Title

Landscape genetics, adaptive diversity and population structure in Phaseolus vulgaris

Journal

New Phytologist, 209(4)

ISSN

0028-646X

Authors

Rodriguez, Monica
Rau, Domenico
Bitocchi, Elena
et al.

Publication Date

2016-03-01

DOI

10.1111/nph.13713

Peer reviewed

New Phytologist

Landscape genetics, adaptive diversity, and population structure in *P. vulgaris*

Journal:	*New Phytologist*
Manuscript ID:	NPH-MS-2015-19711
Manuscript Type:	MS - Regular Manuscript
Date Submitted by the Author:	06-May-2015

Complete List of Authors:	Rodriguez, Monica; Università degli Studi di Sassari, Dipartimento di Agraria Rau, Domenico; Università degli Studi di Sassari, Dipartimento di Agraria Bitocchi, Elena; Università Politecnica delle Marche, Dipartimento di Scienze Agrarie, Alimentari ed Ambientali Bellucci, Elisa; Università Politecnica delle Marche, Dipartimento di Scienze Agrarie, Alimentari ed Ambientali Biagetti, Eleonora; Università Politecnica delle Marche, Dipartimento di Scienze Agrarie, Alimentari ed Ambientali Carboni, Andrea; Consiglio per la ricerca e la sperimentazione in agricoltura (CRA-CIN), Centro di Ricerca per le Colture Industriali Gepts, Paul; University of California, Department of Plant Sciences / MS1 Nanni, Laura; Università Politecnica delle Marche, Dipartimento di Scienze Agrarie, Alimentari ed Ambientali Papa, Roberto; Università Politecnica delle Marche, Dipartimento di Scienze Agrarie, Alimentari ed Ambientali Attene, Giovanna; Università degli Studi di Sassari, Dipartimento di Agraria; Università degli Studi di Sassari, Centro per la Conservazione e Valorizzazione della Biodiversità Vegetale

Key Words:	Phaseolus vulgaris, wild accessions, landraces, SNP genotyping, genetic diversity, landscape genetics, domestication

Landscape genetics, adaptive diversity, and population structure in *P. vulgaris*

Monica Rodriguez[1,2§], Domenico Rau[1§], Elena Bitocchi[3], Elisa Bellucci[3], Eleonora Biagetti[3], Andrea Carboni[4], Paul Gepts[5], Laura Nanni[3], Roberto Papa[3], Giovanna Attene[1,2]*

[1] Dipartimento di Agraria, Università degli Studi di Sassari, Via E. de Nicola, 07100 Sassari, Italy

[2] Centro per la Conservazione e Valorizzazione della Biodiversità Vegetale, Università degli Studi di Sassari, Surigheddu, 07040 Alghero, Italy

[3] Dipartimento di Scienze Agrarie, Alimentari ed Ambientali, Università Politecnica delle Marche, Via Brecce Bianche, 60131, Ancona, Italy

[4] Centro di Ricerca per le Colture Industriali (CRA-CIN), Consiglio per la Ricerca e la Sperimentazione in Agricoltura, via di Corticella, 133, 40128 Bologna, Italy

[5] Department of Plant Sciences / MS1, University of California, 1 Shields Avenue, Davis, CA 95616, USA

[§] These authors contributed equally to this study

*Corresponding author: Giovanna Attene

Tel. +39-07-9229225

E-mail: attene@unss.it

Word counts

Main body of the text (Introduction, Materials and Methods, Results, Discussion and Acknowledgments): 6248

Introduction: 631

Materials and Methods: 1371

Results: 2537

Discussion: 1745

Acknowledgments: 0

Number of Figures (all in colour): 11

Number of Tables: 1

Supplementary Information: Supplementary Tables submitted in a unique Excel file (Suppl Tables.xlsx): Table S1, Table S2, Table S3, Table S4, Table S5, Table S6; Supplementary Figures: Fig. S1, Fig. S2, Fig. S3, Fig. S4, Fig. S5, Fig. S6, Fig. S7, Fig. S8, Fig. S9, Fig. S10, Fig. S11; Supplementary Notes: Note S1.

Summary

- We studied the organisation of the genetic variation of the common bean (*Phaseolus vulgaris*) in its centres of domestication.
- We used 131 single nucleotide polymorphisms to investigate 417 wild common bean accessions, including Mesoamerican and Andean genotypes, and we compared these to a representative sample of 160 domesticated genotypes, for a total of 577 accessions.
- By analysing the genetic spatial patterns of wild common bean, we have documented the existence of several genetic groups and the occurrence of variable levels of diversity in Mesoamerica and the Andes. Moreover, using a landscape genetics approach, we demonstrate that both demographic processes and selection for adaptation are responsible for the observed genetic structure.
- We show that the study of correlations between markers and ecological variables at a continental scale can help in the identification of genes involved in local adaptation. Also, we located the putative area of common bean domestication in Mesoamerica, in the Oaxaca Valley, and in the Andes, in southern Bolivia-northern Argentina. These observations are of paramount importance for the conservation and exploitation of the genetic diversity preserved within this species and other plant genetic resources.

Key words *Phaseolus vulgaris*, wild accessions, landraces, SNP genotyping, genetic diversity, landscape genetics, domestication

Introduction

The common bean (*Phaseolus vulgaris* L.) represents the most important food legume for direct use, and based on the current trends in population growth, its consumption can be expected to increase (Bellucci *et al.*, 2014a). Thus, for common bean breeding, it will be of primary importance to obtain improved varieties that can face compelling future challenges, such as climate change, sustainability, and food security.

Wild *P. vulgaris* has a Mesoamerican origin and its subsequent independent expansions to South America gave rise to the following wild gene pools: two in the Andes (Bitocchi *et al.*, 2012; Desiderio *et al.*, 2013); one in the northern Andes (i.e., Ecuador and northern Peru) that is characterised by a specific seed storage protein, phaseolin type I (the 'Inca') that is not present in the other gene pools (Kami *et al.*, 1995); and one further south (i.e., southern Peru, Bolivia and Argentina; Kami *et al.*, 1995; Bitocchi *et al.*, 2012). These have been extensively investigated using phenotypic, biochemical and genetic data that have shown the higher diversity and stronger population structure of the Mesoamerican gene pool with respect to the Andean gene pool (Gepts *et al.*, 1986; Singh, 1989; Lynch *et al.*, 1992; Kwak & Gepts, 2009; Cortés *et al.*, 2011; Desiderio *et al.*, 2013; Goretti *et al.*, 2014; Bellucci *et al.*, 2014b; Schmutz *et al.*, 2014).

In Central and South America, wild *P. vulgaris* underwent two independent domestication events that led to the domesticated Mesoamerican and Andean gene pools. This offers a unique scenario to study the domestication process (Bitocchi *et al.*, 2013). The domestication bottleneck was stronger in the Mesoamerican than the Andean gene pool, probably because loss of diversity occurred in the Andes before domestication (Bitocchi *et al.*, 2012; Bellucci *et al.*, 2014a; Schmutz *et al.*, 2014). Although domestication of the common bean has been the subject of different studies, the definitive geographical localisation of these events remains controversial (Beebe *et al.*, 2001; Chacón *et al.*, 2005; Kwak *et al.*, 2009; Bitocchi *et al.*, 2013). The areas suggested as domestication sites are the Lerma Santiago Basin (Kwak and Gepts, 2009), and more recently, the Oaxaca Valley (Bitocchi *et al.*, 2013) in Mesoamerica, and southern Peru (Chacón *et al.*, 2007) and southern Bolivia and northern Argentina (Bitocchi *et al.*, 2013) in South America.

To achieve efficient management and deployment of genetic resources, the need to decipher the population structure, crop history and adaptation is a fundamental prerequisite (Diamond & Bellwood, 2003; Kovach *et al.,* 2007; van Zonneveld *et al.*, 2014). In this regard,

the analysis of molecular data in combination with phenotypic and spatial data can be particularly useful. Indeed, a description of the distribution of genetic diversity and its relation to geographical and/or ecological information can provide fundamental insights into evolutionary history, natural selection, adaptation, and the process of domestication (Papa & Gepts, 2003; Papa *et al.*, 2007; Eckert *et al.*, 2010; Rodriguez *et al.*, 2013; Kraft *et al.*, 2014). Indeed, comparisons of genetic and spatial data with archaeobotanical and palaeobiolinguistic data have recently been shown to be useful for tracing back the geographical origins of domesticated pepper (Kraft *et al.,* 2014).

In the present study we used 131 single nucleotide polymorphism (SNP) markers to analyse the spatial distribution of the genetic diversity of a large collection of 577 *P. vulgaris* accessions that included wild and domesticated forms of both the Mesoamerican and Andean gene pools. With particular reference to the Mesoamerican gene pool, we addressed three subtasks: (a) determination of the population structure of wild *P. vulgaris* in the Mesoamerican centre of diversity, while also disentangling the role of geographical and ecological factors in the shaping of the genetic differentiation; (b) detection of loci under selection at a continental scale; and (c) identification of the most likely domestication sites of the common bean.

Materials and methods

Plant materials

In the present study, we analysed 577 *P. vulgaris* accessions subdivided into 435 accessions that belong to the Mesoamerican gene pool (335 wild [MW]; 100 domesticated [MD]), 128 accessions from the Andean gene pool (68 wild [AW]; 60 domesticated [AD]), and 14 wild accessions from northern Peru–Ecuador characterised by the phaseolin type I (PhI) ancestral seed storage protein in *Phaseolus* (Kami *et al.*, 1995). Each accession was a single-seed-descent homozygote individual donated by a gene bank or collected *in-situ* by different donors, and these were multiplied when necessary in a greenhouse under self-reproduction. The list of the accessions and their passport information and donors are given in Table S1, and the sampling sites are indicated in Figure S1.

These accessions encompass the wide geographical distribution of *P. vulgaris* in America. Membership to either one of the two gene pools was determined according to the passport data and based on previous molecular diversity studies (Angioi *et al.*, 2009; Rossi *et al.*, 2009; Nanni *et al.*, 2011; Bitocchi *et al.*, 2012; Desiderio *et al.*, 2013; Bitocchi *et al.*, 2013).

SNP selection and genotyping

The investigated SNPs were from Cortés *et al.* (2011) and Goretti *et al.* (2014). They were mainly from gene regions that are putatively involved in adaptation to both biotic and abiotic stress. Considering the complex population structure of *P. vulgaris*, the SNP set was developed to include both wild and domesticated individuals from the Mesoamerican, Andean and PhI gene pools, to limit possible ascertainment bias (Clark *et al.*, 2005; Goretti *et al.*, 2014).

The list of the loci, their putative functions, and the SNP codes is given in Table S2. Overall, 100 genes were analysed, with 148 SNPs identified with KASPar© genotyping. Based on the alignment of the sequences to the *P. vulgaris* genome, each SNP was also flagged as coding/ non-coding and synonymous/ non-synonymous.

The genomic DNA of each plant was extracted from young leaves (Doyle & Doyle, 1987). Genotyping was performed using KBioscience (Hoddesdon, UK, http://www.lgcgenomics.com/genotyping/).

Data analysis

Diversity statistics

The descriptive diversity statistics, which included the number of polymorphic markers, the mean number of alleles (N_a), the mean effective number of alleles (N_e), and the unbiased expected heterozygosity (H_e; Nei, 1978), were calculated using PopGene 1.32 (Yeh *et al.*, 1997).

To compare the levels of diversity of the wild and domesticated beans, we estimated the relative loss of gene diversity (ΔH) for both the Mesoamerican and Andean gene pools. We used the *ad-hoc* statistic $\Delta H = 1 - (H_d/H_w)$, where H_d and H_w are the genetic diversity in the domesticated and wild accessions, respectively (Vigouroux *et al.*, 2002).

