# **ORIGINAL ARTICLE**

# An integrated approach to the characterization of two autochthonous lentil (*Lens culinaris*) landraces of Molise (south-central Italy)

www.nature.com/hdy

GS Scippa<sup>1</sup>, D Trupiano<sup>1</sup>, M Rocco<sup>2</sup>, V Viscosi<sup>1</sup>, M Di Michele<sup>3</sup>, A D'Andrea<sup>1</sup> and D Chiatante<sup>4</sup> <sup>1</sup>Dipartimento di Scienze e Tecnologie per l'Ambiente e il Territorio, University of Molise, Pesche, Italy; <sup>2</sup>Dipartimento di Scienze Biologiche e Ambientali, University of Sannio, Benevento, Italy; <sup>3</sup>Laboratory of Analytical Techniques and Proteomics, 'John Paul II' Center for High Technology, Research and Education in Biomedical Sciences, Catholic University, Campobasso, Italy and <sup>4</sup>Dipartimento di Scienze Chimiche ed Ambientali, University of Insubria, Como, Italy

Plant biodiversity must be safeguarded because it constitutes a resource of genes that may be used, for instance, in breeding programs. Lentil (*Lens culinaris* Medik.) is one of the most ancient crops of the Mediterranean region. Extensive differentiation of *L. culinaris* over millennia has resulted in a myriad of different landraces. However, in more recent times many landraces have disappeared consequent to environmental and socioeconomic changes. To promote the survival of endangered lentil landraces, we have investigated the genetic relationship between two ancient landrace cultivated in Capracotta and Conca Casale (Molise, south-central Italy) and widely spread commercial varieties using an integrated approach consisting of studies at morphological, DNA and protein level. Seeds of these two landraces were collected from local farmers and conserved in the Molise germoplasm bank. The two local landraces were well differentiated from each other, and the Conca Casale landrace was separated from the commercial varieties at morphological, protein and DNA level. The Capracotta landrace, was well separated from the commercial varieties, except Castelluccio di Norcia, at DNA level showing a more complex and heterogeneous segregation at morphological and biochemical level. The correlation between morphological, DNA and protein data, illustrates that proteomics is a powerful tool with which to complement the analysis of biodiversity in ecotypes of a single plant species and to identify physiological and/or environmental markers.

*Heredity* (2008) **101**, 136–144; doi:10.1038/hdy.2008.39; published online 14 May 2008

Keywords: biodiversity conservation; plant germoplasm lentil landraces; Lens culinaris M.; proteomic; genetic variation

# Introduction

Lentil (*Lens culinaris* Medik.) is an ancient crop that originated in the Near East (Zohary, 1999) and subsequently spread throughout the Mediterranean Basin and central Asia (Cubero and Malbon, 1984; Lev-Yadun *et al.*, 2000). Over time, local constraints produced wide diversity within the *L. culinaris* species, resulting in a myriad of different landraces (Erskine, 1997). However, the cultivation of lentils has progressively decreased in industrialized countries in favour of more remunerative crops. Consequently, several local populations have disappeared and those still being cultivated are at a high risk of severe genetic erosion (Ladizinsky, 1993; Piergiovanni, 2000).

Various local landraces have evolved in several Italian regions thanks to an optimal combination of climate, soil and moisture, although modern agricultural practises are eroding the genetic richness of this plant species. Several

Correspondence: Professor GS Scippa, Dipartimento di Scienze e Tecnologie per l'Ambiente e il Territorio, University of Molise, C.da Fonte Lappone, Pesche (IS) 86090, Italy.

E-mail: scippa@unimol.it

Received 15 January 2008; revised 26 March 2008; accepted 4 April 2008; published online 14 May 2008

landraces are on the verge of extinction; others are grown only by elderly farmers, mainly in marginal lands, for their own consumption and, occasionally, for local markets (Piergiovanni, 2000). Unlike modern cultivars selected for their performance in specific environmental conditions, local landraces have a high genetic variability. Having evolved adaptive gene complexes conserved by genetic linkage or natural or human selection, they are highly adapted to different environmental conditions.

An understanding of genetic variation within the lentil germoplasm and of the genetic relationship between landraces is important because: (1) a broad genetic diversity can accelerate the genetic improvement of crops and (2) the identification of genetic variation is an effective method of stratifying and sampling variation in germoplasm collections and can help to define priorities for conservation programs.

Thus far, genetic variation of the lentil germoplasm has been evaluated based on morpho-physiological traits (Erskine *et al.*, 1989; Lazaro *et al.*, 2001; Tullu *et al.*, 2001; Yan'kov *et al.*, 2001), isozymes (De La Rosa and Jouve, 1992; Rodriguez *et al.*, 1999), seed storage proteins (De La Rosa and Jouve, 1992, Echeverrigaray *et al.*, 1998, Piergiovanni and Taranto, 2005) or such molecular markers as random amplified polymorphic DNA (Sharma *et al.*, 1995) and ISSR (inter-simple sequence repeat) (Alvarez *et al.*, 1997; Sonnante and Pignone, 2001). More recently, proteomics, an innovative approach involving the comparison of several hundred to several thousand gene products revealed by two-dimensional gel electrophoresis of protein extracts, has been successfully applied to investigations of natural variations within plant species populations. Indeed, protein spots resolved by two-dimensional gel electrophoresis are de-facto genetic and physiological markers (Damerval *et al.*, 1994; De Vienne *et al.*, 1996) that can be used to assess genetic variability and to establish genetic distances and phylogenetic relationships between lines, species and genus (Thiellement *et al.*, 1999).

