Theor Appl Genet (2014) 127:703–713
DOI 10.1007/s00122-013-2252-0

ORIGINAL PAPER

EST-SNP discovery and dense genetic mapping in lentil (*Lens culinaris* Medik.) enable candidate gene selection for boron tolerance

Sukhjiwan Kaur · Noel O. I. Cogan · Amber Stephens · Dianne Noy · Mirella Butsch · John W. Forster · Michael Materne

Received: 5 August 2013 / Accepted: 13 December 2013 / Published online: 27 December 2013

Abstract

Key message **Large-scale SNP discovery and dense genetic mapping in a lentil intraspecific cross permitted identification of a single chromosomal region controlling tolerance to boron toxicity, an important breeding objective.**

Abstract Lentil (*Lens culinaris* Medik.) is a highly nutritious food legume crop that is cultivated world-wide. Until recently, lentil has been considered a genomic 'orphan' crop, limiting the feasibility of marker-assisted selection strategies in breeding programs. The present study reports on the identification of single-nucleotide polymorphisms (SNPs) from transcriptome sequencing data, utilisation of expressed sequence tag (EST)-derived simple sequence repeat (SSR) and SNP markers for construction of a gene-based genetic linkage map, and identification of markers in close linkage to major QTLs for tolerance to boron (B) toxicity. A total of 2,956 high-quality SNP markers were identified from a lentil EST database. Sub-sets of 546 SSRs and 768 SNPs were further used for genetic mapping of an intraspecific mapping population (Cassab × ILL2024) that exhibits segregation for B tolerance. Comparative analysis of the lentil linkage map with the sequenced genomes of *Medicago truncatula* Gaertn., soybean (*Glycine max* [L.] Merr.) and *Lotus japonicus* L. indicated blocks of conserved macrosynteny, as well as a number of rearrangements. A single genomic region was found to be associated with variation for B tolerance in lentil, based on evaluation performed over 2 years. Comparison of flanking markers to genome sequences of model species (*M. truncatula*, soybean and *Arabidopsis thaliana*) identified candidate genes that are functionally associated with B tolerance, and could potentially be used for diagnostic marker development in lentil.

Communicated by A. J. Bervillé.

Electronic supplementary material The online version of this article (doi:10.1007/s00122-013-2252-0) contains supplementary material, which is available to authorized users.

S. Kaur · N. O. I. Cogan · A. Stephens · J. W. Forster (✉)
Department of Environment and Primary Industries, Biosciences Research Division, AgriBio, Centre for AgriBioscience, 5 Ring Road, La Trobe University Research and Development Park, Bundoora, VIC 3083, Australia
e-mail: john.forster@depi.vic.gov.au

D. Noy · M. Butsch · M. Materne
Department of Environment and Primary Industries, Biosciences Research Division, Grains Innovation Park, Horsham, VIC 3401, Australia

J. W. Forster
La Trobe University, Bundoora, VIC 3086, Australia

Introduction

Lentil (*Lens culinaris* ssp. *culinaris)* is widely cultivated throughout the Indian subcontinent, Middle East, northern Africa, southern Europe, North and South America and Australia. World production of lentil is estimated at 4.4 million metric tonnes from c. 4.2 million hectares, with an average yield of 950 kg/ha (FAOSTAT 2011). Lentil is generally grown in rotation with cereals to break disease cycles and to fix atmospheric nitrogen, thus reducing demand for artificial nitrogen fertilizers (Rubeena et al. 2003; Tanyolac et al. 2010). In general, legumes are more sensitive to environmental challenges than cereals, and several abiotic stresses such as salinity, boron (B) toxicity, cold, drought and heat adversely affect global yields of lentil (Muehlbauer et al. 2006). B toxicity is a

major problem in several major production zones, including southern Australia, India, Pakistan, Iraq, Peru and USA (Yau and Ryan 2008). High concentrations of B occur at depths between 40 and 100 cm in the soil profile, and inhibit both crop growth and grain yield (Cartwright et al. 1984; Nuttall et al. 2003). Amelioration of B toxicity through soil modification is not a feasible solution from an economic or logistical perspective, and so identification of more tolerant cultivars is considered the best remedy. To date, several resistant cultivars have been identified within lentil breeding programs through phenotypic screening (Yau and Erskine 2000; Hobson 2007), but no information on genomic regions associated with B tolerance in lentil is available. In contrast, chromosomal regions governing B tolerance have been identified in wheat and barley (Jefferies et al. 1999, 2000) through linkage mapping and QTL analysis of recombinant inbred (RIL) populations, and on this basis two genes governing B toxicity tolerance were cloned from barley (Hayes and Reid 2004; Schnurbusch et al. 2010). To understand the mechanism of B tolerance in lentil, therefore, the equivalent genomic regions controlling the trait must be identified and well-characterised.

The size of the lentil genome is large, with a DNA content 4.2 pg/1C or 4,063 Mbp/1C (Arumuganathan and Earle 1991), as compared to that of many other closely related legume species such as *M. truncatula* (500 Mbp), *L. japonicus* (470 Mbp) and chickpea (740 Mbp), and so generation of a reference genome assembly for *Lens* species remains a significant undertaking. However, due to close taxonomic relationships with the model legume species *M. truncatula* and *L. japonicus*, lentil can benefit from comparative genomic studies to perform candidate gene selection. Numerous studies have confirmed that comparative genomics can confirm phylogenetic relationships between species and determine patterns of chromosomal evolution and macrosyntenic relationships (Bennetzen 2000; Nelson et al. 2006; Phan et al. 2006; Ellwood et al. 2008). More importantly, comprehensive comparative genomics can facilitate bidirectional use of genomic resources between different legume species, and help to reduce cost and increase efficiency in genetic research as well as crop breeding. Fine-scale genetic mapping may accelerate candidate gene discovery through closer resolution of conserved syntenic blocks between species under comparison.

Dense genetic linkage maps provide an important resource for plant genetic studies and practical breeding. The availability of molecular genetic marker-based maps has facilitated gene tagging, marker-assisted selection and the positional cloning of genes for many crop species. The first genetic linkage map of lentil was developed from an interspecific cross between *L. culinaris* and *L. orientalis*, and was based on restriction fragment length polymorphisms (RFLPs) (Havey and Muehlbauer 1989). Until recently, the majority of available genetic linkage maps for lentil were interspecific in nature, and based on RFLPs or dominant-type anonymous markers such as amplified fragment length polymorphisms (AFLPs) and randomly amplified polymorphic DNAs (RAPDs) (Eujayl et al. 1998; Rubeena et al. 2003; Durán et al. 2004; Hamwieh et al. 2005; Tullu et al. 2008; Tanyolac et al. 2010; Gupta et al. 2012). This situation was largely due to the unavailability of highly informative and often gene-associated markers such as SSRs and SNPs from *Lens* species. Due to recent advances in sequencing and genotyping technologies, it has become possible to develop a large marker resource for understudied crop species such as lentil at an acceptable price. Recently, a number of transcriptome studies for lentil have generated EST databases, and a large number of EST-SSRs and SNPs have been made available (Kaur et al. 2011; Sharpe et al. 2013). EST-SSR markers target nucleotide diversity in genic regions, and hence permit the possibility of diagnostic marker development for MAS. Until relatively recently, SSRs have provided the most effective molecular marker technology; however, SNPs represent the most abundant form of genetic variation and are ideally suited to high levels of marker assay multiplexing, which enables generation of high-density genetic linkage maps as well as the enhancement of marker density for detection of quantitative trait loci (QTLs) and high-throughput MAS (Choudhary et al. 2012). A dense SNP-based intraspecific genetic linkage map of lentil has been recently published, which will assist in future identification of genes for qualitative and quantitative agronomic traits (Sharpe et al. 2013).

