### Title

Signatures of Environmental Adaptation During Range Expansion of Wild Common Bean (Phaseolus vulgaris)




### Authors

Ariani, Andrea
Gepts, Paul



### Publication Date

2019

### DOI

10.1101/571042

Peer reviewed

# Signatures of potential environmental adaptation during range expansion of wild common bean (*Phaseolus vulgaris*) using an integrated landscape genomics approach

Andrea Ariani and Paul Gepts*

Department of Plant Sciences, Section of Crop and Ecosystem Sciences,

University of California, Davis, CA 95616-8780

*Correspondence**:** plgepts@ucdavis.edu

**Abstract**

**Background:** Landscape genomics integrates population genetics with landscape ecology, allowing the identification of the putative molecular determinants involved in plant environmental adaptation across their natural geographic and ecological range. Wild *Phaseolus vulgaris*, the progenitor of common bean (*P. vulgaris*), has a remarkably extended distribution from northern Mexico to northwestern Argentina. Earlier research has shown that this distribution represents a range expansion from Mesoamerica to the southern Andes through several discrete migration events and that the species colonized areas with different temperature and rainfall compared to its core area of origin.

**Results:** In the current study, we applied a landscape genomics approach to a collection of 246 wild common bean accessions representative of its broad geographical and climatic distribution and genotyped for ~20K SNPs. We applied two different but complementary approaches for identifying loci putatively involved in environmental adaptation: i) an outlier-detection method that identifies loci showing strong differentiation between sub-populations; ii) an association method based on the identification of loci associated with bio-climatic variables. This integrated approach allowed the identification of several genes showing signature of selection across the different natural sub-populations of this species, as well as genes associated with specific bio-climatic variables related to temperature and precipitation.

**Conclusions:** The current study demonstrates the feasibility of landscape genomics approach for a preliminary identification of novel candidate genes involved in environmental adaptation in *P. vulgaris*. As a resource for

broadening the genetic diversity of the domesticated gene pool of this species, the genes identified constitute potential molecular markers and introgression targets for the breeding improvement of domesticated common bean.

**Keywords:** Landscape genomics; Crop Wild Relatives (CWRs); climate adaptation; GWAS; natural selection; domestication

## Background

Climate change represents one of the primary threats for food security worldwide, but especially in developing countries that rely heavily on agricultural production from smallholder farmers [1,2]. Indeed, several studies have highlighted a predominant role of climate change in reducing agricultural productivity and increasing inter-annual variability in crop yields, thus directly affecting food availability and stability [3,4].

The increase in average temperatures, along with the higher frequency and intensity of extreme weather conditions, will require the development of new plant varieties adapted to this changing environment in order to meet future food security needs [5,6]. However, the development of new varieties requires the introduction of genetic diversity into breeding programs to find the correct combinations of favorable alleles in a specific crop [7]. The genetic variability available in domesticated plants is generally low due to the bottleneck effect induced by domestication and subsequent selection during variety improvement (Zamir 2001; Ford-Lloyd et al. 2011; Gepts 2014), thus new sources of genetic diversity need to be introduced into breeding programs. Crop Wild Relatives (CWRs) represent a large, and mostly unexploited, source of genetic diversity readily available for plant improvement under climate change [7,8,10,11]. However, the use of CWRs in breeding programs for improving stress resistance in domesticated species could be hindered by the lack of knowledge of the genetic determinants of resistance and also by difficulties in phenotyping a large number of individuals under agricultural conditions [11–13]. One possible solution for overcoming these difficulties is the integration of environmental and genotypic datasets to understand the

genetic basis of natural selection in wild populations, an approach known as ‘landscape genomics’ [14,15]. In addition, this approach offers both theoretical and practical applications since it strengthens the understanding of plant natural adaptation but allows also the identification of molecular markers that could be readily applicable for breeding improvement of domesticated plants [16].