157

158 ***Population structure analysis***

159 To investigate the population structure, we used STRUCTURE 2.3.4 (Pritchard *et al.*, 2000),
160 which assigns each individual to different groups according to a membership coefficient (q_i).
161 The admixture model was run using the options 'correlated allele frequencies among
162 populations' and 'infer the degree of admixture (α) by the data'. For each K (number of
163 hypothetical populations), 20 runs (burn-in length, 100,000; iterations, 200,000) were
164 carried out, and the most likely number of K was determined using the ΔK statistic (Evanno
165 *et al.,* 2005), as implemented in STRUCTURE Harvester (Earl & vonHoldt, 2011). The genetic
166 structure obtained was then compared with the results from a neighbour-joining tree based
167 on the pairwise differences between individuals and using 10^3 bootstrap replications (MEGA
168 5.2; Tamura *et al.*, 2011) and those from principal component analysis (PCA) (EIGENSOFT
169 6.0.1; Patterson *et al.*, 2006; Price *et al.*, 2006). The genetic distances among the genetic
170 groups were determined using the F_{ST} statistics (Wright, 1951), and their significance was
171 tested using 10^5 permutations (Arlequin 3.5.1.2; Excoffier & Lischer, 2010).

172 The genetic structure obtained with the nuclear SNP data was compared with that
173 previously obtained for chloroplast simple sequence repeats (cpSSRs). This was possible for
174 83 accessions that were shared between the present study and that of Desiderio *et al.*
175 (2013). The associations between the genetic groups obtained and the different marker
176 systems were calculated using JMP 7.0 (SAS Institute Inc, 2007).

177

178 ***Variations among groups for seed weight***

179 Seed weights were also available for 457 accessions (http://isa.ciat.cgiar.org). Associations
180 between genetic groups and seed weight were therefore investigated by ANOVA, using JMP
181 7.0.

182

183 ***Geographical distribution of SNP variation***

184 The associations between the geographical (km) and genetic distances among the different
185 accessions were determined according to the Mantel statistic, using GenAlex 6.5 (Peakall &
186 Smouse, 2012), and tested by permutations (10^3 replicates). The Mantel test was performed
187 for the entire sample and for the Mesoamerican and Andean areas separately.

To further investigate the spatial patterns of genetic variability, we used multivariate analysis to detect global and local structuring (Jombart *et al.*, 2008), which was implemented in the adegenet R package (http://www.r-project.org/). The test statistic used in both procedures is the maximum of t values, denoted max(t). When global and local patterns are present, the observed max(t) is higher than the simulated values. When global structures are present, proximal individuals are more genetically similar than non-neighbour spatial groups; i.e., more than expected from a random distribution. When local structures are present, proximal individuals are more genetically dissimilar than non-neighbour spatial groups. The significance of max(t) was determined using the Monte Carlo procedure. When significant global or local structures were detected, the SGS software version 1.0d (Degen *et al.*, 2001) was used to design the autocorrelogram, by plotting the Moran index (I) against the geographical distance classes. The Moran index can have negative (or positive) values that indicate negative (or positive) spatial autocorrelation. These range from −1 (perfect dispersion) to +1 (perfect correlation). A zero value indicates a random spatial pattern. We set 10 distance classes at nearly 450 km each, to guarantee at least 1,000 pairwise comparisons in each class. The significances of the I values were assessed by randomly permuting the multilocus genotypes over the spatial coordinates of the samplings (500 times).

To visualise the spatial distributions of the genetic groups identified by Structure, we used the kriging method implemented in R (http://membres-timc.imag.fr/Olivier.Francois/plot.membership.r) that spatially interpolates the membership coefficients (q_i).

Finally, spatial analysis was accomplished using an individual-centred approach (Manel *et al.*, 2007). For each of the 310 geo-referenced individuals, we defined a circular neighbourhood of 100-km radius and used the individuals included in each circular neighbourhood to calculate the unbiased gene diversity, H_e (Nei, 1978). The mean size of each neighbourhood was 40.6 individuals; 83.3% of the neighbourhoods included more than 10 individuals. Moreover, the correlation between H_e and neighbourhood size was not significant ($r = 0.040$, $n = 299$, $P = 0.482$). We interpolated the neighbourhood diversity data by applying the kriging method, and the maps were designed using the map tools implemented in different R packages, such as ‘maps’, ‘maptools’, ‘rworldmap’ (http://cran.r-project.org/).

Disentangling the geographical and ecological effects on genetic structure

Associations between the genetic structure and geographical and ecological data were also investigated. Using DIVA-GIS 7.5 (http://www.diva-gis.org/), we extracted the ecological data for each of the 310 geo-referenced accessions from free access databases (Scheldeman & van Zonneveld, 2010).

The extracted ecological data were 3-monthly variables (minimum and maximum temperatures, and precipitation) for a total of 36 variables, and 19 bioclimatic variables (Table S3). We performed PCA on the 55 ecological variables, using JMP 7.0. We then studied the relationships between the genetic structure and the ecological PCAs (ePCAs).

To disentangle the potential roles of these latter factors on the genetic differentiation, we first used the partial Mantel test implemented in Arlequin 3.5.1.2 to calculate the partial correlations between genetic *versus* geographical and ecological distance matrices (Smouse *et al.*, 1986). Pairwise accession distance matrices were obtained using GenAlex with the SNP data, or the geographical coordinates, or the ePCA eigenvalues. As several studies have indicated that the partial Mantel test can be flawed in cases where the data are autocorrelated, we also used the method proposed by Guillot *et al.* (2014), implemented in R and kindly provided by these authors. This method is based on an explicit spatial model, known as a spatial generalised linear mixed model (SGLMM), and it allows quantification of the correlations between genotypes and environmental variables. It best suits datasets at a continental scale, with large enough genetic variation and with spatial autocorrelation, as in the present case.

Results

Genetic diversity in *P. vulgaris*

The SNP frequency spectra obtained for all of the gene pools investigated indicated overall that ascertainment bias did not significantly affect our analysis (Fig. S2). Among the 148 SNPs used for the genotyping, seven were monomorphic, eight showed >5% missing data, and two showed >44% heterozygosity (Table S2). Therefore we used 131 SNPs to perform the analyses.

By genetic diversity analysis (Table 1), we detected higher variability of the Mesoamerican gene pool (H_e = 0.284) compared to the Andean gene pool (H_e = 0.126). Based on Wilcoxon non-parametric tests ($P < 10^{-2}$), both the wild and domesticated forms of the Mesoamerican gene pool – i.e. MW and MD, respectively – show significantly higher gene diversity (H_e = 0.260, H_e = 0.157, respectively) than the wild and domesticated forms of the Andean gene pool – i.e., AW and AD, respectively – (H_e = 0.120, H_e = 0.089, respectively). Moreover, the diversity loss between the wild and domesticated forms is higher in the Mesoamerican gene pool (ΔH = 0.396) than in the Andean gene pool (ΔH = 0.261). The loss of alleles (ΔN_a, ΔN_e) follows the same trend, although less clear-cut differences are observed. The PhI accessions show the lowest genetic diversity (H_e = 0.074).

Genetic structure in *P. vulgaris*

Structure analysis of the 577 accessions of *P. vulgaris* indicates K = 2 as the uppermost hierarchical level of the genetic structure, while there are secondary peaks at K = 3 and K = 6 (Fig. S3a). The first partition at K = 2 splits the Mesoamerican and Andean gene pools, with the PhI accessions in an intermediate position (Fig. 1a). At K = 3, the MW and MD accessions are separated (Fig. 1a). At K = 6 the Mesoamerican gene pool is additionally subdivided into four genetic groups, and a net differentiation of the PhI group from the Andean and the Mesoamerican gene pools is seen. No subdivisions are observed within the Andean gene pool (Fig. 1a).

To further investigate the substructures, we performed separate analyses for the Mesoamerican and Andean accessions. On the basis of the Evanno method (Fig. S3b), the results at K = 4 are shown in Figure 1b. The four genetic groups are: MW1, MW3 and MW4, which contain the MW accessions, and M2, which is mainly constituted by the MD accessions. According to this subdivision, 98 accessions (22.2%) are admixed ($q_i < 0.70$). The MW1 group is mainly constituted by wild accessions from outside Mexico (i.e., Honduras, Guatemala, Costa Rica, Colombia, El Salvador) and from Chiapas (Mexico). The MW3 group is mainly constituted by Mexican accessions from Jalisco and Colima, while the MW4 group is constituted mainly by accessions from Morelos. The M2 genetic group is constituted by four wild accessions (hereafter indicated as MW2) and four weedy and 90 domesticated accessions (hereafter indicated as MD2).

The main subdivision of the Andean gene pool is observed at K = 2 (Fig. S3c) for which the AW and AD forms are neatly distinguished. At K = 4, both the wild and domesticated groups are further divided into sub-groups, AW1 and AW2 respectively (Fig. 1b). AW1 is constituted by accessions mainly from Argentina and Bolivia, and AW2 is constituted by accessions mainly from Peru. AD1 contains more accessions (46) than AD2 (3). A total of 22 individuals (17.2%) are admixed ($q_i < 0.7$).

The genetic diversity of MW1 and MW3 are higher (H_e = 0.205, 0.254, respectively) than for M2 and MW4 (H_e = 0.165, 0.148, respectively) based on Wilcoxon non-parametric tests ($P < 10^{-2}$). Among the AW groups, AW2 has higher diversity (He = 0.103) compared to AW1 (He = 0.059). AW1 has H_e values similar to AD1 (0.039). The AD2 group shows the highest diversity (He = 0.260), despite this estimate only being based on three accessions.

Within the Mesoamerican gene pool, similar F_{ST} distances emerge among the MW1, MW3 and MW4 groups, which vary between 0.227 (MW1-MW3) and 0.361 (MW1-MW4) (Table S4). Among the four MW groups, MW3 is the closest (F_{ST} = 0.383) to the M2 group (mostly domesticated genotypes), with an F_{ST} between MW1 and M2 of 0.468, and between MW4 and M2 of 0.532.

Within the Andean gene pool, AW1 is the wild group nearest to the AD1 group (F_{ST} = 0.458), which contains most of the domesticated accessions, while the AW2 group is the farthest (F_{ST} = 0.479 from AD1).

Genetic diversity heat map for wild *P. vulgaris*

Figure 2 shows the topography of the genetic variation of the MW *P. vulgaris*, as obtained using the individual-centred approach. High levels of diversity are observed across Mexico starting from the state of Oaxaca to Durango with a notably depression of genetic diversity in central Mexico, in the regions of Guerrero, Morelos, Puebla and Estado de Mexico. Low diversity is also observed in Guatemala, Costa Rica and Colombia, and particularly in the Honduras.

In the Andes, a major diversity hotspot is located on the central-northern coast of Peru, while the remaining areas (i.e., Argentina, Brazil, Bolivia) show lower diversity levels (Fig. S4).

Chloroplast and nuclear structure comparisons

We found a significant association (R^2 = 0.33, χ^2 = 79.6, P < 10^{-3}) between the groups obtained using cpSSRs (C1, C2, C3; Desiderio *et al.*, 2013) and the groups detected in this study using nuclear SNP markers (Fig. S5).

The chloroplast C1 group is mainly associated with the Andean gene pool, while the C2 and C3 groups are mainly associated with the Mesoamerican gene pool. In particular, C2 is mainly associated with genotypes from MW2 and MW1. The C3 group, which was suggested to be representative of an ancestral chloroplast genome, includes genotypes from all of the genetic groups, except for the AW1 group, with prevalence of the MW3 group.

Associations among genetic groups and seed weight

The Mesoamerican accessions show lower mean 100-seed weights (6.9 g, 27.7 g, for the MW and MD forms, respectively) than the Andean accessions (10.7 g, 46.5 g. for the AW and AD forms, respectively) (P < 10^{-3}). Within gene pools, the domesticated accessions show significantly higher 100-seed weights than the weedy and wild accessions (P < 10^{-3}; Fig. 3).

The 100-seed weights of the wild genetic groups were also significantly different (P < 10^{-3}; Fig. 3). In Mesoamerica, the highest 100-seed weight is seen for the MW2 group (9.8 g), and the lowest for the MW4 group (4.8 g), with MW1 in an intermediate position (7.7 g). MW3 shows a 100-seed weight (5.6 g) that is not significantly different from MW4 (P < 0.05).

In the Andes, the AW2 group shows significantly higher 100-seed weight (12.1 g) than the AW1 group (9.7 g; P < 10^{-3}).

Relationships among individuals

The neighbour-joining analysis highlights the distinction between the Mesoamerican and Andean gene pools, with the PhI pool in between (Fig. S6a). The Mesoamerican genotypes are separated into four clusters that correspond to the MW1-MW4 groups identified by the Structure analysis (Fig. S6b). The Andean accessions are separated into three main clusters (Fig. S4c), which also correspond to the AW1, AW2 and AD groups identified by Structure.