Different methods sample genetic variation at different levels, and have different powers of genetic resolution: 'neutral' DNA markers like microsatellites, widely used to study genetic diversity among populations, describe relations between species in terms of time divergence ('molecular clock') (Thiellement *et al.*, 1999), whereas phenotypic markers like morpho-physiological traits can provide information about adaptive responses to macroenvironmental conditions (David *et al.*, 1997). Obviously, a study involving all these techniques would provide more information about genetic variation than each technique used alone.

This study had two main goals: (1) to investigate the genetic relationship between two local lentil landraces severely threatened by genetic erosion, using an integrated approach and (2) to explore the potential of proteomics as a tool with which to investigate natural variation within and between lentil populations.

# Materials and methods

### Plant materials

In this study, we compared Molise lentil landraces (six from Conca Casale and seven from Capracotta—two small villages in the Molise region of central-south Italy) with five others: the widely marketed Turkish red lentils and Canadian lentils, the Castelluccio da Norcia ecotype, which is the only one in Italy to be awarded the IGP (geographically protected brand), and with two other autochthonous landraces from the central Appenine area, namely, Rascino and Colfiorito lentils.

*L. culinaris* Medik. seeds of autochthonous landraces were (Conca Casale and Capracotta) collected from local farmers, and stored in the Molise germoplasm bank at the University of Molise (Pesche), and commercially available varieties were used for the molecular, morphological and biochemical investigations.

### Morphological analysis

Eight morphological parameters were measured for each lentil seed population: area (mm<sup>2</sup>), perimeter (mm), major axis length (mm), minor axis length (mm) and roundness were measured from digital images by using the software Image tool V2.1; 100-seed weight, 100-seed volume and density ( $gml^{-1}$ ) were measured with a precision balance and a graduated cylinder.

One hundred seeds of each landrace population and each commercial variety were used for the morphological analysis. In order to process the data set using multivariate statistical analyses, we standardized the basic data matrix as follows: the variables were transformed to obtain a mean of zero and a variance of one. A standardized matrix of 18 populations using eight morphological variables was subjected to principal component analysis (PCA) and cluster analysis, which was performed using the unweighted pair group method with arithmetic (UPGMA) mean method (Sokal and Minchener, 1958).

# Protein analysis

Total protein was extracted according to the method of Rabilloud (2000) with minor modifications. Independent samples (2.0 g of dry seeds) were finely powdered in liquid N2 using a mortar, and the resulting powder was suspended in 10 ml of buffer A (10% trichloroacetic acid or tricarboxylic acid, 0.07% β-mercaptoethanol in cold acetone at -20 °C) and then filtered through Miracloth. Proteins were precipitated at -20 °C for 4 h, centrifuged at 35000 g for 15 min and rinsed twice with cold acetone, 0.07%  $\beta$ -mercaptoethanol for 4 h at -20 °C. The pellet was recovered by centrifuging at 35000 g for 15 min, dried under vacuum and solubilized in  $300\,\mu$ l of L<sub>2</sub> buffer (urea 7 M, thiourea 2 M, Chaps 4%, Triton X-100 1%, Tris 20 mM, dithiothreitol (DTT) 1% and ampholine 3-10 0.2% and 5-7 0.15%). Protein concentration was estimated according to Peterson (Lowry's method modified) using bovine serum albumin as standard (Peterson, 1977).

For two-dimensional gel electrophoresis, protein samples were focused on immobilized pII gradient (IPG) strips and then subjected to SDS-polyacrylamide gel electrophoresis (PAGE) electrophoresis. IPG strips (17 cm pH 4–7, Bio-Rad ReadyStrip; Bio-Rad, Hercules, CA, USA) were rehydrated overnight with 300 µl of isoelectrofocusing (IEF) buffer containing 600 µg of total proteins. Proteins were focused using a Protean IEF Cell (Bio-Rad) at  $12 \degree C$ , applying 250 V (90 min), 500 V(90 min), 1000 V (180 min) and 8000 V for a total of 55 kVh. After focusing, proteins were reduced by incubating the IPG strips with 1% w/v DTT for 15 min, and alkylated with 2.5% w/v iodoacetamide in 10 ml of 50 mM Tris-HCl pH 8.8, 6 M urea, 30% w/v glycerol, 2% w/v SDS and a dash of bromophenol blue, for 15 min. Two-dimensional electrophoresis was carried out using a Protean apparatus (Bio-Rad) and 12% polyacrylamide gels ( $17 \text{ cm} \times 1.5 \text{ mm}$ ) in 25 mM Tris pH 8.3, 1.92 M glycine and 1% w/v SDS, with 120V applied for 12h. Each sample was run in triplicate. Protein spots were annotated only if detected in two out of three gels.

Standard proteins (Bio-Rad) were used to estimate the molecular weight of the protein spots. Gels were fixed for 1 h with a solution of acetic acid 7% and methanol 40% and stained overnight with Brilliant blue G-Colloidal Concentrate (Sigma Aldrich, St Louis, MO, USA), and scanned using a Chemi Doc (Bio-Rad). Image analysis was performed using the PDQuest software 8.0 (Bio-Rad). Spot detection and matching between gels were performed automatically, followed by manual verification. After normalization of the spot densities against the whole-gel densities, the percentage volume of each spot was averaged for the three different replicates for each gel. For qualitative analysis, we considered the presence or absence of spot in each proteome map. PCA and cluster analysis (UPGMA) were carried out on

Jaccard's distance matrix computed on the presence/ absence matrix of proteins for the qualitative data analysis. The same multivariate statistical analyses were carried out on the standardized matrix of the abundance (relative volume) of protein spots (quantitative data analysis).

