Journal of Heredity 2011:102(3):315–329
doi:10.1093/jhered/esr015
Advance Access publication March 30, 2011

Leveraging Genomic Resources of Model Species for the Assessment of Diversity and Phylogeny in Wild and Domesticated Lentil

Fida Alo, Bonnie J. Furman, Eduard Akhunov, Jan Dvorak, and Paul Gepts

From the International Center for Agriculture in the Dry Areas (ICARDA), Aleppo, Syria (Alo); the Department of Plant Sciences/MS1, University of California, Davis, CA 95616-8780 (Furman, Akhunov, Dvorak, and Gepts); the International Center for Maize and Wheat Improvement, Mexico, DF, Mexico (Furman); and the Department of Plant Pathology, Kansas State University, Manhattan, KS (Akhunov).

Address correspondence to Paul Gepts at the address above, or e-mail: plgepts@ucdavis.edu.

Abstract

Advances in comparative genomics have provided significant opportunities for analysis of genetic diversity in species with limited genomic resources, such as the genus *Lens*. *Medicago truncatula* expressed sequence tags (ESTs) were aligned with the *Arabidopsis thaliana* genome sequence to identify conserved exon sequences and splice sites in the ESTs. Conserved primers (CPs) based on *M. truncatula* EST sequences flanking one or more introns were then designed. A total of 22% of the CPs produced polymerase chain reaction amplicons in lentil and were used to sequence amplicons in 175 wild and 133 domesticated lentil accessions. Analysis of the sequences confirmed that *L. nigricans* and *L. ervoides* are well-defined species at the DNA sequence level. *Lens culinaris* subsp. *odemensis*, *L. culinaris* subsp. *tomentosus*, and *L. lamottei* may constitute a single taxon pending verification with crossability experiments. *Lens culinaris* subsp. *orientalis* is the progenitor of domesticated lentil, *L. culinaris* subsp. *culinaris* (as proposed before), but a more specific area of origin can be suggested in southern Turkey. We were also able to detect the divergence, following domestication, of the domesticated gene pool into overlapping large-seeded (megasperma) and small-seeded (microsperma) groups. Lentil domestication led to a loss of genetic diversity of approximately 40%. The approach followed in this research has allowed us to rapidly exploit sequence information from model plant species for the study of genetic diversity of a crop such as lentil with limited genomic resources.

Key words: *diversity, domestication,* Lens, *phylogeny, single nucleotide polymorphism*

The domestication of lentil in the Fertile Crescentdates back to the very beginning of agriculture in that region (Ladizinsky 1979; Webb and Hawtin 1981; Harlan 1995; Zohary and Hopf 2000). Lentil is therefore one of the oldest domesticated grain legumes. The archaeological record places lentil domestication in Syria and Turkey approximately 8500 BC (Cubero 1981). Lentil remains dating back to 8500–7500 BC and 6700 BC were found at Tel Mureybit, Syria and Çayonü, Turkey, respectively (Cubero 1981; Zohary and Hopf 2000). Lentil cultivation spread throughout the Mediterranean region, Asia, and Europe during the Bronze Age (Youngman 1968). The domesticated form of lentil is *Lens culinaris* Medik. subsp. *culinaris* (Fabaceae, Subfamily: Papilionaceae, Webb and Hawtin 1981). *Lens culinaris* subsp. *culinaris* is conspecific with its 3 putative wild progenitors: *L. culinaris* Medikus subsp. *orientalis* (Boiss.) Ponert; *L. culinaris* Medikus subsp. *odemensis* (Ladizinsky) Ferguson, Maxted, van Slageren, and Robertson; and *L. culinaris* subsp. *tomentosus* (Ladizinsky) Ferguson, Maxted, van Slageren and Robertson. In addition, there are 3 wild species within the genus *Lens*: *L. nigricans*, *L. ervoides*, and *L. lamottei* (Ferguson et al. 2000). All *Lens* species are diploid ($2n = 2x = 14$; Sharma et al. 1995). *Lens culinaris* subsp. *orientalis* is fully cross-compatible with domesticated lentil (Muehlbauer and Slinkard 1981; Robertson and Erskine 1997) and has been proposed as the putative ancestor of the domesticated lentil (Barulina 1930; Mayer and Soltis 1994). Understanding of species relationships in the genus and diversity of the taxa is important for both basic biological reasons and for the improvement of lentil cultivars.

Traditional species identification in this genus based on morphological characters has been unreliable as all 7 taxa are

morphologically very similar to each other and very few traits can be considered as diagnostic (Galasso 2003). Hence, there has been a focus on molecular analysis to dissect inter- and intraspecific relationships. Previous studies of molecular diversity species relationships within *Lens* have generally shown quite low levels of polymorphism. The outcomes of analyses with several types of DNA-based markers show that *Lens* species can be readily distinguished and that *L. culinaris* subsp. *orientalis* is the presumed ancestor of domesticated lentil (e.g., nuclear anonymous restriction fragment length polymorphisms [RFLPs]: Havey and Muehlbauer 1989; rDNA internal transcribed spacers: Mayer and Bagga 2002; Sonnante et al. 2003; amplified fragment length polymorphisms: Sharma et al. 1996). The close molecular relationship between *L. culinaris* subsp. *culinaris* and *L. culinaris* subsp. *orientalis* confirms earlier observation on morphological traits, chromosomal arrangement, and crossability (Ladizinsky 1999). However, further information on the genetic diversity of the genus *Lens*, in general, and on the domestication process of lentil, in particular, is needed to better utilize genetic diversity for conservation and breeding purposes in this crop.

Because lentil lags behind other legumes and cereals in economic significance and has a large genome (4063 Mbp/1C, Arumuganathan and Earle 1991), limited genomic resources have been developed for it so far (but see Tullu et al. 2008; Mustafa et al. 2009; Scippa et al. 2010). However, more extensive genomic resources exist for several other species in the Fabaceae, such as soybean, *Medicago truncatula*, *Lotus japonica*, and *Phaseolus vulgaris*, and the more distantly related *Arabidopsis thaliana* (Brassicaceae). Because it is not currently feasible to develop genomic resources for all economically important species, strategies are needed for leveraging resources of model species for genomics of distantly related species for which such resources do not exist.

For instance, wheat represents a species with limited genomic resources within the grass family. To sequence large numbers of gene fragments in wheat and develop single-nucleotide polymorphisms (SNPs), wheat expressed sequence tags (ESTs) were compared with the rice genomic sequence to identify conserved exonic sequences flanking introns in the ESTs (Blake et al. 2004). Thousands of conserved primers (CPs) for polymerase chain reaction (PCR) and sequencing of gene fragments were developed for wheat and its relatives (You et al. 2009). Because the CPs were based on detection of conserved sequences by family wide sequence comparisons, they are expected to be usable in many other grass species beside wheat and the tribe Triticeae.

Here we report on a test of the assumption that CPs developed by this strategy can be used for genomics of species across the legume family. Specifically, ESTs of *M. truncatula* and genome sequence of *A. thaliana* were used for the development of CPs, which were then used for study of nucleotide diversity and phylogenetic relationships in wild and domesticated lentil. *M. truncatula* (Family: Fabaceae, Subfamily: Papilionoideae; Clade: Hologalegina [cool-season legumes]; Tribe: Trifolieae) is a relatively close relative of lentil (Family: Fabaceae; Subfamily: Papilionoideae; Clade: Hologalegina [cool-season legumes]; Tribe: Fabeae); the 2 genera diverged some 24 million years ago (Ma) (Lavin et al. 2005). *Arabidopsis thaliana* (Brassicaceae, Order: Brassicales) is classified into the Malvideae, which is a sister clade of the Fabideae, which includes the order Fabales. We reasoned that exonic regions that are conserved between *M. truncatula* and *A. thaliana* are very likely conserved across the entire Fabaceae, including the genus *Lens*.

The use of model species is predicated in part on their potential role as a source of genomic information for species of lesser or more local importance that include many crops of nutritional importance to humans. For example, information from model species can be used to identify candidate genes underlying important agronomic traits in target species (e.g., Gepts 1993; Horn et al. 2005; Kwak et al. 2008; McClean et al. 2008, 2010). Alternatively, sequence information from model species can assist in developing markers based on conserved exon sequences that bracket more variable intron sequences. This approach was first developed in mammalians (so-called CATS [comparative anchor-tagged sequences] markers; e.g., Adjei et al. 2005) and was later applied in plants such as the legumes (e.g., Fredslund et al. 2006; Hougaard et al. 2008).

In research presented here, we used the intron-bracketing marker-development approach to develop information on the organization of genetic diversity in lentil. To this day, there is limited genomic information available for this species. We show here that using sequence information from 2 model species—one in the same botanical family, *M. truncatula*, and the other in a more distant family (*A. thaliana*, *Brassicaceae*)—we can obtain sequence information within lentil. Using this information, variation for SNPs was studied in this crop species. Although SNPs have been shown to reveal high levels of genetic diversity, the use of direct sequencing to detect SNPs generally requires prior sequence information for the target species. The study presented here demonstrates a methodology that does not require sequence information for the target species (lentil in this case) but only ESTs for a related species in the same botanical family (*M. truncatula* in this case); thus, it makes genomic resources of model species accessible to many species for which such resources have not yet been developed.