The PCA plot confirms major subdivision between the Mesoamerican and Andean gene pools captured by PC1 (Figure S7). The MD accessions separate from MW mainly along PC2, where the closer relationship between the MW1 and M2 groups is also confirmed. When PC3 is considered, the MW3 group is better separated than the other MW groups.

Landscape genetics approach

Spatial structure of the genetic variation

The Mantel test performed considering the Mesoamerican and Adean wild accessions shows significant and positive correlation between the genetic and geographical distances ($r = 0.69$, $P < 10^{-3}$). This was confirmed when the Mesoamerican ($r = 0.27$, $P < 10^{-3}$) and Andean ($r = 0.55$, $P < 10^{-3}$) gene pools were analysed separately. Additionally, the max(t) test shows that both overall (Fig. S8a) and for Mesoamerica (Fig. 4a), proximal individuals are more genetically similar than distant individuals ($P < 10^{-4}$). However, when considering only the Andean data, the test was marginally non-significant ($P = 0.06$; Fig. S8b).

Autocorrelograms showed that when considering classes of increasing geographical distances, the Moran's I decreases, passing from positive to negative values with 11 and nine I values that reach significance ($P < 0.05$) when all or only the Mesoamerican data are considered, respectively (Figs. 4b, S8c). Consistent with the other tests, for the Andean gene pool, there are significant I values ($P < 0.05$) only for the first three and last two distance classes (Fig. S8d).

Figure 5 shows that the wild genetic groups obtained from the structure analysis are essentially subdivided according to their geographicl origin. In detail, MW1 is mainly distributed from Colombia to Chiapas (Mexico); MW2 is widely distributed from Guanajuato (Mexico) to Costa Rica; MW3 is mainly located across the regions of Durango, Jalisco and Guerrero; and MW4 is prevalently located across the Morelos and Puebla regions. The PhI group is localised in Ecuador-northern Peru, while AW1 and AW2 are localised in Peru and Argentina, respectively.

Associations between genetic groups and ecological variables

To study the associations between the genetic groups and ecological variables, we concentrated on the Mesoamerican gene pool, as its large sample size allows greater precision.

We detected strong correlation structure among the 55 climatic variables, as five ecological principal components (ePCAs) capture 95% of the total variance, and the first two ePCAs reach 77.4% (49.7%, 27.7%, for ePCA1 and ePCA2, respectively). The remaining three ePCAs explain 9.3% (ePCA3), 5.8% (ePCA4), and 2.5% (ePCA5) of the total variance. ePCA1 is

positively correlated with 24 variables (adopting a threshold of $r > 0.8$ and $P < 10^{-4}$), which are prevalently represented by the maximum and mean temperatures, in particular during the wettest and warmest quarter of the year (Table S5). ePCA2 correlates with 10 variables ($r > 0.8$, $P < 10^{-4}$), which include annual precipitation and minimum temperatures of the coldest months (Table S5).

When the relationships among accessions were studied as a function of the first two ePCAs, accessions belonging to MW1 tended to separate from the others along ePCA2, with the other accessions intermixed (Fig. 6). On average, individuals from the MW3 group show the lowest ePCA2, followed by individuals from MW4 and from MW1, which defines a north (MW3-MW4)-to-south (MW1) pattern of variation. This also indicates that the three genetic groups of wild bean might be adapted to different ranges of ecological conditions, with MW1 covering the widest range. Individuals in the MW2 group, which is mainly weedy accessions, are also scattered. The associations between genetic distances and ePCA2 absolute differences among individuals are confirmed by the Mantel test ($r = 0.242$, $P < 10^{-2}$). No significant associations emerged with ePCA1.

Disentangling the effects of geography from ecology in shaping genetic patterns

Partial Mantel tests show that the geographical distances and ePCA1 cumulatively explain 5.4% of the SNP genetic variance. Partial correlation is significant with geography ($R^2 = 0.057$, $P < 10^{-2}$), but not with ecology ($R^2 = 0.000$, n.s.). A further 10.8% of the SNP genetic variance is cumulatively explained by geography and ePCA2. In this case, the effect of the ecology on genetic distances is almost three-fold higher than that of geography ($R^2 = 0.082$, $P < 10^{-2}$ and $R^2 = 0.026$, $P < 10^{-2}$, respectively).

The search for non-neutral correlations between single marker loci and ecological variables was performed using the eigenvalues of the first five ePCAs. While no loci are associated with ePCA1, six loci show significant associations with ePCA2 (Fig. 7). Moreover, seven loci are associated with ePCA3, seven with ePCA4, and nine with ePCA5 (Fig. S9). Overall, a total of 26 loci (19.8%) are found to be characterised by a signature of selection (log Bayes factor > 0), of which seven (5.4%) show very strong statistical support (log Bayes factor > 3) (Table S6).

We therefore then removed the SNP under selection and the non-synonymous to obtain a 'putatively neutral' dataset that was used to re-calculate the genetic diversity

411 statistics, re-map the diversity levels, and re-infer the population structure that might better
412 reflect only the demographic history of the common bean (see Supplementary Note for
413 details).

414 The H_e levels observed with the neutral dataset were lower than those with the
415 complete dataset, with a stronger reduction in diversity due to domestication (ΔH) for both
416 the Mesoamerican and Andean gene pools (Supplementary Note). The 'neutral' genetic
417 structure overall confirmed that which was obtained with the complete dataset with a novel
418 outcome: the PhI gene pool is closer to the Mesoamerican than to the Andean gene pool
419 (Supplementary Note). The diversity heat-maps were re-designed and the locations of peaks
420 and valleys of diversity confirmed with a cleaner distinction between high and low diversity
421 areas (Supplementary Note).

422 The results from the s structure analysis with the 'neutral' dataset reveal five genetic
423 groups on the Mesoamerican sample: $MW1_N$, $MW2_N$, $MW3_N$, $MW4_N$ and $MW5_N$ (Fig. 8). The
424 $M2_N$ and $MW4_N$ groups correspond substantially to the M2 and MW4 groups, respectively
425 (Fig. 1, 8). The $MW1_N$ group includes Colombian genotypes from the MW1 group, and $MW5_N$
426 is mainly constituted by accessions from Guatemala, Honduras and Costa Rica. $MW3_N$
427 essentially corresponds to MW3, except for the now missing accessions from Oaxaca and
428 Chiapas. The fifth genetic group, $MW5_N$, includes accessions from Guatemala, Honduras and
429 Costa Rica, which were previously included in MW1, and accessions from Oaxaca and
430 Chiapas from MW3. The gene diversity of the five groups is higher for $MW5_N$ and $MW3_N$ (H_e
431 = 0.203, 0.177, respectively) than for $MW1_N$ (H_e = 0.063). $M2_N$ shows levels of diversity that
432 are similar to $MW4_N$ (H_e = 0.111, 0.103, respectively).

433 The neighbour-joining tree and PCA also show five groups (Fig. 9, and Supplementary
434 Note). The $MW5_N$ group is closer (F_{ST} = 0.620) to the $M2_N$ cluster, which mainly contains
435 domesticated accessions ($MD2_N$) and a few wild accessions ($MW2_N$), followed by the $MW3_N$
436 group (F_{ST} = 0.627). In particular, a $MW5_N$ sub-cluster that contains two genotypes from
437 Durango and four from Oaxaca is the closest to the $M2_N$ group (Fig. 9).

438 The genetic groups obtained using the putatively 'neutral' dataset were also
439 compared with the chloroplast groups found by Desiderio *et al.* (2013), and a significant
440 association (R^2 = 0.20, $P < 10^{-2}$) was again observed (Fig. S10). In particular, most of the
441 genotypes from the $MW5_N$ group are attributed to the C2 chloroplast group, and a small

fraction is associated with the C3 ancestral plastidial type. The $MW3_N$ group is associated with the C3 ancestral plastidial type.

The five genetic groups showed significantly different mean 100-seed weights (one-way ANOVA, $P < 10^{-3}$; Fig. 10). The $MW2_N$ and $MW1_N$ groups show the highest 100-seed weights (9.4 g, 8.7 g, respectively), and $MW3_N$ and $MW4_N$ show the lowest 100-seed weights (4.9 g, 4.5 g, respectively). $MW5_N$ has an intermediate value (6.3 g), which is statistically not different from $MW1_N$. However, the four accessions from Oaxaca, which are the closest to the domesticated genotypes, show a relatively low mean 100-seed weight (4.3 g). When spatial autocorrelation analysis is performed using the 'neutral' dataset, we still observe a negative correlation between I values and geographical distances (Fig. S11a). We observe the same pattern also when we use only the 26 loci under selection (Fig. S11b).

Discussion

In the present study, this analysis of a very large collection has allowed us to gain insights into the structure and distribution of the genetic diversity of the wild common bean in Mesoamerica at an unprecedented high resolution.

Structure of the *P. vulgaris* genetic diversity

The MW gene pool of *P. vulgaris* is divided into four genetic groups that show well-defined geographical distribution except for the MW2 group, which shows a more scattered distribution. This group is also the closest to the domesticated genotypes (MD2), which might be explained by introgression from the domesticated gene pool (Papa & Gepts, 2003).

The genetic distances among the groups detected in the present study are on average higher than in previous studies, especially when compared to microsatellite data (Kwak & Gepts, 2009). This might be because the different markers have different mutation rates, as also for the sampling of individuals and loci. Nonetheless, the relationships depicted among the genetic groups are in line with those from previous studies (e.g. Kwak & Gepts, 2009; Bitocchi *et al.*, 2012, 2013; Desiderio *et al.*, 2013; Schmutz *et al.*, 2014).

The geographical distribution of the SNP genetic groups is largely in agreement with that observed by analysis of non-recombining sequences (Bitocchi *et al.*, 2012), except for

MW1, which is here restricted to central America and Colombia, but is more widespread based on sequence data. Such a difference might be due to recombination between unlinked SNPs that followed the ancient migration from Mexico.

Regarding the Andes, the population structure and the genetic diversity of the wild bean are very low compared to those observed in Mesoamerica, which is most likely the consequence of the Mesoamerican origin of the wild beans (Bitocchi *et al.*, 2012). The genetic diversity is further reduced in the domesticated forms as a consequence of the sequential bottleneck that this gene pool underwent, as noted by Bitocchi *et al.* (2013). Nonetheless, it is worth noting the presence of two well-defined groups in the Andean gene pool, AW1 and AW2, that were also geographically based.

Distribution of the genetic diversity of wild common bean

The diversity map reveals high levels of diversity all across Mexico, from the state of Oaxaca to the Guanajuato and Durango regions (Fig. 2). The high levels of diversity of these areas are also usually characterised by high chloroplast diversity and the occurrence of the ancestral plastidial types (Chacón *et al.*, 2007; Desiderio *et al.*, 2013), which reinforces the hypothesis that Mesoamerica represents the cradle of diversity of *P. vulgaris* (Bitocchi *et al.*, 2013).

A main striking exception is however observed in the area that appears as a diversity 'desert' in Figure 2. This area is located across Guerrero, Morelos, Puebla and Estado de Mexico, where a well-defined genetic group, MW4, is located. Several hypotheses can be made to explain such an observation. First, it can be hypothesised that selection for local adaptation occurred in this area, which is characterised by a very dry climate (http://koeppen-geiger.vu-wien.ac.at/shifts.htm). However such 'selection hypothesis' is hampered by the observation that this diversity 'desert' is more accentuated when only putatively neutral SNPs are used. Secondly, this area was subjected to agricultural intensification that started with the Formative period (1500 BC to 100 AD) (Siebe, 2000; Plunket & Uruñuela, 2012), which might have caused the genetic assimilation of the wild population of this area (Papa & Gepts, 2003). However, the genetic data does not appear to be supportive of this hypothesis that would imply a similar genetic background to that of the domesticated gene pool. Finally, we note that in this area, there is the volcanic front of the Trans-Mexican Volcanic Belt. Within this front, evidence of numerous volcanic events of

varying intensities has been reported for Sierra de Chichinautzin and the region surrounding Popocatépetl Volcano (Plunket & Uruñuela, 1998; Márquez *et al.*, 1999; Siebe *et al.*, 2004). All this would suggest that the low genetic diversity of the population from this area (MW4) is due to selection by a genetic bottleneck caused by the volcanic activities, while being independent of the origin or spread of agriculture. However, it is important to consider that these explanations are not mutually exclusive.