# **DNA** analysis

We used the ISSR markers to analyses genetic relationships between the different lentil landrace populations at DNA level. Five seeds for each landrace population were germinated in a growth chamber under controlled conditions (25 °C), and total genomic DNA was extracted from the youngest leaf of 10-day old seedlings using the adapted hexadecyltrimethylammonium bromide (CTAB) method reported by Taylor *et al.* (1995). DNA was quantified by measuring the absorbance at 260 nm on a spectrophotometer, and individual stock concentrations were adjusted to  $10 \text{ ng }\mu l^{-1}$  for PCR. A total of 10 ISSR primers were used for the analysis (Table 1; Supplementary material), 6 (1–6) primers were designed as reported elsewhere (Sonnante and Pignone, 2001), and the remaining 4 (7–10) were chosen randomly.

Amplification reactions were carried out in volumes of  $25 \,\mu$ l containing 50 ng template DNA, 1 unit of GOTaq DNA polymerase (Promega Inc., Madison, WI, USA),

 
 Table 1 Structure matrix extracted from the principal component analysis performed on eight morphological variables

Variable	PC1	PC2
Area (mm <sup>2</sup> )	0.988	-0.139
Perimeter (mm)	0.995	-0.082
Major axis length (mm)	0.996	-0.077
Minor axis length (mm)	0.994	-0.080
Roundness	0.559	0.794
100-seed weight	0.803	-0.576
100-seed volume	0.988	-0.070
Density $(g m l^{-1})$	0.590	0.785

Values indicate the correlation coefficients between the first two principal components and the morphological characters.

0.25 mM each deoxyribonucleotide triphosphate (Promega Inc.) and 1µM primer (Diatech-Operon Technologies, Huntsville, AL, USA), in  $1 \times$  reaction buffer (50 mM KCl, 1.5 mM MgCl<sub>2</sub>, 10 mM Tris-HCl pH 9.0). PCR reactions were run in a Thermalcycler (Applied Biosystem GeneAmp 2700) under the following conditions: 1 min at 94 °C, for initial denaturation, 35 cycles of 30 s at 94 °C (denaturation), 1 min at annealing temperature (Table 1), 2 min at 72 °C (extension), followed by 10 min at 72 °C for final extension of the single strands. ISSR amplified fragments were resolved on a 1.5% agarose gel stained with ethidium bromide and visualized under UV light. Gels were scanned with the Bio-Rad Gel Imaging System (Bio-Rad) and amplification profiles were analysed with the Quantity-One Band Analysis software (Bio-Rad). Bands were scored as presence or absence and the raw data were processed to obtain genetic distance matrix of Nei (1972) (among populations) by means of the GeneAlEx 6.0 programme (Peakall and Smouse, 2006). A hierarchical partition of genetic variation among and within populations was obtained by means of the analysis of molecular variance (AMOVA) (Excoffier et al., 1992). Subsequently, we used PCA to order the location of the samples in relation to genetic distances, and cluster analysis (UPGMA) to produce a dendrogram.

# Results

# Morphological data analysis

In our morphological analysis, we examined eight different seed characters (see 'Materials and methods'), of the Conca Casale and Capracotta lentil populations, and of the commercial lentil varieties (Castelluccio di Norcia, Turca rossa, Canadese, Rascino and Colfiorito). For each character average and standard deviation was calculated (Table 2 ; Supplementary material). We also carried out a PCA on a standardized matrix of 18 populations for eight morphological variables. In the scatter plot of PC1 and PC2 scores (Figure 1a), which accounted for 77.85 and 20.28% of the total variance respectively, the Conca Casale populations were well

Table 2 Structure matrix extracted from the principal component analysis performed on 71 variables

Protein	PC1	PC2									
Pr5	0.087	0.125	Pr42	-0.370	-0.515	Pr96	0.838	-0.017	Pr154	-0.007	0.688
Pr7	0.827	-0.115	Pr43	0.202	-0.160	Pr101	0.579	-0.042	Pr156	0.488	0.131
Pr8	0.934	-0.097	Pr46	0.316	0.032	Pr102	-0.250	-0.360	Pr161	0.357	0.626
Pr16	0.149	0.089	Pr47	0.367	-0.016	Pr105	0.274	0.133	Pr162	0.617	0.363
Pr17	0.421	0.146	Pr49	-0.275	-0.521	Pr106	0.669	0.012	Pr163	0.257	0.339
Pr18	0.678	0.254	Pr54	0.067	0.785	Pr108	0.915	-0.192	Pr165	0.314	0.359
Pr21	-0.453	-0.473	Pr56	0.189	0.153	Pr114	0.540	0.477	Pr166	0.122	0.172
Pr22	0.772	-0.079	Pr59	-0.382	-0.526	Pr115	0.039	0.328	Pr167	0.846	0.160
Pr23	-0.040	0.098	Pr60	-0.386	-0.348	Pr117	0.074	-0.282	Pr170	0.148	0.620
Pr25	0.344	0.026	Pr63	0.882	-0.145	Pr118	0.864	-0.085	Pr174	0.079	0.577
Pr28	0.144	0.053	Pr64	-0.020	0.511	Pr124	0.104	0.298	Pr175	-0.056	0.266
Pr31	0.836	-0.145	Pr74	0.847	-0.020	Pr127	0.082	0.795	Pr178	-0.045	-0.073
Pr32	0.056	0.329	Pr76	0.403	0.061	Pr128	-0.186	-0.456	Pr179	-0.235	-0.359
Pr34	0.409	0.795	Pr81	0.568	0.434	Pr130	0.494	0.342	Pr180	-0.415	-0.541
Pr38	-0.381	-0.448	Pr86	0.626	-0.002	Pr134	-0.063	0.804	Pr181	-0.038	0.350
Pr39	-0.399	-0.532	Pr93	0.926	-0.087	Pr136	0.652	0.517	Pr183	-0.366	-0.523
Pr40	0.016	0.064	Pr94	0.596	-0.108	Pr139	0.246	0.068	Pr189	0.408	0.780
Pr41	0.613	0.097	Pr95	0.560	0.268	Pr140	0.416	0.663			

Values indicate the correlation coefficients between the first two principal components and the differentially expressed protein spots.