Several methods have been developed for identifying signature of natural selection (e.g., selective sweeps) in natural populations. These methods can be divided mostly in outlier-detection methods, which identify hard-selection sweeps, and association methods, that identify soft-selection sweeps [14,17]. Outlier-detection methods are based on population differentiation analysis and aim at identifying loci with drastic differences in allele frequencies between populations, as measured by $F_{st}$ [18,19]. Although based on the assumption that alleles fixed within sub-populations could confer an evolutionary advantage in the ecological niche occupied [20,21], these methods do not take directly into account climatic data and could be biased by complex population structure and/or demography [22]. On the other hand, association methods directly correlate genotypic with environmental data and rely on the assumption that variations of allele frequencies across environmental gradients are possible signature of local adaptation [23]. The theory beneath environmental association methods are practically the same as that used in Genome Wide Association Studies (GWAS) [24] Both approaches employ mixed model association approaches for correcting for the confounding effects that could be introduced by population structure and relatedness in the sample [25].

Common bean (*P. vulgaris*) is an essential staple crop providing most of proteins and micronutrients in the diet of the majority of the population in several developing countries [26]. The regular consumption of this crop provides several health benefits, like reducing the risks of heart disease, obesity, and diabetes [27]. Its cultivation improves agricultural sustainability thanks to its nitrogen-fixing ability [28]. Common bean shows a surprisingly high genetic diversity, with the presence of at least three geographically isolated and divergent wild gene pools located in 1) Mesoamerica and the northern Andes (MW); 2) the Central Andes (Ecuador and northern Peru; PhI); and 3) the Southern Andes (southern Peru, Bolivia, and northwestern Argentina; AW) [29–31]. Common bean was domesticated independently in Mexico and the Southern Andes, producing locally-adapted varieties and landraces with specific characteristics [32–38]. The intermediate gene pool in the Central Andes was not domesticated [39,40]. This wild group has been recently identified as a cryptic sister species of *P. vulgaris,* named *Phaseolus debouckii*, which was disseminated from the center of origin of this species in Mesoamerica and remained geographically isolated from the other wild gene pools of this species [41,42].

Wild common bean is an annual vine plants that is found from the state of Chihuahua in northern Mexico (approx. 35° N. Lat.) to the Córdoba province in Argentina (approx. 35° S. Lat.), encompassing almost 70 latitudinal degree or about 10,000 km [43,44]. This species grows in both tropical and sub-tropical environments across the Americas at elevations between 500 and 2,000 m a.s.l. with annual rainfall from 500 to 1,800 ml [43,44,12]. This broad geographic and ecological distribution suggests the existence of genotypes

adapted to a wide variety of environmental conditions, which could be useful donors of abiotic stress resistance for improving domesticated common bean production under climate change [44,45].

Future projection of climate changes under different models predict a reduction of suitability for common bean production in areas where this plant is an essential staple crop and also a source of household income, hence endangering food security and increasing rural poverty in already susceptible areas of the world [46]. For this reason, it is essential to understand the molecular mechanisms involved in wild common bean adaptation to different environments and to identify molecular markers that could be useful in breeding improvement of this crop. The application of landscape genomics approaches in wild common bean could help address these issues, as demonstrated previously in several other plant species like soybean, barley, *Medicago truncatula*, maize, and *Brachypodium* [16,47–50].

In the current study, we applied a landscape genomics approach to understand environmental adaptation to a dataset comprised of 246 wild common beans genotyped for ~20K previously developed SNPs [51]. A similar analysis was performed previously in this species using 148 SNPs located in genes putatively involved in adaptation to biotic or abiotic stresses [52]. However, the higher number of markers developed in this study and the broader and more even distribution across the genome of these markers, results in a more comprehensive and precise analysis of environmental adaptation in this species. In addition, the genes identified as associated with environmental variables can be validated and applied in the future for domesticated common bean breeding improvement.

## Materials and Methods

### Plant material and genotype data

A panel of 246 wild *P. vulgaris* accessions, previously genotyped with a Genotyping-By-Sequencing (GBS) protocol using the *Cvi*AII restriction enzyme [51], was analyzed. The panel was representative of the ecological and geographic distribution of this species and included 157 genotypes of the Mesoamerican (MW), 77 of the Southern Andes (AW), and 12 of the Central Andes (Northern Peru-Ecuador; PhI) gene pools. The SNPs considered in this study were those with a Minor Allele Frequency (MAF) ≥ 0.05 and less than 20% missing data. The list of the accessions sequenced, with gene pool information and geographic coordinates, is available in Additional File 1, while genotyping data in VCF format are available in Additional File 2. The seeds were provided by the International Center of Tropical Agriculture (CIAT, Cali, Colombia) and the United States Department of Agriculture Western Regional Plant Introduction Station (Pullman, WA).

