

# HOW CAN THE GENOMIC REVOLUTION HELP IMPROVE LUPINS

Rhonda Foley<sup>1,2</sup>, Ling-ling Gao<sup>1,2</sup>, Judith Lichtenzveig<sup>1,3</sup>, Elaine Smith<sup>1</sup>, Bu-jun Shi<sup>4</sup>,  
Craig Atkins<sup>2,5</sup>, Ben Rosen<sup>6</sup>, Noelia Carrasquilla-Garcia<sup>6</sup>, Andrew D. Farmer<sup>7</sup>,  
Varma Penmetsa<sup>6</sup>, Doug Cook<sup>6</sup> and Karam B. Singh<sup>1,2,\*</sup>

<sup>1</sup>CSIRO Plant Industry, Western Australia

<sup>2</sup>Centre for Food and Genomic Medicine, Western Australia

<sup>3</sup>Australian Centre for Necrotrophic Fungal Pathogens, Murdoch University, Western Australia

<sup>4</sup>Australian Centre for Plant Functional Genomics, University of Adelaide, South Australia

<sup>5</sup>School of Plant Biology, University of Western Australia, Western Australia

<sup>6</sup>Department of Plant Pathology, UC Davis, Davis, California, USA

<sup>7</sup>National Centre for Genome Resources, Santa Fe, New Mexico, USA

Corresponding author email: karam.singh@csiro.au

## ABSTRACT

Over the last few years there has been a growing effort to apply the power of plant genomics to key legume species. We are using the opportunities offered by the legume genomic revolution to progress three areas of research in narrow-leaved lupin (*Lupinus angustifolius*). The first two areas of research form part of a large international project on comparative genomics of legume resistance gene homologs and conserved orthologous gene markers. Under this international collaboration, we are using knowledge of genome content and structure in the model legumes, especially *Medicago truncatula* to isolate low copy orthologous genes (COS markers) and resistance gene homologs (RGH) in *M. truncatula*, and thereby drive synteny analysis and the cloning and mapping of the majority of resistance genes in narrow-leaved lupin (NLL). The third area of research focuses on seed storage protein genes and forms part of the West Australian Centre for Food and Genomic Medicine (CFGM). The aim is to identify and characterise the seed storage protein gene complement in NLL. In the longer term we hope to be able to manipulate the seed storage protein profile to increase the content of desirable proteins and reduce undesirable proteins in the grain. As such, we have an active interest in helping the lupin community develop suitable reverse genetic platforms to accelerate lupin crop improvement.

## KEYWORDS

legume genomics, lupins, reverse genetics, plant resistance genes, legume seed storage proteins

## INTRODUCTION

Lupins form a diverse group of plants and the genus *Lupinus* consists of between 200-600 species, originating from South America, western North America, Africa and the Mediterranean region (Ainouche *et al.* 1999). A number of species have been

domesticated. Of particular importance from an agricultural point of view, is narrow-leaved lupin (NLL), a major grain crop in WA, which accounts for about 80% of world production. NLL are an important part of sustainable farming systems, reducing the need for nitrogenous fertiliser and boosting cereal yields. They are unusual among legume crops in being domesticated for little more than 50 years. Consequently only a limited suite of traits have been removed from the wild progenitors and many agronomic limitations remain. Similarly, the potential end uses for lupin grain and the breeding of cultivars for specific, high-value markets including their use as functional foods or food components have yet to be fully developed. Although a greater diversity of high yielding cultivars with a broader range of traits that favour adaptation to more diverse cultural conditions are now available, low returns to growers have jeopardised the sustainability of lupins in farming systems. Thus, in addition to further improvement in yield potential through greater disease resistance and more favourable agronomic traits, a current breeding imperative is to increase farm gate prices through more effective cultural practices and improved grain quality. The wide range of seed composition among lupin species (*L. angustifolius*, *L. albus*, *L. mutabilis* and *L. luteus*) indicates that substantial changes to the nutritional or nutraceutical qualities in NLL should be feasible. So while there are many challenges to be overcome, there are also some exciting opportunities for lupin researchers. In this paper we explore how advances in genomics can help improve lupins.