This study describes: the development of SNP markers from an EST database developed from six different lentil genotypes (Kaur et al. 2011); generation of a gene-based genetic linkage map, based exclusively on SSRs and SNPs, from the lentil cross Cassab × ILL2024, which segregates for B tolerance; B tolerance QTL identification using data collected from glasshouse-based nursery screens over 2 years in Victoria, Australia; and analysis of syntenic relationships between lentil and other species including the legumes *M. truncatula*, *L. japonicus* and soybean to permit selection of candidate genes for B tolerance.

Materials and methods

Plant material and DNA extraction

Crosses were made between single genotypes of Cassab (sensitive to B toxicity) and ILL2024 (tolerant to B toxicity) at VDEPI-Horsham in the year 2002, and F_2 generation progeny were produced. Single seed descent was

undertaken from F_2 progeny-derived genotypes for four generations in the glasshouse and a total of 126 F_6 RILs were subsequently generated.

Plants were grown under glasshouse conditions at 20 ± 2 °C under a 16/8-h (light/dark) photoperiod. Genomic DNA was extracted from young leaves using the DNeasy® 96 Plant Kit (QIAGEN, Hilden, Germany) according to the manufacturer's instructions. Approximately, 8–10 leaflets per sample were used for each extraction, and were ground using a Mixer Mill 300 (Retsch®, Haan, Germany). DNA was resuspended in milliQ water to a concentration of 10 ng/µl and stored at −20 °C until further use.

SNP discovery and validation

All EST contigs as described by Kaur et al. (2011) (Genebank accession numbers JI846297–JI861594) were used as a reference for SNP prediction using NextGENe software v1.96 (Softgenetics, State College, Pennsylvania, USA). These contigs were generated as a result of de novo assembly of 454 transcriptome data generated from six different cultivars of lentil: Indianhead, Northfield, Digger, ILL2024, ILL7537 and ILL6788. Quality-trimmed reads from all six lentil cultivars were reference aligned against the EST contig database using the 'Alignment' tool and 'SNP/Indel discovery' applications. All variants detected using the above parameters were further filtered in the following manner to obtain a set of high-confidence SNPs. Any base variants arising within a genotype and all insertion-deletion (indel) mutations were excluded, along with those for which any other sequence variation was present within 20 bp on either side. Finally, any base variants with total coverage of <6 sequence reads were also removed from the list.

A sub-set of SNPs was validated on a set of lentil genotypes, including mapping family parents, using a direct Sanger sequencing approach. Forward and reverse primer pairs were designed corresponding to the sequences flanking these SNPs using Primer3 software and Sequencher v4.10.1 (Gene Codes Corporation, MI, USA). PCR amplification was carried out in an 12.5 µl reaction containing 20 ng template DNA, 1× PCR buffer, 15 pmol of each primer, 0.2 mM of each dNTP and 0.1 µl HotStar Immolase (Bioline). PCR conditions included a hot start at 95 °C for 10 min, followed by 35 cycles of 94 °C for 30 s, 46–50 °C for 30 s and 72 °C for 30 s, and a final elongation step of 72 °C for 10 min. PCR products were desalted and purified using 2.5 µl of shrimp alkaline phosphatase (10× SAP) and 0.1 µl of exonuclease I (at 20 unit/µl) per PCR.

Cleaned PCR products were analysed using a sequencing primer and BigDye® Terminator V3.1 sequencing chemistry following manufacturer's instructions. Finalised sequencing PCR products were purified by ethanol precipitation and resuspended in 12 µl Hi-Di formamide for sequencing using an ABI3730xl (Life Technologies, Foster City, California, USA) capillary electrophoresis platform, according to the manufacturer's instructions. Sequence analysis and assembly of the resulting electropherograms were performed in Sequencher v4.10.1, and SNP validation was confirmed visually (Gene Codes Corporation, MI, USA).

SSR and SNP genotyping

Lentil EST-SSR primer pairs (Kaur et al. 2011) and previously described genomic DNA-derived SSRs (Hamwieh et al. 2005) were selected for genotypic analysis. All forward primers were synthesised with addition of an M13 sequence at the 5′-terminus, to enable fluorescent tail addition through the PCR amplification process (Schuelke 2000). PCR conditions included a hot start at 95 °C for 10 min, followed by 10 cycles of 94 °C for 30 s, 60–50 °C for 30 s and 72 °C for 30 s, followed by 25 cycles of 94 °C for 30 s, 50 °C for 30 s and 72 °C for 30 s and a final elongation step of 72 °C for 10 min. PCR products were separated using the ABI3730xl platform according to the manufacturers' instructions with addition of the ABI GeneScan LIZ500 size standard, and amplification product sizes were determined using the GeneMapper® v3.7 software (Applied Biosystems).

For SNP genotyping, a preliminary set of target loci was selected and flanking sequences 100 bp both upstream and downstream from each target SNP were sent to Illumina for assay design. A final set of SNPs was further selected on the basis of designability rank score (>0.6) for GoldenGate™ oligonucleotide pooled assay (OPA) development. A total of 250 ng of genomic DNA from each genotype was used for amplification, after which PCR products were hybridised to bead chips via the address sequence for detection on an Illumina iSCAN Reader. On the basis of obtained fluorescence, allele call data were viewed graphically as a scatter plot for each marker assayed using GenomeStudio software v2011.1 with a GeneCall threshold of 0.20.

Genetic map construction

Data obtained from SNP and SSR genotyping were tested for goodness-of-fit to the expected Mendelian ratio of 1:1 using Chi square (v2) analysis ($P < 0.05$). Linkage analysis was performed using Map Manager software version QTXb19 (Manly et al. 2001). Linkage groups (LGs) were established at an LOD score of 6.0 and a recombination fraction (q) of 0.25. The marker order of each LG was verified using the 'ripple' command. The Kosambi mapping function (Kosambi 1944) was used to convert the recombination fractions into additive

genetic distance [centiMorgans (cM)]. MAPCHART (v2.1) was used for map presentation (Voorips, 2002, http://www.biometris.wur.nl/uk/Software/MapChart/). When possible, LGs were attributed to existing map nomenclature based on location of previously available markers (Phan et al. 2007). All LGs were also compared to the recently published SNP-based lentil genetic linkage maps of Sharpe et al. (2013).

