mutant *leuc* is used to differentiate the domesticated crop by its white flowers and seeds from the bitter, blue-flowered, dark-seeded wild populations which may grow in the same region (Gladstones 1977).

L. angustifolius is a member of the Genistoid clade of the Fabaceae, which is believed to have diverged from the other crop legume species about 56.4 million years ago (Lavin et al. 2005). *L. angustifolius* is diploid with $2n = 40$ chromosomes and an estimated DNA content of 2C = 1.89–2.07 pg (Hajdera et al. 2003; Naganowska et al. 2003). In these respects, *L. angustifolius* is similar to the soybean genome that shares the same number of chromosomes and approximately the same DNA content (Bennett et al. 1982). Furthermore, like soybean, *L. angustifolius* is thought to have undergone polyploidy in the past, most likely early in the formation of *Lupinus* genus (Atkins et al. 1998; Gupta et al. 1996; Naganowska et al. 2003; Shoemaker et al. 1996; Wolko and Weeden 1989; Yan et al. 2004). The chromosomes of *L. angustifolius* are small and morphologically uniform making classical cytogenetic analysis difficult for this species; however, fluorescent in situ hybridisation has been used to differentiate 5 of 20 chromosome pairs (Hajdera et al. 2003; Naganowska and Kaczmarek 2005; Naganowska and Zielinska 2002, 2004). Previous genetic linkage maps of *L. angustifolius* were based primarily on anonymous dominant genetic markers such as amplified fragment length polymorphism (AFLP), random amplified polymorphic DNA (RAPD) and microsatellite-anchored fragment length polymorphism (MFLP) markers (Boersma et al. 2005; Brien et al. 1999; Wolko and Weeden 1994). Such maps are less useful than gene-based maps because they are cultivar-specific and offer no opportunity to compare the genome of *L. angustifolius* with those of other legume species. A valuable resource for lupin genetic research and legume evolutionary studies would therefore be a genetic map of *L. angustifolius* made using codominant markers transferable between lupin crosses and across the legume family.

Comparative genetic mapping is well established in many families of crop species, most notably for cereals and crucifers (Devos 2005; Lagercrantz and Lydiate 1996; Moore et al. 1995) but also more recently for legumes (Choi et al. 2004b; Yan et al. 2004). Classically, comparative mapping has involved the use of the same set of genomic or cDNA probes as restriction fragment length polymorphism (RFLP) markers in related species to identify regions of marker loci that show conservation of linkage (i.e. synteny) and order (i.e. collinearity: Lagercrantz and Lydiate 1996; Moore et al. 1995). More recently, RFLP maps of less well-characterised species (such as most crop species) have been aligned to the genome sequence of model species using the sequences of the RFLP probes (Lukens et al. 2003). A third comparative mapping approach is to use polymerase chain reaction (PCR) markers designed for use across related species. In this approach, oligonucleotide primers are designed to conserve DNA sequences such as gene exons, that encompass polymorphic regions such as introns or microsatellites (Lyons et al. 1997; Varshney et al. 2005).

In this study, we used a variety of sequenced RFLP markers, PCR-based, cross-species markers and AFLP markers to develop the first gene-based map of *L. angustifolius* including key domestication traits. This map was then used to assess the extent of conserved synteny between *L. angustifolius* and the model legume *M. truncatula*.

Materials and methods

Genetic population development

The seed parent was the breeding line 83A:476, a sister line of *L. angustifolius* cv. Wonga (Cowling 1999), and the pollen parent was *L. angustifolius*, accession number P27255, originally collected from a natural population in Morocco (Gladstones and Crosbie 1978). The breeding system of *L. angustifolius* is highly self-pollinating, and the F_1 is fully self-fertile. Both 83A:476 (an F_5-derived

line) and P27255 were expected to be homozygous at most loci. The parents showed different phenotypes at the six key domestication loci used in the breeding program in Australia (Cowling 1999), and P27255 was blue-flowered, late, bitter, hard-seeded, with shattering pods and a non-pigmented cortex at the base of the stem (alleles *Leuc*, *ku*, *Iuc*, *Moll*, *Ta* and *Le*).

An F_1 individual (number 97L380) from this cross was grown over summer in 1997–1998 in a glasshouse, and 141 F_2 seeds harvested, scarified by hand and sown in a quarantine mesh field house at South Perth, Western Australia, in June 1998. F_2 plants were assessed for cotyledon and flower colour (pigmented types classified as *Leuc/-* and white types as *leuc/leuc*), flowering date (early types classified as *Ku/-*, and late types as *ku/ku*), presences of alkaloids (bitter types classified as *Iuc/-*, and sweet types as *iuc/iuc*), pigmented pod walls (white pod walls classified as *Le/-* and pigmented pod walls classified as *le/le*). The pod shattering gene, *Ta*, could not be reliably genotyped in the F2/F3 generations. Of the 136 numbered F_2 plants that emerged, F_3 seed was harvested from 121 survivors.

Up to 40 seeds from each F_2-derived F_3 family were assessed for swelling ability in water (non-permeable types were classified as *Moll/-*, and permeable types as *moll/moll*), and sown (after scarification if necessary) in a quarantine mesh field house at Shenton Park, Western Australia, in June 1999. F_3 family rows were again assessed for segregation of alleles at domestication loci *Leuc*, *Ku*, *Iuc* and *Le* for comparison with the previous year's results. F_4 seed was harvested in bulk from 105 surviving rows, and tested for swelling ability (*Moll*) to confirm the previous year's results.

A single seed was taken at random from the F_4 bulk of each F_2-derived family to commence single seed descent to the F_8 to produce recombinant inbred lines (RIL). In total, 93 F_8-derived RIL progeny were developed from the original 136 F_2 plants. F_8 seeds (one per RIL) were sown in a quarantine mesh field house at Shenton Park, Western Australia, in June 2003 and plants were assessed for domestication traits as described previously. F_8-derived F_9 seed was harvested from each F_8 plant, and grown in rows in 2004 to confirm domestication trait genotypes assessed in 2003 and to provide leaf tissue for DNA sampling. As in the F2/F3 generations, the *Ta* trait could not be reliably assessed in the F8/F9 generations.

RFLP markers

The DNA extraction, restriction enzyme digestion, gel electrophoresis, alkaline transfer and Southern blot hybridisation were carried out according to the methods of Sharpe et al. (1995), except that *Eco*RI-digested lambda DNA was used as a size marker and positive hybridisation control. PCR inserts from a total of 199 partial (3'-end) cDNA clones from a developing seed library of *L. angustifolius* (Merrit cultivar) and 87 soybean genomic clones (Biogenetic Services Inc., SD, USA) were radiolabelled using ^{32}P-dCTP, hybridised to Southern blots of DNA from 83A:476, P27255 and a soybean DNA control using moderate stringency (1× SSC, 65°C) and visualised on Super RX film (Fuji Hanimex, Perth, Australia). The DNA samples were digested separately with each of the three restriction enzymes (*Eco*RI, *Eco*RV and *Hin*dIII). Markers showing polymorphism were then hybridised to *Eco*RI, *Eco*RV or *Hin*dIII-digested DNA of the mapping population. Plasmids containing the lupin cDNA inserts used in this study were sequenced using BigDye v3.1 Terminator Cycle Sequencing Ready Reaction Mix (Applied Biosystems [AB], Foster City, California) and using an AB Prism 3730 DNA Sequencer. The resulting DNA sequences were edited using Vector NTI (Invitrogen, Carlsbad, California). These lupin cDNA sequences along with soybean genomic clone sequences obtained from the National Center for Biotechnology Information (NCBI) genomic survey sequence database were queried in September 2005 against the NCBI non-redundant database using BLASTn and BLASTx with a threshold expected value $< 1\times e^{-20}$ in order to predict the function of the gene segments in the cDNA and genomic clones.