## LEGUME GENOMICS

There has been a rapidly growing effort to apply powerful plant genomic approaches to key legume species with a major aim of generating resources that will be of use not only in these species but in facilitating crop improvement for other legume species. Most of the progress to date has been in the two model legumes, *Medicago truncatula* and *Lotus japonicus* and the major grain legumes, pea and soybean. However, recent

breakthroughs in plant genomics, particularly in the area of genome sequencing are seeing genomics extend to a number of other legume crops such as common bean, chickpea, cowpea, lotus, peanut and pigeon pea. In the case of *M. truncatula* and *L. japonicus*, there are powerful genomic resources to elucidate the functions of genes important for many aspects of legume biology (see Sato *et al.* 2007). In addition, to the genome sequences, these include excellent informational resources, extensive genetic maps, large EST databases, and excellent reverse genetic resources including insertional, fast neutron and TILLING mutant populations. Similar resources are being developed for other legumes with pea and soybean being at the most advanced stages. This explosion of effort offers in our view significant opportunities to help improve various lupin species.

#### LUPIN GENOMICS

In contrast to the situations detailed above for some legumes, there is very little genomic information available for either NLL or other lupin spp. Genetic maps have been generated for NLL and *L. albus* (Boersma *et al.* 2005; Nelson *et al.* 2006; Phan *et al.* 2007), and while there is scope to improve the coverage in these maps, they have suggested regions of conserved syntony with *M. truncatula*, soybean and *Arabidopsis* (Nelson *et al.* 2006; Phan *et al.* 2007). In addition, there has been a limited number of ESTs sequenced and submitted to the genomic databases; these are related to P nutrition and root development, symbiosis with *Rhizobium* spp. and a small number of sequences from cDNA libraries raised in relation to grain development. A NLL bacterial artificial chromosome (BAC) library from cultivar Sonet has been constructed (Kasprzak, 2006). Transformation technology is in place for NLL with transgenic plants of cvs Unicrop and Merrit having been generated using *Agrobacterium*-mediated gene transfer to cotyledonary node apices (Pigeaire *et al.* 1997). In addition, enhanced methionine levels and increased nutritional value of seeds of transgenic of cv. Warrarrah were produced by expressing a sunflower seed albumin gene (Molvig *et al.* 1997). To our knowledge there are no comprehensive reverse genetic platforms yet available for NLL, nor any major genome sequencing initiatives.

#### DEVELOPMENT OF GENOMIC TOOLS IN NARROW-LEAFED LUPIN TO FACILITATE MOLECULAR BREEDING, TACKLE DISEASES AND IMPROVE GRAINS

The plant biotechnology group at CSIRO, WA in collaboration with partners in Australia and overseas is interested in exploiting the opportunities offered by the legume genomic revolution to help provide genomic resources for NLL and to progress research related to disease and resistance gene homologs, seed development and seed storage protein genes, and syntony between lupin and other characterised legume genomes. As part of this activity we have constructed a NLL BAC library. BAC libraries have become invaluable tools in modern plant research. Such libraries

have significantly contributed to the genetic and genomic studies of a wide range of model or economically important plant species. In legumes, for example, BAC libraries are essential resources for the whole-genome sequencing of *M. truncatula* ([www.medicago.org/genome](http://www.medicago.org/genome)), *Lotus japonicus* (<http://www.kazusa.or.jp/lotus/>) and soybean (*Glycine max* L.) ([http://genome.purdue.edu/isgc/tsukuba07/ISGC\\_report\\_Apr2007.htm](http://genome.purdue.edu/isgc/tsukuba07/ISGC_report_Apr2007.htm)). Using the BAC-end sequencing strategy, a large number of simple sequence repeat (SSR) markers have been identified and these in turn have provided integral tools for developing the genetic and physical maps of specific legumes such as soybean, white clover (*Trifolium repens* L.) and chickpea (*Cicer arietinum* L.) (Lichtenzweig *et al.* 2004; Rajesh *et al.* 2004; Song *et al.* 2004; Wu *et al.* 2004; Febrer *et al.* 2007; Shultz *et al.* 2007; Shoemaker *et al.* 2008). These BAC libraries have also facilitated identification of clones associated with resistance against pathogens, for example against cyst nematode in soybean (Tomkins *et al.* 2007).

We have constructed a large BAC library in NLL cv. Tanjil that has good insert sizes and represents 11-12 haploid genome equivalents. The library complements the existing NLL BAC library made by Kasprzak *et al.* (2006) by using a different restriction enzyme (*Bam*HI instead of *Hind*III) and a different genotype. Together, these libraries should provide comprehensive BAC resources for the lupin research community and help facilitate development of advanced breeding tools.