Evaluation of B toxicity and QTL analysis

All individuals from the Cassab × ILL2024 RIL mapping population were screened for response to B-induced stress applied at the seedling stage. Screening was undertaken by sowing three plants of each RIL-derived line at equidistant spacing in PVC cores using sandy clay soil. Two B rates were applied in soluble form as boric acid: 0 ppm (control) and 10 ppm. B toxicity was measured as percentage necrosis, as well as visual assessment of leaf and stem necrotic symptoms on a 1–10 scale performed over two trials conducted in years 2008 and 2009. Phenotyping data were analysed to estimate means after adjustments for any spatial effects within the trial. Models were fitted using residual maximum likelihood (REML) as implemented in GenStat (GenStat Committee 2002 and previous releases). Best linear unbiased predictions (BLUP) analysis was used to calculate narrow-sense heritability. Means of symptom rating and percentage necrosis from each data set were used to construct distribution histograms to interpret Mendelian inheritance for the trait.

QTL detection was conducted using marker regression, simple interval mapping (SIM) and composite interval mapping (CIM) in QTL Cartographer v2.5 (Wang et al. 2012). For CIM, significance levels for log-of-odds (LOD) thresholds were determined using 1,000 permutations.

Comparative genomics and selection of candidate genes

Gene sequences underlying all SNP and SSR marker loci assigned to the linkage map were BLAST analysed against the *M. truncatula* (Mt 3.5), *L. japonicus* (Lj 2.5 pseudomolecules) and soybean (Glyma1 genome assembly) genomes to detect significant matches at threshold E values $<10^{-10}$. For the purpose of graphic preparation, cM distances on the lentil linkage groups were scaled by factors calculated on the basis of the genome size of the species under comparison, to provide matching chromosomal lengths in base pairs. Visualisation of macrosyntenic blocks was performed with Circos (http://circos.ca/). Genomic regions in the fully-sequenced species (*M. truncatula*, soybean, *Lotus japonicus* and *Arabidopsis thaliana*) immediately flanking and between the sequences underlying SNP loci associated with the detected QTL region were compared to identify potential candidate genes involved in B toxicity tolerance.

Results

SNP discovery and validation

The quality trimmed transcriptome reads obtained from six different cultivars of lentil were aligned against the assembled EST contigs. A total of 24,781 base variants were identified, and a SNP frequency of 2.24 per kb (0.37 per kb per genotype) was observed. After filtering the initial set of base variants for indels and heterogeneity within a genotype, a set of 6,165 putative SNPs was generated. This subset of 6,165 base variants was further filtered on the basis of coverage (<6 reads) and presence of any other variation within 20 bp of flanking sequence on either side to obtain a final set of 2,956 predicted SNPs. Subsequently, 863 SNPs were advanced to assay design and scoring based on the polymorphism rate across multiple mapping populations. A final sub-set of 768 SNPs was selected based on predicted assay performance and efficiency of polymorphism detection (ESM 1). Prior to synthesis of the 768-plex SNP-OPA, a sub-set of 48 SNP loci was evaluated by Sanger sequencing, of which 33 (69 %) were successfully detected in Cassab and ILL2024.

Polymorphic markers for map construction

A total of 546 publicly available SSR markers (516 lentil EST-SSRs and 30 genomic DNA-derived SSRs) were screened on the mapping family parents for polymorphism detection (ESM 2). A total of 21 of the genomic DNA-derived SSRs (70 %) detected polymorphism in the mapping population and produced 33 marker alleles. In

Table 1 Total number of markers (SSRs and SNPs) analysed, screened for polymorphisms and assigned to genetic map locations

Marker type	Total No. of markers screened	No. of polymorphic markers	No. of mapped loci
Genomic DNA-derived SSRs	30	21	20
EST-SSRs	516	40	37
SNPs	768	264	261
Total markers	1,326	325	318

contrast, only 40 of the EST-SSRs markers (8 %) detected polymorphism (Table 1). The set of 73 segregating marker alleles were used to genotype the full progeny set, of which 12 failed to conform to the expected Mendelian ratio and were hence excluded from the final dataset that was used for linkage mapping.

A total of 508 of 768 SNP markers (66 %) amplified successfully, of which 271 were polymorphic in the Cassab × ILL2024 mapping population. As the SNP-OPA was designed to be polymorphic in multiple lentil mapping populations (data unpublished), a substantial proportion of markers was expected to be monomorphic in Cassab × ILL2024, as was observed. Of the 271 polymorphic SNP markers, 7 exhibited significant segregation distortion and were hence excluded. A final set of 325 markers (264 SNPs and 61 SSRs) were used for map construction (Table 1).

Linkage map construction and QTL analysis

A dense genetic linkage map was generated from the Cassab × ILL2024 mapping population (Fig. 1; Table 2). The proportion of marker loci amenable to mapping which coalesced into LGs was 97.5 %. A total of 318 markers, including 57 SSRs and 261 SNPs, were attributed to 10 LGs covering 1,178 cM with an average marker density of 1 locus per 3.7 cM (ESM 3), while 7 remained unlinked. The genetic map spanned seven well-supported LGs, matching the fundamental chromosome number for this species, and three satellites (consisting of 3–7 markers each), which presumably represent sets of markers that were unable to fuse with existing LGs. The major LGs were identified on the basis of genomic DNA-derived SSR locus location in comparison to prior studies. All sequences underlying the genic SNP and SSR markers were also compared to the *M. truncatula* genome to assist definition of the various LGs.

B toxicity was recorded as symptom score (on a 1–10 scale) as well as percentage necrosis in the two different years of evaluation. On the basis of the frequency distribution, a best-fitted model was calculated as representing the effect of single gene segregation (1 [tolerant]: 1 [susceptible]) (ESM 4). Narrow-sense heritability values for each measurement ranged from 0.7 to 0.8.

Marker regression analysis identified three markers (SNP_60000240, SNP_20000246 and SNP_20002998) that were associated with variation for the trait. SIM identified one major genomic region that was significantly associated with B tolerance on LG4.2 in the 7.5 cM interval between the markers SNP_20002998 and SNP_20000246. Further analysis using CIM confirmed the presence of QTLs in this genomic region with an average LOD score value of 22.5 (LOD threshold = 19.2) for the year 2008 data and 15.1 (LOD threshold = 10.4) for the year 2009 data. The QTL of large magnitude contained within this single genomic region explained 71 and 52 % of the total variance (V_p) for the two temporal environments (2008 and 2009), respectively.