Material and Methods

Plant Material

A total of 308 accessions were used in this study (Table 1). A list of individual accessions, their passport data, and some environmental variables are provided as Supplementary Table 1. Seeds of 175 wild and 133 domesticated *Lens* types were included in this study and were kindly provided by the International Center for Agriculture Research in Dry Areas (ICARDA), Aleppo, Syria. After planting of seeds in

Table 1 *Lens* species, geographic origins, and number of accessions used in this study

Taxon	Geographic origin	No. of accessions
Lens culinaris subsp. *culinaris* (CUL)	Africa	1
	Asian Peninsula	4
	Central Asia and Caucasus	7
	Eastern Europe	9
	North Africa	17
	North America	2
	South America	4
	South Asia	7
	West Asia	74
	Western Europe	7
	Total	132
*Lens culinaris*subsp. *odemensis* (ODE)	West Asia	12
*Lens culinaris*subsp. *orientalis* (ORI)	Central Asia and Caucasus	10
	Eastern Europe	1
	West Asia	72
	Total	83
Lens culinaris subsp. *tomentosus* (TOM)	West Asia	8
Lens ervoides (ERV)	Eastern Europe	15
	West Asia	28
	Western Europe	2
	Total	45
Lens lamottei (LAM)	West Asia	1
	Western Europe	3
	Total	4
Lens nigricans (NIG)	Eastern Europe	9
	West Asia	7
	Western Europe	8
	Total	24
General total		308

a growth chamber, approximately 1 g of leaf tissue was harvested from each accession and freeze-dried in liquid nitrogen. DNA was extracted using the modified hexadecyltrimethylammonium bromide procedure (Murray and Thompson 1980). DNA extraction was carried out in the Genetic Resources Unit of ICARDA in Syria. DNA samples were quantified using a spectrophotometer at wavelength 260 nm and checked by electrophoresis in 1% w/v agarose gels. DNA sequencing was performed in the Department of Plant Sciences at the University of California, Davis.

Design, Selection, and Testing of PCR Primers

CPs were developed based on sequence comparisons between *M. truncatula* (a model species for the legume family) and *A. thaliana*. Sequences anchored in neighboring exons were detected and used to design primers for sequencing of introns and bordering exonic regions of *Lens* accessions. *Medicago truncatula* ESTs were downloaded from the *M. truncatula* database (http://plantta.jcvi.org/). The ESTs were compared with *A. thaliana* genomic sequence to detect the exon/exon junctions (http://www.ncbi.nlm.nih.gov/BLAST). Approximately 100 primer combinations were designed using the GeneTools primer design program (Beisvag et al. 2006) and tested using PCR on lentil. Of these, a total of 22 primer combinations successfully amplified lentil genomic DNA and were chosen for analysis (Supplementary Table 2).

PCR was performed in a 40 µl reaction volume including 150 ng genomic DNA, 0.2 U/µl AmpliTaq Gold DNA polymerase (Applied Biosystems, Foster City, CA), reaction buffer containing 1.5 mM $MgCl_2$ and supplemented with 2 mM dNTPs and 20 pmol of each primer. Thermal cycling were performed in a Perkin Elmer 9700 (Applied Biosystems) using the following touchdown regime: 98 °C for 5 min, 10 cycles of 10 s 96 °C, 5 s at 63 °C and 58 °C, and 2 min at 72 °C. Annealing was carried out starting at 63 °C for the first cycle and decreasing by 0.5 °C per cycle until the annealing temperature was reached. This was followed by 35 cycles of 20 s 96 °C, 20 s at 58 °C, and 2 min at 72 °C for extension, ending with 10 min at 72 °C, and storage at 4 °C.

Five microliters of PCR products were treated with 3 µl of ExoSAP-IT reagent (Exonuclease I and Shrimp Alkaline Phosphatase). Three microliters of the treated PCR product was then sequenced from both ends using the same primers as used in the PCR amplification with the BigDye Terminator v3.1 cycle sequencing kit (Applied Biosystems). Thermal cycling conditions included 40 cycles with 5 min at 98 °C and 1 min at 94 °C, 15 s at 50 °C, 4 min at 60 °C, sequence reactions were precipitated using 3 M Na acetate and 2.5 vols of 100% ethanol. Precipitates were washed once with 70% ethanol, dried, and dissolved in 12 µl of Hi-Di™

Formamide (Applied BioSystems - Life Technologies, Foster City, California). PCR products amplified from genomic DNA templates were sequenced in both directions using an ABI PRISM 3700 sequencer (Applied Biosystems). Base calling and quality trimming of ABI chromatograms was performed with Phred (Ewing et al. 1998).

Sequence Analysis

Vector NTI V.9 was used to align the contig assembly. This software package assigns quality scores from ABI trace files and maintains association between individual sequences within contigs and their ABI files. We found this helpful in distinguishing between sequence polymorphisms and sequencing errors generated by poor-quality data. To identify the individual SNPs, we used the software BioEdit V.5.0.6 (Hall 1999). The resulting individual sequences were concatenated to create a single sequence for further analyses. To distinguish exon and intron regions, we first identified these regions for each gene locus. We then compared lentil genes and the ESTs from *M. truncatula* using the website http://blast.ncbi.nlm.nih.gov/Blast.cgi to detect the exons and introns of the lentil sequences.

STRUCTURE, Sequence Diversity, Linkage Disequilibrium and Neutrality Analyses

STRUCTURE analyses (Pritchard et al. 2000) were conducted with version 2.3.1 of the software, downloaded from http://pritch.bsd.uchicago.edu/software/structure_v.2.3.1.html. Because STRUCTURE does not deal with tightly linked markers (such as SNPs within a gene sequence), we used haplotypes defined by SNPs as markers instead of individual SNPs. Preset K values ranged from K2 to K15. For each K level, 10 independent runs were conducted. For each run, the length of the burn-in period was 10 000 and the number of Markov chain Monte Carlo repetitions was 50 000 per run. The analysis was conducted assuming admixture and correlated allele frequencies. The most probable partition of our sample was determined using the Evanno et al. (2005) test by submitting the STRUCTURE results files to the Structure Harvester web site (http://taylor0.biology.ucla.edu/struct_harvest/). The final number corresponded to the K value with the highest value of the ad hoc statistic Delta K. Membership in the various groups for the chosen K value was determined by the run with the highest likelihood.

Measures of genetic diversity, linkage disequilibrium (LD), and departures from neutrality (Tajima; Fu and Li) were determined using DnaSP (Librado and Rozas 2009). Nucleotide diversity π is the average number of nucleotide differences per site between 2 sequences. It is expressed as

$$\hat{\pi} = \sum_{i<j} \frac{\pi_{ij}}{n_i},$$

where π_{ij} is the proportion of different nucleotides between the ith and jth DNA sequences and n_c the total number of sequence comparisons $[n(n-1)/2]$ (Nei 1987). Nucleotide diversity θ is the number of polymorphic nucleotide sites per nucleotide site and is expressed as $\theta = N_e\mu$, where N_e is the effective population size and μ is the mutation rate per nucleotide site and per generation (Watterson 1975).

Species relationships were elucidated by the neighbor-joining (NJ) algorithm using MEGA software version 4.0 (Tamura et al. 2007). Bootstrap analysis was carried out using10 000 randomizations (Felsenstein 1985).

Results

Sequence Analyses

A total of 22 genes (Supplementary Table 2) were sequenced giving a total of 12 817 base pairs per accession (Supplementary Table 3). The number of base pairs per locus ranged from 265 to 1082, with an average size of 583. The average number of polymorphic sites per gene was 41, ranging from 11 to 122 (Supplementary Table 3). The lengths of sequenced intronic regions ranged from 0 to 748 base pairs, with an average size of 349. The number of polymorphic sites per intron ranged from 0 to 109, with an average number of 34 polymorphic sites per intron. The length of sequenced exonic regions ranged from 99 to 444 bp, with an average size of 217 bp. The number of polymorphic sites per exon ranged from 0 to 22. The average number of polymorphic sites per exon was 7 (Supplementary Table 3). Not surprisingly, our results showed, therefore, that nucleotide diversity was 4–5 times greater in introns compared with exons (Supplementary Table 4).

STRUCTURE Analyses

The sample of 308 *Lens* accessions was subjected to 2 rounds of STRUCTURE analysis, using version 2.3.1 (Pritchard et al. 2000) according to the parameters described in Materials and Methods. In a first round, an analysis was conducted on the sample omitting prior information on taxon membership. The Evanno et al. (2005) Delta K test showed a marked peak at $K = 8$ (Supplementary Figure 1), suggesting that the sample studied actually consisted of 8 subgroups. A bar graph of posterior membership probabilities showed well-separated subgroups, with a low level of admixture, as posterior membership probabilities were well above 90% for a majority of the members in each group (Figure 1). Two pairs of subgroups were exceptions to this pattern, namely K2 versus K6 (domesticated) and K5 versus K7 (wild). Each of these pairs showed reciprocal admixture (Figure 1).