Our immediate plans for the BAC library are to help characterise lupin resistance genes and facilitate an investigation of lupin seed storage proteins (SSP) for comparative genomic studies of legume crops. The SSP and resistance gene projects are described in the following sections. As part of our legume comparative genomic studies, we are, in the first instance, BAC-end sequencing 5000 randomly selected BAC clones. Analysis of the sequences will help to identify additional molecular markers, such as SSR markers, which will be valuable in the development of a more advanced genetic map and help allow genome comparisons between lupin and other better characterised legume species, for example *M. truncatula* and soybean.

#### ANALYSIS OF CONSERVED ORTHOLOGOUS GENES IN NARROW-LEAFED LUPIN

Based on comparison of transcript and whole genome data sets for *M. truncatula*, *L. japonicus* and *G. max*, we designed and tested a set of 1,440 primer pairs that target a set of 1,369 conserved genes. These primer pairs are being used to survey allelic diversity, especially single nucleotide polymorphisms (SNPs), in intron sequences of orthologous genes across the Papilionoid subfamily of legumes. Lupin is of particular interest, because from a phylogenetic standpoint it represents the most basal major crop species within the Papilionoideae. Our current efforts have generated high

quality sequence data for ~1,000 of the target orthologs in *Lupinus angustifolius* cultivar Tanjil, the same genome used for construction of the BAC library. Ongoing analysis will survey additional accessions of *Lupinus*, with the goal of developing a SNP dataset that will greatly enhance the gene-based molecular marker resource in lupin.

#### **ISOLATION OF RESISTANCE GENE HOMOLOGS IN NARROW-LEAFED LUPIN**

Biotic stresses continue to represent one of the most significant challenges to crop productivity in many legumes including lupins. Despite the considerable progress and success in breeding for disease resistance and in ameliorating severity through widening rotations and integrated management tools, there is a constant threat from a range of other, mainly fungal, diseases that threaten production. By using knowledge of resistance gene analogs in *M. truncatula*, we aim to drive the cloning and mapping (genetically and physically) of the majority of resistance genes in lupin. We are targeting the isolation of nucleotide binding site-leucine-rich repeat (NBS-LRR) proteins using a high-throughput genomic approach. Degenerate markers, developed from resistance gene homologs in *M. truncatula*, provide the basis to methodically survey resistance gene content in NLL. To date, 544 degenerate primer combinations have been used to isolate resistance genes from lupin. Sequence analyses have allowed these resistance genes to be categorised into the TIR-NBS-LRR family or the CC-NBS-LRR family and further analysis has allowed the resistance genes to be further sub-grouped within these groups. This proposal is part of a large, international project on comparative genomics of legume disease resistance gene homologs. The project will also investigate the major evolutionary forces driving the structural diversification of NBS-LRR resistance gene homologs across the Fabaceae by targeting characterisation of resistance gene homologs from more basal lineages.

#### **USING GENOMICS TO CHARACTERISE SEED DEVELOPMENT IN NARROW-LEAFED LUPIN**

There is growing interest to use genomics to study legume seed development (see for example Brandon *et al.* 2007). The lupin grain and lupin flour have a unique composition, being high in protein (~45% by weight) and dietary fibre (~30%), low in fat (4-5%) with virtually no starch: making lupin very attractive from a human health perspective. Recent work conducted in CFGM has shown that lupin-enriched bread can increase self-reported satiety and reduce food intake at a subsequent meal (Lee *et al.* 2006). We are interested in lupin seed development with a focus on SSP genes as these are likely to be the major constituents responsible for satiety and related dietary benefits, as well as potentially contributing to problems associated with allergenicity. SSP gene families in plants typically consist of a number of specific sub-families, each comprising 30 or so closely related members. To determine which family members are important for the

dietary benefits of lupins and which may be contributing to allergenicity problems, we need to identify and characterise the SSP complement in lupins. This will be achieved in the first instance by identifying these proteins from EST sequences derived from seed-specific cDNA libraries. However, SSP family members are often closely related and clustered in the genome and so a more extensive isolation of lupin SSP will be achieved through screening the NLL BAC library that we have constructed.

As a start to analysing seed development in NLL we have made two large cDNA libraries from an early and a later stage of seed development. Pilot sequencing of both libraries suggested the second was more likely to contain large numbers of SSP ESTs, consistent with the accumulation of these proteins towards the end of seed development. We have now sequenced over 1000 ESTs from the two libraries and are currently using bioinformatic approaches to classify these sequences and identify SSP genes. This endeavour is greatly facilitated by the genome sequence research in other legumes as well as the excellent genomics work being carried out on legume seed development, particularly in pea and soybean. Our plan is to sequence another ~2,000 ESTs, primarily from the later stage library to gain a more complete picture of gene expression in NLL seeds.