In the Andes, the wild genetic group that shows the highest diversity (AW2) and is located in the centre of Peru, was also associated with the occurrence of all of the Andean plastidial types, including the ancestral C3 (Fig. 4 from Desiderio *et al.*, 2013) (Fig. S4, Fig. S5). This thus indicates that this area contains a wealth of genetic diversity for the Andean common bean.

Landscape genetics

Spatial analysis of genetic variations in Mesoamerica revealed that there are global structures for both the putatively 'neutral' and 'non-neutral' datasets; i.e., genetic distances between individuals are significantly correlated with geographical distances. This pattern also indicates that migration and drift effects are superimposed on a selection effect in the same direction. This means that the existence of well-defined wild genetic groups is the result of limited long-range gene flow, together with divergent selection due to local adaptation. This is also supported by the association between genetic and ecological data and by the scan for signatures of selection, which show 26 loci (19.8%) with selection signatures, where seven (5.3%) show very strong probability levels (log Bayes factor > 3). However, the proportion of loci under selection might be overestimated, as our data are relative to a panel of sequences this was enriched for genes that are *a-priori* putatively involved in adaptation.

Nonetheless, some of the genes under selection are involved in responses to environmental stress (Kavar *et al.*, 2008; Mao *et al.*, 2010; Rapala-Kozik *et al.*, 2012; Krause *et al.*, 2013), as cold acclimation or chilling susceptibility (Liu *et al.*, 2007; Alcázar *et al.*, 2011; Zhang *et al.*, 2011), or in the adaptation to different conditions of light and temperature, and to drought stress responses (Green *et al.*, 1991; Bocobza *et al.*, 2013). Four of these loci (Table S2) are also in common with those under selection during domestication (Schmutz *et al.*, 2014). This might either suggest that these loci are subject to selection or that they are

marking regions under selection. Indeed, considering the level of inbreeding of *P. vulgaris*, hitchhiking might also have a role here. However, it can be noted that very low levels of linkage disequilibrium (pairwise linkage disequilibrium: 3.4%, average r^2 = 0.04) were previously detected within the Mesoamerican wild gene pool (Rossi *et al.*, 2009).

All this indicates that for the first time in bean, the study of correlations between markers and ecological variables at a continental scale can help in the identification of genes that are involved in local adaptation, as has also been shown for other plants and for animals (Hancock *et al.*, 2011a, 2011b).

This is relevant for both evolutionary genetics, which addresses the relative importance of neutral *versus* adaptive processes, and for strengthening the scientific basis for germplasm conservation and its use in plant breeding.

Domestication sites of common bean

To unravel the role of the Mesoamerican and Andean areas characterised by different genetic diversity patterns for common bean domestication, we compared the genetic evidence with phenotypic and ecological data, and we discuss here these results with the aid of previous archaeological and glottochronological studies. A similar approach was used, for example, to study the origin and dispersal of domesticated rice (Kovach *et al.*, 2007) and to determine the origin of the domesticated chilli pepper (Kraft *et al.*, 2014).

For Mesoamerica, our data and their comparison with additional evidence from archaeology and linguistic information (Kaplan & Lynch, 1999; Brown *et al.*, 2014), indicate that the Oaxaca Valley is the region where domestication of the common bean took place. In support of this, the lowest genetic distance from the domesticated form is observed for the $MW5_N$ group, followed by $MW3_N$ and $MW4_N$ (Fig. 9). The $MW5_N$ group is mainly constituted by individuals from the south of Mexico and from Central America, and it is characterised by the highest gene diversity. Within this group some accessions from Oaxaca are the closest to the domesticated accessions (Fig. 9). The low 100-seed weight of these accessions also indicates that it is unlikely that they derived from hybridisation with domesticated types. Our data are thus also in agreement with Bitocchi *et al.* (2013).

The presence within the Oaxaca area of archaeological sites with common bean macro-remains from 2100-2300 cal BP (Kaplan & Lynch, 1999) indicates the early occurrence of domestication in this area. At the same time, glottochronological studies have shown that

570 this includes the homeland sites of the Zapotecan, Mixtec-Cuicatec, and Popolocan proto-
571 languages, for which ancient bean words can be reconstructed from 3149 to 3036 years BP.
572 Even though a gap exists between the palaeo-biolinguistic reconstructed data and the
573 estimated onset of domestication, the relevance of this species for the speakers of this
574 language has been shown (Brown *et al.*, 2014). All these data together support the Oaxaca
575 Valley as the domestication area for common bean.

576 The southern Lerma-Santiago basin has been previously suggested as a putative
577 domestication site for common bean (Kwak *et al.*, 2009). This region corresponds to the
578 distribution area of the $MW3_N$ genetic group, which is the second wild genetic group to be
579 closer to the domesticated form and which also shows a low mean 100-seed weight.
580 However, in contrast to the Oaxaca region, this area does not have archaeological sites with
581 bean remains. In this regard, glottochronological data have been recently found (Brown *et*
582 *al.*, 2014) that have suggested that the oldest word for beans is included in the Otopamean
583 proto-language, which was spoken around 3,600 years BP in a region that coincided with the
584 easternmost area of the domestication site suggested by Kwak *et al.* (2009) (Figure 11).
585 Thus, considering the available information, the Oaxaca Valley is the most likely origin of
586 common bean domestication in Mesoamerica, although further genetic and
587 archaeobotanical research is needed to shed light on the origin of domestication in
588 Mesoamerica.

589 In the Andes, our data show that the wild accessions from Argentina-Bolivia (AW1)
590 are genetically more similar to the Andean domesticated forms (Fig. S6c). These accessions
591 also show a lower 100-seed weight when compared to the AW2 accessions. These data point
592 towards the region from northern Argentina and southern Bolivia as the one associated with
593 the Andean domestication process (Fig. S4), and they are consistent with the data from
594 previous genetic (Beebe *et al.*, 2001; Bitocchi *et al.*, 2013), archaeological (Tarrago, 1980),
595 and glottochronological (Brown *et al.*, 2014) studies.
596
597

598 **References**

599

600 **Alcázar R, Cuevas JC, Planas J, Zarza X, Bortolotti C, Carrasco P, Salinas J, Tiburcio AF,**
601 **Altabella T. 2011.** Integration of polyamines in the cold acclimation response. *Plant*
602 *Science* **180**: 31-38.
603 **Angioi SA, Desiderio F, Rau D, Bitocchi E, Attene G, Papa R. 2009.** Development and use of
604 chloroplast microsatellites in *Phaseolus* spp. and other legumes. *Plant Biology* **11**: 598-
605 612.
606 **Beebe S, Rengifo J, Gaitan E, Duque MC, Tohme J. 2001.** Diversity and origin of Andean
607 landraces of common bean. *Crop Science* **41**: 854–862.
608 **Bellucci E, Bitocchi E, Rau D, Rodriguez M, Biagetti E, Giardini A, Attene G, Nanni L, Papa R**
609 **2014a.** Genomics of origin, domestication and evolution of *Phaseolus vulgaris*. In: R.
610 Tuberosa, A. Graner, E. Frison eds. *Genomics of Plant Genetic Resources*: Springer
611 Netherlands, 483-507.
612 **Bellucci E, Bitocchi E, Ferrarini A, Benazzo A, Biagetti E, Klie S, Minio A, Rau D, Rodriguez M,**
613 **Panziera A *et al.* 2014b.** Decreased nucleotide and expression diversity and modified
614 coexpression patterns characterize domestication in the common bean. *The Plant Cell*
615 *Online, tpc-114*.
616 **Bitocchi E, Bellucci E, Giardini A, Rau D, Rodriguez M, Biagetti E, Santilocchi R, Spagnoletti**
617 **Zeuli P, Gioia T *et al.* 2013.** Molecular analysis of the parallel domestication of the
618 common bean (*Phaseolus vulgaris*) in Mesoamerica and the Andes. *New Phytologist*
619 **197**: 300-313.
620 **Bitocchi E, Nanni L, Bellucci E, Rossi M, Giardini A, Zeuli PS, Logozzo G, Stougaard J,**
621 **McClean P, Attene G, Papa R. 2012.** Mesoamerican origin of the common bean
622 (*Phaseolus vulgaris* L.) is revealed by sequence data. *Proceedings of the National*
623 *Academy of Science, USAs* **109**: E788-E796.
624 **Bocobza SE, Malitsky S, Araújo WL, Nunes-Nesi A, Meir S, Shapira M, Fernie AR, Aharoni A.**
625 **2013.** Orchestration of thiamin biosynthesis and central metabolism by combined action
626 of the thiamin pyrophosphate riboswitch and the circadian clock in *arabidopsis*. *Plant*
627 *Cell* **25**: 288-307.
628 **Brown CH, Clement CR, Epps P, Luedeling E, Wichmann S. 2014.** The paleobiolinguistics of
629 the common bean (*Phaseolus vulgaris* L.). *Ethnobiology Letters* **5:** 104-115.

Chacón SMI, Pickersgill B, Debouck DG. 2005. Domestication patterns in common bean (*Phaseolus vulgaris* L.) and the origin of the Mesoamerican and Andean cultivated races. *Theoretical and Applied Genetics* **110**: 432-444.

Chacón SMI, Pickersgill B, Debouck DG, Arias JS. 2007. Phylogeographic analysis of the chloroplast DNA variation in wild common bean (*Phaseolus vulgaris* L.) in the Americas. *Plant Systematics and Evolution* **266**: 175-195.

Clark AG, Hubisz MJ, Bustamante CD, Williamson SH, Nielsen R. 2005. Ascertainment bias in studies of human genomewide polymorphism. *Genome Research* **15**:1496–1502.

Cortés AJ, Chavarro MC, Blair MW. 2011. SNP marker diversity in common bean (*Phaseolus vulgaris* L.). *Theoretical and Applied Genetics* **123**: 827-845.

Degen B, Petit R, Kremer A. 2001. SGS-spatial genetic software: a computer program for analysis of spatial genetic and phenotypic structures of individuals and populations. *Journal of Heredity* **92**: 447-448.

Desiderio F, Bitocchi E, Bellucci E, Rau D, Rodriguez M, Attene G, Papa R, Nanni L. 2013. Chloroplast microsatellite diversity in *Phaseolus vulgaris*. *Frontiers in plant science, 3*.

Diamond J, Bellwood P. 2003. Farmers and their languages: the first expansions. *Science* **300**: 597-603

Doyle JJ, Doyle JL. 1987. A rapid DNA isolation procedure for small quantities of fresh leaf tissue. *Phytochemical Bulletin* **19**: 11-15.

Earl DA, vonHoldt BM. 2011. Structure harvester: a website and program for visualizing structure output and implementing the Evanno method. *Conservation Genetics Resources* **4**: 359-361.

Eckert AJ, van Heerwaarden J, Wegrzyn JL, Nelson CD, Ross-Ibarra J, González-Martínez SC, Neale DB. 2010. Patterns of population structure and environmental associations to aridity across the range of loblolly pine (*Pinus taeda* L., *pinaceae*). *Genetics* **185**: 969-982.

Evanno G, Regnaut S, Goudet J. 2005. Detecting the number of clusters of individuals using the software STRUCTURE: a simulation study. Molecular Ecology **14**: 2611-2620.

Excoffier L, Lischer HEL. 2010. Arlequin suite version 3.5: a new series of programs to perform population genetics analyses under Linux and Windows. *Molecular Ecology Resources* **10**: 564-567.

