Single Nucleotide Polymorphism Markers Associated with Seed Quality Characteristics of Cultivated Lentil

Hamid Khazaei, Michael Fedoruk, Carolyn T. Caron, Albert Vandenberg, and Kirstin E. Bett*

ABSTRACT
The dimensions of lentil (Lens culinaris Medik.) seeds are important quality parameters that are major determinants of market preference, cooking time, and post-harvest milling quality. Knowledge of the genetic control of traits related to seed dimensions would be useful for crop improvement. The principal aim of this study was to identify single nucleotide polymorphism (SNP) markers linked to genes that control seed diameter, seed thickness, and seed plumpness. Association mapping analysis with SNP markers was used to study the seed dimensions of 138 diverse cultivated lentil accessions grown at two locations in Saskatchewan, Canada, in 2011 and 2012. Six marker–trait associations were shown to be significant for the studied seed dimension characteristics. Two SNP markers closely associated with seed diameter across locations and years identified in previous work were validated in this study. Three additional marker–seed thickness associations were identified. Using the association mapping strategy, we confirmed the presence of two genomic regions controlling seed diameter and plumpness. This information can be used worldwide as a resource for lentil seed quality improvement programs.

Core Ideas
- Only a few genomic regions control seed shape in lentil.
- Single nucleotide polymorphism markers associated with seed dimensions were identified.
- Lentil seed size is not dependent on cotyledon color.

Cultivated lentil is globally known as a quick-cooking and nutritious staple pulse crop. The seeds have a wide range of coat colors and patterns, and the cotyledons can be red, yellow, or green. These are primary determinants of consumer preferences for whole seed consumption. Equally important market characteristics are the size (mass and density) and dimensions of the lentil seeds, which influence market class, cooking time, and dehulling efficiency (e.g., Erskine et al., 1991; Wang, 2008; Subedi et al., 2017). Although seed size is a quantitative trait in lentil (Abbo et al., 1991; Verma et al., 2015), cultivated lentils have historically been classified into two main groups: microperma (small-seeded) and macrosperma (large-seeded) (Barulina 1930). Size can also be described by seed mass, which is influenced by the diameter and the thickness of the lens-shaped seeds. Most red cotyledon lentil types are small-seeded (15–50 mg seed weight) and are primarily consumed in dehulled form in South Asia and the Middle East. Green lentils typically have a seed weight of 30 to 75 mg with yellow cotyledons and pale green seed coats, and are consumed mainly in Europe and South Asia (Erskine, 1996; Muehlbauer, 2009). Seed size is measured by seed dimensional characteristics such as seed
diameter, seed thickness, and seed plumpness (the ratio of seed thickness to seed diameter).

Screening segregating populations for seed size and shape is time-consuming, involving the use of round hole and slotted sieves or low-throughput image analysis methods to determine both seed diameter and plumpness. Having molecular mapping populations have been assessed to search for genomic regions controlling seed dimensions using different molecular marker systems. In their first attempts, Abbo et al. (1991) and Fratini et al. (2007) used interspecies crosses \[L. culinaris \times Lens orientalis (Boiss.) Schmalh. and L. culinaris \times Lens ervoides Brign., and L. culinaris \times L. orientalis, respectively\] for their genetic populations, which makes detected markers less suitable for marker-assisted selection within domesticated germplasm. Later, simple sequence repeat markers linked to seed size were reported in cultivated lentil populations (Saha et al., 2013; Verma et al., 2015), although few quantitative trait loci (QTLs) were identified. In previous studies, seed weight was used as the main phenotypic assessment method, which does not distinguish differences in seed dimensions such as seed thickness and plumpness from seed diameter. Three major genomic regions were identified for seed dimension related traits on a gene-based SNP-based genetic linkage map (Fedoruk et al., 2013) derived from a cross between a small red cultivar and a large green breeding line, both from a Canadian breeding program. In that study, there was a major QTL for seed diameter coincident with the \(Y_c\) locus, which controls red/yellow cotyledon color (Slinkard, 1978) and several others associated with thickness and plumpness. In this study, we used an association mapping approach to validate those SNP markers and to identify any additional markers associated with seed dimension-related traits (seed diameter, thickness, and plumpness) in diverse cultivated lentil germplasm.