Design of gene-based PCR markers

Lupinus spp.-derived cross-species markers

Intron targeted amplified polymorphic sequence (ITAP) primers were designed to anneal to conserved adjacent exon sequences. Using the bioinformatics resource at LegumeDB (Moolhuijzen et al. 2006), *Lupinus* spp. EST sequences available from the NCBI dbEST in September 2004 were individually queried against *M. truncatula* EST subsets of NCBI databases. ESTs with hits to *M. truncatula* genes at significant values $< e^{-20}$ were retained. Putative orthologous pairs were identified if the reciprocal best fit of the *M. truncatula* gene was the same *Lupinus* EST. ITAP oligonucleotide primers were designed to anneal in exon sequences conserved between *M. truncatula* and *Lupinus* spp., and to amplify across intron regions predicted by GlimmerM (Majoros et al. 2003). *L. angustifolius* parental DNA of the mapping population was then amplified with each primer pair and purified products

were sequenced directly as described above. DNA polymorphisms were identified by aligning parental sequences in Vector NTI.

Lupinus spp.-derived SSR markers

Lupinus spp. ESTs from NCBI (September 2004) that were not developed into ITAP markers were screened for microsatellite (or Simple Sequence Repeat, SSR) motifs. Primers flanking the SSR motifs were then designed to anneal and conserve DNA sequence flanking SSR motifs as described for ITAP markers. Parental *L. angustifolius* PCR amplicons were assessed for length polymorphism by polyacrylamide gel electrophoresis. Markers that did not show length polymorphism and gave single bands were then direct-sequenced in order to identify sequence polymorphism as for the ITAP markers described above.

M. truncatula cross-species markers

One hundred and forty-four *M. truncatula* PCR markers designed for cross-legume use (Choi et al. 2004a; b, and H.-K. Choi, personal communication 2003) were tested for PCR amplification in *L. angustifolius*. Temperature gradient PCR was used to identify the optimum annealing temperature for each of the markers in *L. angustifolius*, and for *M. truncatula* positive controls. Markers that amplified single bands in *L. angustifolius* were used to amplify the parents of the mapping population, which were sequenced and analysed for polymorphism as described for the ITAP markers above.

M. truncatula SSR markers

A set of 360 primer pairs was selected from *M. truncatula* genomic sequences containing short repetitive motifs in the intron (information on primer sequences received from Professor Thierry Huguet UMR CNRS/INRA, France). PCR conditions were optimised on parental lines of the mapping population by the teNSmperature gradient or standard touch-down PCR methods.

Amplification and allele resolution of gene-based PCR markers

Genomic DNA from the parents and 93 RILs were used as templates for PCR marker amplification using the following basic PCR protocol with minor variations: 50–100 ng of genomic DNA template, 1× PCR reaction buffer, 2 mM $MgCl_2$, 0.25 mM of each dNTP, 10 pmol of each primer and 1 unit of Taq DNA polymerase. The thermocycling conditions (with minor variations) were: 5 min initial denaturation step followed by 35 cycles of 94°C for 30 s, marker-specific annealing temperatures for 30 s and 72°C for 60 s, then a final extension step of 7 min at 72°C. Where restriction enzymes recognising differences in DNA sequence were available, markers were run as cleaved amplified polymorphic sequence (CAPS: Konieczny and Ausubel 1993). SNPs for which no restriction enzyme was available were detected using the SNaPshot kit (AB) and analysed on an AB3730 capillary sequencer. Large fragment size polymorphisms and CAPS markers were resolved by agarose gel electrophoresis and visualised using ethidium bromide staining. Small fragment size polymorphisms were resolved by native polyacrylamide gel electrophoresis or by AB3730 capillary sequencer using fluorescently labelled primers and GeneScan^{TM} −500 LIZ® Size Standard (AB).

AFLP markers

AFLP markers were generated as described by Vos et al. (1995). Genomic DNA samples were digested using *Mse*I and *Eco*RI restriction endonucleases, adapters ligated (Invitrogen Australia Pvt Ltd, Mt Waverley, Australia) followed by pre-amplification of 5 ng of 10-fold diluted, restricted/ligated template DNA. Pre-selective amplification was carried out using 12.5 pmol of each primer pair (*Eco*RI-A: GACTGCG TACCAATT CA and *Mse*I-C: GATGAGTCCTGAG TAAC) combined in a PCR Master Mix^{TM} (Promega Corp, Madison, WI) or in a reaction mixture containing 1 U of Taq DNA polymerase, 200 mM of each dNTP and 2.5 mM $MgCl_2$. Two thermal cycling profiles were used: the first was 94°C for 5 min, followed by 30 cycles at 94°C for 30 s, 56°C for 60 s and 72°C for 60 s, then a final extension for 10 min at 72°C. The second profile was 72°C for 2 min, 20 cycles of 60 s at 94°C, 30 s at 56°C and 2 min at 72°C, then held for 30 min at 60°C. Selective amplification was carried out using two selective nucleotides added to the 3' ends of the pre-selective primers. The *Mse*I selective primers were: M47 (*Mse*I-C+AA), M48 (*Mse*I-C+AC), M49 (*Mse*I-C+AT), M59 (*Mse*I-C+AA), M60 (*Mse*I-C+AA), M61 (*Mse*I-C+AA) and M62 (*Mse*I-C+AA). The *Eco*RI selective primers were labelled with fluorescent dyes: E32 (*Eco*RI-A+AC), E35 (*Eco*RI-A+CA), E37 (*Eco*RI-A+CG), E38 (*Eco*RI-A+CT) and E41 (*Eco*RI-A+GG). The combinations of *Mse*I and *Eco*RI selective primers were chosen on the basis of polymorphisms found during AFLP mapping in a subset of the F_2 population (Brien et al. 1999). Selective amplification reactions contained 5 µl of 20-fold diluted pre-amplified

template DNA, 1 pmol *Eco*RI selective primer and 5 pmol *Mse*I selective primer in a PCR reaction mixture (as above) followed by standard touch down PCR conditions at 30 cycles of 94°C for 60 s, 65↓56°C for 30 s, 72°C for 2 min and 60°C for 30 min. The AFLP products were resolved with an AB3730 DNA Sequencer using a GeneScan^{TM}-500 LIZ® Size Standard (AB). Data were analysed using the method of Rinehart (2004) and by the automated allele calling parameters in GeneMapper® (AB). Three sequence-characterised amplified region (SCAR) markers were developed using standard cloning and sequencing procedures from AFLP bands identified in the F_2 generation to be linked to domestication traits (unpublished data). An additional SCAR marker previously reported by Brien et al. (1999) was also used in this study.

Linkage mapping

The mapping was conducted in two stages with the initial formation of a framework map using gene-based, codominant markers, followed by the addition of AFLP loci to add density. Scoring data from each marker were analysed with the aid of MapManager QTX version b20 (Manly et al. 2001). The most likely order of gene-based markers was checked by eye and with the help of the RIPPLE function. The marker scoring data were proof-read by re-scoring markers with particular attention to double recombinants which often indicate erroneous scoring. AFLP markers were tested separately for normal segregation ratios using the χ^2 test (one degree of freedom) and only those markers with $P > 0.01$ were incorporated into the framework map using the DISTRIBUTE function at a higher significance threshold (LOD4).