We are also using RT-PCR to determine the RNA abundance and cell types within the lupin seed where the seed storage proteins are expressed. In a complimentary approach we also plan to raise antibodies against the major SSP. These antibodies will provide valuable tools for proteomic studies on lupin seed extracts and provide a more in depth analysis of SSP expression in lupin grains. In the long term the work conducted here will lay the platform for conventional and molecular breeding efforts to select for desirable, and reduce undesirable, end products in grains and in fractions used for food ingredients.

#### **DEVELOPMENT OF ADDITIONAL GENOMIC TOOLS IN NARROW-LEAFED LUPIN**

NLL is a recently domesticated crop and consequently still has a number of problems associated with it. However, the upside for a 'primitive' crop like NLL is that there is great potential to improve their value through modern breeding. In our view this would include using reverse genetic platforms such as Fast Neutron mutagenesis and/or TILLING (Targeted Induced Local Lesions In Genomes), which allows specific targeting of genes associated with undesirable traits in order to reduce/eliminate their activity. An added bonus for these platforms, compared to other reverse genetic platforms, is that there are no transgenic plants involved, and therefore no GMO issues that could delay the uptake of this research into breeding programs in certain parts of the world. Both approaches can also be performed in an elite cultivar, thereby facilitating the incorporation of new traits into current cultivars. The

TILLING approach can generate 'loss of function' mutants as well as alleles with new, enhanced functions.

We are interested in exploiting fast neutron and/or TILLING platforms to generate more favourable grain quality and improve other traits in lupins. Some of these traits would be targeted at enhancing the value of lupins to the food processing and feed industry which will lead to increased returns to growers, processors and marketers of lupin grain. As a first step we have begun the development of large amounts of a pure population of NLL seed from a commercial cultivar, suitable for fast neutron and/or EMS mutagenesis. Because it is the SSPs in lupin that impart nutritional qualities, health benefits and potentially allergenic properties, the most obvious target genes are those encoding this group of proteins and these will be a major focus of our research.

A second area where genomics could have a big impact on lupin improvement is through genome sequencing. With the rapid pace of development in this area, it should be feasible in the not too distant future to sequence 3-4 key lupin species such as *L. angustifolius*, *L. albus*, *L. mutabilis* and *L. luteus*. The resulting information will help to shed light on some of the major differences between these species, for example how *L. mutabilis* is able to produce much more oil in the grain than the other major commercial lupin species.

In conclusion, we believe that a coordinated and well funded effort over the next decade to harness the power of plant genomics for lupin improvement with strong links to lupin breeding programs will lead to significant improvements in quality and productivity as well as benefits for human health.

### ACKNOWLEDGEMENTS

We thank Prof Peter Langridge for facilitating the construction of the NLL BAC library at the ACPFG. The research in the authors laboratories related to this paper is supported in part by grants from the WA State Government to CA and KS through CFGM, by the NSF to DC and VP and by the ACPFG through ARC/GRDC funding for BS.