**MATERIALS AND METHODS**

**Plant Material**

A set of 138 cultivated lentil genotypes was collected from various sources that included breeding lines from the Crop Development Centre collection at the University of Saskatchewan, Canada; genebank accessions from the USDA-ARS; and genebank accessions from ICARDA (Supplemental Table S1). The germplasm set exhibited relatively large variation for seed plumpness as determined by seed diameter and thickness, and for seed coat color and pattern (Fig. 1). All three cotyledon color types (69, 67, and 2 accessions with red, yellow, and green cotyledons, respectively) were represented.

**Experimental Conditions**

Accessions were sown in the last week of May in both 2011 and 2012 at two locations in Saskatchewan, Canada: Preston research plots (52.14°N, 106.62°W) at the University of Saskatchewan, and the Saskatchewan Pulse Grower farm (52.05°N, 106.41°W). The experimental design was a randomized complete block, with three and six replicates in 2011 and 2012, respectively. The in-season precipitation (May–August) was nearly twofold higher in 2012 (375 mm) than in 2011 (201 mm); however, the mean in-season temperature was similar during both growing seasons (15.5 and 15.7°C in 2011 and 2012, respectively) (Supplemental Fig. S1). Experimental plots were seeded with 60 seeds in three rows in 1-m² plots with 0.25 m distance between rows. Seeds were harvested at the end of the season.

**Seed Dimension Measurements**

In 2012, phenotypic data were collected from a single location (Preston), since many plots grew poorly at the Saskatchewan Pulse Grower location because of above average rainfall which resulted in intermittent flooding and low seed yield. Seed diameter and seed thickness for each plot were determined using a 50-g sample of mature seeds following the method described by Hossain et al. (2010) as modified by Fedoruk et al. (2013). In brief, seed diameter was measured by passing the sample through a set of nine stacked round-hole sieves that decreased from 6.7 mm (17/64˝) to 3.6 mm (9/64˝) in 0.25 mm (1/64˝) increments. The same samples were then passed through a stacked set of seven slotted sieves that decreased from 3.2 mm (8/64˝) down to 2.0 mm (5/64˝) in 0.2 mm (0.5/64˝) increments to determine seed thickness. The seed diameter (\(D\)) and seed thickness (\(ST\)) for each sample were then calculated using the following formulæ:

\[
\text{Seed Dimension Measurements}
\]

\[
\text{Indianhead} \quad \text{CDC Greenland} \quad \text{CDC Maxim} \quad \text{CDC Astrix} \quad \text{CDC Rosie} \quad \text{I.LL 80} \quad \text{I.LL 1139} \quad \text{CDC QG-3} \quad \text{CDC KR-1} \quad \text{CDC QG-2}
\]

Fig. 1. Sample of variation in seed diameter and plumpness characterized in a set of diverse cultivated lentil lines (ticks showing units in mm).
% on sieve = \frac{\text{mass on sieve}}{\text{mass of total sample}} \times 100; \ [1]  \\
\[ D = \sum \frac{\% \text{ on round sieve}}{100} \times \text{hole size}; \ [2]  \\
\[ ST = \sum \frac{\% \text{ on slotted sieve}}{100} \times \text{slot size}; \ [3]  \\
\text{seed plumpness} = \frac{ST}{D}. \ [4]

Cotyledon color of each accession was visually confirmed (yellow, red, or green) on harvested, mature, dehulled seeds.

Genotyping and Association Mapping

A total of 1150 SNP markers were used for genotyping the 138 accessions as described in Khazaei et al. (2017). All SNP and genotyping information is available through the KnowPulse web portal (http://knowpulse.usask.ca/portal/project/Lc1536-Golden-Gate-Assay, accessed 9 Nov. 2017, and http://knowpulse.usask.ca/portal/chado/genotype/Lens, accessed 9 Nov. 2017) as well as in Supplemental Table S6. The Q-matrix and K-matrix calculations are described in Khazaei et al. (2017). TASSEL version 5.2.31 (Bradbury et al., 2007) was used to perform the association mapping analysis. Associations between SNP markers and seed dimension related traits were analyzed with a mixed linear model (MLM) by incorporating genotypes, phenotypes, the Q-matrix (Q), and the kinship matrix (K) [MLM (Q + K)]. The Bonferroni-corrected thresholds at \( \alpha = 1 \) as a cut-off and \( \alpha = 0.5 \) were calculated as \( 8.69 \times 10^{-4} \) and \( 4.35 \times 10^{-4} \), corresponding to \(-\log_{10} P(\alpha n^{-1}) \) values of 3.06 and 3.36, respectively.