### Spatial analysis

Spatial analyses were conducted within the R statistical environment (www.r-project.org) using the dismo package and its dependencies (raster and sp). The geographic coordinates of the individuals analyzed in this study were used for retrieving the 19 bio-climatic summary variables from the WorldClim database (http://www.worldclim.org/). The data were downloaded at a 30-second resolution (approximately 0.86 $km^2$ at the equator). In order to identify a subset of bio-climatic variables that best summarizes our dataset, we performed a

Principal Component Analysis (PCA) on the scaled and centered variables using the ChemometricsWithR package [53]. We then selected the first two variables with the highest positive and negative loading in the first four principal components (PC1 to PC4) (Additional File 3). Since some of the selected bio-climatic variables showed a high correlation (Additional File 4), we decided to pick only one of the correlated variables for further analysis. The final bio-climatic variables analyzed in this study were: bio_3 (Isothermality), bio_5 (Max Temperature of Warmest Month), bio_6 (Minimum Temperature of Coldest Month), bio_7 (Temperature Annual Range), bio_12 (Annual precipitation), bio_14 (Precipitation of Driest Month), and bio_18 (Precipitation of Warmest Quarter). In addition to the above-mentioned bio-climatic variables, we included also annual Potential EvapoTranspiration (PET) downloaded from the Global Aridity and PET Database (http://www.cgiar-csi.org/data/global-aridity-and-pet-database).

**Genome scans for selection and association analysis**

Genome scans for selection (i.e., hard selective sweeps) were performed on the final set of SNPs using the pcadapt R package [54], an algorithm able to detect population structure and outlier loci by performing a PCA analysis on SNPgenotypic data. The best number of sub-populations was inferred by visually evaluating the scree plot of eigenvalues for the different principal components (K); the genomic scans for selection were performed for K in the range 2-5. The p-values obtained by this analysis were corrected using the Bonferroni method and only SNPs with a corrected p-value ≤ 0.001 were considered as significant.

Association analysis (i.e., soft selective sweeps) was performed separately for each of the seven selected bio-climatic variables and annual PET. For this analysis, we used the LFMM algorithm [55] implemented in the LEA R package [56]. This method was developed specifically for identifying signature of environmental selection in genomic data and is able to efficiently correct for population history and isolation-by-distance (IBD). In order to correct for spurious association determined by population structure or IBD, the number of latent factors (i.e., populations) needs to be decided *a priori* and subsequently evaluated using the genomic inflation factor parameter. Since LFMM is based on Monte Carlo Markov Chain (MCMC) sampling, we ran it multiple times for each association analysis and then averaged the p-values (as suggested in the software documentation). To identify the best number of populations (K) for association with each bio-climatic variable, we performed three runs of the program with K in the range 4-10 and estimated the inflation factor from these runs [57]. Plots of the inflation factor for different values of K (Additional File 5) showed that the best inflation factor for reducing False Discovery Rate (FDR) (i.e., closest to 1) was six for Bio12, Bio14, and Bio5, and 7 for Bio6, Bio18, Bio7, Bio3, and PET. Based on this preliminary screening, we re-ran the program with the best number of K for 10 times with 10,000 MCMC iterations and a burn-in period of 1,000. The p-values where then averaged across the different runs and corrected using the Bonferroni method. SNPs with a corrected p-value ≤ 0.05 were considered as significant.

**Identification of putatively selected genes**

The distance between significant SNPs, identified by genome scans or association analysis based on the *P. vulgaris* v1.0 genome annotation (https://phytozome.jgi.doe.gov/pz/portal.html), was evaluated using the GenomicRanges/rtracklayer packages or R [58,59]. Only genes within 5 Kb of a significant SNPs were chosen as putatively selected genes. This 5 Kb upper limit was selected based on the genotyping approach used in this study (that did not allow a full coverage of the genome), but also took into account the presence of possible regulatory regions immediately adjacent to gene sequences [60].

**Candidate genes evaluation across genetic groups**

For clustering individuals based on genetic groups and visualize allele frequency variations across clusters, we applied a K-means clustering approach using the first 5 PCs obtained from pcadapt analysis. As best number of clusters we selected K=5, as suggested by the scree plot of the eigenvalues obtained with pcadapt. The clustering analysis was performed using the python scikit-learn library [61]. For each genetic cluster we calculate allele frequencies for SNPs tagging candidate genes using VCFtools [62], and plot them on genetic maps using R.