Characterization of two autochthonous lentil landraces GS Scippa et al



**Figure 1** (a) Scatter plot of the first two principal components from the principal component analysis (PCA) performed on Euclidean distance matrix computed among 18 populations of *Lens culinaris* Medik., using eight morphological features. (b) Dendrogram computed from the Euclidean distance matrix among 18 populations of the *L. culinaris*, using eight morphological features. Hierarchical clustering was performed using unweighted pair group method with arithmetic (UPGMA) mean method and average linkage criterion for linkage.

separated from all other populations. Along PC1, Capracotta population 5 was totally distinct from the other populations, and the differences were in terms of size. In fact, area, perimeter and the lengths of the major and minor axes were the morphological variables more closely correlated to the PC1. Also weight and volume differentiated population 5 from the others along this component (Table 1). Along PC2 are the differences in roundness and density (Table 1) that distinguish the Conca Casale populations from the other populations.

We examined the same standardized matrix of morphological features with cluster analysis, and the resulting dendrogram revealed three main clusters (Figure 1b). The first cluster is divided into subcluster 1a, which is related to the Conca Casale populations, and cluster 1b, which is related to Capracotta populations 4 and 6 and the Rascino population. A comparison of the results of cluster analysis with the PCA results showed that the three latter populations and the Conca Casale populations were related along PC1, which implies they are similar in size. Cluster 2 consists of all commercial populations and the Capracotta populations, except for population 5, which constitutes the third cluster and differs from the others in size.

### Protein analysis

We examined the genetic relationships between and within lentil populations by comparing their seed proteomes. Total proteins extracted as described under 'Materials and methods', were resolved by twodimensional electrophoresis and stained with colloidal Coomassie blue. In a preliminary two-dimensional analysis, where total seed proteins were isoelectrofocused on a 3-10 IGP strip, we found that the majority of proteins were concentrated in the range of pH between 4 and 7 (data not shown). By resolving the protein on a 4-7 IGP, the maps obtained for each population were highly reproducible, with an average of approximately 193 well-resolved spots, without streaking. For each sample, the triplicate gels were first matched to create an average gel containing spots observed at least twice in the three gels. Using the PDQuest software 8.0 (Bio-Rad), we then matched the



**Figure 2** Two-dimensional polyacrylamide gel electrophoresis (PAGE) reference map of total lentil seed proteins from the Capracotta 1 population. Two-dimensional PAGE conditions: pH 4–7 IPG (first dimension) and 12% SDS–PAGE (second dimension). Protein spots are visualized by colloidal Coomassie blue staining. Using the PDQuest software 8.0 (Bio-Rad), the average gel of each population was compared with the reference map gel to identify the differentially expressed protein spots (arrows).

average gel of each population with a master gel. The map of Capracotta population 1 was chosen as master gel because it had the highest number and the best resolution of spots. Conca Casale population 5 was excluded from the biochemical analysis because of the poor quality of proteome maps. As reported in Figure 2, the statistical analysis carried out with the PDQuest software showed that at least 71 proteins were differentially expressed among the different landrace populations. To analyse qualitative biochemical data, we applied a PCA using Jaccard's distance matrix (Jaccard, 1908). The scatter plot of PC1 and PC2 (accounting for 38.43 and 12.35% of total variance, respectively), and the scatter plot of PC1 and PC3 (accounting for 8.55%) are shown in Figures 3a and b. Two main groups were recognizable along PC1: because of negative loading of 139



△ Conca Casale

**Figure 3** Scatter plots of the first three principal components from the principal component analysis (PCA) performed on Jaccard's distance matrix computed on the presence/absence of 193 proteins (qualitative data) in the 17 populations of *Lens culinaris*. (a) Distribution along PC1 and PC2 (b) Distribution along PC1 and PC3.



**Figure 4** Dendrogram built from the distance matrix calculated according to the Jaccard index on all the 193 spots of the 17 populations of *Lens culinaris*, using qualitative biochemical data (presence/absence). Hierarchical clustering was performed using unweighted pair group method with arithmetic (UPGMA) mean method and average linkage criterion for linkage.

the axis, the populations of Capracotta were grouped with the four commercial populations; whereas because of positive loading, all the Conca Casale populations were grouped with populations 3 and 4 of Capracotta and with the Rascino population (Figure 3a). PC2 distinguished Capracotta population 5, which was previously morphologically different from all the other populations (Figure 3a). Finally, PC3 separated the Conca Casale populations from the three populations (populations 3 and 4 of Capracotta and the Rascino population) that were similar along PC1 (Figure 3b).

Cluster analysis yielded similar results (Figure 4). Three clusters were identified: cluster 1 comprised the four commercial populations and all Capracotta populations, which were grouped because of negative loading of PC1, whereas cluster 2 was formed by Capracotta population 5, which was separated from other populations along PC2. Finally, the third cluster comprised the Conca Casale population, the Rascino population and Capracotta populations 3 and 4. This cluster was separated from all other populations because of positive loading of PC1.

The quantitative biochemical analysis was based on the relative abundance (% vol) of protein spots in the map of each population. Thus, we analysed also the quantitative biochemical matrix of 17 populations and 193 proteins by means of PCA. The matrix of quantitative data was subjected to PCA and all cases were projected on the first two principal components (Figure 5a), which explained 31.92 and 10.69% of total variance, respectively. Along PC1, Capracotta populations 1 and 2 were distinct from other populations; whereas, because of the negative loadings of PC2, all Conca Casale populations and Capracotta population 5 were distinct from all other populations because of positive loadings (Figure 5a).