Comparison of genetic linkage maps with genomes of closely related species

All LGs obtained from the genetic linkage map were compared against the chromosomes of the related plant species *M. truncatula*, soybean and *L. japonicus*, to understand the extent of colinearity and rearrangement (Fig. 2). From the current linkage map of Cassab × ILL2024, corresponding sequences were available for 298 (37 from EST-SSR and 261 from SNPs) of the 318 mapped loci. A total of 243 unique loci of lentil showed significant matches to different chromosomal positions on the *M. truncatula* (Mt v3.5) chromosomes. On LG1, 26 loci showed similarity with genomic regions on Mt Chr7, 20 with Mt Chr4 and 14 with Mt Chr8. Similarly, loci from LG2 showed maximum matches to Mt Chr4 and Mt Chr8, followed by LG3 with Mt Chr1, LG4 with Mt Chr2 and Mt Chr6, LG5 with Mt Chr1 and Mt Chr5, LG6 with Mt Chr3 and Mt Chr7, LG7 with MtChr7 (ESM 5). As the fundamental chromosome numbers of *M. truncatula* and lentil differ, it is evident from the comparison that some lentil LGs correspond to fusions of Mt chromosomes (e.g. Mt Chrs 6 and 2 corresponded to LG4). A large number of rearrangements and inversions were also observed: for instance, the lower halves of LGs 1 and 2.2 were inverted when compared to Mt Chr7 and 8, respectively, and LG3 showed some inversions in comparison with Mt Chr1. However, in general, each lentil LG displayed substantial levels of macrosynteny to one or more Mt chromosome (Fig. 2a).

In the comparison of lentil with soybean, 227 unique loci matched to different *G. max* chromosomes. Each linkage group of lentil showed hits with more than one soybean chromosome (Fig. 2b). As soybean is a palaeopolyploid, the number of chromosomes is much higher than that of lentil, and a significant number of fusions and rearrangements were, therefore, observed during comparison. A smaller number of matches (178 unique loci) were observed in the *L. japonicus* genome (Fig. 2c). As the chromosome number of lentil is higher ($2x = 14$) than that of *L. japonicus* ($2x = 12$), multiple LGs showed matches to individual chromosomes of the model species. For example, lentil LGs 6 and 7 showed significant matches to Lj Chr 1.

Selection of candidate genes for diagnostic marker development

The contig sequences underlying both of the B tolerance QTL-linked SNPs (SNP_20002998 and SNP_20000246)

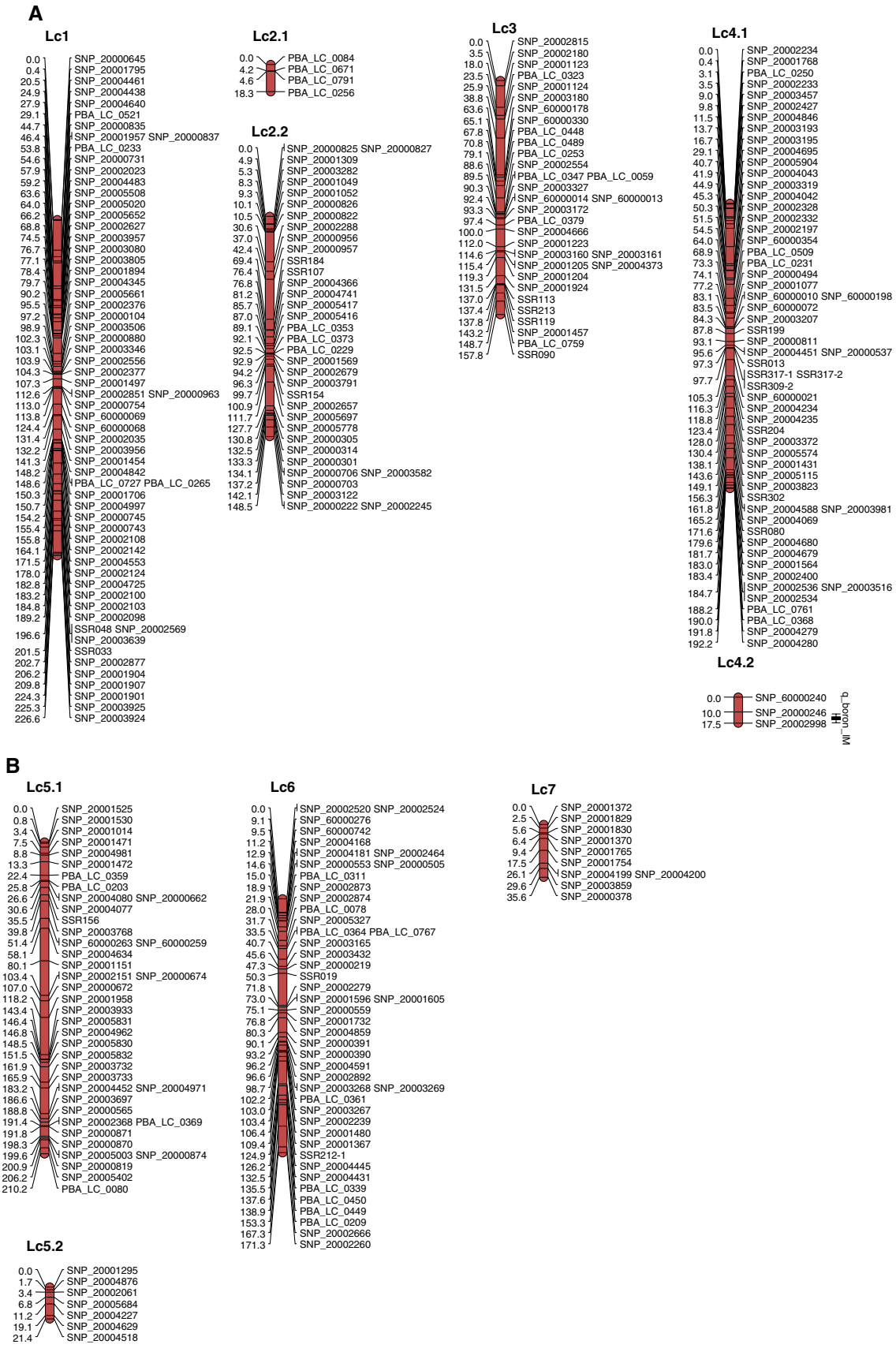

Fig. 1 Genetic linkage map derived from the Cassab (B-susceptible) × ILL2024 (B-tolerant) mapping population, depicting the single genomic region involved in B tolerance in lentil. **a** LGs 1–4; **b** LGs 5–7

Table 2 Linkage map statistics for the Cassab × ILL2024 RIL population-derived genetic map