# Results

## Bio-climatic data analysis

The bio-climatic variables downloaded from the WorldClim database concern mostly temperature and rainfall during the year. These bio-climatic variables were developed for generating biologically informative variables useful for species distribution modeling and landscape genomics approaches. In our analyses, the 19 bio-climatic variables analyzed showed a great degree of correlation, in particular for similar variables like bio_14 (precipitation of the driest month) and bio_17 (precipitation of the driest quarter), or bio_13 (precipitation of the wettest month) with bio_16 (precipitation of the wettest quarter) (Additional File 4).

The loading plot on the first two PCs showed some correlations between bio-climatic variables and principal components, as well as strong correlations between some of the bio-climatic variables analyzed (**Fig. 1A**). In particular, bio_12 (annual precipitation) and bio_4 (temperature seasonality) showed a strong correlation with PC1. On the other hand, bio_5 (max temperature of the warmest month), bio_8 (mean temperature of the wettest quarter), and bio_10 (mean temperature of the warmest quarter) showed a strong correlation with PC2. Interestingly, most of the variables related to precipitation (bio_12, bio_14, bio_16, bio_17, bio_18, and bio_19) were positively correlated with PC1, the variables related to seasonal variation (bio_2, bio_4, bio_7, and bio_15) were negatively correlated with PC1, while the variables related to temperature (bio_1, bio_5, bio_8, bio_9, bio_10, and bio_11) were negatively correlated with PC2.

In addition, this PCA on the bio-climatic variables for the genotypes analyzed showed that the first two principal components (PC1 and PC2) explained 75% of the variance (**Fig. 1**), while PC1 to PC4 explained cumulatively > 90% of the variance (Additional File 6). A plot of PC1 vs. PC2 showed some differences in the distribution of the different gene pools of wild common bean in the PC dimensional space. In particular, the majority of genotypes from the Mesoamerican (MW1 to MW3) and Intermediate (PhI) gene pools were distributed towards the positive part of PC1, while the Andean group were located in the negative part of this axis (**Fig. 1A**). Given the origin of the genus *Phaseolus* in the Mesoamerican area (with local descendants represented by MW1 and MW2), three range expansions characterize this species: 1) PhI, which established wild populations on the western slope of the Andes in Ecuador and northern Peru; 2) AW, encompassing wild populations in the southern Andes; and 3) MW3, a more recent and perhaps ongoing dissemination to Central America and the eastern slope of the northern Andes [51]. Inspection of Fig. 1A and Additional File 3 shows that the distribution of the PhI group, which resulted from the earliest range expansion event, correlates - on bioPC3 - with Isothermality (bio_3), Temperature Seasonality (bio_4), bio_13 (Precipitation of the Wettest Month), and bio_18 (Precipitation of the Warmest Quarter), consistent with a dispersal to an equatorial region. In contrast, the predominant distribution of the southern Andean accessions (AW) in the upper left quadrant of Fig. 1 is consistent with earlier observations that the populations of this gene pool are distributed in cooler and drier locations, as shown by correlations with bio_6 (Minimum Temperature of the Coldest Month), bio_9 (Mean Temperature of the Driest Quarter, bio_11 (Mean

Temperature of the Coldest Quarter) and bio_1 (Annual Mean Temperature). This dissemination occurred with a concomitant lower potential evapotranspiration [51]. Dispersal of the MW3 group (Fig. 1) increased Isothermality (bio_3) and decreased Seasonality (bio_4) and Precipitation Seasonality (bio_15); it also increased Precipitation during the Driest Month (bio_14) and Driest Quarter (bio_17).

**Genome scan of selection**

An analysis of the scree plot of the PCA analysis conducted on SNP data showed that the majority of the variance could be explained by the first principal component (15%), even though PC2 to PC5 also explained a considerable amount of variance in the data (**Fig. 2A)**. On the other hand, after PC5, no significant increase in the explained variance could be detected. This pattern of the scree plot is representative of a possible range expansion of this species across the Americas, as hypothesized by a prior evolutionary analysis of this same collection [51]. Visual inspection of p-value distribution for genome scans for K=2 and K=3 showed a large proportion of low and high p-values, while for K=4 and K=5 the distribution of p-values was more uniform, especially for K=5 (Additional File 7). For this reason, we selected K=5 for further genome scan analysis.