661 **Gepts P, Osborn TC, Rashka K, Bliss FA. 1986.** Phaseolin-protein variability in wild forms and
662 landraces of the common bean (*Phaseolus vulgaris*): evidence for multiple centers of
663 domestication. *Economic Botany* **40**: 451-468.
664 **Goretti D, Bitocchi E, Bellucci E, Rodriguez M, Rau D, Gioia T, Attene G, McClean P, Nanni L,**
665 **Papa R. 2014.** Development of single nucleotide polymorphisms in *Phaseolus vulgaris*
666 and related *Phaseolus* spp. *Molecular Breeding* **33**: 531-544.
667 **Green BR, Pichersky E, Kloppstech K. 1991.** Chlorophyll *a/b*-binding proteins: an extended
668 family. *Trends in biochemical sciences* **16**: 181-186.
669 **Guillot G, Vitalis R, Rouzic Al, Gautier M. 2014.** Detecting correlations between allele
670 frequencies and environmental variables as a signature of selection. A fast
671 computational approach for genome-wide studies. *Spatial Statistics* **8**: 145-155.
672 **Hancock AM, Brachi B, Faure N, Horton MW, Jarymowycz LB, Sperone FG, Toomajian C,**
673 **Roux F, Bergelson J. 2011a.** Adaptation to climate across the *Arabidopsis thaliana*
674 genome. *Science* **334**: 83-86.
675 **Hancock AM, Witonsky DB, Alkorta-Aranburu G, Beall CM, Gebremedhin A, Sukernik R,**
676 **Utermann G, Pritchard JK, Coop G, Di Rienzo A. 2011b.** Adaptations to climate-
677 mediated selective pressures in humans. *PLoS Genetics* 7: e1001375.
678 **Jombart T, Devillard S, Dufour AB, Pontier D. 2008.** Revealing cryptic spatial patterns in
679 genetic variability by a new multivariate method. *Heredity* **101**: 92-103.
680 **Kami J, Velásquez VB, Debouck DG, Gepts P. 1995.** Identification of presumed ancestral
681 DNA sequences of phaseolin in *Phaseolus vulgaris*. *Proceedings of the National Academy*
682 *of Sciences, USA* **92**: 1101-1104.
683 **Kaplan L, Lynch T. 1999.** *Phaseolus* (fabaceae) in archaeology: AMS radiocarbon dates and
684 their significance for pre-Columbian agriculture. *Economic Botany* **53**: 261-272.
685 **Kavar T, Maras M, Kidrič M, Šuštar-Vozlič J, Meglič V. 2008.** Identification of genes
686 involved in the response of leaves of *Phaseolus vulgaris* to drought stress. *Molecular*
687 *Breeding* **21**: 159-172.
688 **Kovach MJ, Sweeney MT, McCouch SR. 2007.** New insights into the history of rice
689 domestication. *Trends in Genetics* **23**: 578-587.
690 **Kraft KH, Brown CH, Nabhan GP, Luedeling E, Ruiz JDJL, Coppens d'Eeckenbrugge G,**
691 **Hijmans RJ, Gepts P. 2014.** Multiple lines of evidence for the origin of domesticated chili

692 pepper, *Capsicum annuum*, in Mexico. *Proceedings of the National Academy of Sciences,*
693 *USA* **111**: 6165-6170.
694 **Krause C, Richter S, Knöll C, Jürgens G. 2013.** Plant secretome - from cellular process to
695 biological activity. *Biochimica et Biophysica Acta (BBA)-Proteins and Proteomics* **1834**:
696 2429-2441.
697 **Kwak M, Gepts P. 2009.** Structure of genetic diversity in the two major gene pools of
698 common bean (*Phaseolus vulgaris* L., fabaceae). *Theoretical and Applied Genetics* **118**:
699 979-992.
700 **Kwak M, Kami JA, Gepts P. 2009.** The putative Mesoamerican domestication center of
701 *Phaseolus vulgaris* is located in the Llerma-Santiago basin of Mexico. *Crop Science* **49**:
702 554-563.
703 **Liu J-H, Kitashiba H, Wang J, Ban Y, Moriguchi T. 2007.** Polyamines and their ability to
704 provide environmental stress tolerance to plants. *Plant Biotechnology* **24**: 117-126.
705 **Lynch J, González A, Tohme JM, García JA. 1992.** Variation in characters related to leaf
706 photosynthesis in wild bean populations. *Crop Science* **32**: 633-640.
707 **Manel S, Berthoud F, Bellemain E, Gaudeul M, Luikart G, Swenson JE, Waits LP, Taberlet P,**
708 **Consortium I. 2007.** A new individual-based spatial approach for identifying genetic
709 discontinuities in natural populations. *Molecular Ecology* **16**: 2031-2043.
710 **Mao X, Zhang H, Tian S, Chang X, Jing R. 2010.** Tasnrk2.4, an snf1-type serine/threonine
711 protein kinase of wheat (*Triticum aestivum* L.), confers enhanced multistress tolerance
712 in arabidopsis. *Journal of experimental botany* **61**: 683-696.
713 **Márquez A, Verma SP, Anguita F, Oyarzun R, Brandle JL. 1999.** Tectonics and volcanism of
714 Sierra Chichinautzin: extension to the front of the central Trans-Mexican Volcanic Belt.
715 *Journal of Volcanology and Geothermal Research* **93**: 125-150.
716 **Nanni L, Bitocchi E, Bellucci E, Rossi M, Rau D, Attene G, Gepts P, Papa R. 2011.** Nucleotide
717 diversity of a genomic sequence similar to shatterproof (pvshp1) in domesticated and
718 wild common bean (*Phaseolus vulgaris* L.). *Theoretical and Applied Genetics* **123**: 1341-
719 1357.
720 **Nei M. 1978.** Estimation of average heterozygosity and genetic distance from a small
721 number of individuals. *Genetics* **89**: 583-590.

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. *Theoretical and Applied Genetics* **106**: 239-250.

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. *Annals of Botany* **100**: 1039-1051.

Patterson N, Price AL, Reich D. 2006. Population structure and eigenanalysis. *PLoS genetics* 2.12: e190.

Peakall R, Smouse PE. 2012. Genalex 6.5: Genetic analysis in Excel. Population genetic software for teaching and research—an update. *Bioinformatics* **28**: 2537-2539.

Piperno DR, Flannery K. 2001. The earliest archeological maize (*Zea mays* L.) from highland Mexico: new accelerator mass spectrometry dates and their implications. *Proceedings of the National Academy of Science, USA* **98**: 2101-2103.

Piperno DR, Ranere AJ, Holst I, Iriarte J, Dickau R. 2009. Starch grain and phytolith evidence for early ninth millennium B.P. maize from the Central Balsas River Valley, Mexico. *Proceedings of the National Academy of Science, USA* **106**: 5019-5024.

Plunket P, Uruñuela G. 1998. Preclassic household patterns preserved under volcanic ash at Tetimpa, Puebla, Mexico. *Latin American Antiquity* **9**: 287-309.

Plunket P, Uruñuela G. 2012. Where East meets West: the formative in Mexico's central highlands. *Journal of Archaeological Research* **20**: 1-51.

Price AL, Patterson NJ, Plenge RM, Weinblatt ME, Shadick NA, Reich D 2006. Principal components analysis corrects for stratification in genome-wide association studies. *Nature genetics* **38(8)**: 904-909.

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

Rapala-Kozik M, Wolak N, Kujda M, Banas AK. 2012. The up-regulation of thiamine (vitamin b1) biosynthesis in *Arabidopsis thaliana* seedlings under salt and osmotic stress conditions is mediated by abscisic acid at the early stages of this stress response. *BMC plant biology* **12**: 2.

Rodriguez M, Rau D, Angioi SA, Bellucci E, Bitocchi E, Nanni L, Knüpffer H, Negri V, Papa R, Attene G. 2013. European *Phaseolus coccineus* L. Landraces: Population structure and adaptation, as revealed by cpSSRs and phenotypic analyses. *PLoS ONE* **8**: e57337.

Rossi M, Bitocchi E, Bellucci E, Nanni L, Rau D, Attene G, Papa R. 2009. Linkage disequilibrium and population structure in wild and domesticated populations of *Phaseolus vulgaris* L. *Evolutionary Applications* **2**: 504-522.

Scheldeman X, van Zonneveld M. 2010 *Training Manual on Spatial Analysis of Plant Diversity and Distribution.* Available online (accessed 6 October, 2011). *www.Bioversityinternational.Org/training/training_materials/gis_manual/gis_download.Html.* Rome: Bioversity International.

Siebe C, Rodriguez-Lara V, Schaaf P, Abrams M. 2004. Radiocarbon ages of Holocene Pelado, Guespalapa, and Chichinautzin scoria cones south of Mexico City: implications for archaeology and future hazards. *Bulletin of Volcanolgy* **66**: 203–225.

Siebe C. 2000. Age and archaeological implications of Xitle Volcano, southwestern basin of Mexico city. *Journal of Volcanology and Geothermal Research* **104**: 45-64.

Singh S. 1989. Patterns of variation in cultivated common bean (*Phaseolus vulgaris*, fabaceae). *Economic Botany* **43**: 39-57.

Smouse PE, Long JC, Sokal RR. 1986. Multiple regression and correlation extensions of the mantel test of matrix correspondence. *Systematic zoology* **35**: 627-632.

Schmutz JPE, McClean PE, Mamidi S, Wu GA, Cannon SB, Grimwood J, Jenkins J, Shu S, Song Q, Chavarro C *et al.* 2014. A reference genome for common bean and genome-wide analysis of dual domestications. *Nature genetics* **46(7):** 707-713.

Tamura K, Peterson D, Peterson N, Stecher G, Nei M, Kumar S. 2011. Mega5: Molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. *Molecular Biology and Evolution* **28**: 2731-2739.

Tarrago MN. 1980. El proceso de agriculturización en el Noroeste Argentino, zona Valliserrana. *Actas del V Congreso Nacional Arqueologı´a Argentina, vol. 1*. San Juan, Argentina: Instituto de Investigaciones Arqueológicas y Museo, Universidad de San Juan, 181–217.

van Heerwaarden J, Doebley J, Briggs WH, Glaubitz JC, Goodman MM, Gonzalez JDJS, Ross-Ibarra J. 2011. Genetic signals of origin, spread, and introgression in a large sample of maize landraces. *Proceedings of the National Academy of Sciences, USA* **108**: 1088-1092.

van Zonneveld M, Dawson I, Thomas E, Scheldeman X, Etten J, Loo J, Hormaza JI 2014. Application of molecular markers in spatial analysis to optimize *in-situ* conservation of

plant genetic resources. In: R. Tuberosa, A. Graner, E. Frison eds. *Genomics of Plant Genetic Resources*: Springer Netherlands, 67-91.

Vigouroux Y, McMullen M, Hittinger CT, Houchins K, Schulz L, Kresovich S, Matsuoka Y, Doebley J. 2002. Identifying genes of agronomic importance in maize by screening microsatellites for evidence of selection during domestication. *Proceedings of the National Academy of Sciences, USA* **99**: 9650-9655.

Wright S. 1951. The genetical structure of populations. *Annals of Eugenics* **15**: 323-353.

Yeh FC, Yang RC, Boyle TBJ. 1997. Popgen version 1.32: the user-friendly shareware for population genetic analysis. *Canada, University of Alberta. Molecular Biology and Biotechnology Center*.

Zhang Y, Wang C, Hu H, Yang L. 2011. Cloning and expression of three fatty-acid desaturase genes from cold-sensitive lima bean (*Phaseolus lunatus* L.). *Biotechnology letters* **33**: 395-401.

798 **Figure Legends:**

799

800 **Figure 1.** Results of the Structure analysis. (a) Results at K = 2, K = 3 and K = 6, based on 131
801 SNPs across all of the 577 *P. vulgaris* accessions. (b) Results of the analyses performed
802 separately for the Mesoamerican and Andean gene pools, at K = 4. The wild Mesoamerican
803 and Andean accessions are ordered according their country of origin, from the North to the
804 South. The Mexican regions are specified when they include more than three accessions.
805 MW, Mesoamerican wild; MD, Mesoamerican domesticated; PhI, accessions with phaseolin
806 type I; AW, Andean wild; AD, Andean domesticated; DU, Durango; NA, Nayarit; JA, Jalisco;
807 CO, Colima; MI, Michoacan; GN, Guanajuato; GR, Guerrero; MO, Morelos; PU, Puebla; OA,
808 Oaxaca; CH, Chiapas; GU, Guatemala; ES, El Salvador; HO, Honduras; CR, Costa Rica; CL,
809 Colombia; PE, Peru; BZ, Brazil; BO, Bolivia; AR, Argentina. The colour and code of each
810 genetic group are also specified in the Figure.

811

812 **Figure 2**. Genetic diversity heat map of the wild common bean in Mesoamerica. The map
813 was drawn by interpolation and is based on an individual-centred approach. Colour keys:
814 from low (blue) to high (red) diversity levels.

815

816 **Figure 3.** Differences among the genetic groups for mean 100-seed weights. Groups that do
817 not share the same letter are statistically different ($P < 0.05$). W, wild genotypes; D,
818 domesticated genotypes; Wee, weedy genotypes; MW1, MW2, MW3, MW4, Mesoamerican
819 wild groups; AW1, AW2, Andean wild groups.