Statistical Analysis

To determine if there were significant differences in seed diameter between lines with yellow and red cotyledon color, the two groups were subjected to a \( t \)-test with the R statistical package (R Development Core Team, 2016) after testing for normality. The datasets were subjected to analysis of variance for the seed dimension characters. Broad-sense heritability was calculated as the ratio of genotypic variance to phenotypic variance with SAS version 9.4 (SAS Institute, 2015).

RESULTS

Phenotypic Diversity

On average, seed diameter ranged from 3.60 to 5.92 mm and seed thickness varied from 2.03 to 2.50 mm, whereas the calculated seed plumpness rating ranged from 0.38 to 0.62 mm (Table 1). ANOVA showed there was a high genotypic effect for the seed dimension traits. The site–year (environment) effect was significant for seed thickness and plumpness but not diameter. The accession \( \times \) environment interaction was significant for all seed dimension traits (Table 2). Larger variation existed for seed diameter and seed plumpness than for seed thickness. The estimated heritabilities for seed diameter and seed plumpness were very high and that for thickness was less (Table 2). A positive correlation was observed between seed diameter and seed thickness (\( r = 0.454 \), Fig. 2). Seed diameter and seed plumpness were highly negatively correlated (\( r = -0.940 \), \( P < 0.001 \)) (Supplemental Table S2). While there were slight differences between the yellow and red cotyledon groups for seed diameter and plumpness (Table 1), it is evident that various combinations of seed size and cotyledon color exist (Fig. 2; Supplemental Table S1) and, therefore, shape is not being dictated by the cotyledon color.

Population Structure and Genetic Relationships

The STRUCTURE analysis revealed two groups (\( K = 2 \), Supplemental Table S1) which coincided with the accession’s breeding pedigree or history and geographical origin. One group was dominated by Canadian germplasm, and the other consisted of germplasm from a mix of other origins. About 72% of the pairwise relationship

<p>| Table 1. Range, mean (± SD) of seed dimension-related traits (diameter, thickness, and plumpness) in 138 cultivated lentil germplasm over locations and years. |
|---------------------------------|-----------------|-----------------|-----------------|-----------------|</p>
<table>
<thead>
<tr>
<th>Year</th>
<th>Location</th>
<th>Seed diameter (mm)</th>
<th>Seed thickness (mm)</th>
<th>Seed plumpness ratio</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Range</td>
<td>Mean (± SD)</td>
<td>Range</td>
</tr>
<tr>
<td>2011</td>
<td>Preston</td>
<td>3.62–5.94</td>
<td>4.41 (1.13)</td>
<td>2.01–2.50</td>
</tr>
<tr>
<td></td>
<td>SPG†</td>
<td>3.57–5.97</td>
<td>4.41 (0.91)</td>
<td>2.04–2.56</td>
</tr>
<tr>
<td>2012</td>
<td>Preston</td>
<td>3.61–6.09</td>
<td>4.43 (0.67)</td>
<td>2.01–2.60</td>
</tr>
<tr>
<td>2011–2012</td>
<td>–</td>
<td>3.60–5.92</td>
<td>4.42 (0.63)</td>
<td>2.03–2.50</td>
</tr>
<tr>
<td>Cotyledon color</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Red (n = 69)</td>
<td></td>
<td>3.60–5.92</td>
<td>4.25 (0.54)</td>
<td>2.06–2.46</td>
</tr>
<tr>
<td>Yellow (n = 67)</td>
<td></td>
<td>3.33–5.80</td>
<td>4.58 (0.69)</td>
<td>2.03–2.50</td>
</tr>
<tr>
<td>t-test</td>
<td></td>
<td>3.19*</td>
<td></td>
<td>11.1 ns</td>
</tr>
</tbody>
</table>

* Significant at the 0.05 level of probability.  
† SPG, Saskatchewan Pulse Grower; ns, nonsignificant.
estimates were genetically unrelated (values close to 0, Supplemental Table S3).