Comparative mapping with *M. truncatula*

M. truncatula pseudochromosomes 1–8 were constructed from phases 1, 2 and 3 BAC sequences (http://mtgenome.ucdavis.edu) and positioned on the *M. truncatula* genetic map using marker, fingerprinting (using FingerPrint Contig software) and overlap information. Markers mapped in *L. angustifolius* were located on the *M. truncatula* map by aligning marker sequences with the *M. truncatula* pseudochromosomes using BLASTN with an expected value $< 1e^{-20}$, hsp identity $\geq 60\%$ and hsp length > 50 nt. The results were displayed using CMap (http://www.gmod.org/cmap/), and GridMap 3.0 (http://cbr.jic.ac.uk/dicks/software/Grid_Map/) programs. A Perl script provided by Martin Trick (personal communication 2004; martin.trick@bbsrc.ac.uk) and described by Nelson and Lydiate (2006) was modified to parse the data output from CMap and previously parsed BLAST results into a format suitable for GridMap. The *M. truncatula* genetic map (Choi et al. 2004a) was referred to for the approximate positions of markers that had been genetically mapped in *M. truncatula* but not yet positioned on the physical map.

Results

Scoring of domestication traits

The domestication trait phenotypes segregated in the ratios expected for single gene traits in the F_2 and F_8-derived RILs with the exception *Moll*, which had an excess of hard-seeded lines over soft-seeded lines in both the F_2 and RIL populations and *Iuc*, which had an excess of bitter over sweet types in the RIL population (Table 1). There were loss of plants in this population at the F_2 stage (due to insect damage) and in the F_2-derived F_3 plots (due to herbicide damage). Hence, the number of RILs recovered in the F_8 was only 93 compared with a possible 141 in the F_2. Double recessive progeny for each domestication gene in the F_2, as confirmed in F_3 and F_4 families, gave rise as expected to uniform double recessive F_8-derived RIL progeny with the exception of one line (97L380-118) which was bitter (*Iuc*) in the RILs and uniformly sweet (*iuc/iuc*) in the F_2 and F_2-derived families.

Genetic marker development

Three strategies were used to generate genetic markers: (1) RFLP markers using lupin cDNA and soybean genomic clones, (2) SNP and SSR polymorphisms in genic regions assayed by PCR, and (3) AFLP markers. RFLP and gene-based markers were used as core markers to form linkage groups and AFLP loci were used to add density to less-populated genomic regions. Table 2 summarises the efficiency of each marker type in terms of the number mapped in *L. angustifolius* and the usefulness of mapped markers for comparative mapping to *M. truncatula*.

RFLP markers

Of the 199 lupin cDNA probes tested, 99 detected at least one polymorphic locus between the parents of the mapping population. The GenBank accession numbers for each of the polymorphic markers, along with putative gene functions are provided in S1. The 99 polymorphic

Table 1 Segregation of domestication genes in the F_2 and F_8-derived RIL progeny

Locus	Leucospermus		Iucundis		Mollis		Ku		Lentus	
Genotype	*Leuc/-*	*leuc/ leuc*	*Iuc/-*	*iuc/ iuc*	*Moll/-*	*moll/ moll*	*Ku/-*	*ku/ ku*	*Le/-*	le/le
F_2	87	26	86	25	95	18	77	23	85	20
χ^2 (3:1)	0.239 NS		0.363 NS		4.959*		0.172 NS		1.984 NS	
F_8 RILs	53	41	59	35	58	36	42	50	52	41
χ^2 (1:1)	1.532 NS		6.128*		5.149*		0.696 NS		1.301 NS	

F_2 data are presented only for those lines whose genotypes were confirmed in F_2-derived F_3 and F_4 families

NS Not significant ($P > 0.05$)

*Significant ($P < 0.05$)

Table 2 Summary of markers used for mapping in *L. angustifolius* and comparative mapping with *M. truncatula*

Marker type	Screened	Loci mapped in *L. angustifolius*[a]	Loci used for comparative mapping[b]
Lupinus RFLP	199	139 (70%)	45 (32%)
Soy RFLP	87	24 (28%)	11 (46%)
Lupinus ITAP	280	85 (30%)	63 (74%)
Lupinus SSR	28	14 (50%)	4 (19%)
M. truncatula SSR	360	13 (4%)	7 (54%)
M. truncatula cross-species	144	23 (16%)	17 (74%)
AFLP	276	75 (27%)	0 (0%)
AFLP-derived SCAR	–	4	0 (–)
Total	1,374	377	147

[a] Percentage figures in parentheses are the proportion of mapped loci compared to the number of markers screened

[b] Percentage figures in parentheses are the proportion of mapped loci that were used for comparative mapping

probe-enzyme combinations detected 139 loci in the mapping population: an average of 1.4 loci per probe. Most lupin probes detected additional monomorphic bands, many of which probably represented additional loci that could not be mapped in this population.

Of 87 soybean genomic RFLP probes screened, 77 hybridised to lupin DNA and the remaining 10 hybridised only to the soybean DNA control. Of the 77 soybean probes that hybridised to lupin DNA, 14 detected 24 polymorphic loci in the mapping population: an average of 1.7 loci per probe. As with the lupin RFLP probes, most soybean probes also detected additional monomorphic bands that could not be mapped in this population. The GenBank accession numbers for each of the polymorphic markers, along with putative gene functions (when predicted coding sequences were present) are provided in S2.

Gene-based PCR markers

Table S3 provides detailed information on each polymorphic, gene-based PCR marker used to develop the genetic map. From 2,492 *Lupinus* EST sequences screened, 280 ITAP primer pairs were designed in conserved exons and spanning introns. Of these 280 primer pairs, 230 amplified single bands in *L. angustifolius*, while 28 amplified more than one band. Single amplicons were sequenced directly, resulting in 75 SNPs that were used to genotype the mapping population using CAPS and SNP assays (S3). Nine primer pairs detected length polymorphisms and one a dominant polymorphism. In total, 84 ITAPs were used to genotype the mapping population resulting in 85 mapped loci.

From the remaining 2,192 *Lupinus* EST sequences, 28 with microsatellite repeat motifs were selected as template sequences for designing primer pairs located in conserved exon regions and spanning the microsatellite repeat sequences. The *Lupinus*-derived SSR markers were highly efficient at detecting polymorphisms in the lupin cross with 14 loci mapped from 28 primer pairs tested (Table 2). Although most of these polymorphisms arose from variable numbers of microsatellite motif repeats, three of the markers detected SNPs that were assayed as CAPS markers (S3).

The *M. truncatula* cross-species markers were relatively inefficient in detecting polymorphisms in the lupin parents: 144 primer pairs tested resulted in only 23 mapped loci (16%; Table 2). This was due to a combination of a relatively low amplification success rate of these markers in *L. angustifolius* (87 markers produced ≥1 band on an agarose gel, 60% of the total), inability to directly sequence PCR products due to the amplification of > 1 locus (25 markers, 17%), and sequence monomorphism (38 markers, 26%).