820

821 **Figure 4.** Results of spatial structure analysis. (a) Results of the global test, showingthe
822 distribution of the simulated values. Sim, simulated values. The observed value is indicated
823 by a segment that ends with a black diamond, and is larger than all of the simulated values,
824 which indicates the presence of spatial structure ($P < 10^{-4}$). (b) Results of the autocorrelation
825 analysis performed in Mesoamerica. $L_{95\%}$, lower limit; Obs, observed values; $U_{95\%}$, upper
826 limit.

827

828 **Figure 5.** Geographical distribution of the genetic groups identified by Structure when all of
829 accessions are considered. The maps were obtained by interpolation of the Structure

membership coefficients (q_i). (a) Results for K = 6. (b) Results for the three wild Andean groups. Colour keys are the same as those used in Figure 1b.

Figure 6. Relationships among the Mesoamerican wild bean accessions as a function of the first two ecological principal components (ePCA1, ePCA2). The analysis was obtained from the original 55 ecological variables. MW1, MW2, MW3, MW4: Mesoamerican wild genetic groups based on Structure analysis. The 95% density ellipses are calculated for each group, except the MW2 group, which includes only six individuals.

Figure 7. Correlation between the SNP polymorphism environmental (ePCA1, ePCA2) data in Mesoamerica with the SGLMM approach. Loci that show an 'unsually high' correlation with environmental data are indicated with orange dots (0 < log[BF] <3) and red dots (log[BF] > 3).

Figure 8. Results of Structure analysis at K = 5, based on the putatively 'neutral' dataset for the Mesoamerican accessions. The accessions are ordered according to their country of origin, from North to South. The Mexican regions are specified when they include more than three accessions. MW, Mesoamerican wild; MD, Mesoamerican domesticated; DU, Durango; NA, Nayarit; JA, Jalisco; CO, Colima; MI, Michoacan; GN, Guanajuato; GR, Guerrero; MO, Morelos; PU, Puebla; OA, Oaxaca; CH, Chiapas; GU, Guatemala; ES, El Salvador; HO, Honduras; CR, Costa Rica; CL, Colombia. The colour and code of each genetic group are also specified in the Figure.

Figure 9. Results of the neighbour-joining analysis performed on the genotypes with qi > 0.70, excluding the weedy accessions, and considering the putatively 'neutral' ($_N$) SNP dataset. The accessions are coloured according to their membership to the specific genetic groups (see also Fig. 7). $MW1_N$, $MW2_N$, $MW3_N$, $MW4_N$, $MW5_N$: Mesoamerican wild accessions from the different genetic groups based on the Structure analysis; $MD2_N$, Mesoamerican domesticated group.

Figure 10. Differences among the genetic groups obtained using the putatively 'neutral' ($_N$) dataset for 100-seed weights. Groups that do not share the same letter are statistically different (P < 0.05). W, wild genotypes; D, domesticated genotypes; Wee, weedy genotypes;

862 $MW1_N$, $MW2_N$, $MW3_N$, $MW4_N$, $MW5_N$: Mesoamerican wild groups. Within $MW5_N$, the 100-
863 seed weight of the genotypes from Oaxaca is also shown.
864
865 **Figure 11.** Map showing the genetic, archaeological and glottochronological information for
866 the Mesoamerican wild common bean. Orange-red areas, genetic diversity hot-spots, as pin-
867 pointed in the present study; green area, 'desert' of diversity overlapping with the Trans-
868 Mexican Volcanic Belt (Plunket and Uruñela 1998; Marquez *et al.*, 1999; Siebe *et al.*, 2004);
869 light-green dots, wild accessions from Oaxaca that are closer to the Mesoamerican
870 domesticated gene pool; yellow dots, Mesoamerican wild accessions closest to the
871 domesticated gene pool, according to Bitocchi *et al.* (2013); blue triangles: G, Guilá Naquitz
872 Cave (Oaxaca State) archaeological site where common bean macro-remains were dated c.
873 2100 cal BP (Kaplan & Lynch, 1999); T, Tehuacán Valley (Puebla State) where the common
874 bean and maize macro-remains were dated c. 2300 cal BP and c. 6300 cal BP, respectively
875 (Kaplan & Lynch, 1999; Piperno & Flannery, 2001); orange triangle: X, Xihuatoxtla Shelter
876 (Guerrero State), where the oldest maize records were dated c. 8700 cal BP (Piperno *et al.*,
877 2009); azure dashed-line area, Mesoamerican common bean domestication, as suggested by
878 Kwak *et al.* (2009); orange dashed-line area, maize domestication site (Matsuoka *et al.*,
879 2002; Piperno *et al.*, 2009; van Heerwaarden *et al.*, 2011); blue circles, homelands of the
880 language families for which a 'bean' term has been posited: Oto, Otopamean 3654 BP; Pop,
881 Popolocan 3036 BP; Mix, Mixtec-Culcatec 3140 BP; Zap, Zapotecan 3149 BP (Brown *et al.*,
882 2014).

Tables

Table 1. Genetic diversity of the different groups of *Phaseolus* accessions, as estimated by the SNPs analysis.

Population	Genotypes (n)	Polymorphic SNPs (n)	Na	Ne	H_E	ΔH
Mesoamerican gene pool	435	126	1.962	1.490	0.284	0.396
MW	335	119	1.908	1.440	0.260	
MD	100	112	1.855	1.234	0.157	
Andean gene pool	128	125	1.939	1.803	0.126	0.261
AW	68	114	1.870	1.161	0.120	
AD	60	95	1.725	1.121	0.089	
PhI	14	66	1.504	1.086	0.074	
Whole sample - *P. vulgaris*	577	131	2.000	1.624	0.360	

MW, Mesoamerican wild; MD, Mesoamerican domesticated; AW, Andean wild; AD, Andean domesticated; PhI, Ecuador-northern Peru wild group; Na, mean number of alleles; Ne, mean effective number of alleles; He, unbiased expected heterozygosity (Nei, 1978); ΔH, diversity variation between wild and domesticated forms within the same gene pool. When Δ is positive, the diversity of the wild groups is higher than the domesticated groups.

***New Phytologist* Supporting Information**

Supplementary Figures.

Fig. S1. Collection sites of the wild *P. vulgaris* accessions used in the present study.

Fig. S2. Site frequency spectra. The proportion of SNPs with minor allele frequencies (MAF) within the overall sample (ALL), the wild and domesticated groups collected in Mesoamerica (MW, MD, respectively) and in the Andes (AW, AD, respectively), and within the northern Peru-Ecuador group (PhI).

Fig. S3. Estimation of the number of genetic groups (K) calculated according to the delta K value (ΔK) of Evanno *et al.* (2005). The data are shown for the complete dataset (*P. vulgaris*), and within the Mesoamerican (Meso) and Andean (Andes) samples separately.

Fig. S4. Genetic diversity heat map of the wild common bean in the Andes. The map was drawn by interpolation and based on an individual-centred approach. Colour keys: from low (blue) to high (red) diversity levels. The map also shows the genetic, archaeological and glottochronological information for the Andean wild common bean. Light-blue dots, wild accessions closest to the domesticated gene pool, according to Bitocchi *et al.* (2013); the orange triangle (H) indicates Huachichocana (Jujuy Province, Argentina; Tarrago, 1980), the site where the common bean archaeological remains were found; red circle (M) indicates the homelands of the language families for which a 'bean' term has been posited: Matacoan, 2404 BP (Brown *et al.*, 2014).

Fig. S5. Comparison between the chloroplast (cpSSRs) and nuclear (SNPs) genetic structures. The contingency Table shows the association between the plastidial groups (C1, C2, C3; Desiderio *et al.*, 2013) and the nuclear genetic groups. AW1, AW2: Andean wild groups; MW1, MW2, MW3, MW4: Mesoamerican wild groups; PhI: wild group from Ecuador-northern Peru.

Figure S6. Neighbour-joining tree that illustrates the relationships among the genotypes with $q_i > 0.70$, excluding the weedy accessions. The data are shown from (a) the overall *P. vulgaris* dataset; (b) the Mesoamerican dataset; and (c) the Andean dataset. The accessions are coloured according to their membership to the specific genetic groups (see also Fig. 7). MW1, MW2, MW3, MW4: Mesoamerican wild genetic groups; MD2: Mesoamerican domesticated group; PhI, wild group from Ecuador-northern Peru; AW1, AW2: Andean wild groups; AD1, AD2: Andean domesticated groups.

Fig. S7. PCA performed across the individuals with qi > 0.70, excluding the weedy accessions. The accessions are coloured according to their membership to the specific genetic groups (see also Fig. 7). MW1, MW2, MW3, MW4: Mesoamerican wild genetic groups; MD2: Mesoamerican domesticated group; PhI, wild group from Ecuador-northern Peru; AW1, AW2: Andean wild groups; AD1, AD2: Andean domesticated groups.

Fig. S8. Results of the spatial structure analysis. The global test was performed (a) overall (*P. vulgaris*), and (b) in the Andes. The distributions of the simulated values are shown. Sim, simulated values. The observed value is indicated by a segment that ends with a black diamond, and it is larger than all of the simulated values, which indicates the presence of spatial structure ($P < 10^{-4}$). The autocorrelation analysis was performed (c) overall, and (d) in the Andes. $L_{95\%}$, lower limit; Obs, observed values; $U_{95\%}$, upper limit.

Figure S9. Correlation between SNP polymorphism environmental (ePCA3, ePCA4, ePCA5) data in Mesoamerica with the SGLMM approach. Loci that show an 'unusually high' correlation with environmental data are indicated with orange dots (0 < log[BF] < 3) and red dots (log[BF] > 3).

Fig. S10. Comparison between the chloroplast (cpSSRs) and nuclear (SNPs) genetic structures. The contingency Table shows the association between the plastidial groups (C1, C2, C3; Desiderio *et al.*, 2013) and the nuclear genetic groups, as obtained from Structure analysis on the putatively 'neutral ($_N$) dataset. $MW1_N$, $MW2_N$, $MW3_N$, $MW4_N$, $MW5_N$: Mesoamerican wild groups.

Fig. S11. Results of the autocorrelation analysis on the Mesoamerican wild gene pool. The analysis was performed on (a) the the 'neutral' dataset, and (b) loci under selection.

Supplementary Tables (submitted as a unique excel file):

Table S1. Accessions used for the analyses. Passport data and results of structure analyses are reported. Accessions in common with Bitocchi et al. (2013) and Desiderio et al. (2013), and relative results of the genetic structure are specified.

Table S2. Loci used for the SNPs detection. Hypothetical gene function, when available, is indicated. Further details are available in references 1 (Goretti et al., 2013) and 2 (Cortès et al., 2011).

Table S3. Environmental variables used for the spatial analysis of the genetic diversity.

Table S4. F_{ST} values among the genetic groups, as obtained from the Structure analysis.

958 **Table S5.** Ecological variables associated with ePCA1 and ePCA2.

959 **Table S6**. SNPs and their relative loci that are under putative selection.

960

961 **Supplementary Notes**

962 **Note S1.** Genetic structure and diversity analyses with the putatively neutral dataset.