From a comparison of the PCA and PDQuest results (Figure 2), the protein spots that characterize the different populations may be identified. In fact, as reported in Table 2, among the 71 differentially expressed protein spots, 8 distinguish populations 1 and 2 of Capracotta from the other populations along PC1 (Pr7, Pr8, Pr31, Pr63, Pr74, Pr108, Pr118, Pr167); 5 protein spots characterize the Capracotta population 5 (the macrosperma) along PC2 (Pr34, Pr54, Pr127, Pr134, Pr189); 12 protein spots separate all the Conca Casale populations from the others along both PC1 and PC2 (Pr21, Pr22, Pr38, Pr39, Pr42, Pr49, Pr59, Pr60, Pr178, Pr179, Pr180, Pr183); 8 protein spots characterize the Conca Casale population only along PC1 (Pr23, Pr64, Pr102, Pr128, Pr134, Pr154, Pr175, Pr181), whereas an additional 15 discriminate the same population only along PC2 (Pr7, Pr8, Pr31, Pr43, Pr47, Pr63, Pr74, Pr86, Pr93, Pr94, Pr96, Pr101, Pr108, Pr117, Pr118). The remaining 48 protein spots discriminate the Capracotta varieties (3, 4, 6, 7) from the commercial varieties (Table 2 ; Figures 2 and 5).

Cluster analysis (UPGMA) of the same quantitative biochemical data revealed three main clusters (Figure 5b). The first cluster consisted of all Conca Casale populations, which were separated from other populations along PC2. Cluster 2 consisted of all commercial populations and the Capracotta populations in which populations 1 and 2 formed a subcluster. Population 5 of Capracotta formed the third cluster.

### DNA data analysis

Genetic relations among the autochthonous lentil landraces and commercial varieties were evaluated by using the ISSR markers that differentiate closely related genotypes (Zietkiewicz *et al.*, 1994). Of the 10 ISSR primers initially tested, 7 (Table 1 ; Supplementary material; ISSR 1, 2, 3, 5, 6, 9, 10) yielded satisfactory polymorphic and reproducible amplification profiles and were thus used to analyse 90 samples belonging to 18 populations (13 autochthonous, and 5 commercial). Each primer amplified an average of 40 major scorable DNA

Characterization of two autochthonous lentil landraces GS Scippa et al



**Figure 5** (a) Scatter plot of the first two principal components from the principal component analysis (PCA) performed on the quantitative biochemical matrix of 17 populations of *Lens culinaris* and 193 proteins. (b) Dendrogram built from the quantitative biochemical matrix of 17 populations of *L. culinaris* and 193 proteins. Hierarchical clustering was performed using unweighted pair group method with arithmetic (UPGMA) mean method and average linkage criterion for linkage.

 Table 3 Summary table of hierarchical analysis of molecular variance

Summary AMOVA table	d.f.	SS	Variance	Variance (%)	PhiPT	Р
Among species Within species Total	17 23 40	536.57 96.33 632.390	12.03 4.19 16.22	74 26	0.742	< 0.01

AMOVA, analysis of molecular variance; d.f., degrees of freedom; *P*, relative probability; SS, sum of squares.

markers, but only the polymorphic marker, accounting for the 70%, was considered for further analysis.

The AMOVA showed that most (74%) of the total genetic variability was among populations; the remaining 26% of total diversity was within populations. The PhiPT indicated a significant genetic differentiation among populations (Table 3).

PCA was performed on Nei's genetic distance and Figure 6a shows the scatter plot of PC1 and PC2 scores, which accounted for 53.15 and 17.50%, respectively of total variance. Along PC1, Conca Casale populations were well separated from all other populations. Along PC2, the populations of Capracotta were separated from all others, apart from the Castelluccio di Norcia population. The Turca Rossa population was well differentiated from the others.

The dendrogram (Figure 6b) constructed using the UPGMA confirmed PCA results. In fact, four main clusters were identified: all Conca Casale populations were grouped in cluster 1, and all the Capracotta populations were grouped in cluster 3 together with the Castelluccio di Norcia population. All commercial populations were grouped in cluster 2, except Turca rossa that formed cluster 4.

# Correlation between matrices

To evaluate the congruency of the results obtained with the different procedures, and to verify if our integrated approach of analyses at morphological, DNA and protein level is a valid tool with which to assess genetic variation, we examined the correlation between the matrices of the data sets using the Mantel's test (Mantel, 1967). We carried out Mantel's test using the Numerical Taxonomy System (NTSYS pc version 2.2) software, and three genetic distance matrices (Nei, 1972): dissimilarity matrices of morphological traits (Euclidean distance), Jaccard's distance matrix of qualitative biochemical data and Pearson's distance matrix of quantitative biochemical data. As shown in Table 4, the genetic distance was closely correlated with morphological data, with quantitative biochemical data and with qualitative biochemical data. Moreover, a three-way Mantel test (Smouse et al., 1986) confirmed a close correlation (0.763; P = 0.001) among matrices of DNA markers, morphological and qualitative protein data. Significant correlations were also identified between geographical distance and distance matrices of DNA markers (0.472; P = 0.002), quantitative protein (0.486; P = 0.001) and qualitative protein (0.509; P = 0.001) data, in Italian populations of L. culinaris.