Inspection of the composition of each of the 8 groups revealed a general correspondence with taxonomic boundaries and status (wild or domesticated) (Table 2). All groups, except K3, were composed primarily of a single taxon. In contrast, the K3 group included all taxa, except domesticated *Lens*. Nevertheless, there was a majority taxon within K3, namely *L. odemensis*. *Lens culinaris* subsp. *orientalis*, the nearest wild relative of *L. culinaris* subsp. *culinaris*, was included in 5 K groups. In 3 of these (K4, K5, and K7), it was the majority component. K4, K5, and K7 showed

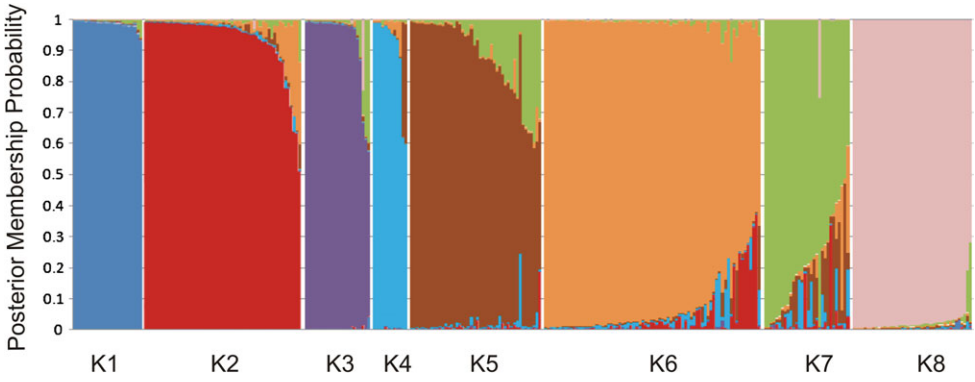

Figure 1. STRUCTURE analysis of DNA sequence haplotype variation among 308 *Lens* accessions.

a differential geographic distribution (Figure 2). Group K4 (*orientalis*-I) was distributed in Central Asia, between Turkmenistan and Tadjikistan. Its distribution did not overlap with those of the K5 (*orientalis*-II) and K7 (*orientalis*-III) subgroups. The distribution of the latter groups overlapped in the Levant, that is, Israel, Lebanon, Palestine, western Syria, and part of southern Turkey. Subgroup K5 was partly distributed in eastern Turkey and Syria as well, with one additional accession each from Armenia and Iran. In contrast, subgroup K7 included additional accessions from western and northern Turkey (Figure 2).

Domesticated lentil (*L. culinaris* subsp. *culinaris*) was included in 2 K groups, K2 (in which it was the only taxon represented) and K6 (in which there was a single exception, an *L. ervoides* accession) (Table 2). These 2 groups had different, although overlapping, distributions (Figure 3). K2 was distributed primarily in southwestern Asia, the Mediterranean and in northeastern Africa (along the Nile valley from Egypt to Ethiopia). In contrast, K6 was distributed in southwestern and southern Asia.

Relationship between K Group Membership and Other Traits

Lens orientalis

Because of their importance in the present study, the K groups containing wild and domesticated *L. culinaris* were subjected to an additional round of STRUCTURE analysis. Groups K4-*orientalis*-I, K5-*orientalis*-II, and K7-*orientalis*-III, the groups in which *orientalis* entries formed a majority were subjected to a joint STRUCTURE analysis. According to the Delta K analysis, 3 subgroups could be identified (Supplementary Figure 2).

The results showed that the K4 group remained separate from the K5 and K7 groups, consistent with its geographic isolation. The K5 and K7 groups were redistributed into 2 new groups, K5K7.1 and K5K7.2, respectively. However, neither of these 2 new groups included members with posterior membership probabilities above 70% (Supplementary Figure 3). Thus, this second round of STRUCTURE analysis did not provide a better separation among *orientalis* accessions belonging to the K5 and K7 groups but confirmed the distinctness of the Central Asian *orientalis* accessions compared with their Southwestern Asian relatives.

The availability of passport data for some *orientalis* accessions provided us with the opportunity to compare local environmental data among the K4, K5, and K7 groups (Table 3). First, the data showed the distinctness of the environment in which the K4 group was distributed compared with those of the K5 and K7 groups. On average, K4 accessions grew at higher altitude, in an environment with a 3–4 °C cooler maximum and minimum temperature and overall less aridity (at least compared with the K7 group). The environmental differences between K5 and K7 subgroups were much smaller. The K5 group grew in environments that had on average a 1 °C lower maximum temperature and were slightly less arid (Table 3).

Table 2 Taxonomic composition of the K groups identified by the STRUCTURE analysis

	K group number and color								
Lens taxon	K1	K2	K3	K4	K5	K6	K7	K8	Total
culinaris		55				75	1	1	132
odemensis			12						12
orientalis			2	10	43		26	2	83
tomentosus			5		3				8
ervoides	1		1	2		1	1	39	45
lamottei			2				2		4
nigricans	23		1						24
Total	24	55	23	12	46	76	30	42	308

Lens culinaris

The domesticated members included in the K2 and K6 subgroups had a geographic distribution that was reminiscent of the subdivision identified previously on the basis of seed weight, namely the *megasperma* (larger seeded) and *microsperma* (smaller seeded) groups (e.g., Chahota et al. 2007). To determine whether the K2–K6 subdivision corresponded to the *megasperma–microsperma* distinction, seed weight data were obtained for accessions included in the 2 subgroups (K2: $n = 55$; K6: $n = 75$). Subgroup averages

Figure 2. (**A**) Distribution of the 3 main *L. culinaris* subsp. *orientalis* groups identified by the STRUCTURE analysis. Color symbols: K4-*orientalis*-I (teal), K5-*orientalis*-II (red), and K7-*orientalis*-III (green). (**B**) Three K7 *orientalis* accessions most closely related to domesticated lentils in southern Turkey (see also Figure 5). Based on Google Map data (2011 Google).

(±standard deviation) were similar: 3.93 g/100 seeds ± 1.14 for K2 and 3.68 g/100 seeds ± 1.22 for K6. This difference was not statistically significant according to a two-tailed ($P = 0.23$) or a one-tailed *t*-test ($P = 0.11$). A closer examination of the seed weight distribution revealed, however, a difference between the 2 groups (Figure 4). Whereas the K6 group was more abundant among smaller seed weight classes, the K2 group was represented more frequently in larger seed weight classes. This frequency distribution difference was statistically significant according to a chi-square test ($\chi^2 = 27.13$, 7 degrees of freedom; $P = 0.0003$). Thus, the distributions of K2 and K6 seed weights showed considerable overlap, but K6 accessions generally included a higher frequency of smaller seeds. The K2 and K6 subgroups, therefore, correspond approximately to the *megasperma* (K2-*culinaris*-M) and *microsperma* (K6-*culinaris*-m) groups, respectively.

Sequence Divergence and Diversity in Relation to STRUCTURE K Groups

After alignment and concatenation as described in the Materials and Methods, the sequence diversity and divergence were analyzed within and among K groups identified by STRUCTURE (Pritchard et al. 2000). Using the MEGA software, version 4.1 (Tamura et al. 2007; available at http://www.megasoftware.net/), the number of base substitutions per site was calculated using the Kimura 2-parameter method (Kimura-2, henceforth). All positions containing gaps were eliminated from the data set, resulting in a total of 11803 positions in the data set.

The values of the Kimura-2 distances among K groups were generally as expected (Table 4). For example, *L. nigricans* was the most distant taxon to *L. culinaris*, followed by *L. ervoides* (Table 4 and Supplementary Figure 4). Divergence among the other taxa, including wild and domesticated lentils, was less marked. Whereas the overall nucleotide diversity (π) was 0.00892, there was also variation among K groups with π ranging from 0.00248 in K2 (*culinaris*-M) to 0.00751 in K7 (*orientalis*-III) (Table 5). The most diverse group (as measured by π) was K3, which included all wild taxa represented in this study. The 2 *orientalis* groups most likely involved in lentil domestication (K5, *orientalis*-II and K7, *orientalis*-III) had the next highest diversity, possibly due to the large number of accessions

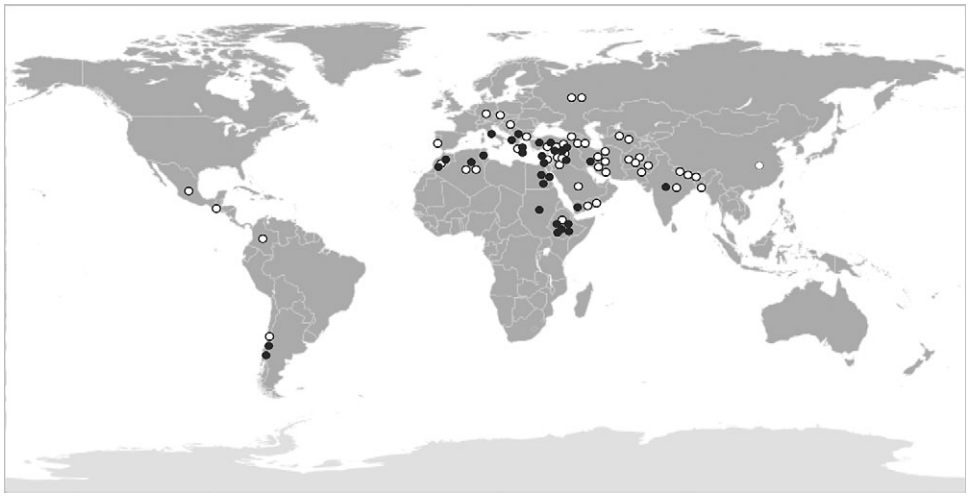

Figure 3. Geographic distribution of STRUCTURE groups containing domesticated lentil accessions. Circles: black: K2; white: K6.

they included. In contrast, the 2 domesticated groups (K2, *culinaris*-M and K6, *culinaris*-m) had some of the lowest diversity levels compared with other K groups, in spite of their larger membership (Table 5).

An analysis of LD in the entire set of 308 sequences was carried out using only parsimony informative polymorphic sites (652 sites of a total of 11 803 sites after excluding sites with gaps). There were no significant pairwise comparisons identified using Fisher's Exact test. Thus, LD appears to be minimal to nonexistent in the entire set. Furthermore, tests of departure from neutrality failed to reveal signatures from selection. Tajima's D was -0.99 (not significant [ns], $P > 0.10$). Fu and Li's D^* and F^* statistics were -1.95 (ns, $0.10 > P > 0.05$) and -1.69 (ns, $P > 0.10$), respectively.