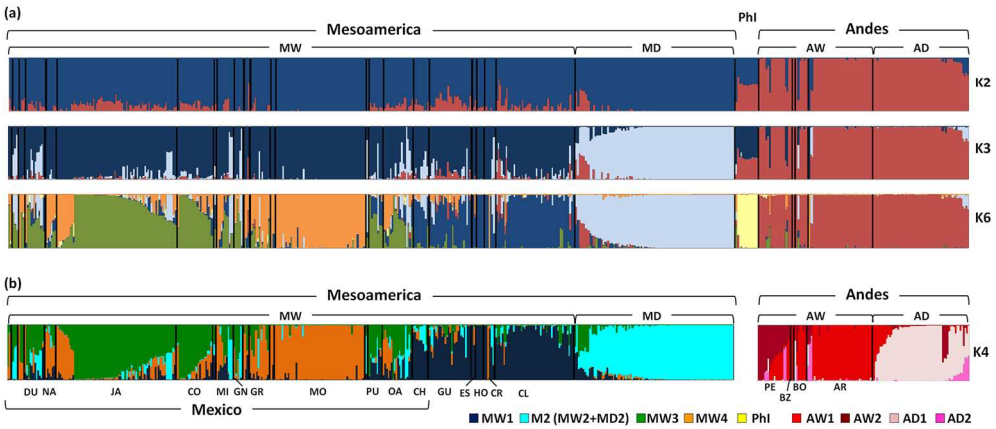


Figure 1. Results of the Structure analysis. (a) Results at K = 2, K = 3 and K = 6, based on 131 SNPs across all of the 577 P. vulgaris accessions. (b) Results of the analyses performed separately for the Mesoamerican and Andean gene pools, at K = 4. The wild Mesoamerican and Andean accessions are ordered according their country of origin, from the North to the South. The Mexican regions are specified when they include more than three accessions. MW, Mesoamerican wild; MD, Mesoamerican domesticated; PhI, accessions with phaseolin type I; AW, Andean wild; AD, Andean domesticated; DU, Durango; NA, Nayarit; JA, Jalisco; CO, Colima; MI, Michoacan; GN, Guanajuato; GR, Guerrero; MO, Morelos; PU, Puebla; OA, Oaxaca; CH, Chiapas; GU, Guatemala; ES, El Salvador; HO, Honduras; CR, Costa Rica; CL, Colombia; PE, Peru; BZ, Brazil; BO, Bolivia; AR, Argentina. The colour and code of each genetic group are also specified in the Figure.
169x79mm (300 x 300 DPI)

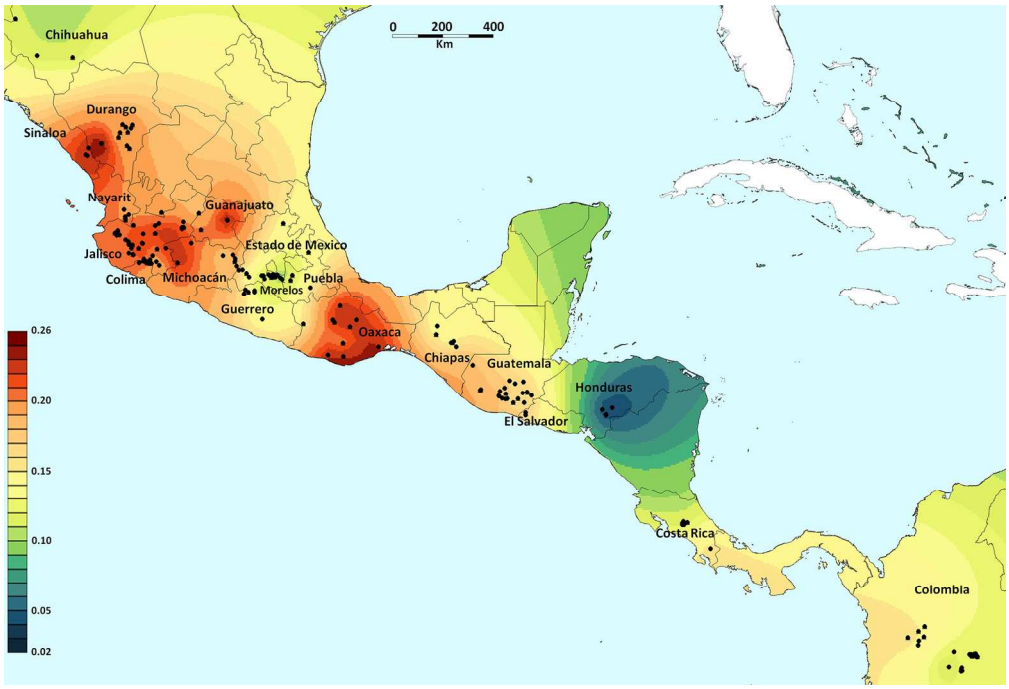


Figure 2. Genetic diversity heat map of the wild common bean in Mesoamerica. The map was drawn by interpolation and is based on an individual-centred approach. Colour keys: from low (blue) to high (red) diversity levels.
169x114mm (300 x 300 DPI)

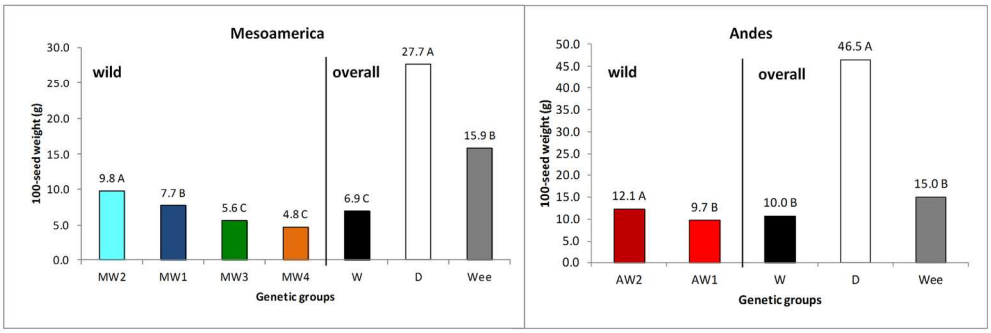


Figure 3. Differences among the genetic groups for mean 100-seed weights. Groups that do not share the same letter are statistically different ($P < 0.05$). W, wild genotypes; D, domesticated genotypes; Wee, weedy genotypes; MW1, MW2, MW3, MW4, Mesoamerican wild groups; AW1, AW2, Andean wild groups.
169x58mm (300 x 300 DPI)

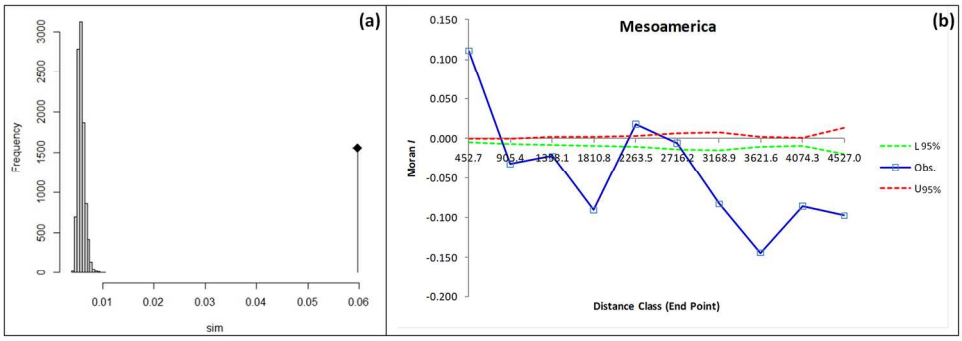
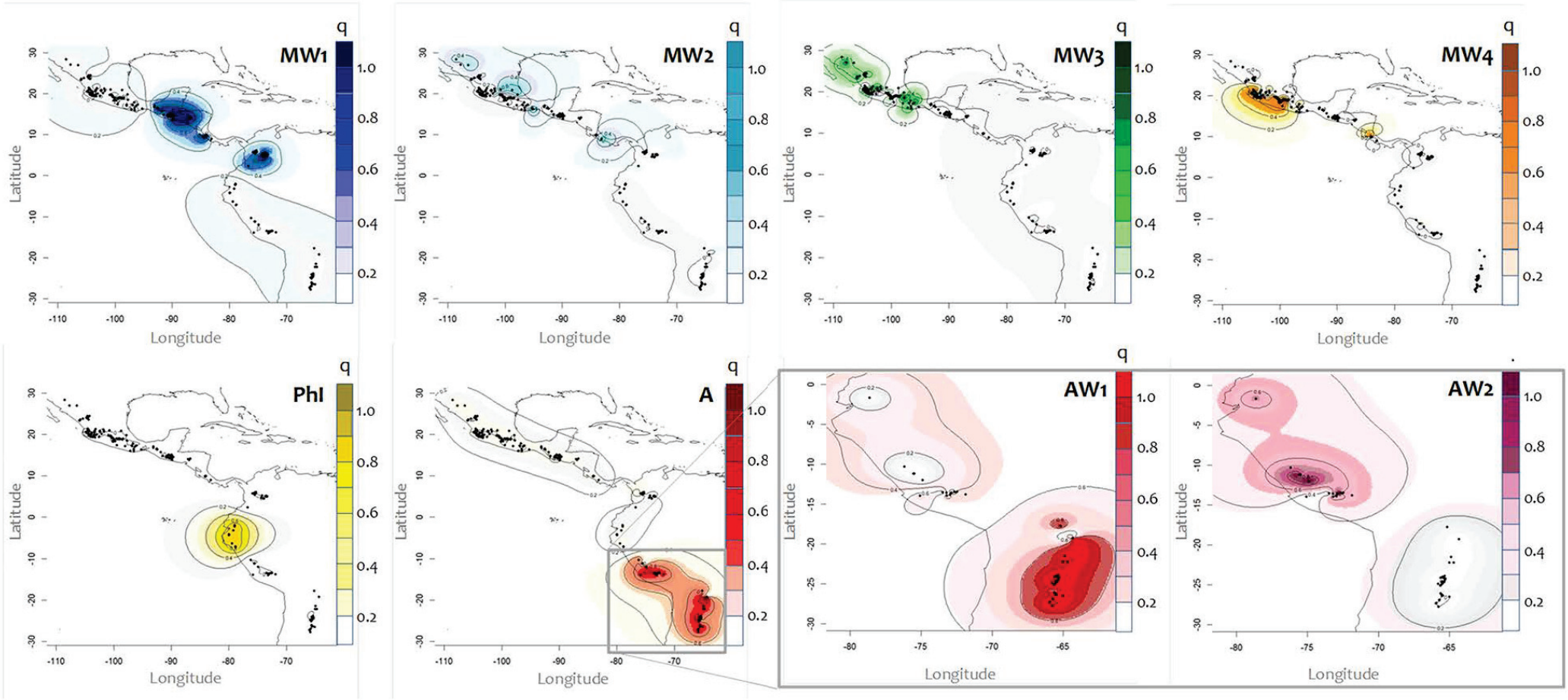


Figure 4. Results of spatial structure analysis. (a) Results of the global test, showingthe distribution of the simulated values. Sim, simulated values. The observed value is indicated by a segment that ends with a black diamond, and is larger than all of the simulated values, which indicates the presence of spatial structure (P < 10-4). (b) Results of the autocorrelation analysis performed in Mesoamerica. L95%, lower limit; Obs, observed values; U95%, upper limit.
169x66mm (300 x 300 DPI)

Fig. 5



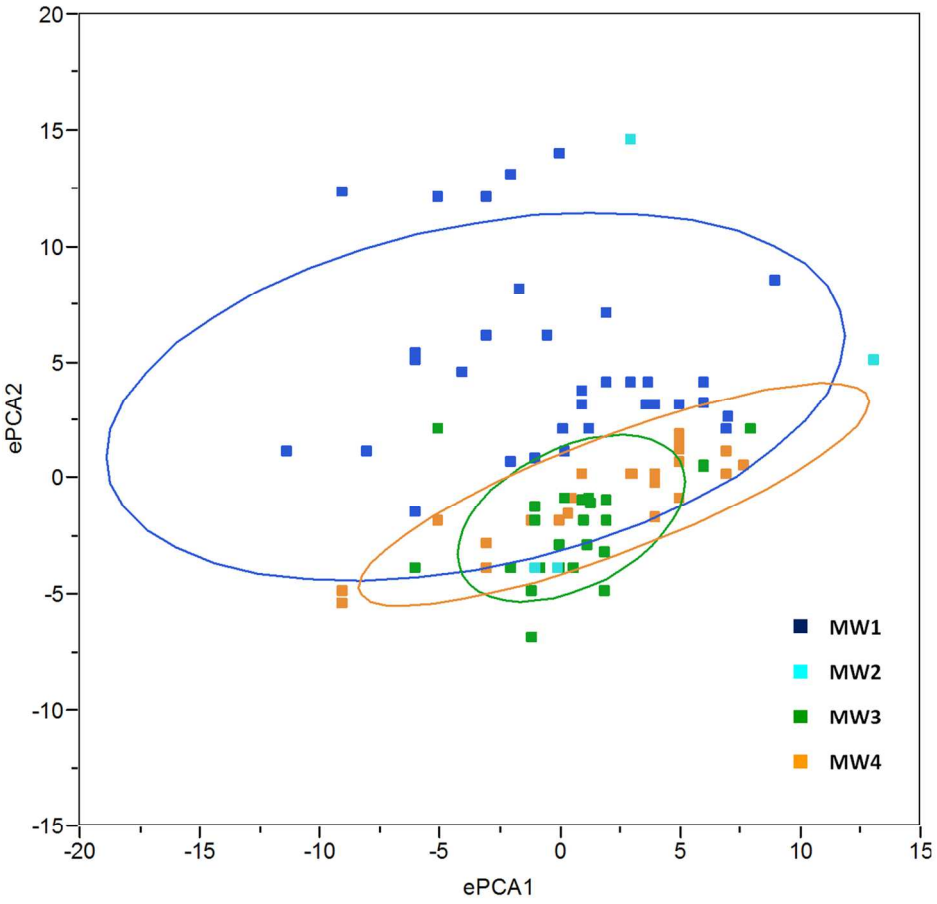


Figure 6. Relationships among the Mesoamerican wild bean accessions as a function of the first two ecological principal components (ePCA1, ePCA2). The analysis was obtained from the original 55 ecological variables. MW1, MW2, MW3, MW4: Mesoamerican wild genetic groups based on Structure analysis. The 95% density ellipses are calculated for each group, except the MW2 group, which includes only six individuals.
169x161mm (300 x 300 DPI)