A plot of genetic PCA analysis performed with the pcadapt algorithm was able to discriminate between the different wild gene pools of this species (**Fig. 2B**). In particular, the MW1-MW3 and PhI groups were mostly localized on the positive part of PC1, while the AW gene pool was localized towards the

negative end of this PC. Interestingly, PC1 mostly differentiated MW vs. AW, while PC2 and PC3 clearly separated the MW+AW groups from the PhI (**Fig. 3**).

The genome scan analysis with K=5 identified 84 significant variants (Bonferroni-corrected p-value ≤ 0.001) distributed throughout the 11 chromosomes of common bean (**Fig. 4**), tagging 70 annotated genes. The highest number of tagged genes were identified on chromosomes Pv02 and Pv04 with 15 and 11 genes, respectively. The genes identified as selected by genome scan analysis were mostly related to plant development (17 genes), hormone response (10 genes), ion homeostasis (5 genes), and response to stress (9 genes) (**Table 1**). Among the genes identified, we found several of them related to drought and/or abscisic acid (ABA) response like Phvul.002G331700, a homolog of the *Arabidopsis* KUP6 involved in potassium uptake transporter and stomata movement; Phvul.002G143100, a glycine-rich domain protein (GRP) involved in auxin signaling and stress response; Phvul.004G102800, homologous to *Arabidopsis* SLAH3 involved in ABA response; Phvul.008G161000, homolog of *Arabidopsis* CAO, a gene related to chlorophyll biosynthesis and ABA signaling; and Phvul.009G050600, a gene annotated as an importin β protein involved in ABA and drought response in *Arabidopsis*.

## Genome-wide association analysis

A genome-wide association analysis identified 49 genes associated with the bio-climatic variables selected for this analysis. Except for the Bio18 (Precipitation of Warmest Quarter) variable, for which no associations were detected, the other variables were associated with at least one gene. The bio-climatic variables with the highest number of associated genes were Bio3 (Isothermality) with 29 genes, and Bio12 (Annual precipitation) with 11 genes (**Table 2**). The associated genes were located in all 11 common-bean reference genome chromosomes, except for chromosome Pv06 where there were no significantly associated SNPs. Interestingly, some of the genes were associated with more than one bio-climatic variables (**Fig. 5**), suggesting the possibility that they could be related to multiple environmental stimuli. Among the genes significantly associated with one or more bio-climatic variables, we found several of them related to hormone response, ion homeostasis, plant development, metabolism and response to stress, in particular drought (**Table 2**). There was no overlap between the genes identified by genome scan and association analysis. Among the genes identified, we found some interesting candidates probably involved in stress resistance, like Phvul.001G034400, a homolog of *Arabidopsis* KEA6 involved in potassium homeostasis; Phvul.010G155000, homologous to an *Arabidopsis* phospholipase D α 1 (PLDα1) involved in ABA signaling; Phvul.010G035200 homolog of a cytokinin responsive factor homologous of *Arabidopsis*; and Phvul.008G161700, homologous to an *Arabidopsis* thioredoxin involved in ROS signaling.

## Candidate gene allele distributions

To evaluate the geographic distribution of alleles in candidate genes identified by genome scan and association analysis, we clustered the genotypes into groups with a K-means clustering approach on the molecular PCs calculated with pcadapt. The advantage of a K-means clustering approach, over a standard population structure analysis, is that it clearly assigns individuals to specific clusters. The K-means clustering approach identified three clusters for the MW group, with two clusters (MW1 and MW2) located in Mexico and another (MW3) in Central America and Colombia, plus one cluster each for the intermediate (PhI) and the Andean (AW) group (Additional File 8). Interestingly, the clustering results closely resembled those obtained in a previous study with more advanced population structure approaches (Additional File 1) [51].

The allele frequency distribution of the candidate genes identified by genome scan showed drastic differentiation between the genetic groups identified (**Fig. 6**), as expected from the assumptions of the genome scan approach, with some alleles being private for just one of the genetic group (like the alternative alleles for GRP and CAO that are present only in the AW group). On the other hand, the genes identified by association analysis showed a wide variety of allele frequencies distribution across the different genetic groups (**Fig. 7**), even though some genes had only a single allele in some of the populations (like the reference allele for PLD and TRX in the PhI and AW group). In general, the genes identified by association analysis showed a higher variation of allele frequencies in the different MW groups.