The *M. truncatula* SSR markers showed relatively high amplification efficiency with 264 markers out of

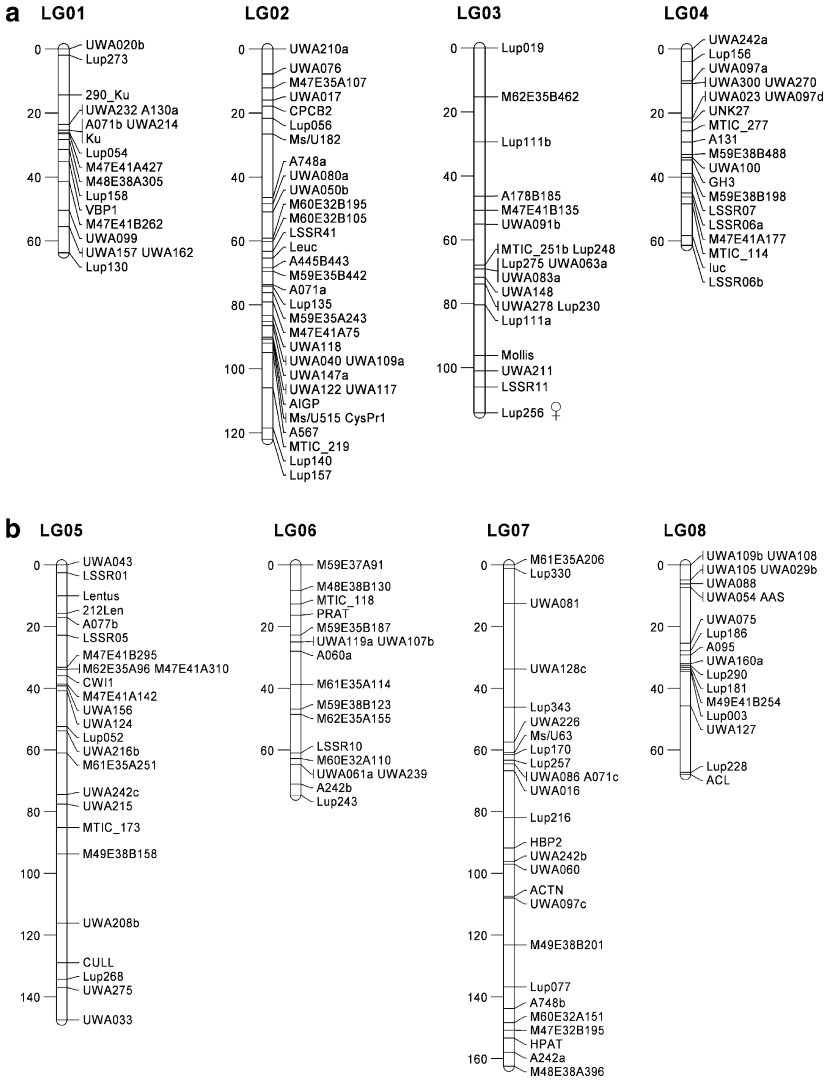


Fig. 1 A genetic linkage map of *L. angustifolius* based on 382 loci in 20 major linkage groups (LG01–LG20) plus 2 triplets, 3 pairs and 11 unlinked loci. Each *vertical bar* represents an individual linkage group drawn to scale with Kosambi centiMorgan values on the left side at regular intervals. Locus names are shown on the right side of each linkage group with *connecting lines* indicating the position of each locus on the linkage group

360 primer pairs tested (73%; Table 2), amplifying one or more DNA products. However, the proportion of markers showing visible band size polymorphisms on agarose gels was very low with just 12 primer pairs (4%) detecting 13 length polymorphisms.

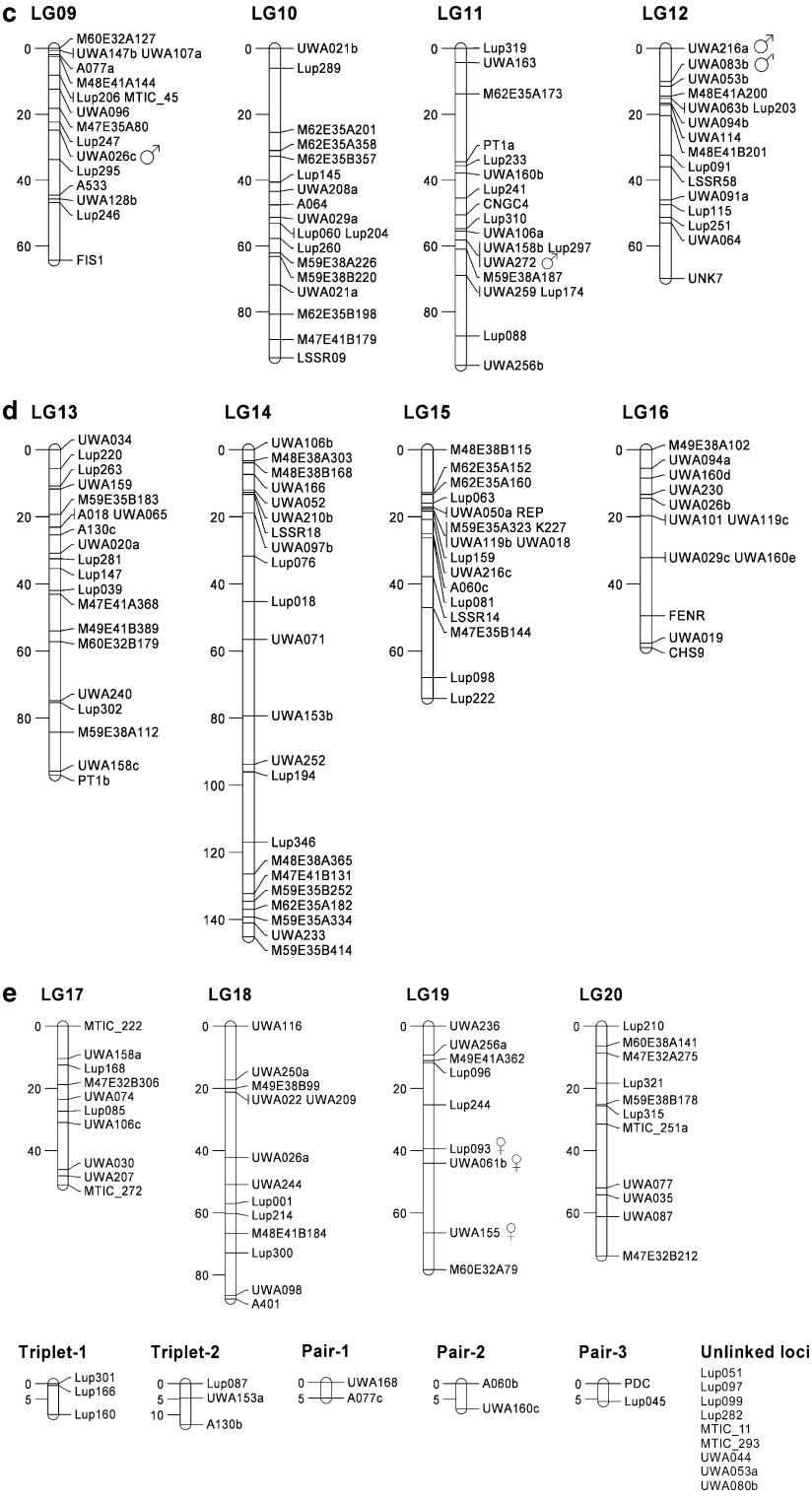
AFLP markers

From 15 selective primer combinations, 276 bands from either the female or the male parent appeared to segregate in a dominant fashion (i.e. presence or absence) in the mapping population. Of these segregating bands, 235 had segregation ratios in the mapping population that conformed to the expected 1:1 ratio of parental alleles at $\chi^2 < 6.635$ ($P > 0.01$), and these were selected for further analysis in the mapping procedure described below.