LG	Predicted Mt chromosome	Corresponding LG in Sharpe et al. (2013)	Total No. of markers (SSRs and SNPs)	Length of LG (cM)	Average marker density (cM)
1	4/7/8	LG4/6/7	79	226.6	2.9
2.1	4	LG4	4	18.3	4.6
2.2	4/8	LG4/7	36	148.5	4.1
3	1/3	LG1/3	33	157.8	4.8
4.1	2/6	LG2	59	192.2	3.3
4.2	2	LG2	3	17.5	5.8
5.1	1/5	LG5	41	210.2	5.1
5.2	5	LG5	7	21.4	3.1
6	3	LG3	46	171.3	3.7
7	7	LG6	10	35.6	3.6

were compared with *M. truncatula*, soybean and *A. thaliana* coding sequences and whole genomes to define genomic regions containing candidate genes for putative trait-associated function. A total of 3 genes with a function related to B transport were identified from the three comparator species on the basis of annotation (Fig. 3): Boron transporter 2 (Medtr2102630.1) on Mt Chr2; BOR4, anion exchanger (Glyma15g01390.1) on Gm Chr15; and BOR4, anion exchanger (AT1G15460.1) on At Chr1. Each of these potential candidate genes has a function related to B transport or anion exchange.

Discussion

Significance of gene-associated markers

EST-SSRs have been widely used for genetic applications such as linkage mapping, population structure and marker-assisted plant breeding, due to their genic nature and high levels of cross-species transferability (Ellis and Burke 2007; Kaur et al. 2011, 2012). However, rapid progress in next-generation sequencing and related bioinformatics analysis, along with decreasing price points for implementation, has facilitated large-scale discovery and implementation of SNPs for various model and non-model plant species. As a consequence, SNPs provide the current marker of choice of molecular plant breeding applications, due to ubiquity within the genome and suitability for automated detection. SNP resources for lentil have only recently been developed from both cultivated and non-domesticated cultivars (Sharpe et al. 2013). The present study describes the discovery of a large collection of additional SNPs for potential use in lentil breeding programs. The average SNP frequency was 0.37 per kb between two genotypes, which is similar to previously reported values for *L. culinaris* (0.21) and *L. ervoides* (0.31) (Sharpe et al. 2013). These values are lower than for other related legume species such as soybean (2.7 SNPs per kb: Choi et al. 2007), but higher than for chickpea (0.043 SNPs per kb: Agarwal et al. 2012). These results indicate a very low level of genetic diversity present within cultivated lentil germplasm and, therefore, a much larger marker resource is required to assist in genomic studies.

Features of genetic linkage map

The Cassab × ILL2024 population was predominantly genotyped with EST-derived SSRs and SNPs (Kaur et al. 2011), along with a small number of genomic DNA-derived SSRs (Hamwieh et al. 2005). A highly significant difference in the polymorphism rate was observed between genomic (70 %) and EST-SSRs (8 %), consistent with similar studies (Chabane et al. 2005; Mattioni et al. 2010). However, EST-SSR markers present some advantages over genomic DNA-derived SSRs, as described above. Almost one-third of the SNP loci, which were generated from a range of cultivated genotypes, were observed to be polymorphic in the mapping population and were used in concert with SSRs to generate a high-density linkage map for lentil. The cumulative map length was larger than that reported for the most recently described SNP-based map of lentil (Sharpe et al. 2013), even though the number of markers used in the present study was comparatively smaller. High-resolution genetic linkage maps provide useful information for a number of applications such as anchoring and assembly of scaffolds and contigs derived from genome sequencing efforts, trait-dissection studies, and identification of candidate genes to assist in development of diagnostic markers for potential incorporation into breeding programs. Therefore, the SNP-based linkage maps described here and previously (Sharpe et al. 2013) could prove valuable for ordering of scaffolds derived from sequencing of lentil genomic DNA into pseudomolecules.

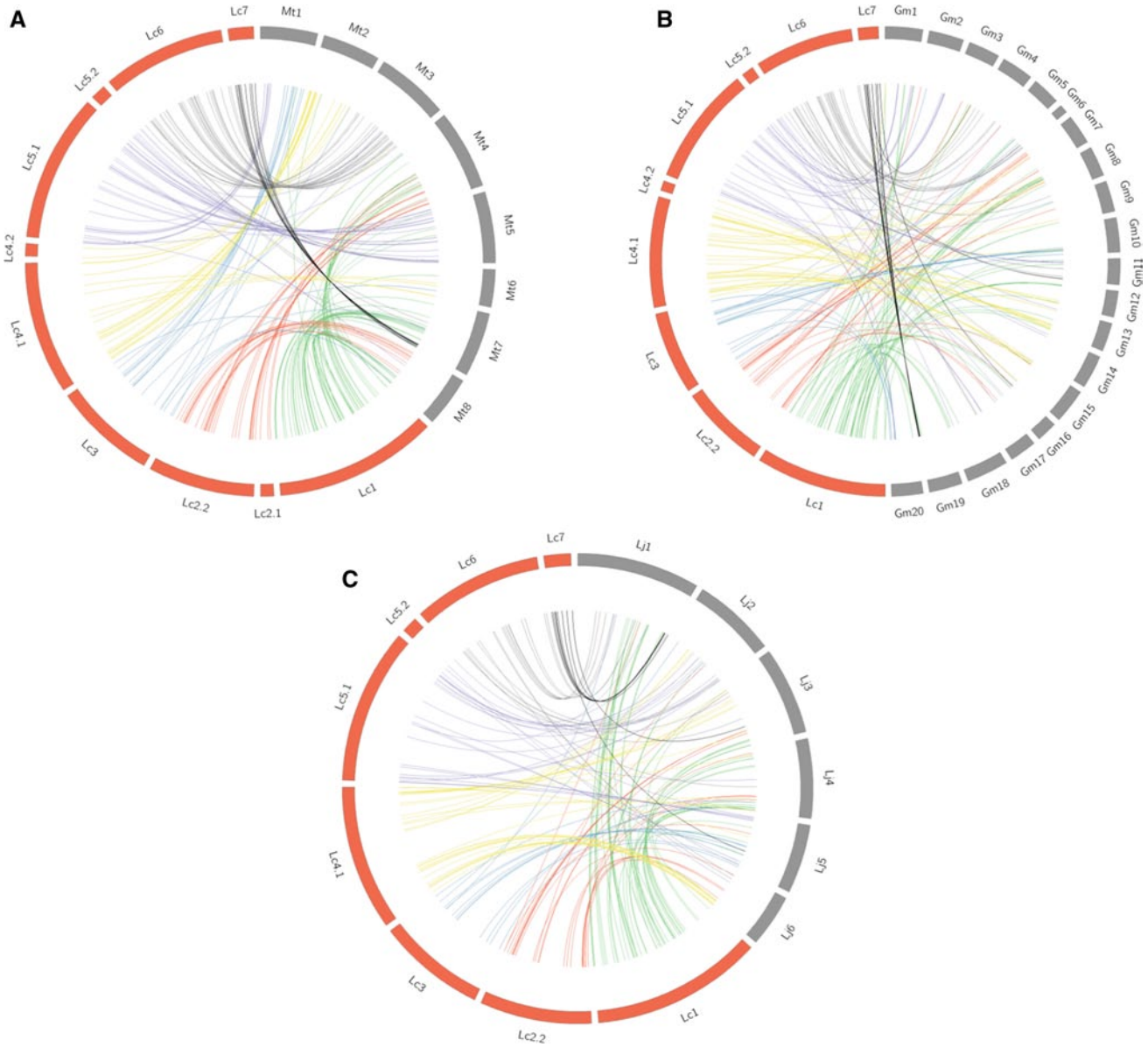

Fig. 2 Macrosyntenic relationships between the lentil genome and those of four other species: *M. truncatula* (**a**), soybean (**b**) and *L. japonicus* (**c**)

