A DENSE REFERENCE MAP OF THE *LUPINUS ANGUSTIFOLIUS* L. GENOME: A FOUNDATION FOR BUILDING LUPIN GENOME RESEARCH

Matthew Nelson[1*], Jeff Boersma[2,3†], Magdalena Chudy[4], Karolina Lesniewska[4], Simon Ellwood[5], Huyen Phan[5], Paula Moolhuijzen[6], Matthew Bellgard[6], Richard Oliver[5], Wojciech Święcicki[4], Bogdan Wolko[4] and Wallace Cowling[1]

[1]School of Plant Biology, Faculty of Natural and Agricultural Sciences, The University of Western Australia, 35 Stirling Highway, Crawley WA 6009, Australia
[2]School of Earth and Geographical Sciences, Faculty of Natural and Agricultural Sciences, The University of Western Australia, 35 Stirling Highway, Crawley WA 6009, Australia
[3]Department of Agriculture and Food, Western Australia, South Perth WA 6151, Australia
[4]Institute of Plant Genetics, Polish Academy of Sciences, 60-479 Poznan, Poland
[5]Australian Centre for Necrotrophic Fungal Pathogens, WA State Agricultural Biotechnology Centre, Division of Health Sciences, Murdoch University, Perth WA 6150, Australia
[6]Centre for Comparative Genomics, Murdoch University, Perth WA 6150, Australia
[†]Current address: Department of Plant Agriculture, Crop Science Bldg., University of Guelph, Guelph, Ontario, N1G 2W1, Canada

*Corresponding author: mnelson@plants.uwa.edu.au

ABSTRACT

A genetic map provides a foundation on which genomic research can be built, linking together otherwise disconnected studies and forming the basis on which to build physical and cytogenetic maps and conduct whole genome sequencing. As part of a wide collaborative effort, we developed a dense genetic map of the narrow-leafed lupin (*Lupinus angustifolius* L.) comprising 1073 molecular markers, 6 domestication trait loci and an Anthracnose resistance locus distributed across 20 major linkage groups. This map will be used to investigate the extent of genome conservation between narrow-leafed lupin and the model legume species, *Medicago truncatula* and *Lotus japonicus*. We discuss the potential applications of this reference map in lupin genetic and genomic research.

KEYWORDS

molecular markers, linkage mapping, syntony, *Lotus japonicus, Medicago truncatula*

INTRODUCTION

Chromosome research arguably began with the first microscopic descriptions of chromosomes in the mid-1800s, but the true significance of these cellular features as vehicles of inheritance was not recognised until the beginning of the 1900s. The first genetic linkage map was developed in Drosophila by Sturtevant (1913), but the first extensive genetic maps for plant species (tomato and maize) were developed over twenty years later (MacArthur, 1934; Emerson *et al.* 1935). Early genetic maps required a large investment of time and resources because their construction involved the synthesis of many separate studies each of which involved segregation of a few gross morphological markers. It was not until the advent of DNA markers in the 1980s that sufficient numbers of markers became available to easily develop genetic maps in single mapping populations. Since then, genetic maps have been developed for most crop species including narrow-leafed lupin (*Lupinus angustifolius* L.) (Boersma *et al.* 2005; Nelson *et al.* 2006) and white lupin (*L. albus* L.) (Phan *et al.* 2007). Genetic mapping has many uses in basic genetic research including understanding the inheritance of traits of interest to research and agriculture, understanding chromosome structure and behaviour, understanding population structure and diversity, and forming a foundation on which to build a physical map, a prerequisite for whole genome sequencing.

In order to fully exploit a genetic map, it should represent as closely as possible the entire nuclear genome of the species. The lupin species with the most developed linkage map is narrow-leafed lupin, the most important grain legume in Australia. The two major genetic maps so far produced were developed using two somewhat different subsets of the same population of recombinant inbred lines derived from a cross between domesticated and wild lupin parents (Table 1). The two genetic maps differed somewhat in their characteristics, as summarised in Table 1. The strength of the Boersma *et al.* (2005) map included the larger number of loci that were used and the stronger statistical support for the linkage groups. The main strength of the Nelson *et al.* paper was the use of primarily locus-specific, gene-based markers that allowed a comparative analysis with the draft genome of the model legume species, *Medicago truncatula* – the first assessment of syntony involving a lupin species.