# Discussion

A wide variety of methods has been used to investigate genetic similarities and relations among L. culinaris Medik. landraces. Morphological and phenological studies have revealed relevant differences in Italian lentil populations (Gallo et al., 1997), whereas few studies have evaluated genetic variation at molecular (Sonnante and Pignone, 2001) or biochemical level (Senatore *et al.*, 1992). Morphological and phenological traits are often controlled by multiple genes and are subjected to the action of environmental factors, and differences between clones or closely related species are not always absolute (Ahmad et al., 1996). However, patterns of variation, analysed at genome level, may be influenced by the genetic markers used (Sonnante and Pignone, 2001). Different methods have different powers of genetic resolution and provide different information. In this study we used a combination of morphological, genomic and proteomic analyses to characterize two autochthonous lentil landraces and establish their relationships with several widely used commercial varieties.

Both PCA and cluster analysis of the morphological traits of lentil seeds separated the 18 populations into

Characterization of two autochthonous lentil landraces GS Scippa et al



**Figure 6** (a) Scatter plot of the first two principal components from the principal component analysis (PCA) performed on Nei's genetic distance (1972) computed among 18 populations of *Lens culinaris*, using 10 inter-simple sequence repeat (ISSR) primers. (b) Dendrogram computed from the matrix of Nei's genetic distances (1972) among 18 populations of the *L. culinaris*, using 10 ISSR primers. Hierarchical clustering was performed using unweighted pair group method with arithmetic (UPGMA) mean method and average linkage criterion for linkage.

Table 4 Mantel's test of correlation among distance matrices

	Genetic	Morphological
Genetic	1	0.641**
Morphological	0.641**	1
Proteins qualitative data	0.841**	0.550**
Proteins quantitative data	0.628**	0.527**

\*\**P* < 0.001.

three main groups: group 1 represented by six populations from Conca Casale (subgroup 1a), two populations from Capracotta and the commercial variety Rascino (subgroup 1b); group 2 consisted of four populations from Capracotta and the commercial varieties Colfiorito, Castelluccio, Canadese and Turca Rossa; group 3 consisted of population 5 from Capracotta. Area, perimeter, major and minor axes were the morphological variables that distinguished, along PC1, population 5 from Capracotta from the other populations, including those from central Appenine areas.

On the basis of seed size, cultivated lentils are classified as *macrosperma* and *microsperma*, which are considered subspecies, races or varieties (Cubero and Malbon, 1984). Our results show that Capracotta population 5 can be classified *macrosperma*, whereas all the other populations we studied are *microsperma*. Lentil populations of Conca Casale had very homogeneous morphological traits and their reduced density and roundness clearly distinguished them from the other *microspermae* along PC2.

In the first half of the 20th century, *microsperma* and *macrosperma* lentils were cultivated in marginal areas of the Appenninic ridge, however changes in land use over the past 50 years resulted in the disappearance of many of these populations (Piergiovanni, 2000). Although the increase in seed size is one of the most conspicuous changes resulting from domestication, and the large-seeded types are considered more 'advanced' (Zohary, 1976), *macrosperma* landraces are not commercially successful and have thus become the endangered landrace species. An overview of recent cultivation of lentils in Italy identified 63 populations mainly in Appennine

and sub-Appennine marginal regions (Laghetti *et al.*, 1996; Hammer *et al.*, 1999). The list included the *microspermae* of Capracotta, but neither the *macrosperma* of Capracotta nor the *microsperma* landrace of Conca Casale.

The results of the quantitative (continuous variation in spot intensity) and qualitative (presence/absence) biochemical analysis of total seed proteins separated by two-dimensional electrophoresis confirmed our morphological data. In both cases, the PCA differentiated the *macrosperma* population of Capracotta from all other populations, including the *microspermae* lentils from Capracotta. In the case of *microspermae*, the Conca Casale populations had a homogeneous quantitative and qualitative protein composition and clustered together. Differently, the Capracotta populations were more heterogeneous and clustered with the commercial lentil varieties depending on their qualitative or quantitative composition.

Earlier studies investigated the genetic variability within and among different lentil populations using isozymes (De La Rosa and Jouve, 1992; Rodriguez et al., 1999) or seed storage proteins (De La Rosa and Jouve, 1992; Echeverrigaray et al., 1998). In our study, we used two-dimensional electrophoresis to evaluate genetic variability. Unlike one-dimensional electrophoresis, which separates proteins according to their molecular weight, two-dimensional electrophoresis resolve proteins based on isoelectric point and molecular weight, allowing the examination of a broad spectrum of proteins and, consequently, a substantially larger number of protein-encoding loci. Moreover, twodimensional electrophoresis is a high-resolution technique that separates thousands of genetic products (protein spots) on a single gel and detects isomorphs, polymorphisms and changes such as post-translational Modifications (that is phosphorylation, glycosylation, acetylation and methylation) induced for instance, by precise ecological situations experienced by individuals (David et al., 1997; Chevalier et al., 2004). Furthermore, two-dimensional electrophoresis coupled with mass spectrometry are the key components of the current proteomics technology, which has been successfully used

140

in various phylogenetic studies (Mosquera et al., 2003; Chevalier et al., 2004). In addition, proteomic approach represents a powerful tool with which to identify 'physiological markers', and to determine if differences between populations are due to adaptation to particular environments (David et al., 1997). In our study, we compared the two-dimensional maps of the different populations to assess genetic relationships among lentil species and to evaluate if differences between and within populations are stochastic or are specific protein markers. Our results show that at least 71 protein spots are qualitatively and/or quantitatively differentially expressed among the different landraces. Furthermore, the correlation between the differentially expressed protein spots and the PC scores shows the ones that have greater major weight in the population discrimination. These protein spots recognized on the proteomic maps will be analysed by means of amino-acid sequencing to characterize further and to determine their physiological function.