Similar approaches have been taken for recently sequenced genomes of closely related species such as chickpea (Varshney et al. 2013) and soybean (Hyten et al. 2010).

Comparative genomics

Comparative genomics studies in legumes have provided strong evidence for correlations between phylogenetic distances and incidences of chromosome rearrangements (Phan et al. 2006), and also revealed a high degree of conserved macrosynteny between members of closely related taxonomic groups. Such relationships have previously been reported between the genomes of legume species such as lentil and field pea (Weeden et al. 1992), lentil and *M. truncatula* (Phan et al. 2006), and to some extent between *M. truncatula*, field pea, mung bean, soybean and common bean (Choi et al. 2004). The present study compared the syntenic relationships between lentil and other species based on gene sequence-associated genetic markers. Overall, direct and simple relationships were demonstrated with the *M. truncatula* genome, although due to the influence of evolutionary rearrangements, matches were obtained in some instances to multiple Mt chromosomes, as observed in previous studies (Phan et al. 2006; Sharpe et al. 2013). However, some rearrangements and non-colinear relationships were also detected.

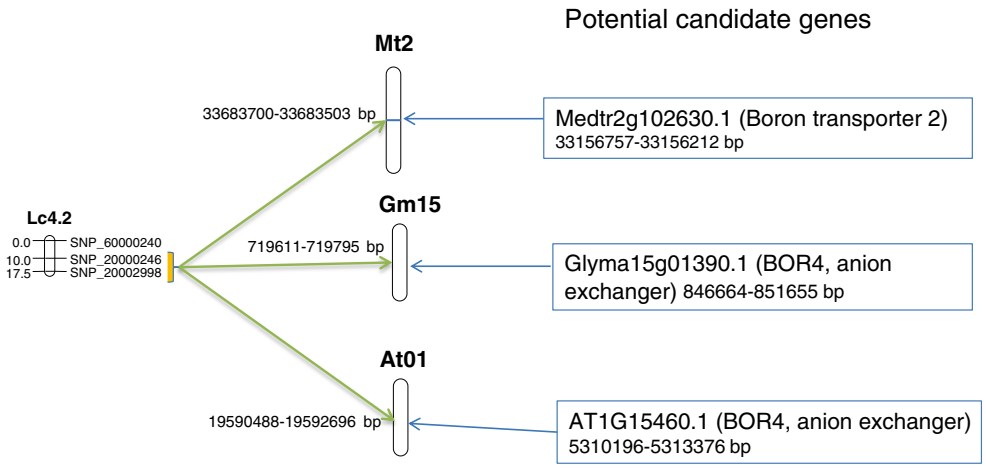

Fig. 3 Candidate gene prediction based on comparison of genomic intervals defined by sequences underpinning SNP sequences flanking the B tolerance QTL-containing genomic region in lentil to the genomes of *M. truncatula*, soybean and *A. thaliana*

Similar relationship patterns were observed for comparisons to the genomes of soybean and *L. japonicus*, which are novel to the present study. Compared to *M. truncatula*, a slightly smaller number of marker loci (16 % less) showed matches to the soybean genome, presumably reflecting a larger phylogenetic distance between these species which are located within different clades of the Papilinoideae subfamily (Doyle and Luckow 2003; Lavin et al. 2005). In this context, it is somewhat surprising that a lower proportion of similarity matches was observed to the *L. japonicus* genome as compared to soybean, given that a number of studies have consistently revealed a relatively closer phylogenetic relationship between lentil and *L. japonicus* than between lentil and soybean (Doyle and Luckow 2003; Lavin et al. 2005; Cannon et al. 2009). Both, however, are more distantly related to lentil than *M. truncatula*. Nonetheless, the observed relationships are consistent with prior knowledge of macrosynteny between *M. truncatula* and *L. japonicus*: for example, lentil LG1 exhibited matches with Lj Chr1, Lj Chr3 and Lj Chr4, which correspondingly show high similarity to Mt Chr4, Mt Chr7 and Mt Chr8 (Choi et al. 2004).

Although comparative genomics based on a relatively low number of genome-wide distributed markers has provided valuable information, caution must be applied when interpreting relationships over increasing taxonomic distance. As lentil and *M. truncatula* are both located within the inverted-repeat-loss clade (IRLC) of the galegoid sub-division of the Papilionoideae, sequence divergence between orthologous sequences in these two genomes would be expected to be minimised. In contrast, *L. japonicus* is located within the robinioid clade of the galegoid sub-division and, as described previously, should provide the next-closest comparison. Soybean is located in a separate papilionoid sub-division, the millettiod clade (Lavin et al. 2005; Cannon et al. 2009), and, therefore, provides the most distant comparator in this study. Apart from potential error in discriminating between orthologous and paralogous sequences within relatively simple genomes, comparison with soybean is further complicated by a palaeopolyploid constitution, arising from multiple whole genome duplication events (Schmutz et al. 2010). In the present study, this effect was evident from multiple matches of lentil candidate genes to presumptive duplicated regions of the soybean genome (data not shown).

Despite these concerns, assessment of conserved synteny based on several comparator genomes remains valuable for several reasons. Firstly, the degree of completion and confidence in accurate genome assembly differs between various species, and may change over time. Although *M. truncatula* provides the most obvious model genome for lentil, some apparent gaps have been identified in the genome assembly (Hand et al. 2010), and such problems have been mitigated by inclusion of *L. japonicus* and *A. thaliana* in microsyntenic comparisons. Multiple comparisons of the kind described here will hence increase confidence, especially for candidate gene identification (see below). Secondly, the most closely related species may not always display trait-specific variation for a particular character of interest, despite the likely presence of similar gene structures. In this instance, a broad survey of comparator species may assist identification of putative conserved QTLs, increasing the credibility of candidate gene selection. In the particular case of boron tolerance, comparative analysis between *M. truncatula* and *A. thaliana* has proven effective, despite the large phylogenetic divergence between the two species (Bogacki et al. 2013).