Sequence Divergence and Diversity in Relation to Domestication of Lentil

Because of a more specific interest in the domestication process, a more detailed analysis of the relationships among 5 K groups was conducted. The 5 groups included wild and domesticated *Lens culinaris*: K4 (*orientalis*-I), K5 (*orientalis*-II), K7 (*orientalis*-III), K2 (*culinaris*-M), and K6 (*culinaris*-m). The K4 group (*orientalis*-I from Central Asia) was most likely not involved in lentil domestication because the DNA sequence distances of K4 to the 2 domesticated subgroups K2 and K6 were larger than those of K5 (*orientalis*-II) and K7 (*orientalis*-III) to K2 and K6 (Table 6). On the basis of the same distance, the K2 subgroup (*culinaris*-M) may have originated either from K5 or K7 as current data, based on concatenated sequences, show similar distances between K2 and the 2 *orientalis* subgroups. In contrast, the K6 subgroup (*culinaris*-m) may have originated from K7 as the K2–K7 distance is smaller than the K2–K5 distance (Table 6). An analysis of the remaining groups—K5 and K7 for *orientalis* and K2 and K6 for *culinaris*—by an NJ tree further confirmed the separation of the K2 and K 6 groups (Figure 5). However, the K5 and K7 groups were more difficult to distinguish in this analysis. Thus, it is difficult to determinate at this stage whether the large- and small-seeded domesticated races resulted from separate domestications or, alternatively, from a single domestication followed by divergence post-domestication. Further analyses are needed to distinguish among these different domestication scenarios. Based on this observation, the K4 subgroup was not considered in further analysis and discussions about the domestication process in *L. culinaris*.

Table 3 Ecological distribution parameters—as described by Furman (2006)—and distinguishing the K4, K5, and K7 *orientalis* groups

K group	*N*	Latitude	Longitude	Altitude (m)	Yearly average maximum temperature (°C)	Yearly average minimum temperature (°C)	Predominant agro-climatic zone[a]	Yearly average potential evapotranspiration (mm)	Yearly average aridity index[b]
K4 (*orientalis*-I)	8	40.69	60.33	1209	17.9	5.7	SH-K-M	1087	0.46
K5 (*orientalis*-II)	19	35.48	38.94	903	20.9	9.2	SA-C-W	1306	0.45
K7 (*orientalis*-III)	17	34.90	35.81	952	21.9	9.6	SH-C-W	1240	0.39

[a] SA: semi-arid moisture regime (aridity index: 0.2–0.5); SH: sub-humid moisture regime (aridity index: 0.5–0.75); C: cool winter type (>0 °C); K: cold winter type (<0 °C); M: Mild summer type (10–20 °C); W: Warm summer type (20–30 °C) (De Pauw et al. 2008).

[b] Aridity index defined as the ratio of yearly rainfall over potential evapotranspiration (De Pauw et al. 2008).

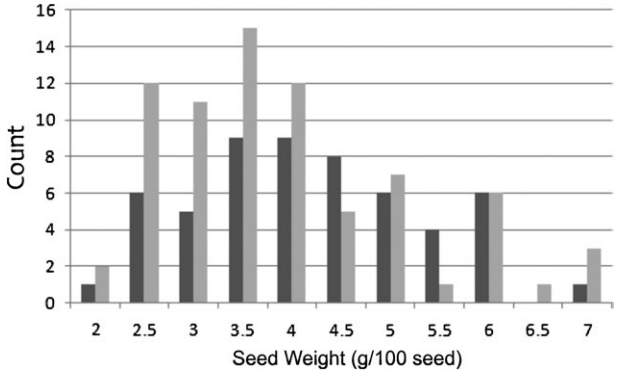

Figure 4. Seed weight distribution differences between STRUCTURE groups K2 (black) and K6 (white).

The 2 remaining *orientalis* groups (K5 and K7) were pooled into a Wild (W) ancestral group and the 2 *culinaris* groups were pooled as well into a single Domesticated (D) group. Comparison of nucleotide diversity (π) in the W and D groups showed a marked reduction (42%) in the latter group (Table 7). Using Watterson's theta (θ_W), there was a 67% decrease in genetic diversity. Surprisingly, however, the number of haplotypes observed was larger in the D group compared with the W group.

Further analyses using the NJ algorithm were conducted to attempt to clarify the relationship among the 2 domesticated groups (K2 and K6) and the 2 *L. culinaris* subsp. *orientalis* groups directly implicated in lentil domestication (K5 and K7). Addition of a randomly chosen *L. nigricans* accession (157nigTURK10) as outgroup did not provide a different picture of the relationship among the 4 K groups (data not shown). In addition, removal of accessions with posterior membership coefficients below 0.95 in their respective K groups—to remove putative hybrids—had the disadvantage of removing most of the K5 or K7 accessions. Removal of accessions with posterior membership coefficients below 0.90 led to larger numbers of remaining K5 and K7 accessions. The resulting NJ tree (Figure 5) showed a separation between the 2 domesticated groups (K2-*culinaris*-M and K6-*culinaris*-m). It also revealed a small group of 3 *orientalis* accessions from the K7 group that were most closely related to the domesticated group and originated from southern Turkey: 297oriTURK7, 298oriTURK7, and 299oriTURK7. The geographic proximity of these 3 populations suggests a possible domestication core area in southern Turkey from the K7 *orientalis* group (Figure 2B).

LD and Neutrality Analysis on Concatenated Sequences

An LD analysis was conducted on the sample that included the W and D groups (Supplementary Table 5). This analysis considered only the 358 parsimony informative polymorphic sites of a total of 11 929 sites. Results showed that 9 797 of 63 903 pairwise comparisons (15%) were significant using Fisher's Exact test, of which 867 (1.4%) were still significant after Bonferroni correction. A separate analysis of the W and D groups shows a 10-fold increase in LD from the W to the D group (0.6 to 6.2%), as measured by the proportion of pairwise locus comparisons that were statistically significant after application of Fisher's Exact test and Bonferroni correction. This increase in LD in the D group relative to the W group is to be expected given the recent domestication of lentil and the prevalence of selection (Bulmer 1971) and drift (Avery and Hill 1979) during and after domestication.

Measures of departure from neutrality were determined for the entire W + D sample and for the W and D samples individually. Results suggest some effects of selection in the current sample based on the variation of 22 genes in 207 accessions (Supplementary Table 5). Only the Fu and Li D^* and F^* parameters were statistically significant in the combined W + D sample, whereas all parameters were nonsignificant in the W and D samples, individually. This observation suggests that no strong selective effect could be observed in this sequence sample.

Discussion

The legume family is one of the largest botanical families with some 20 000 species. Several of these, located mostly in the Papilionoideae subfamily, have economic importance principally as food and feed sources. Because of practical reasons, the development of genomic resources in the legume family has been focused on only 3 model or

Table 4 Kimura 2-parameter net distance (below the diagonal) and standard error (above the diagonal) among K groups, calculated with the MEGA 4.1 program

K group	K group Predominant taxon	K1 *nigricans*	K2 *culinaris*-M	K3 *odemensis* and others	K4 *orientalis*-I	K5 *orientalis*-II	K6 *culinaris*-m	K7 *orientalis*-III	K8 *ervoides*
K1	*nigricans*		0.00100	0.00115	0.00115	0.00103	0.00102	0.00101	0.00100
K2	*culinaris*-M	0.01754		0.00037	0.00042	0.00014	0.00015	0.00013	0.00073
K3	*odemensis*and others	0.01547	0.00491		0.00038	0.00026	0.00038	0.00019	0.00048
K4	*orientalis*-I	0.01737	0.00315	0.00458		0.00032	0.00019	0.00023	0.00065
K5	*orientalis*-II	0.01567	0.00133	0.00301	0.00204		0.00012	0.00004	0.00063
K6	*culinaris*-m	0.01730	0.00122	0.00446	0.00131	0.00131		0.00001	0.00066
K7	*orientalis*-III	0.01520	0.00134	0.00249	0.00148	0.00028	0.00093		0.00058
K8	*ervoides*	0.01613	0.00821	0.00523	0.00736	0.00626	0.00746	0.00549	

Table 5 Comparison of sequence diversity indices among K groups established by STRUCTURE

K group	Number of Sequences	Segregating sites (*S*)	Haplotypes (*h*)	Haplotype diversity (Hd)	Average no. of differences (*K*)	Nucleotide diversity ($\pi \times 10^3$)
K1	24	135	22	0.99275	33.53261	2.85
K2	55	138	52	0.99720	20.23270	2.48
K3	23	323	23	1.00000	84.11067	7.17
K4	12	111	10	0.96970	38.75758	3.28
K5	46	320	43	0.99614	69.04348	5.85
K6	76	173	76	1.00000	44.29965	3.76
K7	30	439	30	1.00000	88.65747	7.51
K8	42	159	42	1.00000	29.59582	2.51
Total	308	910	297	0.99974	105.33625	8.92

reference species (Gepts et al. 2005). These include soybean (*Glycine max*) for the warm-season group, which comprises species such as beans (*Phaseolus* spp.), cowpea (*Vignaunguiculata*), mung bean (*Vignaradiata*), and pigeon pea (*Cajanuscajan*), and *M. truncatula* and *Lotus japonicus* for the cool-season group, which includes alfalfa (*Medicago sativa*), pea (*Pisumsativum*), chickpea (*Cicerarietinum*), and lentil, among others. *Medicago truncatula* is phylogenetically the closest model species to lentil from which it diverged some 24 Ma (Lavin et al. 2005). *Arabidopsis thaliana*, the other model species used in this study, belongs to the Brassicaceae (Eurosid II), which have diverged from the Eurosid I (including the Fabaceae) some 100–110 Ma (Fawcett et al. 2009). Thus, the CP method, as used here, employed 2 model species, *M. truncatula* and *A. thaliana*, that are more distantly related to each other than species within the Poaceae family, as used by Blake et al. (2004). Moreover, CPs were used for PCR and sequencing of a species separated from *M. truncatula*, the source of EST sequence information, by 24 million years of divergence (Lavin et al. 2005). In spite of this, 22% of CPs produced amplicons in *Lens* and could be used as sequencing primers in 308 lentil accessions. Because this strategy does not require prior sequence information for the target species (lentil, in this case), but only ESTs from a related species in the same botanical family (*M. truncatula*, in this case), it makes genomic resources of model species accessible to many species for which such resources have not yet been developed.