**Marker–Trait Associations**

A list of SNP markers significantly associated with seed dimensions at $\alpha = 1.0$ and $\alpha = 0.5$ are listed in Table 3. Two markers were found to be associated with seed diameter: LcC09638p190 and LcC08798p992 were both significant across locations and years. These markers were located on lentil chromosomes 1 and 7, respectively. Marker LcC09638p190 was also associated with seed plumpness, as was marker LcC05070p391 on chromosome 2. For seed thickness, three markers, distributed on chromosomes 1 and 7, were significant. The first two, LcC18352p485 and LcC13880p412, were significant at $\alpha = 1.0$ and the third, LcC10829p367, was significant at $\alpha = 0.5$. All the markers associated with seed thickness and seed plumpness were significant only in one location–year (Table 3).

The list of candidate genes for the significant SNP markers on Table 3, annotated in *Medicago truncatula* Gaertn. (Mt4.0v1; https://phytozome.jgi.doe.gov/pz/portal.html#!search?show=BLAST&method=Org_Mttruncatula, accessed 9 Nov. 2017) and *L. culinaris* (version 1.2, lentil JBrowse, http://knowpulse.usask.ca/portal/jbrowse/Lentil, accessed 9 Nov. 2017), are provided in Supplemental Table S4.

**DISCUSSION**

Barulina (1930) divided cultivated lentil germplasm on the basis of seed diameter into two different gene pools, small and large-seeded, which were suggested to be formed during domestication. The germplasm assessed in this study also separated into two groups but on the basis of geographical regions, not seed size characteristics. The first group consisted of mainly Canadian breeding lines and the second
group included the rest of the accessions (Supplemental Table S1; Khazaei et al., 2016). Sharma et al. (1995) and Alo et al. (2011) also reported no clear classification for different seed size classes in cultivated lentil germplasm. However, an assessment of genetic diversity in global cultivated lentil germplasm has shown that accessions predominantly reflect their agro-ecological adaptation zones (Khazaei et al., 2016).

Seed size and dimensions and their perceived uniformity play important roles in determining the market class and value of lentils. Among seed dimension parameters, seed plumpness has a substantial effect on dehulling outcomes. The plumper or more rounded lentil seeds result in greater dehulling efficiency than thinner, less plump seeds, thus increasing the value of the crop (Erskine et al., 1991; Wang, 2008; Shahin et al., 2012). The results presented here, along with previous reports (Shahin et al., 2012; Fedoruk et al., 2013), indicate a very strong negative relationship between seed diameter and seed plumpness, meaning that smaller lentil seeds are generally plumper than seeds with a larger diameter. This does not mean it is not possible to increase plumpness in larger-seeded material, as there are regions of the genome that are associated with plumpness that are independent of diameter.

Phenotyping for seed dimensions is very time-consuming and labor-intensive, and this limits our ability to look at a broader set of germplasm over more environments. To look at seed shape in an intensive manner within a breeding program is not practical. Breeder-friendly molecular markers would facilitate selection for the ideal seed dimensions as dictated by market preference for specific seed market classes. Seed image analysis technologies can be used effectively to measure seed shape in two dimensions (Firatligil-Durmus et al., 2008; Tanabata et al., 2012) which might result in higher resolution mapping of diameter.

Technologies for measuring plumpness in a high-throughput manner, however, have yet to be developed. Until such time as we can rapidly screen for plumpness through imaging technology, the markers identified in this study and that of Fedoruk et al. (2013) will remain the most efficient method for selecting for plumper lentils.

We identified three new SNP markers and validated four SNP markers previously associated with seed diameter, seed thickness, and seed plumpness in a diverse set of cultivated lentil germplasm. These newly detected markers reinforce the approach of association mapping and linkage mapping as complementary (Sonah et al., 2015; Mammadov et al., 2015). The biparental mapping population (LR-18) used by Fedoruk et al. (2013) was derived from

Table 3. Marker–trait associations for seed diameter, seed thickness, and seed plumpness across locations and years in a panel of 138 diverse cultivated lentil accessions. $R^2$ values (%) are presented in brackets after $P$ values.