## Discussion

Wild common bean (*P. vulgaris*) grows in several areas of Mexico and Central and South America, from northern Mexico to northwestern Argentina across ~70 latitudinal degrees, in different environments with a wide range of altitudes, average temperatures, and rainfall regimes [12,43,44]. Thanks to this exceptional geographic distribution, its complex evolutionary history, and high levels of genetic diversity, this species represents an extraordinary resource for evolutionary studies [29–31,41,42,63,64], but can be also a conceptual framework for testing and validating landscape genomics approaches in wild plant populations and its feasibility for breeding improvement of domesticated crops [16]. In the current study, we identified several genes that could be involved in environmental adaptation in wild common bean by combining genome scan and association analysis. If validated, the genes identified could be useful candidates for improving stress resistance in domesticated common bean.

### Genome scan of selection

Genetic PCA analysis clearly separated the three groups of this species, as observed in previous research. In particular, the Intermediate gene pool was shown again to diverge from the Mesoamerican and Andean gene pools, especially along the PC2 and PC3, further supporting the hypothesis that this gene pool is actually a distinct species of *Phaseolus* [41,42]. A genome scan based on genetic PCA analysis identified several genes with a strong signature of selection (hard-selection sweep) that could be involved in environmental adaptation across the geographical range of this species. The identification of

414 several genes involved in plant development and hormone and stress
415 response, suggests that the different populations of this species adapted to
416 their environment by integrating and adjusting to developmental, hormonal
417 and environmental cues. There are several genes among those identified that
418 could be of interest for improving stress resilience in common bean, like the
419 KUP6 potassium ($K^+$) transporter located on chromosome Pv02
420 (Phvul.002G331700). This gene has been directly linked to drought stress by
421 regulating ABA response and stomata movements in *Arabidopsis* [65]. A
422 homolog of this gene located on chromosome Pv03 (Phvul.003G052900)
423 showed a higher genetic and transcriptional diversity in Mesoamerican
424 domesticated beans than in wild ones [66,67]. Due to the possible role of KUP-
425 like genes in response to drought stress and their identification as selected
426 genes in both wild and domesticated populations of common bean, further
427 studies should focus on the evolution and diversity of this gene family in this
428 species.
429 Another gene identified in the current study and possibly involved in
430 adaptation to drought response in wild common bean is Phvul.004G102800,
431 homolog of SLAH3 of *Arabidopsis*, which was annotated as an S-type anion
432 channel. This type of channels is rapidly regulated by ABA and stimulate
433 stomata closure by inhibiting inward $K^+$ channels, thus reducing $K^+$ influx into
434 guard cells [68,69]. In addition to being involved in drought stress response,
435 this same gene has been recently identified also as related to salinity stress
436 response in *Arabidopsis* by regulating ion homeostasis between root and shoots
437 [70].

438 The chlorophyll alpha oxygenase (Phvul.008G161000), identified as a gene
439 under selection (and a homolog of *Arabidopsis* CAO), has a primary role in the
440 biosynthesis of chlorophyll b [71]. However, *Arabidopsis* mutants for this gene
441 showed a reduction of antioxidant compounds (specifically glutathione) in
442 guard cells and an increased ABA sensitivity in comparison to wild type plants
443 [72], suggesting a possible involvement of this gene in adaptive response to
444 stressful environments.
445 Phvul.002G143100, identified as selected within the different sub-
446 populations of *P. vulgaris,* is annotated as a glycine-rich domain protein,
447 homologous of *Arabidopsis* GRDP2 gene. GRP are a multi-gene superfamily
448 present in several organisms, including plants [73]. This gene family has been
449 associated in plants with several developmental processes and in responses to
450 both biotic and abiotic stresses [74]. A recent study focusing on the
451 characterization of the direct *Arabidopsis* homologs of this gene (ATGRDP2)
452 demonstrated that this gene regulates plant growth and flowering by
453 accumulating higher level of indole-3-acetic acid and improves abiotic stress
454 response [75]. In particular, the over-expression of this gene in transgenic
455 plants increased growth rate and reduced days to flowering. It also increased
456 salt tolerance in comparison to wild-type plants.
457 In addition to the previous genes identified as selected by genome scan
458 analysis and putatively involved in environmental response in plants, was
459 Phvul.009G050600, which is annotated as an importin β-protein homologous to
460 *Arabidopsis* KPNB1. This gene mediates the import of proteins and protein
461 complexes between the cytoplasm and the nucleus and is essential in
462 regulating signal transduction pathways in response to environmental and