These results provide the first extensive sampling of genetic diversity at the DNA sequence level in domesticated and wild lentils. For the sake of brevity, only the π values will be discussed as the θ values provide a similar picture. Overall, our lentil sample of 308 accessions and 22 genes showed a sequence diversity of π of 8.92×10^{-3} (Table 5). Because experiments to measure sequence diversity vary in the number and types of sequences and plant accessions involved as well as the reproductive biology of individual species, it is difficult to compare this value with those of other species; however, it is roughly similar to those observed in other legume species (McClean et al. 2004; Hyten et al. 2006; Jing et al. 2007; McClean and Lee 2007; Gaitán-Solís et al. 2008; Yuan et al. 2008; Chen et al. 2010).

The subset of this lentil sample grouping K5 + K7 (wild *orientalis* directly implicated in domestication) and K2 + K6 (domesticated *L. culinaris*) had on overall value of $\pi = 5.35 \times 10^{-3}$. Domestication induced a reduction in genetic diversity of approximately 40% (Wild: $\pi = 6.7 \times 10^{-3}$; Domesticated: $\pi = 3.91 \times 10^{-3}$). This reduction in genetic diversity confirms earlier observations in lentil (Pinkas et al. 1985; Ferguson et al. 1998). Loss of diversity along with an increase of LD are the most widespread characteristics of crop domestication. Similar observations have been made in many other crop species, such as common bean (Gepts et al. 1986), soybean (Hyten et al. 2006), tomato (Tanksley and McCouch 1997), cereals (Buckler et al. 2001; Wright et al. 2005; Luo et al. 2007; Haudry et al. 2007), purple mombin (Miller and Schaal 2006), and sunflower (Liu and Burke 2006).

Intronic regions are expected to have higher levels of diversity than exonic regions as they are less likely to undergo natural selection (e.g., Kreitman 1983). The higher sequence diversity observed among introns compared with exons in our sample is consistent with results from studies on *Arabidopsis*, soybean, and *Drosophila* (Moriyama and Powell 1996; Kawabe and Miyashita 1999; Kawabe et al. 2000; Kuittinen and Aguade 2000; Zhu et al. 2003). Such results indicate that studying DNA sequence variation in these regions will likely uncover high amounts of genetic diversity not found using other methods and supports the CP approach of designing PCR primers in more conserved exons, which bracket more polymorphic introns.

The NJ dendrogram constructed from the Kimura-2 genetic distance among the 308 lentil accessions indicated that *L. nigricans* and *L. ervoides* were well separated from each other and the other taxa (Supplementary Figure 4). Most accessions belonging to the *L. culinaris* subsp. *odemensis* and

Table 6 Kimura 2-parameter net distance (below the diagonal) and standard error (above the diagonal) among K groups implicated in the lentil domestication process

	K group	K4	K5	K7
K group	Predominant taxon	*orientalis*-I	*orientalis*-II	*orientalis*-III
K2	*culinaris*-M	0.00042	0.00014	0.00013
K6	*culinaris*-m	0.00131	0.00131	0.00010

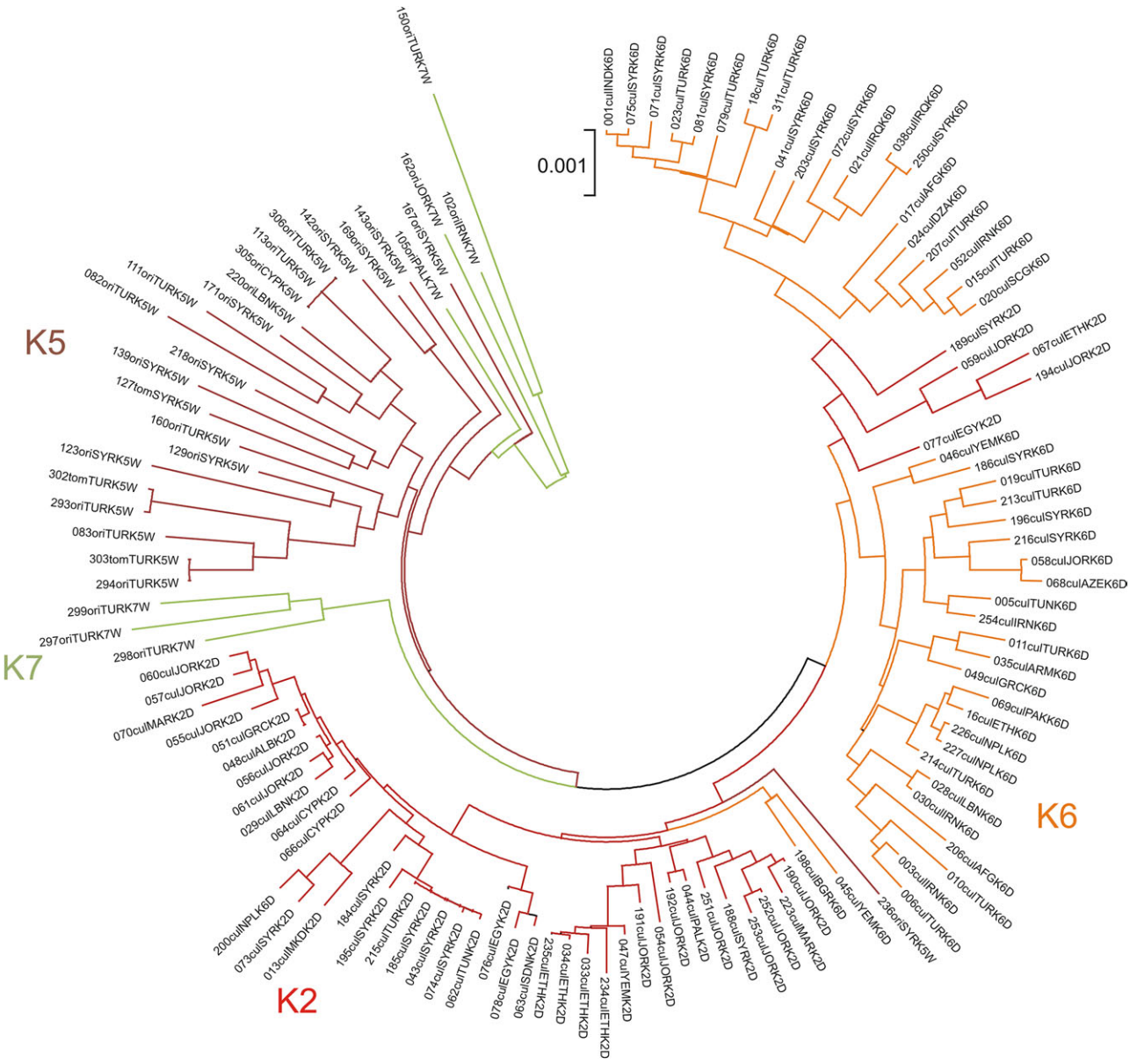

Figure 5. NJ tree of DNA sequence variation in wild (K5 + K7) and domesticated (K2 + K6) lentils, including only accessions with posterior membership probabilities in their respective K groups above 0.90.

tomentosis and *L. lamottei* taxa were grouped in a single clade distinct from the other branches in the tree, suggesting that the current taxonomic status of these taxa should be revisited in combination with crossability analyses to possibly incorporate them into a single taxon, as a single subspecies of *L. culinaris* or as a separate *Lens* species (Supplementary Figure 4). Domesticated lentil was most closely related to *L. culinaris* subsp. *orientalis*, generally supporting the idea that *L. culinaris* subsp. *orientalis* is the wild progenitor of domesticated lentil and that the 2 taxa are subspecies of the same species, *L. culinaris* (Ladizinsky 1999; Ferguson et al. 2000). This conclusion had been reached by previous studies using morphological, cytological, isozyme, RFLP, and chloroplast DNA (cpDNA) data (Havey and Muehlbauer 1989; Sammour 1994; Abo-elwafa et al. 1995; Sharma et al. 1995; Ahmad and McNeil 1996; Ford et al. 1997; Van Oss et al. 1997; Ferguson et al. 2000; Zimniak-Przybylska et al. 2001; Galasso 2003).

The interspersion of the *culinaris* and *orientalis* branches in the NJ tree (Supplementary Figure 4) is probably due to outcrossing events between the 2 conspecific taxa. Average outcrossing levels ranging from 2.2% to 6.6% have been documented in domesticated lentil by Erskine and Muehlbauer (1991). Horneburg (2006) recorded a range of average outcrossing between 0.06% and 5.12% among cultivars. Outcrossing rate varied according to location and

Table 7 Comparison of sequence diversity in wild and domesticated *Lens culinaris*

Population	No. of accessions	No. of polymorphic sites[a]	No. of haplotypes	Average no. of nucleotide differences (*k*)	Nucleotide diversity (π)	Watterson's Theta (θ_W)
Wild (K5 + K7)	76	557	73	80	0.00674	0.00951
Domesticated (K2 + K6)	131	197	128	47	0.00391	0.00313
Total (W + D)	207	583	201	64	0.00535	0.00827

[a] Total number of sites analyzed: 11 929 (excluding gaps).

year as well. Furthermore, variation for outcrossing rate also varied among plants within individual varieties (0–22.2%). Hence, although lentil is a predominantly self-pollinating species, outcrossing is taking place. The allogamy levels may seem low; however, one should remember that these are single-year direct estimates and that the effects of outcrossing are cumulative. In common bean (*P. vulgaris*), another predominantly self-pollinating species with similar outcrossing values, cumulative measures of gene flow between wild and domesticated beans reveal that outcrossing can have a significant effect on the genome. For example, there is a 3-fold higher level of gene flow from domesticated to wild beans than in the opposite direction (Papa and Gepts 2003). Differentiation between the wild and domesticated bean genomes is highly significant around domestication genes but nonsignificant away from these genes (Papa et al. 2005, 2007). Thus, even in predominantly autogamous species, outcrossing is a factor affecting the distribution of genetic diversity within and among populations.