Identification of QTLs and selection of candidate genes

Attempts have been made to understand the mechanism of B toxicity tolerance in cereal species such as barley, and a role for membrane proteins involved in active efflux of B in barley roots has been suggested. One of the models proposed a mechanism similar to the action of a B transporter for xylem loading previously identified in *A. thaliana* (Takano et al. 2002). A gene for such a transporter has been cloned and characterised from barley (Hayes and Reid 2004; Schnurbusch et al. 2010). Different types of B transporters may hence be widespread in higher plants, and provide the major mechanisms for tolerance.

To date, no prior study of the genetics of B toxicity tolerance in lentil has been performed. The current study reports a single gene model, calculated from the frequency distribution pattern of phenotypic scores that was further supported by molecular genetic analysis. Two SNP loci were found to be associated with B tolerance through marker regression, and a single genomic region was detected in the interval between these SNP loci on LG4.2 using phenotypic data from both 2008 and 2009 trials, accounting for a high proportion of V_p. Based on the comparative analysis of the genomic region associated with B tolerance, three candidate genes with B transport-associated functions were identified, one from each comparator species. When aligned at the sequence level, these genes appear to be orthologous to one another and so presumably also to a lentil counterpart. The gene locations lie outside the QTL-containing interval as defined by SIM analysis. Nonetheless, significant LOD values extend beyond the linked marker locations, indicating that flanking regions may also contain plausible candidate gene loci. It is also formally possible that microsynteny with the genomes of other species may not be conserved in this chromosomal region. Isolation and characterisation of the predicted lentil B transporter gene will permit sequence analysis within diverse lentil germplasm to develop and validate a diagnostic marker for use in breeding applications.

Marker-assisted selection for B tolerance

Closely linked molecular markers for important agronomic traits have been demonstrated to be highly applicable to selection for desirable gene variants in different breeding programs. MAS is time-efficient, non-destructive and relatively error-free. The major benefit of marker-assisted selection for B tolerance in lentil would be to combine, or pyramid, genes for tolerance from different sources. As the genomic region on LG4.2 contains a QTL of large magnitude and is located in a relatively small chromosomal interval, strategies based on introgression by backcrossing from donor to recipient genetic backgrounds may be feasibly designed. Transfer of such genomic regions conferring improved B tolerance from selected parental germplasm into elite agronomic backgrounds could significantly improve the grain yield and quality of lentils grown on soil prone to B toxicity.

Acknowledgments This work was supported by funding from the Victorian Department of Environment and Primary Industries and the Grains Research and Development Council, Australia.

Conflict of interest The authors declare that they have no conflict of interest.

Ethical standards The experiments conducted in this study comply with current laws of Australia.

References

Agarwal G, Jhanwar S, Priya P, Singh VK, Saxena MS et al (2012) Comparative analysis of kabuli chickpea transcriptome with desi and wild chickpea provides a rich resource for development of functional markers. PLoS One 7(12):e52443. doi:10.1371/journal.pone.0052443

Arumuganathan K, Earle ED (1991) Nuclear DNA content of some important plant species. Plant Mol Biol Rep 9:208–218

Bennetzen JL (2000) Comparative sequence analysis of plant nuclear genomes: microcolinearity and its many exceptions. Plant Cell 12:1021–1030

Bogacki P, Peck DM, Nair RM, Howie J, Oldach KH (2013) Genetic analysis of tolerance to boron toxicity in the legume *Medicago truncatula*. BMC Genom 13:54

Cannon SB, May GD, Jackson SA (2009) Three sequenced legume genomes and many crop species: rich opportunities for translational genomics. Plant Physiol 151:970–977

Cartwright B, Zarcinas BA, Mayfield AH (1984) Toxic concentrations of B in a red–brown earth at Gladstone, South Australia. Aust J Soil Res 22:261–272

Chabane K, Ablett GA, Cordeiro GM, Valkoun J, Henry RJ (2005) EST versus genomic derived microsatellites for genotyping wild and cultivated barley. Genet Res Crop Evol 52:903–909

Choi HK, Kim DJ, Uhm T, Limpens E, Lim H, Mun JH, Kalo P, Penmesta RV, Seres A, Kulikova O et al (2004) A sequence based genetic map of *Medicago truncatula* and comparison of marker colinearity with *M. sativa*. Genetics 166:1463–1502

Choi I-Y, Hyten DL, Matukumalli LK, Song Q, Chaky JM, Quigley CV, Chase K, Lark KG, Reiter RS, Yoon M-S, Hwang E-Y, Yi S-I et al (2007) A soybean transcript map: gene distribution, haplotype and single-nucleotide polymorphism analysis. Genetics 176:685–696

Choudhary S, Gaur R, Gupta S (2012) EST-derived genic molecular markers: development and utilization for generating an advanced transcript map of chickpea. Theor Appl Genet 124:1449–1462

Doyle JJ, Luckow MA (2003) The rest of the iceberg. Legume diversity and evolution in a phylogenetic context. Plant Physiol 131(900):910

Durán Y, Fratini R, García P, Pérez de la Vega M (2004) An intersubspecific genetic map of *Lens*. Theor Appl Genet 108:1265–1273

Ellis JR, Burke JM (2007) EST-SSRs as a resource for population genetic analyses. Heredity 99:125–132

Ellwood SR, Phan HTT, Jordan M, Hane J, Torres AM, Avila CM, Cruz-Izquierdo S, Oliver RP (2008) Construction of a comparative genetic map in faba bean (*Vicia faba* L.); conservation of genome structure with *Lens culinaris*. BMC Genom 9:380

Eujayl I, Baum M, Powell W, Erskine W, Pehu E (1998) A genetic linkage map of lentil (*Lens* sp.) based on RAPD and AFLP markers using recombinant inbred lines. Theor Appl Genet 97:83–89

Genstat Committee (2002) Genstat Release 6.1. Reference manual, Parts 1–3. VSN Int., Oxford

Gupta M, Verma B, Kumar N, Chahota RK, Rathour R, Sharma SK, Bhatia S, Sharma TR (2012) Construction of intersubspecific molecular genetic map of lentil based on ISSR, RAPD and SSR markers. J Genet 91:3

Hamwieh A, Udupa SM, Choumane W, Sarkar A, Dreyer F, Jung C, Baum M (2005) A genetic linkage map of *Lens* sp. based on microsatellite and AFLP markers and the localization of fusarium vascular wilt resistance. Theor Appl Genet 110:669–677

Hand ML, Cogan NOI, Sawbridge TI, Spangenberg GC, Forster JW (2010) Comparison of homoeolocus organisation in paired BAC clones from white clover (*Trifolium repens* L.) and microcolinearity with model legume species. BMC Plant Biol 10:94