Genetic map construction

An initial framework map of 23 major linkage groups was developed based on 298 gene-based marker loci, 5 domestication trait scores and 4 AFLP-derived SCAR markers. This number of linkage groups was greater than the number expected for *L. angustifolius*, which has a haploid chromosome number of 20. Therefore, additional AFLP markers were placed on the framework map to add density and length to the linkage groups. Of the 235 AFLP bands that segregated in a normal Mendelian fashion, 75 AFLP loci appeared to have low error rates as inferred from the presence of few unexpected double recombinants when integrated into the framework map. These 75 AFLP loci were retained and added to the other markers for a second round of linkage mapping. With these additional data, 20 major linkage groups were

Fig. 1 (Contd.)



resolved matching the known haploid chromosome complement for this species. Figure 1 presents the genetic map for *L. angustifolius* comprising 359 loci in 20 major linkage groups, plus 2 triplets, 3 pairs and 11 unlinked loci (382 loci in total). The map length was 1,846 cM with a mean interval size of 5.3 cM. The major linkage groups varied in length from 51 cM (LG17) to 162.5 cM (LG07; Fig. 1).

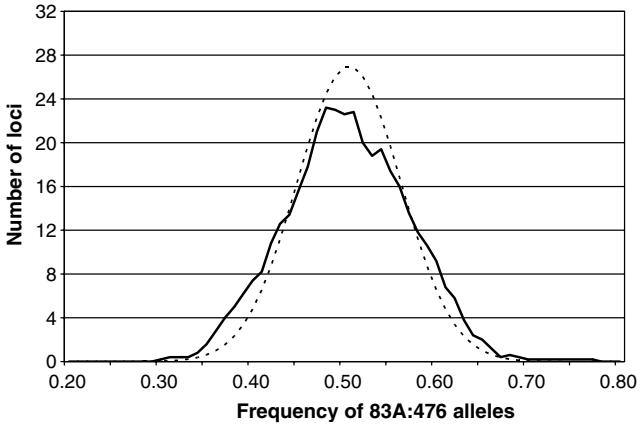


Fig. 2 The distribution of allele frequencies for 382 loci scored in 93 recombinant inbred lines. The graph was plotted using 5% windows taken at 1% intervals. The observed distribution (*solid line*) of 83A:476 (female parent) was compared with the expected binomial distribution (*dashed line*), with $P = 0.5$ and the average number of 84 individuals scored at each locus

Allele frequencies in the RIL population

Both parents of the mapping population appeared to be completely homozygous at all codominant marker loci, so that the interpretation of allele segregation ratios was straightforward. Taking all 382 loci together, the overall allele transmission frequency was 0.499:0.501 (83A:476:P27255 alleles). The allele frequency was calculated at each of the 382 loci to evaluate the degree of deviation from the expected 0.5 transmission frequency for each allele at each locus. The observed frequency distribution for allele frequencies in the 93 RILs was compared with the expected binomial distributions (Fig. 2). There was in general a close correspondence between the observed and expected distributions. However, there were eight loci located on linkage groups LG03, LG09, LG11, LG12 and LG19 whose allele frequencies were significantly skewed ($P < 0.01$) towards one of the parents, and these are indicated by male and female icons on Fig. 1.

Genome duplications within the lupin genome

As described above, there was evidence of duplication in the *L. angustifolius* genome with lupin and soybean RFLP probes detecting an average of 1.4 and 1.7 polymorphic loci per probe, plus additional putative monomorphic loci. Four PCR-based markers (LSSR06, Lup111, MTIC251 and PT1) also detected duplicate loci (S3, and Fig. 1). When the distribution of these duplicate loci was investigated, 68 duplicate pairs of loci were observed between different linkage groups and 5 markers detected duplications within the same

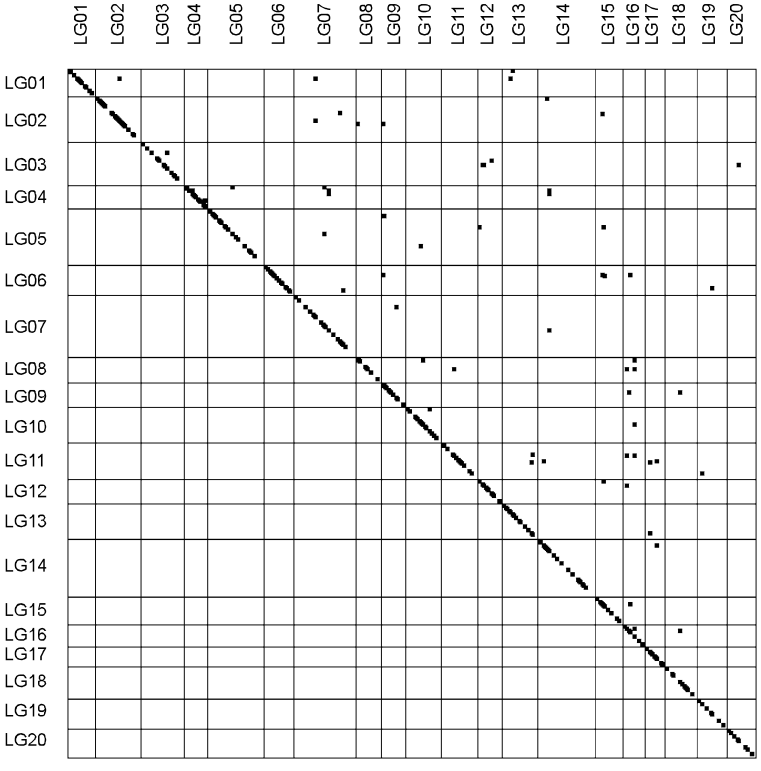


Fig. 3 Duplication within the *L. angustifolius* genome. Twenty linkage groups (LG01–LG20) are plotted on each axis. *Each dot* represents a correspondence between linkage groups. The *dense line* at 45° represents self-correspondences. *Other dots* to the top right hand side represent non-self correspondences between duplicate loci. The reciprocal correspondences to the bottom left of the diagram have been omitted for simplicity

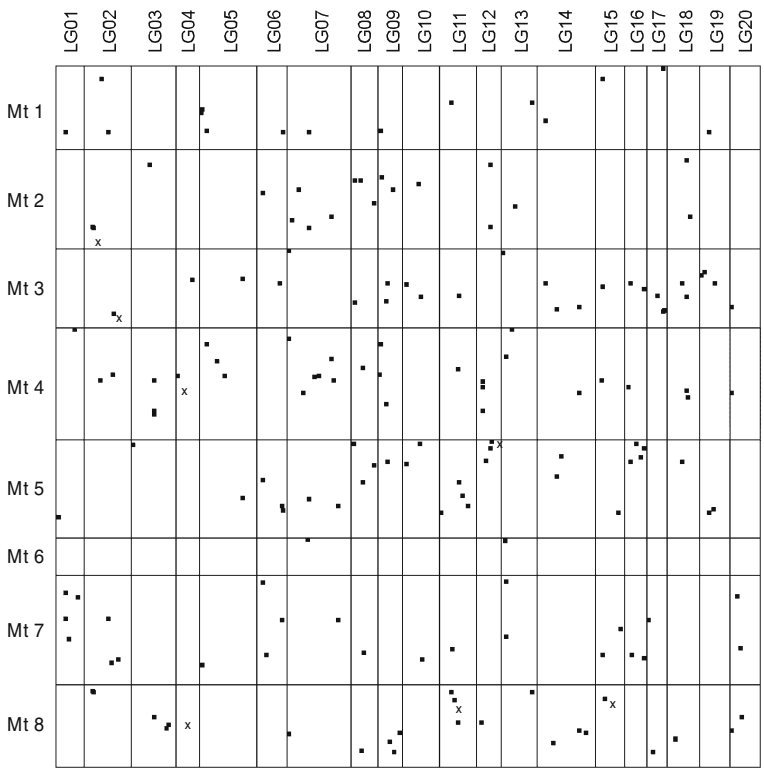


Fig. 4 A global overview of homology between the *L. angustifolius* and *M. truncatula* genomes. *Dots* indicate where mapped markers from *L. angustifolius* found homologous sequences in *M. truncatula*. *Crosses* denote the approximate positions of loci with known genetic map locations in *M. truncatula* but not yet within the sequenced portion of the *M. truncatula* genome

linkage groups (Fig. 3). There was no clear pattern in the distribution of these putative genome duplications.