<table>
<thead>
<tr>
<th>Trait</th>
<th>Marker</th>
<th>Chr</th>
<th>Location (start–end)</th>
<th>2011</th>
<th>2012</th>
<th>2011–2012</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Preston</td>
<td>SPG§</td>
<td>Preston</td>
</tr>
<tr>
<td>Seed diameter</td>
<td>LcC09638p190</td>
<td>LcChr1</td>
<td>212,502,660–212,502,965</td>
<td>$5.35 \times 10^{-4}$ (14)</td>
<td>$8.68 \times 10^{-4}$ (12)</td>
<td>$2.20 \times 10^{-4}$ (14)</td>
</tr>
<tr>
<td></td>
<td>LcC08798p992</td>
<td>LcChr7</td>
<td>75,482,664–75,484,976</td>
<td>–</td>
<td>$6.27 \times 10^{-4}$ (12)</td>
<td>$8.98 \times 10^{-4}$ (11)‡</td>
</tr>
<tr>
<td>Seed thickness</td>
<td>LcC18352p485</td>
<td>LcChr1</td>
<td>256,962,579–256,967,861</td>
<td>$5.30 \times 10^{-4}$ (12)</td>
<td>–</td>
<td>–</td>
</tr>
<tr>
<td></td>
<td>LcC13880p412†</td>
<td>Unknown</td>
<td>–</td>
<td>$6.88 \times 10^{-4}$ (12)</td>
<td>–</td>
<td>–</td>
</tr>
<tr>
<td></td>
<td>LcC10829p367</td>
<td>LcChr7</td>
<td>237,330,709–237,332,609</td>
<td>–</td>
<td>$3.37 \times 10^{-4}$ (13)</td>
<td>–</td>
</tr>
<tr>
<td>Seed plumpness</td>
<td>LcC05070p391†</td>
<td>LcChr2</td>
<td>290,127,335–290,130,364</td>
<td>–</td>
<td>$5.46 \times 10^{-4}$ (12)</td>
<td>–</td>
</tr>
<tr>
<td></td>
<td>LcC09638p190</td>
<td>LcChr1</td>
<td>212,502,660–212,502,965</td>
<td>–</td>
<td>$8.84 \times 10^{-4}$ (11)‡</td>
<td>–</td>
</tr>
</tbody>
</table>

† This marker was not included in Fedoruk et al.’s (2013) genetic map.
‡ Single nucleotide polymorphism markers with $P$-values slightly above the cut-off of $8.69 \times 10^{-4}$ ($\alpha = 1$).
§ SPG, Saskatchewan Pulse Grower; Chr, chromosome.
two Canadian breeding lines and may share the same alleles at the additional seed thickness QTLs identified by the association mapping. These additional markers might also be a result of the increased number of SNP markers used for this study compared with our linkage map study. Unlike segregating populations, the association panels used in association mapping consist of individuals that are not directly related to one another, with a larger number of recombination events that should result in finer resolution mapping. A previous analysis (Fedoruk 2013) on the set of SNP markers used in this study showed that there was a sufficient level of linkage disequilibrium for association mapping ($r^2 = 0.1$ with 5 cM map distance).

Genetic linkage mapping studies in lentil have revealed that cotyledon color is associated with seed diameter and seed weight in certain populations; in other words, seed with larger diameter had yellow cotyledons and smaller seed had red cotyledons (Abbo et al., 1991; Saha et al., 2013; Fedoruk et al., 2013). This could simply have arisen from the limited levels of recombination in biparental populations that are developed between parental lines that are small with red cotyledons and large with yellow cotyledons. Our results clearly show no association between cotyledon color and seed dimension characteristics based on 138 genetically variable lentil lines (Table 1; Fig. 2), suggesting that this is indeed the case. Our results agree with Tullu et al. (2001), who found different combinations of seed size and cotyledon color in lentil.