IN J.A. Palta and J.B. Berger (eds). 2008. 'Lupins for Health and Wealth' Proceedings of the 12th International Lupin Conference, 14-18 Sept. 2008, Fremantle, Western Australia. International Lupin Association, Canterbury, New Zealand. ISBN 0-86476-153-8.

Table 1. Comparing the genetic linkage maps of narrow-leafed lupin of Boersma *et al.* (2005) and Nelson *et al.* (2006).

Features of the two maps	Boersma *et al.* (2005)	Nelson *et al.* (2006)
No. of RILs [1]	89 [13] [2]	93 [17] [2]
No. of loci	522 [3]	382
No. of linkage groups	21	20
Statistical support for groups	LOD ≥ 2.5	LOD ≥ 2.0
Total genetic length	1,543 cM	1,846 cM
Type of marker	Arbitrary, non-genic	Mainly gene-based, locus-specific

[1]RIL = recombinant inbred line, from a total population size of 106 RILs.

[2]Numbers in square brackets indicate number of RILs unique to that genetic map. 76 RILs were shared between both maps.

[3]Markers with unique segregation patterns. Total marker number including redundant markers was 1083.

Direct comparison between the two published maps is limited to just five shared loci (domestication major genes). However, it should be feasible to combine the two genotyping data sets because the two mapping studies shared 76 RILs. This paper reports the development of a new, more complete genetic map of narrow-leafed lupin based on 1083 marker loci, 6 major genes for domestication traits and 1 major gene for Anthracnose resistance, scored in up to 106 RILs. Eighty-one new gene-based markers are included in the new map.

MATERIALS AND METHODS

SOURCES OF MARKERS

Published markers and major gene traits came from several published sources: Boersma *et al.* (2005, 2007a, 2007b, 2007c) and Nelson *et al.* (2006). Three markers and the *Tardus* domestication trait locus came from Boersma *et al.* (manuscript in preparation). A further 81 new gene-based markers were developed by various approaches (Nelson *et al.* manuscript in preparation).

GENETIC MAPPING APPROACH

The software application, MultiPoint 2.1 (MultiQTL Ltd, Haifa, Israel) was used to conduct genetic linkage mapping. This software uses a unique approach to resolving locus order – known as the 'evolutionary optimisation strategy' (Mester *et al.* 2003). The advantage of MultiPoint 2.1 over other linkage analysis packages is that it is able to identify the most reliable subset of markers and to use them to form a 'skeleton map' on which less reliable markers are attached in the most likely genetic intervals. This approach is well suited to situations where there is a high ratio of markers to lines in the mapping population and for collaborative projects where proof-reading of raw genotyping data is made difficult by the dispersed nature of data storage.

Severe segregation distortion (that is, deviation from Mendelian allele segregation ratios) can be problematic in linkage analysis, most notably the occurrence of false linkages between skewed loci. However, significant skewing (Chi-square $P < 0.01$) were reported by both Boersma *et al.* (2005) and Nelson *et al.* (2006) in two subsets of this RIL population. Taking these contrasting requirements into consideration, markers showing extreme segregation distortion ($P < 0.001$) were excluded from further analysis.

Initial clustering by MultiPoint 2.1 was performed at a maximum recombination frequency (rf) of 0.10. Redundant markers (that is, markers showing identical segregation patterns) were removed, leaving the most informative markers (termed 'delegate' markers). Locus order of delegate markers was then determined using the 'order' function. The function 'control of monotony' was used to identify the most informative subset of markers and to temporarily remove markers that introduced ambiguities in locus order. The same set of procedures was carried out serially at gradually increasing maximum recombination frequencies (rf = 0.15, 0.20, 0.25, 0.28, 0.30 and 0.31). At each new clustering cycle, manual inspection of proposed new clusters assisted the identification of valid and invalid clusters. A new cluster was considered valid if its two progenitor clusters were most closely related to each other via their terminal loci. A new cluster was considered **invalid** if its two progenitor clusters were most closely related by **non-terminal** loci. Such invalid clusters were associated with problematic loci that showed either high degree of missing data points and/or severe segregation distortion of alleles towards one of the two founding parents of the population. Only clusters judged valid were permitted during linkage mapping.