Comparative mapping with *M. truncatula*

The narrow-leafed lupin genetic map was compared with the most up-to-date version of *M. truncatula* pseudochromosomes in order to assess the level of conserved synteny. Table 2 includes a summary of the efficiency of each marker type in identifying regions of homology in the *M. truncatula* genome map. Of the gene-based markers, the most efficient marker types were *Lupinus* ITAP and *M. truncatula* cross-species markers where 74% of the loci mapped in *L. angustifolius* could be used for comparative mapping (Table 2). Six RFLP probes (UWA023, UWA097, UWA158, UWA160, UWA270 and UWA300) and one LSSR marker (LSSR18) encoded histone or cysteine proteinase genes that had multiple homologues in the *M. truncatula* genome sequence. These markers were deemed unsuitable for comparative mapping and were therefore excluded from the comparative analysis.

Using this homology search approach, 140 *L. angustifolius* marker loci found 184 homologous loci in *M. truncatula*. Seven *M. truncatula* cross-species markers whose physical locations on the *M. truncatula* genome had not yet been identified but which had already been genetically mapped in *M. truncatula* (Choi et al. 2004a) were placed manually in approximate positions in the *M. truncatula* pseudochromosomes. Figure 4 provides a global view of similarity between the lupin and *M. truncatula* genomes based on 181 homologous loci. We considered a region to have conserved synteny when there were at least three homologous loci mapped in both species. Of the 20 *L. angustifolius* linkage groups, 14 appeared to have regions of conserved synteny with the *M. truncatula*. Of the eight *M. truncatula* pseudochromosomes, only Chromosome 6 had no apparent syntenic regions in *L. angustifolius*. In total, there were 94 markers that fell into the regions of 3 or more conserved markers. Examples of conserved syntenic blocks are detailed in Fig. 5. *Ku*, *Le* and *Moll* domestication genes fell within three such conserved blocks, while *Leuc* was adjacent to another conserved block (Fig. 5). The whole of LG04, including the part harbouring *Iuc*, did not fall within any syntenic blocks. The remaining 87 homologous loci did not fall into conserved homologous blocks.

Discussion

We present the first gene-based map of *L. angustifolius*. A total of 359 predominantly codominant markers

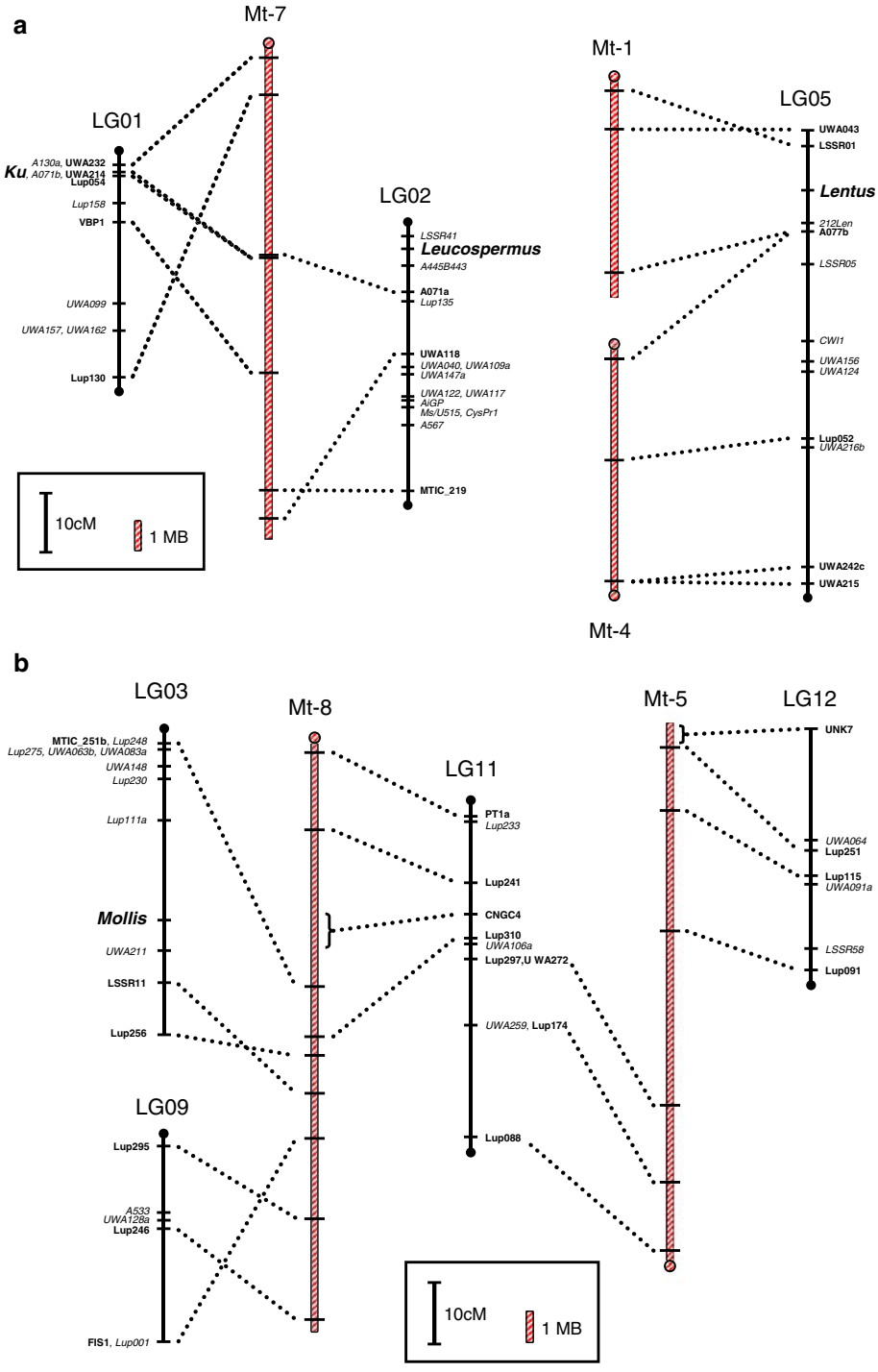


Fig. 5 Detailed examples of conserved synteny between *L. angustifolius* and *M. truncatula*. *L. angustifolius* linkage groups (LG) are represented by *black solid vertical bars* drawn to Kosambi centiMorgan (cM) scale. *M. truncatula* pseudochromosomes (Mt) are represented by *red hatched vertical bars* drawn to mega-base (MB) scale. Homologous loci are shown in *bold* and are connected by *dotted lines* between the *L. angustifolius* linkage groups and *M. truncatula* pseudochromosomes. *L. angustifolius* loci that do not fall within the conserved syntenic blocks are *italicised*. *Circles* at the ends of pseudochromosomes and linkage groups indicate that further DNA sequence/loci extend beyond the end of the syntenic blocks. Domestication gene loci are shown in *large, italicised, bold font*. Loci that have been genetically mapped in *M. truncatula* but not yet within the sequenced potion of the genome are shown in *parentheses* at their approximate locations in the *M. truncatula* pseudochromosomes. Linkage group LG12 is shown in the reverse orientation from Fig. 1. AFLP loci and gene-based markers that detected multiple homologues in *M. truncatula* have been omitted from this figure

resolved into 20 well-supported linkage groups, this matching the cytogenic chromosome number. There were 23 additional loci that either formed triplets or pairs or were unlinked; these loci may join onto the ends of smaller linkage groups as additional markers are screened in this mapping population. The mapping was achieved using an F_8 RIL population that has proved to be well-balanced in terms of allele segregation and was moderately polymorphic at the marker loci assayed. This genetically characterised mapping population now provides an excellent resource for the lupin genetic community.