The three data sets (morphological, qualitative and quantitative protein analysis) gave similar results and were significantly correlated with DNA markers data. In fact, the DNA investigations, carried out with ISSR polymorphisms, confirmed the data obtained with the morphological and protein analyses. In detail, multivariate analyses (AMOVA, PCA, UPGMA) of ISSR data showed that the genetic variation is higher among than within populations, and that the populations of Conca Casale are all grouped in a separate cluster that is distinct from the Capracotta and from all other commercial populations. Moreover, the seven Capracotta populations, including the macrosperma, and the commercial variety from Castelluccio were grouped in a single cluster, well distinguished from the Conca Casale cluster and from the other commercial populations. Thus, even though the morphology and proteome of the Capracotta macrosperma lentil variety (population 5) distinguished it from the other six populations, it does belong to the same genetic group. Furthermore, it is possible that the commercial variety of Castelluccio di Norcia may have the same origin as the landrace from Capracotta.

We next applied Mantel's test (Mantel, 1967) to investigate the correlation between the data with three different methods, and to determine whether proteomics is a valid tool with which to evaluate genetic relationships between and among lentil populations. The results showed significant correlations between the matrices. The genetic distance was highly correlated with morphological data, and with quantitative and qualitative biochemical data. Moreover, when a three-way Mantel's test was performed among DNA, morphological and qualitative biochemical data, the high correlation found confirmed a strong relation among these three data set, which gave similar information about population relationship.

The high correlation between the DNA molecular markers and protein analysis data may indicate that the variability observed at proteomic level could be due to alterations in proteins in response to specific environments and related to allelic variation between populations and/or to epigenetic mechanisms. In fact, qualitative protein variation involves the presence or absence of spots according to the genotype and is often due to such allelic variations resulting in protein isoforms with different isoelectric points or apparent molecular masses (De Vienne *et al.*, 1996), whereas quantitative variation involves continuous variation in spot intensity, which reflects differences in the relative abundance of the protein whose expression can be controlled by several loci (Damerval *et al.*, 1994).

Moreover, there was a significant geographical structure among the Italian lentil populations at both genetic and biochemical level. This suggests that a reduced genetic diversity within populations can be related to geographical isolation of populations, which increased the distinctiveness features in the biochemical composition of the two autochthonous populations of *L. culinaris*.

In conclusion, our study shows that: (1) the lentil populations of Conca Casale are clearly differentiated from the five commercial populations, at morphological, protein and DNA level; (2) all the Capracotta populations, including the *macrosperma*, separate very well, at DNA level, from all the Conca Casale populations and from the commercial varieties, except Castelluccio di Norcia, which seems to have similar origin. However, at morphological and protein level the Capracotta populations show a more heterogeneous segregation, as for instance although the Capracotta macrosperma and microsperma landraces belong to the same genetic group, they have developed different morphological and biochemical traits and (3) the proteomic approach can be considered a powerful tool in the phylogentics and in the identification of the physiological and/or environmental markers that characterize different populations. The strong correlation found between the genomic and proteomic data strongly suggests that protein markers, resolved by two-dimensional electrophoresis, well characterize the different landraces, evidencing differences according to their physiological status. In fact, taken together our results support the use of proteomics as a tool in phylogenetic studies and highlight its potential to identify proteins that may be specific markers of particular environmental conditions. We plan to purify further and identify by mass spectrometry the protein spots that differentiate the various populations in an attempt to cast light on their physiological function.

# Acknowledgements

The authors thank Jean Gilder for text editing and ARSIA Molise (Agenzia Regionale Sviluppo Agricolo) for financial support. Lentil landaraces are conserved at the Molise Germoplasm Bank (www.bancagermoplasmamolise.un. mol.it).

# References

- Ahmad M, McNeil DL, Fautrier AG, Armstrong KF, Paterson AM (1996). Genetic relationships in *Lens* species and parentage determination of their interspecific hybrids using RAPD markers. *Theor Appl Genet* **92**: 1091–1098.
- Alvarez MT, Garcia P, Perez de la Vega M (1997). RAPD polymorphism in Spanish lentil landraces and cultivars. *J Genet Breed* **1**: 91–96.
- Chevalier F, Martin O, Rofidal V, Devauchelle AD, Barteau S, Sommerer N *et al.* (2004). Proteomic investigation of natural variation between *Arabidopsis* ecotypes. *Proteomics* **4**: 1372–1381.

- Cubero A, Malbon CC (1984). The fat cell beta-adrenergic receptor. Purification and characterization of a mammalian beta 1-adrenergic receptor. *J Bio Chem* vol. 259. **2**: 1344–1350.
- Damerval C, Maurice A, Josse J-M, De Vienne D (1994). Quantitative trait loci underlying gene product variation: a novel perspective for analyzing regulation of genome expression. *Genetics* **137**: 289–301.
- David JL, Zivy M, Cardin ML, Brabant P (1997). Protein evolution in dynamical managed population of wheat: adaptative responses to macro-environmental conditions. *Theor Appl Genet* **95**: 932–941.
- De La Rosa L, Jouve N (1992). Genetic variation for isozyme genes and proteins in Spanish primitive cultivars and wild subspecies of genus *Lens. Europhytica* **59**: 181–187.
- De Vienne D, Burstin J, Gerber S, Leonardi A, Le Guilloux M, Murigneux A *et al.* (1996). Two-dimensional electrophoresis of proteins as a source of monogenic and codominant markers for population genetics and mapping the expressed genome. *Heredity* **76**: 166–177.
- Echeverrigaray S, Oliveira AC, Carvalho MTV, Derbyshire E (1998). Evaluation of the relationship between lentil accessions using comparative electrophoresis of seed proteins. *J Genet Breed* **52**: 89–94.
- Erskine W (1997). Lessons for breeders from landraces of lentil. *Euphytica* **93**: 107–112.
- Erskine W, Adham Y, Holly L (1989). Geographic distribution of variation in quantitative traits in a world lentil collection. *Euphytica* **43**: 97–103.
- Excoffier L, Smouse PE, Quattro JM (1992). Analysis of molecular variance inferred from metric distances among DNA haplotypes: application to human mitochondrial DNA restriction data. *Genetic* 131: 479–491.
- Gallo G, Barbera AC, Venora G (1997). Valutazione agronomica di popolazioni di lenticchia (*Lens culinaris* Medik.) in ambienti semi-aridi. In: Proceedings of the 3rd Convegno Nazionale sulla Biodiversità; 16–17 June 1997; Reggio Calabria, Italy. pp 279–288.
- Hammer K, Laghetti G, Perrino P (1999). A checklist of the cultivated plants of Ustica (Italy). *Genet Resour Crop Evol* **46**: 95–106.
- Jaccard P (1908). Nouvelles rescherches sur la distribution florale. *Bull Soc Vaud Sci Nat* **44**: 223–270.