developmental stimuli [76]. In particular, the *Arabidopsis* homolog of Phvul.009G050600 (AtKPNB1) has been directly related to ABA and drought response previously [77].

**Genome wide association analysis**

Association analysis between genotypic data and bio-climatic variables identified several genes significantly associated with one or more bio-climatic variables, putatively involved in plant development, ion homeostasis, and stress response. Among these genes, several could be useful as potential molecular markers for improving abiotic stress in domesticated common bean. As examples, we identified a gene related to potassium homeostasis and annotated as a $K^+$ efflux antiporter (KEA) gene associated with Bio12 (Annual Precipitation) and Bio7 (Temperature Annual Range). Potassium is an essential macronutrient involved in several physiological and developmental processes in all living organism, and in plants this cation is also essential in maintaining plant osmotic potential, cytosolic pH, and stomata movement [78,79]. In addition, variation in $K^+$ homeostasis is one of the first response to several abiotic and biotic stresses in plants, allowing the plants to rapidly respond to stressful conditions [80] and making the KEA gene identified in the current study an interesting candidate gene for further analysis.

Another gene, significantly associated with Bio14 (Precipitation of Driest Month) is Phvul.010G155000, which is annotated as a phospholipase (PLDα1). These genes are involved in the biosynthesis of phosphatidic acid (PA), which is an important signaling molecule in response to several stresses in plants [81]. In particular, PA is involved in the ABA signaling cascade and regulates stomata

closure in plants by directly interacting and blocking ABI1, an inhibitor of ABA response in plants [82]. This gene regulates stomatal closure and ABA-dependent hydrogen peroxide ($H_2O_2$) production in *Vicia faba* as well [83], making this gene an interesting candidate for improving drought response in common bean.

An additional gene, significantly associated with Bio5 (Max Temperature of Warmest Month), is Phvul.010G035200, annotated as a cytokinin response factor homologous of *Arabidopsis* CRF4. Cytokinin is an essential plant hormone involved in growth and developmental processes [84,85], but in recent years it has also been implicated in the response and adaptation to different environmental stresses [86,87]. CRF genes are a class of plant transcription factors responsive to cytokinin that integrate hormonal and environmental signals for adapting plant growth and development in response to the environment [88,89]. The *Arabidopsis* homolog of this gene has been previous related to acclimation to cold temperatures [90]. Since this gene has been associated with temperature variables in wild common bean, it could also be involved in adaptation to temperature variation in this species.

Another gene of interest, Phvul.008G161700, is significantly associated with Bio3 (Isothermality) and is annotated as a thioredoxin protein. These proteins are involved in the regulation of oxidative stress response and in scavenging reactive oxygen species (ROS) in plants [91]. Other than being simple byproducts of cellular metabolism, ROS molecules has been recognized as important signaling molecules that regulate the response to several environmental stresses in plants [92,93]. Due to their ability to control the redox state of the cell, thioredoxin represents a key component of the ROS

513 signal transduction pathways in plants and in the response to environmental
514 stress [94]. Thus, this gene could constitute another interesting candidate gene
515 for improving stress resistance in domesticated common bean.
516
517 **Comparison of genes identified by genome scan and GWAS**
518 Even though the genes identified by outlier-detection methods (hard-selection
519 sweeps) and association methods (soft-selection sweeps) are involved in
520 similar processes, there is no overlap between the candidate genes identified
521 by the two approaches. This could be the direct results of the different
522 assumptions underlying these methods. Indeed, genome scan analysis identify
523 genes that shows drastic variations of allele frequencies between natural
524 subpopulations [14,17]. This approach is independent from bio-climatic
525 variables, thus the SNPs identified as under selection by this analysis could be
526 the results of selective mechanisms not taken into account by association
527 analysis, like soil composition, pathogen pressure and/or competition with other
528 plants. On the other hand, association analyses identify SNPs showing slight
529 variations in allele frequencies across environmental gradients that can
530 increase environmental adaptation in natural populations [14,17]. This
531 selection process usually acts on natural standing variations and favor the
532 presence of multiple alleles and haplotypes, instead of allele fixation within
533 populations [95].
534
535 **Conclusions**
536 In conclusion, landscape genomic analysis of wild common bean genotypes
537 allowed us to identify several genes showing a signature of presumed selection