This outcrossing makes the identification of a better-defined center of domestication more difficult. In an attempt to resolve this situation, STRUCTURE was run on those accessions with a membership coefficient above 90%. The outcome of this analysis (Figure 5) suggests that a putative center of lentil domestication may be located within a defined region in southern Turkey extending from Denizli in the southwest to Antakya in south-central Turkey (Figure 2B). A domestication region in Turkey and northern Syria had already been suggested by Ladizinsky (1999). However, our results allow us to hypothesize a more circumscribed region within Turkey. This region does not overlap with but is adjacent to the presumed core domestication area proposed for some of the founder crops of the Southwest Asian domestication center, that is, the Fertile Crescent (Abbo et al. 2010). The core area, located in the KaracaDağ range near Diyarbakir in southern Turkey, would have been the domestication hearth of einkorn (Heun et al. 1997) and emmer (Özkan et al. 2002; Luo et al. 2007) wheat, chickpea (Abbo et al. 2003), and lentil (Ladizinsky 1999). How can this discrepancy be explained? First, it is possible that the distribution of the wild relatives of the different crop species and of their genetic diversity may have changed over the millennia since domestication. Although data suggest that environmental conditions have remained quite similar in the Holocene in this region (Abbo et al. 2010 and references therein; Willcox 2005), more subtle differential eco-geographic changes among the wild relatives may have obscured any original geographic correspondence in the domestication areas. This situation suggests caution when attempting to locate and compare presumed domestication areas of different crops in a center of origin, such as the Fertile Crescent. This is especially the case of lentil for which the actual number of *L. culinaris* var. *orientalis* accessions analyzed at the molecular level prior to this study is very limited (Mayer and Soltis 1994; Ladizinsky 1999). Of the 4 *L. culinaris* var. *orientalis* accessions analyzed 3 had a cpDNA matching that of the domesticated gene pool, were crossable with a domesticated accession, and showed similar chromosome structure to that of domesticated lentils. The geographic origin of the 3 accessions led Ladizinsky (1999) to propose a domestication origin in central and south Turkey, as well as northwestern Syria. Although germplasm passport information is incomplete, 2 of the *L. culinaris* var. *orientalis* accessions studied by Ladizinsky (1999) may be included in this study based on similar geographic origins: 307oriTUR corresponding to Lo77 (ICARDA IG136671; from the vicinity of Tokat, Turkey) and 143oriSYR corresponding to Lo156 or Lo157 (ICARDA IG72772; from the vicinity of Idlib, Syria). Neither, however, belonged to the proposed ancestral group. The small number of wild accessions analyzed by Mayer and Soltis (1994) and Ladizinsky (1999) cannot preclude a domestication elsewhere. Thus, cpDNA, crossability, and chromosome structure studies should be extended to a broader sample of *L. culinaris* var. *orientalis*, including the 3 accessions identified in the putative domestication area in this article.

Second, a reexamination of results sheds doubt about the existence of a narrow core area within the Fertile Crescent in southern Turkey as the actual center of agricultural origin in this area of the world (Lev-Yadun et al. 2000; Abbo et al. 2010). This core area would be centered on the KaracaDağ region west of Diyarbakir, Turkey. The main arguments that have been advanced in favor of this location include genetic relatedness based on neutral markers, the current distribution of the wild progenitor of some of the 7 so-called founder crops, and the existence of archaeological sites containing archaeobotanical remains in the same general area. Heun et al. (1997) first identified a closer genetic relationship of the *Triticum monococcum* var. *boeoticum* population from the KaracaDağ region with *T. monococcum* var. *monococcum* or einkorn, the domesticated descendant, compared with other *T. monococcum* var. *boeticum* populations. They interpreted this closer relationship as evidence that the

KaracaDağ region is a putative center of domestication of einkorn. However, these same results also show that the KaracaDağ population is an outlier among *T. monococcum* var. *boeticum* populations (see Figure 2A of Heun et al. 1997) and is genetically intermediate between *T. monococcum* var. *boeticum* and var. *monococcum* populations (see Figure 2B,C of Heun et al. 1997). An alternative explanation to the ancestor-descendant relationship is, therefore, that the KaracaDağ population results from outcrossing between wild and domesticated *T. monococcum*. In a later article, Kilian et al. (2007) invoke multiple domestications to account for the high levels of genetic diversity in einkorn. Multiple outcrosses over an extended time could also account for the absence of reduction in genetic diversity in einkorn domestication, which contrasts with the genetic bottleneck that is generally observed in crop domestication (Gepts 2004). As argued earlier in this Discussion, even modest outcrossing levels (as have been observed in einkorn, Zohary and Hopf 2000) can have detectable effects in the distribution of genetic diversity within a genome and among populations. The absence of morphological hybridity symptoms is a weak indication for the absence of outcrossing as selection can act to eliminate such low-fitness traits (Papa et al. 2005).

Genetic relatedness data for emmer wheat (*T. turgidum*) presents contradictory evidence on the domestication origin of this crop (Mori et al. 2003; Özkan et al. 2005; Luo et al. 2007). While nuclear markers point to the KaracaDağ region, chloroplast markers point to the KartalDağı region further west in southern Turkey. It is to be noted that the KartalDağı region is adjacent to the location of the easternmost population proposed in this study as a putative ancestor of domesticated lentil.

Whereas many wild progenitors of Fertile Crescent have an extensive distribution (e.g., barley, pea, wheat), the presumed chickpea progenitor (*Cicer reticulatum*) has a distribution limited to the headwaters of the Tigris and Euphrates (Zohary and Hopf 2000). Because this distribution area includes the KaracaDağ region, Lev-Yadun et al. (2000) used this distribution as an argument in favor of a core domestication area centered in the KaracaDağ region. More recently, Abbo et al. (2010) observed that 6 of 7 of the founder crops had wild relatives growing sympatrically south of the KaracaDağ region around Mardin. Although the presence of a wild ancestor is an obvious necessary condition for domestication, it is not a sufficient condition. Furthermore, one could question the need for domesticating more than one cereal or grain legume (pulse) in the same region, given the contrasting agronomic and nutritional functionalities of the 2 types of crops. Other factors must play a role in determining the actual domestication areas, including behavioral, ecological, cultural, social, and political ones, to drive the transition from hunting–gathering to cultivation and, ultimately, domestication. An important aspect here is the well-documented trade and other sociocultural interactions that existed in the Neolithic in the Fertile Crescent (Willcox 2005; Abbo et al. 2010). Trade and communications would have facilitated the exchange of ideas about cultivation and the introduction of crops, leading up to the establishment of complex crop production systems in fully developed agricultural economies.

Third, a similar situation exists in the Mesoamerican center origin, where bean and maize were presumably domesticated in 2 different, albeit adjacent regions in western Mexico, the Lerma-Santiago (Kwak et al. 2009) and Balsas river basins (Matsuoka et al. 2002), respectively. Given that wild beans and maize have overlapping distributions and can occasionally be observed growing together (Delgado-Salinas et al. 1988) and bean–maize associated cropping is widespread, the question arises as to how the maize-bean (-squash) trilogy became widely established as a consequence of the transition from hunting–gathering to agriculture. Current data suggest that a geographic overlap of domestication area may not be necessary. However, Zizumbo-Villarreal and Colunga-García Marín (2010) have pointed out that the presumed bean and maize domestication areas show commonalities in geological (morpho-tectonic) and ecological (tropical dry forest) contexts. Thus, geographically distinct domestications can lead to a single cropping system, the milpa system in this case.

On the whole, the available evidence suggests that the present data on lentil domestication origin favor a polycentric view (Gebel 2004; Willcox 2005) of crop domestication in the Fertile Crescent and Mesoamerica. Given the multiplicity of factors, it is likely that this crucial transition followed different trajectories in different regions within centers of agricultural regions.

Finally, the method described here has uncovered additional genetic diversity in the genus *Lens* and promises to be a very effective method for studying diversity of germplasm collections and obtaining additional insights into the domestication of lentil. Even with this small number of gene sequences, the NJ tree has clustered accessions into natural groups. The addition of more gene sequences and accessions from neighboring regions holds the promise to further subdivide the clusters and shed light on relationships among the accessions and allow for detection of admixture. Such data, combined with information on past human migration and trade, will provide valuable insights into lentil domestication.

Supplementary Material

Supplementary material can be found at http://www.jhered.oxfordjournals.org/.

Funding

National Science Foundation Plant Genome Research Program (DBI-0321757).

Acknowledgments

The authors gratefully acknowledge the ICARDA Genetic Resources Unit for providing seeds and use of their laboratory. Drs Martin Wojciechowski and Alfonso Delgado-Salinas provided enlightening information on the age of divergence of major taxa. Funding by the U.S. Agency for International Development Linkages program is gratefully acknowledged.

References

Abo-elwafa A, Murai K, Shimada T. 1995. Intra- and inter-specific variations in Lens revealed by RAPD markers. Theor Appl Genet. 90:335–340.

Abbo S, Berger J, Turner NC. 2003. Evolution of cultivated chickpea: four bottlenecks limit diversity and constrain adaptation. Funct Plant Biol. 30:1081–1087.