Havey MH, Muehlbauer FJ (1989) Linkages between restriction fragment length, isozyme and morphological markers in lentil. Theor Appl Genet 77:395–401

Hayes JE, Reid RJ (2004) Boron tolerance in barley is mediated by efflux of boron from the roots. Plant Physiol 136:3376–3382

Hobson KB (2007) Investigation of boron toxicity in lentil. PhD thesis, The University of Melbourne, Melbourne

Hyten DL, Cannon SB, Song Q, Weeks N, Fickus EW, Shoemaker RC, Specht JE, Farmer AD, May GD, Cregan PB (2010) High-throughput SNP discovery through deep resequencing of a reduced representation library to anchor and orient scaffolds in the soybean whole genome sequence. BMC Genom 11:38

Jefferies SP, Barr AR, Karakousis A, Kretschmer JM, Manning S, Chalmers KJ, Nelson JC, Islam AKMR, Langridge P (1999) Mapping of chromosome regions conferring boron toxicity tolerance in barley (*Hordeum vulgare* L.). Theor Appl Genet 98:1293–1303

Jefferies SP, Pallotta MA, Paull JG, Karakousis A, Kretschmer JM, Manning S, Islam AKMR, Langridge P, Chalmers KJ (2000) Mapping and validation of chromosome regions conferring boron toxicity tolerance in wheat (*Triticum aestivum*). Theor Appl Genet 101:767–777

Kaur S, Cogan NOI, Pembleton LW, Shinozuka M, Savin KW, Materne M, Forster JW (2011) Transcriptome sequencing of lentil based on second-generation technology permits large-scale unigene assembly and SSR marker discovery. BMC Genom 12:265

Kaur S, Pembleton L, Cogan N, Savin K, Leonforte T, Paull J, Materne M, Forster J (2012) Transcriptome sequencing of field pea and faba bean for discovery and validation of SSR genetic markers. BMC Genom 13:104

Kosambi DD (1944) The estimation of map distances from recombination values. Ann Eugen. 12:172–175

Lavin M, Herendeen PS, Wojciechowski MF (2005) Evolutionary rates analysis of leguminosae implicates a rapid diversification of lineages during the tertiary. Syst Biol 54:575–594

Manly KF, Cudmore RH, Meer JM (2001) Map Manager QTX, cross-platform software for genetic mapping. Mamm Genome 12:930–932

Mattioni C, Cherubini M, Taurchini D, Villani F, Martin MA (2010) Evaluation of genomic SSRs and EST-SSRs markers in genetic diversity studies in european chestnut populations. Acta Hort 866:151–156

Muehlbauer FJ, Cho S, Sarkar A, McPhee KE, Coyne CJ, Rajesh PN, Ford R (2006) Application of biotechnology in breeding lentil for resistance to biotic and abiotic stress. Euphytica 147:149–165

Nelson M, Phan H, Ellwood S, Moolhuijzen P, Hane J, Williams A, O'Lone C, Fosu-Nyarko J, Scobie M, Cakir M, Jones M, Bellgard M, Ksiarkiewicz M, Wolko B, Barker S, Oliver R, Cowling W (2006) The Wrst gene-based map of *Lupinus angustifolius* L.-location of domestication genes and conserved synteny with *Medicago truncatula*. Theor Appl Genet 113:225–238

Nuttall JG, Armstrong RD, Connor DJ, Matassa VJ (2003) Interrelationships between edaphic factors potentially limiting cereal growth on alkaline soils in north–western Victoria. Aust J Soil Res 41:277–292

Phan HTT, Ellwood SR, Ford R, Thomas S, Oliver RP (2006) Differences in syntenic complexity between *Medicago truncatula* with *Lens culinaris* and *Lupinus albus*. Funct Plant Biol 33:775–782

Phan HTT, Ellwood SR, Hane JK, Ford R, Materne M, Oliver RP (2007) Extensive macrosynteny between *Medicago truncatula* and *Lens culinaris* ssp. *culinaris*. Theor Appl Genet 114:549–558

Rubeena, Ford R, Taylor PWJ (2003) Construction of an intraspecific linkage map of lentil (*Lens culinaris* ssp. *culinaris*). Theor Appl Genet 107:910–916

Schmutz J, Cannon SB, Schlueter J, Ma J, Mitros T, Nelson W, Hyten DL, Song Q, Thelen JJ, Cheng J et al (2010) Genome sequence of the palaeopolyploid soybean. Nature 463:178–183

Schnurbusch T, Hayes J, Hrmova M, Baumann U, Ramesh SA, Tyerman SD, Langridge P, Sutton T (2010) Boron toxicity tolerance in barley through reduced expression of the multifunctional aquaporin HvNIP2;1. Plant Physiol 153:1706–1715

Schuelke M (2000) An economic method for the fluorescent labeling of PCR fragments. Nature Biotech 18:233–234

Sharpe AG, Ramsay L, Sanderson L-A, Fedoruk MJ, Clarke WE, Li R, Kagale S, Vijayan P, Vandenberg A, Bett KE (2013) Ancient orphan crop joins modern era: gene-based SNP discovery and mapping in lentil. BMC Genom 14:192

Takano J, Noguchi K, Yasumori M, Kobayashi M, Gajdos Z, Miwa K, Hayashi H, Yoneyama T, Fujiwara T (2002) *Arabidopsis* boron transporter for xylem loading. Nature 420:337–340

Tanyolac B, Ozatay S, Kahraman A, Muehlbauer F (2010) Linkage mapping of lentil (*Lens culinaris* L.) genome using recombinant inbred lines revealed by AFLP, ISSR, RAPD and some morphological markers. J Agric Biotech Sustainable Dev 2:001–006

Tullu A, Taran B, Warkentin T, Vandenberg A (2008) Construction of an intraspecific linkage map and QTL analysis for earliness and plant height in lentil. Crop Sci 48:2254–2262

Varshney RK, Song C, Saxena RK, Azam S, Yu S, Sharpe AG, Cannon S, Baek J, Rosen BD, Taran B et al (2013) Draft genome sequence of chickpea (*Cicer arietinum* L.) provides a resource for trait improvement. Nature Biotech 31:240–248

Wang S, Basten CJ, Zeng Z-B (2012) Windows QTL Cartographer 2.5. Department of Statistics, North Carolina State University, Raleigh, NC. (http://statgen.ncsu.edu/qtlcart/WQTLCart.htm)

Weeden NF, Muehlbauer FJ, Ladizinsky G (1992) Extensive conservation of linkage relationships between pea and lentil genetic maps. J Hered 83:123–129

Yau SK, Erskine W (2000) Diversity of boron-toxicity tolerance in lentil growth and yield. Genet Resour Crop Ev 47:55–62

Yau SK, Ryan J (2008) Boron toxicity tolerance in crops: a viable alternative to soil amelioration. Crop Sci 48:854–865