A suite of new PCR-based markers designed for cross-legume utility are described. These should be useful for genetic research across all legumes and in particular for developing genetic maps of the other

agriculturally significant lupin species, *L. albus* and *L. luteus* (white and yellow lupin).

We positioned five domestication genes that differentiate domestic from wild forms of *L. angustifolius* on this linkage map. Segregation distortion of two genes, *Moll* and *Iuc*, in the RILs did not disturb mapping of these loci. Two RFLP markers (A071b and UWA214) cosegregated with *Ku* on LG01, the early flowering time gene (Fig. 1). After conversion to SCAR markers, these would enable marker-assisted introgression of new germplasm from wild accessions into domesticated lines. *Leuc* mapped to LG02 in a 5 cM interval defined by flanking microsatellite marker LSSR41 and SCAR marker A445B443 (Fig. 1); these simple PCR markers are suitable for marker-assisted selection of *Leuc* without any further development. For *Moll* (LG03), *Iuc* (LG04) and *Le* (LG05) genes, more closely linked markers are required for marker-assisted selection to be practical in breeding programmes. For *Moll* and *Le* genes, this could feasibly be achieved using a candidate gene or marker approach since these genes fall within, or are immediately adjacent to, conserved syntenic blocks (Fig. 5). Additional gene-based markers are required for this approach to be feasible with *Iuc*, as linkage group 4 currently has few markers that corresponded to orthologous loci in *M. truncatula* (Fig. 4).

This is the first genome-wide study to assess the level of conserved synteny between *L. angustifolius* and *M. truncatula*. As expected from the wide evolutionary distance and differing chromosome numbers between these species, there were numerous breaks in synteny between the two genomes. However, there were also striking examples of conserved synteny that could form the basis for marker-assisted selection in lupin and be used to direct the fine-mapping of important genes located in conserved regions. Interestingly, Chromosome 6 of *M. truncatula* showed very little homology to any of the markers used in this study, which was also previously found in a study between *M. truncatula* and other more closely related legume species (Choi et al. 2004b). Given that the majority of comparative markers used in this present study were newly developed adds further strength to the assertion that Chromosome 6 of *M. truncatula* is sparsely populated by conserved, transcribed genes (Choi et al. 2004a).

RFLP markers provided strong evidence of widespread genome duplication in *L. angustifolius*, possibly arising from a polyploidisation event. If that was the case, the polyploidy event must have been quite ancient because there was little evidence of residual structure in the distribution of duplicated loci, which is generally apparent in more recent polyploidy events such as those seen in soybean or in crucifer species (Nelson and Lydiate 2006; Shoemaker et al. 1996). These findings are supported by isozyme and cytological marker analyses, which suggested that there was a polyploidy event early in the formation of the *Lupinus* genus (Naganowska et al. 2003; Wolko and Weeden 1994). However, we cannot rule out other duplication mechanisms such as segmental duplications or chromosome additions from related species. Developing genetic maps in other lupin species using the same RFLP markers would help clarify the nature of the duplications. Pfeil et al. (2005) have assessed polyploidy in relation to the divergence between the *Glycine max* and *Medicago truncatula* lineages and infer a full-genome polyploidy event in the ancestral lineage of both species. Our research provides a starting point for further examination of polyploidy events early in evolution of the legumes. The development and cross-comparison of genome models in each of the legume crown clades promises significant insight to the molecular events that shaped the phylogeny of this extremely diverse and successful plant family.

Acknowledgements The mapping population originated from research begun by WAC during his employment at the Department of Agriculture Western Australia. Subsequent development of RILs, phenotyping, mapping of markers and comparative mapping was supported by Grains Research and Development Council (GRDC) UWA372, the Australian Research Council (ARC) LP0454871, and the Polish State Committee for Scientifc Research. Salaries of MNN and WAC are supported by a grant from the Export Grains Centre Ltd. The development of Lupinus spp.-derived cross-species marker primers were supported by GRDC UMU105 and ARC DP0559547. We thank Bevan Buirchell for providing seed for the RIL population, Sarah Brien for early assistance in the project and Michael Francki, Craig Atkins and Penny Smith for providing the L. angustifolius cDNA library.

References

Atkins CA, Smith PMC, Gupta S, Jones MGK, Caligari PDS (1998) Genetics, Cytology and Biotechnology. In: Gladstones JS, Atkins CA, Hamblin J (eds) Lupins as crop plants: biology, production and utilization. CAB International, London, pp 67–92

Bennett MD, Smith JB, Heslopharrison JS (1982) Nuclear-DNA Amounts in Angiosperms. Proc R Soc Lond B Biol Sci 216:179–199

Bhardwaj HL (2002) Evaluation of lupin as a new food/feed crop in the Mid-Atlantic region. In: Janick J, Whipkey A (eds) Trends in new crops and new uses. ASHS Press, Aleksandria, pp 115–119

Boersma JG, Pallotta M, Li CD, Buirchell BJ, Sivasithamparam K, Yang HA (2005) Construction of a genetic linkage map using MFLP and identification of molecular markers linked to domestication genes in narrow-leafed lupin (*Lupinus angustifolius* L.). Cell Mol Biol Lett 10:331–344

Brien SJ, Cowling WA, Potter RH, Jones RAC, Jones MGK (1999) A molecular marker for early maturity (*Ku*) and marker-assisted breeding of *Lupinus angustifolius*. In: Santen Ev, Wink M, Weissmann S, Romer P (eds) 9th Inter-