### LITERATURE CITED

- Ainouche, A.-K. and R.J. Bayer. 1999. Phylogenetic relationships in *Lupinus* (Fabaceae: Papilionoideae) based on internal transcribed spacer sequences (ITS) of nuclear ribosomal DNA. *Am. J. Bot.* 86: 590.
- Boersma, J.G., M. Pallotta, C.D. Li, B.J. Buirchell, K. Sivasithamparam and H. Yang 2005. Construction of a genetic linkage map using MFLP and identification of molecular markers linked to domestication genes in narrow-leaved lupin (*Lupinus angustifolius* L.). *Cell. Mol. Biol. Lett.* 10: 331.
- Brandon, H.L., J.A. Wagmaster, T. Kawashima, A.Q. Bui, J.J. Harada and R.B. Goldberg. 2007. Using genomics to study legume seed development. *Plant Physiol.* 144: 562.
- Febrer, M. F. Cheung, C.D. Town, S.B. Cannon, N.D. Young, M.T. Abberton, G. Jenkins and D. Milbourne. 2007. Construction, characterisation, and preliminary BAC-end sequencing analysis of a bacterial artificial chromosome library of white clover (*Trifolium repens* L.). *Genome* 50: 412.
- Kasprzak A, J. Safa , J. Janda, J. Dolezel, B. Wolko and B. Naganowska. 2006. The bacterial artificial chromosome (BAC) library of the narrow-leaved lupin (*Lupinus angustifolius* L.). *Cell. Mol. Biol. Lett.* 11: 396.
- Lee, Y.P., T.A. Mori, S. Sipsa, A. Barden, I.B. Puddey, V. Burke, R.S. Hall, and J.M. Hodgson 2006. Lupin-enriched bread increases satiety and reduces energy intake acutely. *Am. J. Clinical Nutrition* 84: 975.
- Lichtenzveig J, C. Scheuring, J. Dodge, S. Abbo and H.B. Zhang HB. 2004. Construction of BAC and BIBAC libraries and their applications for generation of SSR markers for genome analysis of chickpea, *Cicer arietinum* L. *Theor. Appl. Genet.* 110: 492.
- Molvig, L., L.M. Tabe, B. Eggum, A.E. Moore, S. Craig, D. Spencer, and T.J.V. Higgins. 1997. Enhanced methionine levels and increased nutritive value of seeds of transgenic lupins (*Lupinus angustifolius* L.) expressing a sunflower seed albumin gene. *Proc Natl Acad. Sci.* 94: 8393.
- Nelson, M.N., H.T.T. Phan, S.R. Ellwood, P.M. Moolhuijzen, J. Hane, A. Williams, C.E. O'Lone, J. Fosu-Nyarko, M. Scobie, M. Cakir, M.J.K. Jones, M. Bellgard, M. Ksiarkiewicz, B. Wolko, S.J. Barker, R.P. Oliver and W.A. Cowling. 2006. The first gene-based map of *Lupinus angustifolius* L.-location of domestication genes and conserved syntony with *Medicago truncatula*. *Theor. Appl. Genet.* 113: 225.
- Phan, H.T.T, S.R. Ellwood, K. Adhikari, M.N. Nelson and R.P. Oliver. 2007. The First Genetic and Comparative Map of White Lupin (*Lupinus albus* L.): Identification of QTLs for Anthracnose Resistance and Flowering Time, and a Locus for Alkaloid Content. *DNA Res.* 14: 59.
- Pigeaire, A., D. Abernethy, P.M. Smith, K. Simpson, N. Fletcher, C.Y. Lu, C. A. Atkins and E. Cornish. 1997. Transformation of a grain legume (*Lupinus angustifolius* L.) via *Agrobacterium tumefaciens*-mediated gene transfer to shoot apices. *Mol. Breed.* 3: 341.
- Rajesh, P.N., C. Coyne, K. Meksem, K.D. Sharma, V. Gupta and F.J. Muehlbauer. 2004. Construction of a HindIII Bacterial Artificial Chromosome library and its use in identification of clones associated with disease resistance in chickpea. *Theor. Appl. Genet.* 108: 663.
- Sato, S., Y. Nakamura, E. Asamizu, S. Isobe and S. Tabata. 2007. Genome sequencing and genome resources in model legumes. *Plant Physiol.* 144: 588.
- Shoemaker, R.C., D. Grant, T. Olson, W.C. Warren, R. Wing, Y. Yu, H. Kim, P. Cregan, B. Joseph, M. Futrell-Griggs, W. Nelson, J. Davito, J. Walker, J. Wallis, C., Kremitski, D. Scheer, S.W. Clifton, T. Graves, H. Nguyen, X. Wu, M. Luo, J. Dvorak, R. Nelson, S. Cannon, J. Tomkins, J. Schmutz, G. Stacey and S. Jackson. 2008. Microsatellite discovery from BAC end sequences and genetic mapping to anchor the soybean physical and genetic maps. *Genome* 51: 294.
- Shultz, J.L., S. Kazi, R. Bashir, J.A. Afzal and D.A. Lightfoot. 2007. The development of BAC-end sequence-based microsatellite markers and placement in the physical and genetic maps of soybean. *Theor. Appl. Genet.* 114: 1081.

- Song, Q.J., L.F. Marek, R.C. Shoemaker, K.G. Lark, V.C. Concibido, X. Delannay, J.E. Specht and P.B. Cregan. 2004. A new integrated genetic linkage map of the soybean. *Theor. Appl. Genet.* 109: 122.
- Tomkins, J.P., R. Mahalingam, H. Smith, J.L. Goicoechea, H.T. Knap and R.A. Wing. 1999. A bacterial artificial chromosome library for soybean PI 437654 and identification of clones associated with cyst nematode resistance. *Plant Mol. Biol.* 41: 25.
- Wu, C.C., P. Nimmakayala, F.A. Santos, R. Springman, C. Scheuring, K. Meksem, D.A. Lightfoot and H.B. Zhang. 2004. Construction and characterisation of a soybean bacterial artificial chromosome library and use of multiple complementary libraries for genome physical mapping. *Theor. Appl. Genet.* 109: 1041.