RESULTS AND DISCUSSION

Linkage mapping identified 20 major linkage groups named NLL-01 to NLL-20, 3 small clusters and 9 unlinked markers (Table 2). The total map length was 2361.8 cM, greater than either previous map (Table 1). The sharp division between the main 20 linkage groups and the smaller clusters is a strong indication that each linkage group represents a single chromosome in the narrow-leafed lupin genome, which has a haploid chromosome number of n = 20.

Table 2. Details of the new genetic map of narrow-leafed lupin. New linkage groups are named NLL-01 to NLL-20 in descending length. The equivalent linkage groups in previously published maps by Boersma *et al.* (2005) and Nelson *et al.* (2006) are shown.

Linkage group	Length (cm) [1]	Skeleton markers [2]	Attached markers [3]	Total markers	Boersma *et al.* (2005)	Nelson *et al.* (2006)
NLL-01	168.1	52	42	94	LG1	LG07 (top) and LG18
NLL-02	155.6	34	28	62	LG6 and LG24	LG19 (top) and LG15
NLL-03	153.9	43	22	65	LG11 and LG27	LG02
NLL-04	143.9	34	22	56	LG5 and LG26	LG10
NLL-05	137.1	34	28	62	LG4	LG13
NLL-06	133.4	42	40	82	LG3	LG16 and LG14 (top)
NLL-07	121.8	40	26	66	LG9 and LG25	LG04
NLL-08	121.7	34	21	55	LG13 and LG21	LG05(top)
NLL-09	119.4	32	10	42	LG18, LG28 and LG29	LG11
NLL-10	117.8	29	28	57	LG17 and LG20	LG01, LG19 (bottom) and Triplet-1
NLL-11	112.9	36	42	78	LG2	LG06 and LG20 (bottom)
NLL-12	110.2	20	11	31	LG21 and LG23	LG09
NLL-13	102.9	25	18	43	LG16, LG30 and LG34	LG07 (bottom) and Pair-3
NLL-14	102.9	25	11	36	LG15	LG14 (bottom)
NLL-15	101.6	30	12	42	LG12	LG17 and Triplet-2
NLL-16	100.1	17	17	34	LG19 and LG31	LG05 (bottom) and Pair-2
NLL-17	99.7	27	22	49	LG8	LG03
NLL-18	90.8	29	16	45	LG14	LG12
NLL-19	85.6	22	18	40	LG7	LG08
NLL-20	69.7	27	14	41	LG10	LG20 (top)
Cluster-1	9.5	4	1	5	LG32	Pair-1
Cluster-2	1.8	3	0	3	LG33	-
Cluster-3	1.4	2	0	2	Unlinked	Unlinked
TOTAL	2361.8	641	449	1090		

[1]Genetic lengths of linkage group are expressed in Kosambi centiMorgans (cM).
[2]Skeleton markers are used to order linkage groups.
[3]Attached markers are associated with intervals defined by skeleton markers.

The linkage groups identified in this study were compared to the linkage groups in previous maps by Boersma *et al.* (2005) and Nelson *et al.* (2006) and equivalent linkage groups were identified (Table 2). This comparison gave insight into the different mapping approaches used in those previous studies. The more conservative statistical threshold (LOD = 2.5) used by Boersma *et al.* (2005) was successful in preventing false linkages but failed to integrate small linkage groups (most of which were omitted from the Boersma *et al.* (2005) publication) with larger groups. For example, LG18, LG28 and LG29 of Boersma *et al.* (2005) coalesced to form NLL-09 in the new combined map (Table 2). In contrast, the less conservative statistical threshold (LOD = 2.0) used by Nelson *et al.* (2006) was successful in joining small linkage groups to larger groups, but failed to prevent false linkages occurring. For example, LG14 in Nelson *et al.* 2006 was proven not to be a single linkage group in the new combined map where it was divided between NLL-06 and NLL-11 (Table 2).