national lupin conference– Lupin, an ancient crop for the new millennium. International Lupin Asssociation, Klink/ Muritz, Germany, pp 115–117
Carruthers K, Prithiviraj B, Fe Q, Cloutier D, Martin RC, Smith DL (2000) Intercropping of corn with soybean, lupin and forages: silage yield and quality. J Agron Crop Sci 185:177–185
Choi HK, Kim D, Uhm T, Limpens E, Lim H, Mun JH, Kalo P, Penmetsa RV, Seres A, Kulikova O, Roe BA, Bisseling T, Kiss GB, Cook DR (2004a) A sequence-based genetic map of *Medicago truncatula* and comparison of marker colinearity with *M. sativa*. Genetics 166:1463–1502
Choi HK, Mun JH, Kim DJ, Zhu HY, Baek JM, Mudge J, Roe B, Ellis N, Doyle J, Kiss GB, Young ND, Cook DR (2004b) Estimating genome conservation between crop and model legume species. Proc Natl Acad Sci U S A 101:15289–15294
Cowling WA (1999) Pedigrees and characteristics of narrow-leafed lupin cultivars released in Australia from 1967 to 1998. Bulletin 4365, Agriculture Western Australia, Perth
Cowling WA, Huyghe C, Swiecicki W (1998) Lupin breeding. In: Gladstones JS, Atkins CA, Hamblin J (eds) Lupins as crop plants: biology, production and utilization. CAB International, London, pp 93–120
Devos KM (2005) Updating the 'Crop circle'. Curr Opin Plant Biol 8:155–162
Edwards AC, van Barneveld RJ (1998) Lupins for livestock and fish. In: Gladstones JS, Atkins C, Hamblin J (eds) Lupins as Crop plants: biology, production and utilisation. CAB International, London, pp 385–411
Gladstones JS (1967) Selection for economic characters in *Lupinus angustifolius* L. and *L. digitatus* Forsk. I. Non-shattering pods. Aust J Exp Agric Anim Husbandry 7:360–366
Gladstones JS (1977) The narrow-leafed lupin in Western Australia (*Lupinus angustifolius* L.). Western Australian Department of Agriculture Bulletin 3990
Gladstones JS, Crosbie GB (1978) Lupin wild types introduced into Western Australia to 1973. Technical Bulletin No. 43. Department of Agriculture Western Australia, South Perth, Australia
Gupta S, Buirchell BJ, Cowling WA (1996) Interspecific reproductive barriers and genomic similarity among the rough-seeded *Lupinus* species. Plant Breed 115:123–127
Hajdera I, Siwinska D, Hasterok R, Maluszynska J (2003) Molecular cytogenetic analysis of genome structure in *Lupinus angustifolius* and *Lupinus cosentinii*. Theor Appl Genet 107:988–996
Konieczny A, Ausubel FM (1993) A procedure for mapping arabidopsis mutations using codominant ecotype-specific PCR-based markers. Plant J 4:403–410
Lagercrantz U, Lydiate DJ (1996) Comparative genome mapping in Brassica. Genetics 144:1903–1910
Lavin M, Herendeen PS, Wojciechowki MF (2005) Evolutionary rates analysis of leguminosae implicates a rapid diversification of lineages during the tertiary. Syst Biol 54:575–594
Lukens L, Zou F, Lydiate D, Parkin I, Osborn T (2003) Comparison of a Brassica oleracea genetic map with the genome of Arabidopsis thaliana. Genetics 164:359–372
Lyons LA, Laughlin TF, Copeland NG, Jenkins NA, Womack JE, Obrien SJ (1997) Comparative anchor tagged sequences (CATS) for integrative mapping of mammalian genomes. Nat Genet 15:47–56
Majoros WH, Pertea M, Antonescu C, Salzberg SL (2003) GlimmerM, exonomy and unveil: three *ab initio* eukaryotic genefinders. Nucleic Acids Res 31:3601–3604
Manly KF, Cudmore RH, Meer JM (2001) Map manager QTX, cross-platform software for genetic mapping. Mamm Genome 12:930–932
Mikolajczyk J (1966) Genetic studies in *Lupinus angustifolius*. Part III. Inheritance of the alkaloid content, seed hardness and length of the growing season in blue lupin. Genet Pol 7:181–196
Moolhuijzen PM, Cakir M, Macgregor A, Smith C, Hunter A, Schibeci D, Francki M, Jones M, Appels R, Bellgard M (2006) LegumeDB bioinformatics resource: Comparative genomic analysis and novel cross-genera marker identification in Lupin and pasture legume species. Genome (in press)
Moore G, Devos KM, Wang Z, Gale MD (1995) Cereal genome evolution—grasses, line up and form a circle. Curr Biol 5:737–739
Naganowska B, Kaczmarek A (2005) Repetitive DNA sequences in cytogenetic studies of Lupinus genome. Monograph, 11th international Lupin conference, "Mexico, where Old and New World lupins meet", Guadalajara, Mexico
Naganowska B, Zielinska A (2002) Physical mapping of 18S-25S rDNA and 5S rDNA in *Lupinus* via fluorescent *in situ* hybridization. Cell Mol Biol Lett 7:665–670
Naganowska B, Zielinska A (2004) Localisation of rDNA in the *Lupinus* genome during the cell cycle. J Appl Genet 45(2), 2004, pp 189–193 45:189–193
Naganowska B, Wolko B, Sliwinska E, Kaczmarek Z (2003) Nuclear DNA content variation and species relationships in the genus *Lupinus* (Fabaceae). Ann Bot 92:349–355
Nelson P (1994) The development of the lupin industry in Western Australia and its role in sustainable farming systems. In: Neves Martins JM, Beirão da Costa L (eds) Advances in Lupin Research. ISA Press, Lisbon, pp 565–576
Nelson MN, Lydiate D (2006) New evidence from *Sinapis alba* L. for ancestral triplication in a crucifer genome. Genome (in press)
Petterson DS (1998) Composition and food uses of lupins. In: Gladstones JS, Atkins CA, Hamblin J (eds) Lupins as crop plants: biology, production and utilization. CAB International, London, pp 353–384
Pfeil BE, Schlueter JA, Shoemaker RC, Doyle JJ (2005) Placing paleopolyploidy in relation to taxon divergence: A phylogenetic analysis in legumes using 39 gene families. Syst Biol 54:441–454
Reeves DW (1991) Experiences and prospects for lupin in the South and Southeast. Prospects for Lupin in North America, Symposium Proceedings, University of Minnesota, St. Paul, pp 23–30
Rinehart T (2004) AFLP analysis using GeneMapper software and an Excel macro that aligns and converts output to binary. Biotechniques 37:186–188
Sharpe AG, Parkin IAP, Keith DJ, Lydiate DJ (1995) Frequent nonreciprocal translocations in the amphidiploid genome of oilseed rape (*Brassica napus*). Genome 38:1112–1121
Shoemaker RC, Polzin K, Labate J, Specht J, Brummer EC, Olson T, Young N, Concibido V, Wilcox J, Tamulonis JP, Kochert G, Boerma HR (1996) Genome duplication in soybean (*Glycine* subgenus *soja*). Genetics 144:329–338
Varshney RK, Sigmund R, Borner A, Korzun V, Stein N, Sorrells ME, Langridge P, Graner A (2005) Interspecific transferability and comparative mapping of barley EST-SSR markers in wheat, rye and rice. Plant Sci 168:195–202
Vos P, Hogers R, Bleeker M, Reijans M, van de Lee T, Hornes M, Frijters A, Pot J, Peleman J, Kuiper M, Zabeau M (1995) AFLP: A new technique for DNA fingerprinting. Nucleic Acids Res 23:4407–4414
Wilkins RJ, Jones R (2000) Alternative home grown protein sources for ruminants in the United Kingdom. Anim Feed Sci Technol 85:23–32
Wolko B, Weeden N (1989) Estimation of *Lupinus* genome polyploidy on the basis of isozymic loci number. Genet Pol 30:165–171

Wolko B, Weeden NF (1994) Linkage map of isozyme and RAPD markers for the *Lupinus angustifolius* L. In: Neves Martins JM, Beirão da Costa L (eds) Advances in lupin research. ISA Press, Lisbon, pp 42–49

Yan HH, Mudge J, Kim DJ, Shoemaker RC, Cook DR, Young ND (2004) Comparative physical mapping reveals features of microsynteny between *Glycine max*, *Medicago truncatula*, and *Arabidopsis thaliana*. Genome 47:141–155