Ladizinsky G (1993). Wild lentils. Crit Rev Plant Sci 12: 169-184.

- Laghetti G, Hammer K, Perrino P (1996). Plant genetic resources in Pantelleria and Pelagie archipelagos, Italy: collecting and conservation of local crop germoplasm. *IPGRI Plant Genet Resour Newslett* **108**: 17–25.
- Lazaro A, Ruiz M, De la Rosa L, Martin I (2001). Relationships between agro/morphological characters and climatic parameters in Spanish landraces of lentil (*Lens culinaris* Medik.). *Genet Resour Crop Evol* **48**: 239–249.
- Lev-Yadun S, Gopher A, Abbo S (2000). The cradle of agriculture. *Science* **288**: 1602–1603.
- Mantel NA (1967). The detection of disease clustering and a generalized regression approach. *Cancer Res* **27**: 209–220.
- Mosquera E, López JL, Alvarez G (2003). Genetic variability of the marine mussel *Mytilus galloprovincialis* assessed using two-dimensional electrophoresis. *Heredity* **24**: 432–442.
- Nei M (1972). Genetic distance between populations. *Am Nat* **106**: 283–292.

- Peakall R, Smouse PE (2006). GENEALEX ver.6: genetic analysis in Excel. Population genetic software for teaching and research. *Mol Ecol* **6**: 288–295.
- Peterson GL (1977). A simplification of the protein assay method of Lowry *et al.* which is more generally applicable. *Anal Biochem* **83**: 346–356.
- Piergiovanni AR (2000). The evolution of lentil (*Lens culinaris* Medik.) cultivation in Italy and its effects on the survival of autochthonous populations. *Genet Resour Crop Evol* **47**: 305–314.
- Piergiovanni AR, Taranto G (2005). Assessment of the genetic variation in Italian lentil populations by electrophoresis (SDS–PAGE) of seed storage proteins. *Genet Resour Newslett* **141**: 42–47.
- Rabilloud T (2000). Proteome research: two-dimensional gel electrophoresis and identification methods. Principles and Practice, Springer, Verlag Berlin Heidelberg, Germany 109–110.
- Rodriguez MM, Paredes CM, Bescerra VV (1999). Isoenzymatic diversity in Chilean lentil germoplasm (*Lens culinaris* Medik.). Agricoltura Tecnica 59: 186–195.
- Senatore F, Schettino O, Basso F (1992). Fatty acid and sterol content in eight populations of lentil (*Lens esculenta* Moench). *Rivista Italiana Sostanze Grasse* **69**: 329–331.
- Sharma SK, Dawson IK, Waugh R (1995). Relationships among cultivated and wild lentils revealed by RAPD analysis. *Theor Appl Genet* **91**: 647–654.
- Smouse PE, Long JC, Sokal RR (1986). Multiple regression and correlation extensions of the Mantel Test of Matrix Correspondence. *Syst Zool* 35: 627–632.
- Sonnante G, Pignone D (2001). Assessment of genetic variation in a collection of lentil using molecular tools. *Euphytica* **120**: 301–307.
- Sokal RR, Minchener CD (1958). A Statistical method for evaluating systematic relationships. *Univ Kansas Sci Bull* **28**: 1409–1438.
- Taylor PWJ, Fraser TA, Ko H-L, Henry RJ (1995). RAPD analysis of sugarcane during tissue culture. In: Terzi M, Cella R, Falavigna A (eds). *Current Issues in Plant Molecular and Cellular Biology*. Kluwer Academic Int.: Dordrecht, The Netherlands. pp 241–246.
- Thiellement H, Bahrman N, Damerval C (1999). Proteomics for genetic and physiological studies in plants. *Electrophoresis* **20**: 2013–2026.
- Tullu A, Kusmenoglu I, McPhee KE, Muehlbauer FJ (2001). Characterization of core collection of lentil germoplasm for phenology, morphology, seed and straw yields. *Genet Resour Crop Evol* **48**: 143–152.
- Yan'Kov I, Varlahov M, Andreeva N (2001). Phenotypic Variability for Traits of Aagronomic Interest in the Lentil Germoplasm Collection Maintained in the Vavilov Institute. *Proceedings of the 4th European Conference on Grain Legumes*; 8– 12 July 2001; Cracow, Poland. pp 174–175.
- Zietkiewicz E, Rafalski A, Labuda D (1994). Genome fingerprinting by simple sequence repeat (SSR)-anchored polymerase chain reaction amplification. *Genomics* **20**: 176–183.
- Zohary D (1976). Lentils. In: Simmonds NW (ed). Evolution of Crop Plants. Longman: London. pp 63–64.
- Zohary D (1999). Morphyletic vs. polyphyletic origin of the crops on which agriculture was founded in the Near East. *Genet Res Crop Evol* **46**: 133–142.

Supplementary Information accompanies the paper on Heredity website (http://www.nature.com/hdy)

-