Among the significant markers, the M. truncatula and L. culinaris annotations for the sequence of marker Lc08798p992 was a P450 family 78 protein (Supplemental Table S4). The cytochrome P450 (CYP) family is considered to be one of the largest plant protein families (Adamskia et al., 2009). Work on several plant species has shown that seed development and its size characteristics are regulated by CYP78A such as AtCYP78A5, AtCYP78A6, and AtCYP78A9 in Arabidopsis thaliana (L.) Heynh. and GmCYP78A10 in soybean (Glycine max (L.) Merr.) (Wang et al., 2015). The CYP78A5 in A. thaliana is a maternal regulator of seed size (Adamskia et al., 2009). In rice (Oryza sativa L.) and wheat (Triticum aestivum L.), CYP78A3 and TaCYP78A5 proteins affected seed size and shape (Tang et al., 2016; Ma et al., 2015, 2016). Furthermore, in tomato (Solanum lycopersicum L.) (Chakrabarti et al., 2013) and jatropha curcas L. (Tian et al., 2016), the same protein was involved in the regulation of seed size. None of the other markers are in genes that are annotated as being related to seed shape or size. It is much more common to phenotype seed of other species in terms of weight rather than dimensions. Generally, these species are either spherical, as in A. thaliana and other Brassicaceae, or irregularly shaped, as in the case of chickpea (Cicer arietinum L.) and M. truncatula. There have been QTLs reported for seed weight on M. truncatula chromosomes 2, 4, and 6 (Sankaran et al., 2009) but none of these regions is in the regions that are syntenic with the QTLs that were found in lentil.

Seed development in plant species involves the coordinated growth of embryo, endosperm, and maternal tissue (Adamskia et al., 2009; Venglat et al., 2014; Li and Li, 2015). Generally, in legumes, two phases (cell division and cell expansion) affect seed diameter and seed thickness and, consequently, seed plumpness (Domoney et al., 2006). Cell division, the first developmental phase, depends on the embryo genotype, which controls the embryo cell number. This phase is largely insensitive to environmental factors. The cell expansion phase is affected by a range of factors including environmental regulators, which modulate carbohydrate partitioning. This affects the seed filling and therefore seed thickness, and is highly regulated by the interaction of individual genotypes and the environment during seed filling. In lentil, the pod remains flat while the seed coat develops to maximum size within the pod. After this, the seeds begin to fill, causing the pod to inflate. The seed filling phase takes a longer time, which probably makes it more susceptible to environmental influence for a longer time than during the growth phase when seed diameter is established. The lower estimated heritability for seed thickness and higher site–year effect than for seed diameter is explained by the large impact of environmental factors on this trait. Seed plumpness is a derived trait from seed thickness and diameter. Since it is more highly correlated with seed diameter, this translates to the high heritability of this trait.

**CONCLUSIONS**

Our results confirmed the large variation for seed dimension characteristics in cultivated lentil, which is independent from cotyledon color. The SNP markers for seed diameter and seed plumpness identified in this study via an association study are valuable resources for lentil seed quality improvement programs. The molecular markers that are linked with highly heritable traits such as seed plumpness may be useful for marker-assisted selection, especially for selection at early generations in multiparent crosses and for selecting new parents from diverse germplasm.

**Supplemental Information Available**

Supplemental Fig. S1. Total precipitation (mm) and mean temperature (°C) during the 2011–2012 growing seasons in the Saskatoon area. Statistics were extracted from the Environment Canada website (http://climate.weather.gc.ca/climate_data/daily_data_e.html?StationID=47707, accessed 9 Nov.).

Supplemental Table S1. Descriptions and population structure of lentil accessions used in this study.

Supplemental Table S2. Correlation coefficients of seed dimension-related traits.

Supplemental Table S3. Kinship matrix.

Supplemental Table S4. Candidate genes within syntenic regions for SNP markers associated with seed dimension traits in lentil.

Supplemental Table S5. Phenotypic data.

Supplemental Table S6. Genotypic calls.
Conflict of Interest Disclosure

The authors declare that there is no conflict of interest.

Acknowledgments

This research was supported by the Natural Sciences and Engineering Research Council of Canada, the Saskatchewan Pulse Growers, the Saskatchewan Ministry of Agriculture, and the University of Saskatchewan. The authors thank Rob Stonehouse, Brent Barlow, Alison Sackville, Donna Lindsay, and Lacey-Anne Sanderson for their kind technical support during the experiments.

References


R Development Core Team. 2016. R: A language and environment for sta-

Saha, G.C., A. Sarker, W. Chen, G.J. Vandemark, and F.J. Muehlbauer. 2013. Inheritance and linkage map positions of genes conferring agronomo-