Abbo S, Lev-Yadun S, Gopher A. 2010. Agricultural origins: centers and noncenters; A Near Eastern reappraisal. Critical Rev Plant Sci. 29:317–328.

Adjei S, Sato A, Tanaka S, Kobayashi E, Tanaka K, Namikawa T, Ishikawa A. 2005. Development and characterization of CATS markers for genetic linkage mapping in the house musk shrew, *Suncus murinus*. Exp Anim. 54:173–180.

Ahmad M, McNeil D. 1996. Comparison of crossability, RAPD, SDS-PAGE and morphological markers for revealing genetic relationships within and among *Lens* species. Theor Appl Genet. 93:788–793.

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

Avery PJ, Hill WG. 1979. Variance in quantitative traits due to linked dominant genes and variance in heterozygosity in small populations. Genetics. 91:817–844.

Barulina H. 1930. Lentil of the U.S.S.R. and of other countries. Bull Appl Bot Plant Breed Suppl. 40:1–319.

Beisvag V, Jünge FKR, Bergum H, Jølsum L, Lydersen S, Günther C-C, Ramampiaro H, Langaas M, Sandvik A, Lægreid A. 2006. *GeneTools*—application for functional annotation and statistical hypothesis testing. BMC Bioinformatics. 7:470.

Blake NK, Sherman JD, Dvořák J, Talbert LE. 2004. Genome-specific primer sets for starch biosynthesis genes in wheat. Theor Appl Genet. 109:1295–1302.

Bulmer M. . 1971. The effect of selection on generic variability. Amer. Nat. 105:201–211.

Buckler M, Thornsberry J, Kresovich S. 2001. Molecular diversity, structure, and domestication of grasses. Genet Res. 77:213–218.

Chahota R, Kishore N, Dhiman K, Sharma T, Sharma S. 2007. Predicting transgressive segregants in early generation using single seed descent method-derived micro-macrosperma genepool of lentil (*Lens culinaris* Medikus). Euphytica. 156:305–310.

Chen J, Zhang X, Jing R, Blair M, Mao X, Wang S. 2010. Cloning and genetic diversity analysis of a new *P5CS* gene from common bean (*Phaseolus vulgaris* L.). Theor Appl Genet. 120:1393–1404.

Cubero JI. 1981. Taxonomy, distribution and evolution of the lentil and its wild relatives. In: Witcombe J, Erskine W, editors. Genetic resources and their exploitation chickpeas, faba beans and lentils. Boston (MA): M. Nijhoff& W. Junk Publishers. p. 187–204.

Delgado Salinas A, Bonet A, Gepts P. 1988. The wild relative of *Phaseolus vulgaris* in Middle America. In Gepts P ed. Genetic resources of *Phaseolus* beans. Dordrecht, the Netherlands: Kluwer, p. 163–184.

De Pauw E, Mirghasemi A, Ghaffari A, Nseir B. 2008. Agro-ecological zones of Karkheh river basin. Aleppo (Syria): International Center for Agricultural Research in the Dry Areas, p. viii + 96 pp.

Erskine W, Muehlbauer FJ. 1991. Allozyme and morphological variability, outcrossing rate and core collection formation in lentil germplasm. Theor Appl Genet. 83:119–125.

Evanno G, Regnaut S, Goudet J. 2005. Detecting the number of clusters of individuals using the software STRUCTURE: a simulation study. Mol Ecol. 14:2611–2620.

Ewing B, Hillier L, Wendl M, Green P. 1998. Base-calling of automated sequencer traces using phred. I. Accuracy assessment. Genome Res. 8:175–185.

Fawcett JA, Maere S, Van de Peer Y. 2009. Plants with double genomes might have had a better chance to survive the Cretaceous-Tertiary extinction event. Proc Nat Acad Sci U S A. 106:5737–5742.

Felsenstein J. 1985. Confidence limits on phylogenies: an approach using the bootstrap. Evolution. 39:783–791.

Ferguson ME, Maxted N, Slageren MV, Robertson LD. 2000. A re-assessment of the taxonomy of *Lens* Mill. (Leguminosae, Papilionoideae, Vicieae). Bot J Linn Soc. 133:41–59.

Ferguson ME, Newbury HJ, Maxted N, Ford-Lloyd BV, Robertson LD. 1998. Population genetic structure in *Lens* taxa revealed by isozyme and RAPD analysis. Genet Res Crop Evol. 45:549–559.

Ford R, Pang E, Taylor P. 1997. Diversity analysis and species identification in *Lens* using PCR generated markers. Euphytica. 96:247–255.

Fredslund J, Madsen LH, Hougaard BK, Nielsen AM, Bertioli D, Sandal N, Stougaard J, Schauser L. 2006. A general pipeline for the development of anchor markers for comparative genomics in plants. BMC Genomics. 7:207.

Furman BJ. 2006. Methodology to establish a composite collection: case study in lentil. Plant Genet Res. 4:2–12.

Gaitán-Solís E, Choi I-Y, Quigley C, Cregan P, Tohme J. 2008. Single nucleotide polymorphisms in common bean: their discovery and genotyping using a multiplex detection system. Plant Gen. 1:125–134.

Galasso I. 2003. Distribution of highly repeated DNA sequences in species of the genus *Lens* Miller. Genome. 46:1118–1124.

Gebel HG. 2004. There was no centre: the polycentric evolution of the Near Eastern Neolithic. Neo-Lithics. 1/04:28–32.

Gepts P. 1993. The use of molecular and biochemical markers in crop evolution studies. Evol Biol. 27:51–94.

Gepts P. 2004. Domestication as a long-term selection experiment. Plant Breed Rev. 24(Part 2):1–44.

Gepts P, Beavis WD, Brummer EC, Shoemaker RC, Stalker HT, Weeden NF, Young ND. 2005. Legumes as a model plant family. Genomics for food and feed Report of the Cross-Legume Advances through Genomics Conference. Plant Physiol. 137:1228–1235.

Gepts P, Osborn TC, Rashka K, Bliss FA. 1986. Phaseolin-protein variability in wild forms and landraces of the common bean (*Phaseolus vulgaris*): evidence for multiple centers of domestication. Econ Bot. 40:451–468.

Hall T. 1999. BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. Nucl Acids Symp Ser. 41:95–98.

Harlan J. 1995. Agricultural origins and crop domestication in the Mediterranean region. Diversity. 2:14–16.

Haudry A, Cenci A, Ravel C, Bataillon T, Brunel D, Poncet C, Hochu I, Poirier S, Santoni S, Glémin S, et al. 2007. Grinding up wheat: a massive loss of nucleotide diversity since domestication. Mol Biol Evol. 24:1506–1517.

Havey MJ, Muehlbauer FJ. 1989. Variability for restriction fragment lengths and phylogenies in lentil. Theor Appl Genet. 77:839–843.

Heun M, Schafer-Pregl R, Klawan D, Castagna R, Accerbi M, Borghi B, Salamini F. 1997. Site of einkorn wheat domestication identified by DNA fingerprinting. Science. 278:1312–1314.

Horn R, Lecouls A-C, Callahan A, Dandekar A, Garay L, McCord P, Howad W, Chan H, Verde I, Main D, et al. 2005. Candidate gene database and transcript map for peach, a model species for fruit trees. Theor Appl Genet. 110:1419–1428.

Horneburg B. 2006. Outcrossing in lentil (*Lens culinaris*) depends on cultivar, location and year, and varies within cultivars. Plant Breed. 125:638–640.

Hougaard BK, Madsen LH, Sandal N, Moretzsohn MD, Fredslund J, Schauser L, Nielsen AM, Rohde T, Sato S, Tabata S, et al. 2008. Legume anchor markers link syntenic regions between *Phaseolus vulgaris, Lotus japonicus, Medicago truncatula* and *Arachis*. Genetics. 179:2299–2312.

Hyten DL, Song QJ, Zhu YL, Choi IY, Nelson RL, Costa JM, Specht JE, Shoemaker RC, Cregan PB. 2006. Impacts of genetic bottlenecks on soybean genome diversity. Proc Nat Acad Sci U S A. 103:16666–16671.

Jing R, Johnson R, Seres A, Kiss G, Ambrose MJ, Knox MR, Ellis THN, Flavell AJ. 2007. Gene-based sequence diversity analysis of field pea (*Pisum*). Genetics. 177:2263–2275.

Kawabe A, Miyashita NT. 1999. DNA variation in the basic chitinase locus (ChiB) region of the wild plant *Arabidopsis thaliana*. Genetics. 153:1445–1453.

Kawabe A, Yamane K, Miyashita NT. 2000. DNA polymorphism at the cytosolic phosphoglucose Isomerase (PgiC) locus of the wild plant *Arabidopsis thaliana*. Genetics. 156:1339–1347.

Kilian B, Ozkan H, Walther A, Kohl J, Dagan T, Salamini F, Martin W. 2007. Molecular diversity at 18 loci in 321 wild and 92 domesticate lines reveal no reduction of nucleotide diversity during *Triticum monococcum* (einkorn) domestication: implications for the origin of agriculture. Mol Biol Evol. 24:2657–2668.

Kreitman M. 1983. Nucleotide polymorphism at the alcohol dehydrogenase locus of *Drosophila melanogaster*. Nature. 304:412–417.

Kuittinen H, Aguade M. 2000. Nucleotide variation at the CHALCONE ISOMERASE locus in *Arabidopsis thaliana*. Genetics. 155:863–872.

Kwak M, Velasco DM, Gepts P. 2008. Mapping homologous sequences for determinacy and photoperiod sensitivity in common bean (*Phaseolus vulgaris*). J Hered. 99:283–291.