The new combined genetic map included approximately 400 gene-based markers. Because genes are well conserved between species, these markers can be used to compare the genetic map of narrow-leafed lupin with the genome sequence of model legumes, such as *Medicago truncatula* and *Lotus japonicus* (Nelson *et al.* 2006). This analysis is currently in progress. It is hoped that improved understanding of genome structure conservation between narrow-leafed lupin and these legume species will permit greater access to the genomic resources developed in these model genomes.

Because genes are present in low copy number and are relatively free of repetitive sequences, they can be used as hybridisation probes to screen genomic BAC (bacterial artificial chromosome) libraries and mitotic sections. The utility of screening BAC libraries is for developing improved markers for targeted regions of the genome. As cytogenetic probes, gene-based markers with known genetic map locations can be used to integrate the genetic map with the physical cytogenetic map.

Finally, a robust, marker-rich genetic map is the foundation for building physical maps for genome sequencing. For example, a genetic map provides the framework for orientating and associating BAC contigs. The combined map presented here is highly suitable for such a purpose for three reasons: 1) it is based on a moderately sized population of 'immortal' RILs that have been made available to the lupin genetic research community; 2) the map is densely populated with markers, of which almost 400 have sequence data published on Genbank that can be used to link genetic and physical maps of narrow-leafed lupin; and 3) the linkage groupings and locus orders are strongly supported using the powerful 'evolutionary optimisation strategy' of MultiPoint 2.1.

In summary, we present a new and improved genetic map of narrow-leafed lupin that will serve as a strong foundation for future genomic research in this important grain legume species.

LITERATURE CITED

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-leafed lupin (*Lupinus angustifolius* L.). Cellular and Molecular Biology Letters 10: 331-344.

Boersma, J.G., B.J. Buirchell, K. Sivasithamparam and H. Yang. 2007a. Development of two sequence-specific PCR markers linked to the *le* gene that reduces pod shattering in narrow-leafed Lupin (*Lupinus angustifolius* L.). Genetics and Molecular Biology 30: 623-629.

Boersma, J.G., B.J. Buirchell, K. Sivasithamparam and H. Yang. 2007b. Development of a sequence-specific PCR marker linked to the *Ku* gene which removes the vernalisation requirement in narrow-leafed lupin. Plant Breeding 126: 306-309.

Boersma, J.G., B.J. Buirchell, K. Sivasithamparam and H. Yang. 2007c. Development of a PCR marker tightly linked to *mollis*, the gene that controls seed dormancy in *Lupinus angustifolius* L. Plant Breeding 126: 612-616.

Emerson, R.A., G.W. Beadle and A.C. Fraser. 1935. A summary of linkage studies in maize. Cornell Univ. Agr. Exp. Sta. Memoir 180.

MacArthur, J.W. 1934. Linkage groups in tomato. *J. Genet*. 29: 123-133.

Mester D., Y. Ronin, D. Minkov, E. Nevo and A. Korol. 2003. Constructing large scale genetic maps using evolutionary strategy algorithm. Genetics 165: 2269-2282.

Nelson, M., H. Phan, S. Ellwood, P. Moolhuijzen, J. Hane, A. Williams, C. O'Lone, J. Fosu-Nyarko, M. Scobie, M. Cakir, M. Jones, M. Bellgard, M. Ksiazkiewicz, B. Wolko, S. Barker, R. Oliver and W. Cowling. 2006. The first gene-based map of *Lupinus angustifolius* L.- location of domestication genes and conserved syntony with *Medicago truncatula*. *T*heoretical and Applied Genetics 113: 225-238.

Phan, H.T.T., S.R. Ellwood, K. Adhikai, 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 Research 14: 59-70.

Sturtevant, A.H. 1913. The linear arrangement of six sex-linked factors in *Drosophila*, as shown by their mode of association. Journal of Experimental Zoology 14: 43-59.