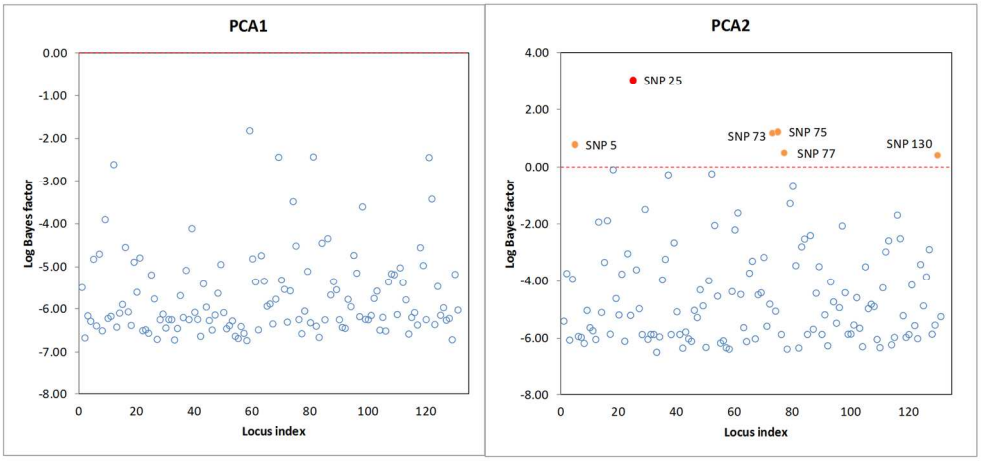


Figure 7. Correlation between the SNP polymorphism environmental (ePCA1, ePCA2) data in Mesoamerica with the SGLMM approach. Loci that show an 'unsually high' correlation with environmental data are indicated with orange dots (0 < log[BF] <3) and red dots (log[BF] > 3).
169x82mm (300 x 300 DPI)

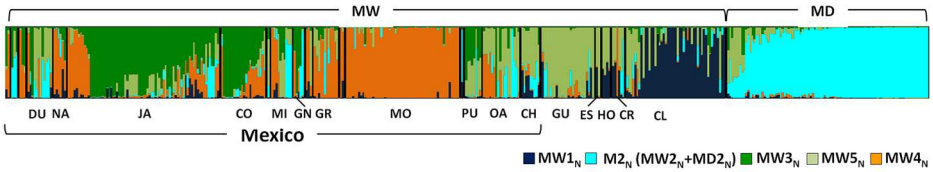


Figure 8. Results of Structure analysis at K = 5, based on the putatively 'neutral' dataset for the Mesoamerican accessions. The accessions are ordered according to their country of origin, from North to South. The Mexican regions are specified when they include more than three accessions. MW, Mesoamerican wild; MD, Mesoamerican domesticated; DU, Durango; NA, Nayarit; JA, Jalisco; CO, Colima; MI, Michoacan; GN, Guanajuato; GR, Guerrero; MO, Morelos; PU, Puebla; OA, Oaxaca; CH, Chiapas; GU, Guatemala; ES, El Salvador; HO, Honduras; CR, Costa Rica; CL, Colombia. The colour and code of each genetic group are also specified in the Figure.
169x49mm (300 x 300 DPI)

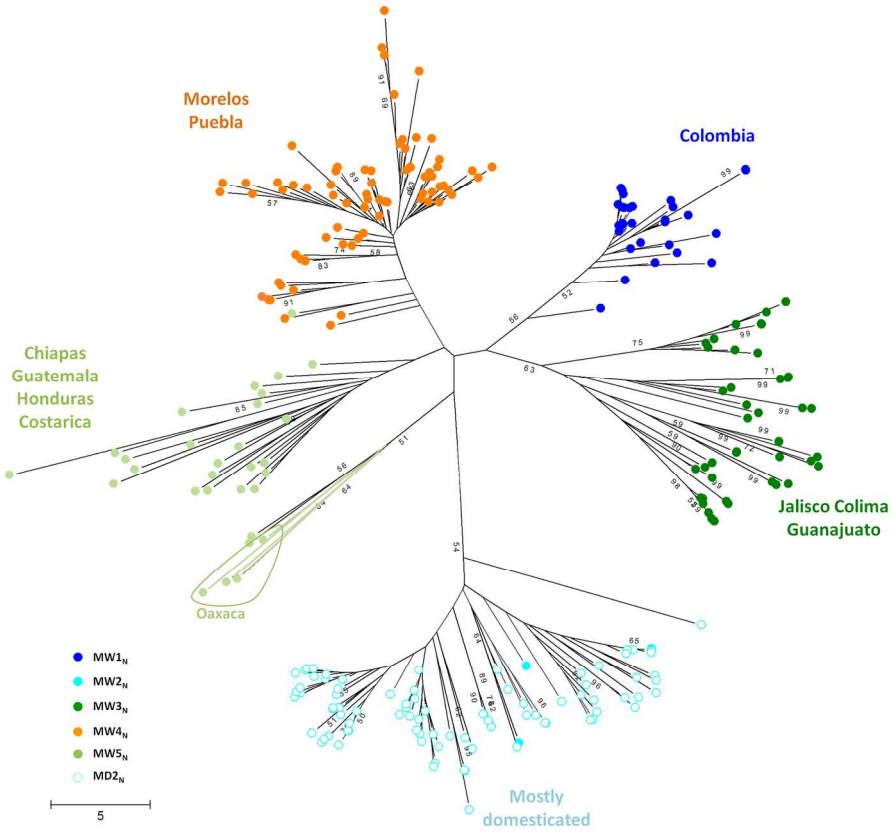


Figure 9. Results of the neighbour-joining analysis performed on the genotypes with qi > 0.70, excluding the weedy accessions, and considering the putatively 'neutral' (N) SNP dataset. The accessions are coloured according to their membership to the specific genetic groups (see also Fig. 7). MW1N, MW2N, MW3N, MW4N, MW5N: Mesoamerican wild accessions from the different genetic groups based on the Structure analysis; MD2N, Mesoamerican domesticated group.
169x146mm (300 x 300 DPI)

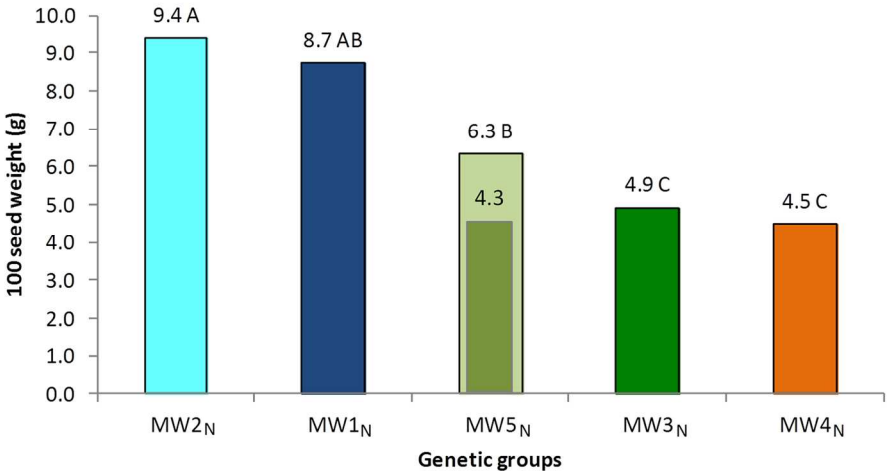


Fig. 10 Differences among the genetic groups obtained using the putatively 'neutral' (N) dataset for 100-seed weights. Groups that do not share the same letter are statistically different ($P < 0.05$). W, wild genotypes; D, domesticated genotypes; Wee, weedy genotypes; MW1N, MW2N, MW3N, MW4N, MW5N: Mesoamerican wild groups. Within MW5N, the 100-seed weight of the genotypes from Oaxaca is also shown.
169x106mm (300 x 300 DPI)

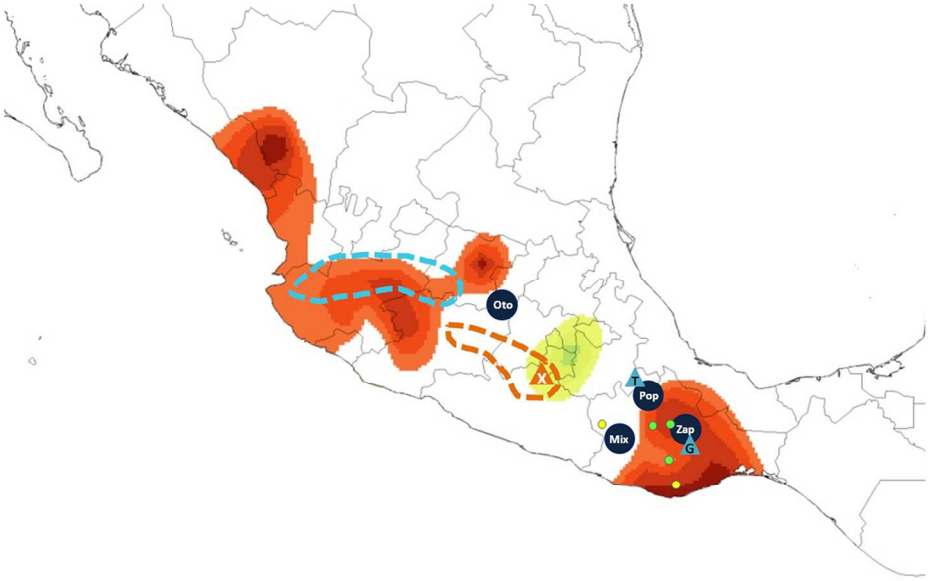


Figure 11. Map showing the genetic, archaeological and glottochronological information for the Mesoamerican wild common bean. Orange-red areas, genetic diversity hot-spots, as pin-pointed in the present study; green area, 'desert' of diversity overlapping with the Trans-Mexican Volcanic Belt (Plunket and Uruñela 1998; Marquez et al., 1999; Siebe et al., 2004); light-green dots, wild accessions from Oaxaca that are closer to the Mesoamerican domesticated gene pool; yellow dots, Mesoamerican wild accessions closest to the domesticated gene pool, according to Bitocchi et al. (2013); blue triangles: G, Guilá Naquitz Cave (Oaxaca State) archaeological site where common bean macro-remains were dated c. 2100 cal BP (Kaplan & Lynch, 1999); T, Tehuacán Valley (Puebla State) where the common bean and maize macro-remains were dated c. 2300 cal BP and c. 6300 cal BP, respectively (Kaplan & Lynch, 1999; Piperno & Flannery, 2001); orange triangle: X, Xihuatoxtla Shelter (Guerrero State), where the oldest maize records were dated c. 8700 cal BP (Piperno et al., 2009); azure dashed-line area, Mesoamerican common bean domestication, as suggested by Kwak et al. (2009); orange dashed-line area, maize domestication site (Matsuoka et al., 2002; Piperno et al., 2009; van Heerwaarden et al., 2011); blue circles, homelands of the language families for which a 'bean' term has been posited: Oto, Otopamean 3654 BP; Pop, Popolocan 3036 BP; Mix, Mixtec-Culcatec 3140 BP; Zap, Zapotecan 3149 BP (Brown et al., 2014).
169x121mm (300 x 300 DPI)