Kwak M, Kami JA, Gepts P. 2009. The putative Mesoamerican domestication center of *Phaseolus vulgaris* is located in the Lerma-Santiago basin of Mexico. Crop Sci. 49:554–563.

Ladizinsky G. 1979. The origin of lentil and its wild genepool. Euphytica. 28:179–187.

Ladizinsky G. 1999. Identification of the lentil's wild genetic stock. Genet Res Crop Evol. 46:115–118.

Lavin M, Herendeen PS, Wojciechowski MF. 2005. Evolutionary rates analysis of Leguminosae implicates a rapid diversification of the major family lineages immediately following an Early Tertiary emergence. Syst Biol. 54:575–594.

Lev-Yadun S, Gopher A, Abbo S. 2000. The cradle of agriculture. Science. 288:162–163.

Librado P, Rozas J. 2009. DnaSP v5: a software for comprehensive analysis of DNA polymorphism data. Bioinformatics. 25:1451–1452.

Liu AZ, Burke JM. 2006. Patterns of nucleotide diversity in wild and cultivated sunflower. Genetics. 173:321–330.

Luo MC, Yang ZL, You F, Kawahara T, Waines J, Dvorak J. 2007. The structure of wild and domesticated emmer wheat populations, gene flow between them, and the site of emmer domestication. Theor Appl Genet. 114:947–959.

Matsuoka Y, Vigouroux Y, Goodman MM, Sanchez GJ, Buckler E, Doebley J. 2002. A single domestication for maize shown by multilocus microsatellite genotyping. Proc Nat Acad Sci U S A. 99:6080–6084.

Mayer MS, Bagga SK. 2002. The phylogeny of *Lens* (Leguminosae): new insight from ITS sequence analysis. Plant Syst Evol. 232:145–154.

Mayer MS, Soltis PS. 1994. Chloroplast DNA phylogeny of *Lens* (Leguminosae): origin and diversity of the cultivated lentil. Theor Appl Genet. 87:773–781.

McClean PE, Lavin M, Gepts P, Jackson SA. 2008. Phaseolus vulgaris: a diploid model for soybean. In: Stacey G, editor. Genetics and genomics of soybean. New York: Springer. p. 55–76.

McClean P, Mamidi S, McConnell M, Chikara S, Lee R. 2010. Synteny mapping between common bean and soybean reveals extensive blocks of shared loci. BMC Genomics. 11:184.

McClean PE, Lee RK. 2007. Genetic architecture of chalcone isomerase non-coding regions in common bean (*Phaseolus vulgaris* L.). Genome. 50:203–214.

McClean PE, Lee RK, Miklas PN. 2004. Sequence diversity analysis of dihydroflavonol 4-reductase intron 1 in common bean. Genome. 47:266–280.

Miller AJ, Schaal BA. 2006. Domestication and the distribution of genetic variation in wild and cultivated populations of the Mesoamerican fruit tree *Spondias purpurea* L. (Anacardiaceae). Mol Ecol. 15:1467–1480.

Mori N, Ishii T, Ishido T, Hirosawa S, Watatani H, Kawahara T, Nesbitt M, Belay G, Takumi S, Ogihara Y, Nakamura C. 2003. Origin of domesticated emmer and common wheat inferred from chloroplast DNA fingerprinting. 10th International Wheat Genetics Symposium. Paestum, Italy, p. 25–28.

Moriyama E, Powell J. 1996. Intraspecific nuclear DNA variation in Drosophila. Mol Biol Evol. 13:261–277.

Muehlbauer F, Slinkard A. 1981. Genetics and breeding methodology. In: Webb C, Hawtin G, editors. Lentils. Farnham Royal (UK): Commonwealth Agricultural Bureau. p. 69–90.

Murray MG, Thompson WF. 1980. Rapid isolation of high molecular weight plant DNA. Nucl Acid Res. 8:4322–4325.

Mustafa BM, Coram TE, Pang ECK, Taylor PWJ, Ford R. 2009. A cDNA microarray approach to decipher lentil (*Lens culinaris*) responses to *Ascochytalentis*, Australasian. Plant Pathol. 38:617–631.

Nei M. 1987. Molecular evolutionary genetics. New York: Columbia University Press.

Özkan H, Brandolini A, Schäfer-Pregl R, Salamini F. 2002. AFLP analysis of a collection of tetraploid wheats indicates the origin of emmer and hard wheat domestication in southeast Turkey. Mol Biol Evol. 19:1797–1801.

Özkan H, Brandolini A, Pozzi C, Effgen S, Wunder J, Salamini F. 2005. A reconsideration of the domestication geography of tetraploid wheats. Theor Appl Genet. 110:1052-1-060.

Papa R, Gepts P. 2003. Asymmetry of gene flow and differential geographical structure of molecular diversity in wild and domesticated common bean (*Phaseolus vulgaris* L.) from Mesoamerica. Theor Appl Genet. 106:239–250.

Papa R, Acosta J, Delgado-Salinas A, Gepts P. 2005. A genome-wide analysis of differentiation between wild and domesticated *Phaseolus vulgaris* from Mesoamerica. Theor Appl Genet. 111:1147–1158.

Papa R, Bellucci E, Rossi M, Leonardi S, Rau D, Gepts P, Nanni L, Attene G. 2007. Tagging the signatures of domestication in common bean (*Phaseolus vulgaris*) by means of pooled DNA samples. Ann Bot. 100:1039–1051.

Pinkas R, Zamir D, Ladizinsky G. 1985. Allozyme divergence and evolution in the genus *Lens*. Plant Syst Evol. 151:131–140.

Pritchard J, Stephens M, Donnelly P. 2000. Inference of population structure using multilocus genotype data. Genetics. 155:945–959.

Robertson L, Erskine W. 1997. Lentil. In: Fuccillo D, Sears L, Stapleton P, editors. Biodiversity in trust. Cambridge: Cambridge University Press. p. 128–138.

Sammour R. 1994. Species relationships in genus *Lens* as indicated by electrophoresis—a reappraisal. Lens News. 21:1–4.

Scippa GS, Rocco M, Ialicicco M, Trupiano D, Viscosi V, Di Michele M, Arena S, Chiatante D, Scaloni A. 2010. The proteome of lentil (*Lens culinaris* Medik.) seeds: discriminating between landraces. Electrophoresis. 31: 497–506.

Sharma S, Knox M, Ellis T. 1996. AFLP analysis of the diversity and phylogeny of *Lens* and its comparison with RAPD analysis. Theor Appl Genet. 93:751–758.

Sharma SK, Dawson IK, Waugh R. 1995. Relationships among cultivated and wild lentils revealed by RAPD analysis. Theor Appl Genet. 91:647–654.

Sonnante G, Galasso I, Pignone D. 2003. ITS sequence analysis and phylogenetic inference in the genus *Lens* Mill. Ann Bot. 91:49–54.

Tamura K, Dudley J, Nei M, Kumar S. 2007. MEGA4: Molecular Evolutionary Genetic Analysis (MEGA) software version 4.0. Mol Biol Evol. 24:1596–1599.

Tanksley S, McCouch S. 1997. Seed banks and molecular maps: unlocking genetic potential from the wild. Science. 277:1063–1066.

Tullu A, Tar'an B, Warkentin T, Vandenberg A. 2008. Construction of an intraspecific linkage map and QTL analysis for earliness and plant height in lentil. Crop Sci. 48:2254–2264.

Van Oss H, Aron Y, Ladizinsky G. 1997. Chloroplast DNA variation and evolution in the genus *Lens* Mill. Theor Appl Genet. 94:452–457.

Watterson GA. 1975. On the number of segregating sites in genetical models without recombination. Theor Pop Biol. 7:256–276.

Webb C, Hawtin G. 1981. Lentils. Farnham Royal (UK): Commonwealth Agricultural Bureaux.

Willcox G. 2005. The distribution, natural habitats and availability of wild cereals in relation to their domestication in the Near East: multiple events, multiple centres. Veg His Archaeobot. 14:534–541.

Wright SI, Bi IV, Schroeder SG, Yamasaki M, Doebley JF, McMullen MD, Gaut BS. 2005. The effects of artificial selection of the maize genome. Science. 308:1310–1314.

You F, Huo N, Gu Y, Lazo G, Dvorak J, Anderson O. 2009. ConservedPrimers 2.0: a high-throughput pipeline for comparative genome referenced intron-flanking PCR primer design and its application in wheat SNP discovery. BMC Bioinformatics. 10:331.

Youngman V. 1968. Lentils-a pulse of the Palouse. Econ Bot. 22:135–139.

Yuan C, Zhou G, Li Y, Wang K, Wang Z, Li X, Chang R, Qiu L. 2008. Cloning and sequence diversity analysis of GmHs1pro-1 in Chinese domesticated and wild soybeans. Molec Breed. 22:593–602.

Zhu YL, Song QJ, Hyten DL, Van Tassell CP, Matukumalli LK, Grimm DR, Hyatt SM, Fickus EW, Young ND, Cregan PB. 2003. Single-nucleotide polymorphisms in soybean. Genetics. 163:1123–1134.

Zizumbo-Villareal D, Colunga García-Marín P. 2010. Origin of agriculture and plant domestication in West Mesoamerica. Genet Res Crop Evol. 57:813–825.

Zimniak-Przybylska Z, Przybylska J, Krajewski P. 2001. Electrophoretic seed globulin patterns and species relationships in the genus *Lens* Miller. J Appl Genet. 42:435–447.

Zohary D, Hopf M. 2000. The domestication of plants in the Old World. 3rd ed. Oxford: Oxford University Press.

Received August 27, 2010; Revised January 28, 2011; Accepted February 10, 2011

Corresponding Editor: Reid Palmer