in this species. It is likely that two methods – genome scan and GWAS - are indeed complementary for understanding local adaptation in wild plant populations, as observed previously in other species [49,96] and are a feasible approach for the preliminary identification of novel candidate genes for adaptation to climatic differences along the exceptionally broad habitat of wild common bean. Further corroboration of the actual role of the candidate genes in adaptation will come from introgression of these genes from wild to domesticated beans and a concurrent phenotypic analysis showing improved performance under stress conditions.

**Abbreviations**

CWRs: Crop Wild Relatives; GWAS: Genome Wide Association Studies; SNPs: Single Nucleotide polymorphisms; GBS: Genotyping-By-Sequencing; MW: Mesoamerican wild gene pool; AW: Andean wild gene pool; PhI: Northern Peru-Ecuador gene pool; PCA: Principal Component Analysis; PET: Potential EvapoTranspiration; LFMM: Latent Factor Mixed Models; IBD: isolation-by-distance; MCMC: Monte Carlo Markov Chain; ABA: abscisic acid; GRP: glycine-rich domain protein; PLDα1: phospholipase D α 1; PA: phosphatidic acid; ROS: reactive oxygen species

**Acknowledgements**

This work used the Vincent J. Coates Genomics Sequencing Laboratory at UC Berkeley, supported by NIH S10 Instrumentation Grants S10RR029668 and S10RR027303. We thank the Genetic Resources Unit of the Centro Internacional de Agricultura Tropical (CIAT, Cali, Colombia) and the Western Regional Plant

Introduction Station of the USDA (Pullman, WA) for providing samples of wild *P. vulgaris* used in this study.

**Funding**

This project was supported by Agriculture and Food Research Initiative (AFRI) Competitive Grant No. 2013-67013-21224 from the USDA National Institute of Food and Agriculture.

**Availability of Data and Materials**

Raw sequencing data are available at the NCBI Sequence Read Archive (http://www.ncbi.nlm.nih.gov/sra) under the accession numbers SRX2771627 and SRX2771628. The variants file and the relative geographical coordinates used for performing the analysis are available as additional files of the current manuscript.

**Authors' contributions**

AA performed the experiment, analyzed the data and wrote the manuscript. PG designed the experiment, supervised the work and wrote the manuscript. All authors approved the final version of the manuscript.

**Ethics approval and consent to participate**

Not applicable

**Consent for publication**

Not applicable

## Competing interests

The authors declare that they have no competing interests.

## References


1. Rippke U, Ramirez-Villegas J, Jarvis A, Vermeulen SJ, Parker L, Mer F, et al. Timescales of transformational climate change adaptation in sub-Saharan African agriculture. Nat Clim Change. 2016;6:605–9.

2. Campbell BM, Vermeulen SJ, Aggarwal PK, Corner-Dolloff C, Girvetz E, Loboguerrero AM, et al. Reducing risks to food security from climate change. Glob Food Secur. 2016;11:34–43.

3. Wheeler T, Braun J von. Climate change impacts on global food security. Science. 2013;341:508–13.

4. Challinor AJ, Watson J, Lobell DB, Howden SM, Smith DR, Chhetri N. A meta-analysis of crop yield under climate change and adaptation. Nat Clim Change. 2014;4:287–91.

5. Lobell DB, Burke MB, Tebaldi C, Mastrandrea MD, Falcon WP, Naylor RL. Prioritizing climate change adaptation needs for food security in 2030. Science. 2008;319:607–10.

6. IPCC. Managing the risks of extreme events and disasters to advance climate change adaptation: special report of the Intergovernmental Panel on Climate Change. Cambridge University Press; 2012.

7. Ford-Lloyd BV, Schmidt M, Armstrong SJ, Barazani O, Engels J, Hadas R, et al. Crop wild relatives—undervalued, underutilized and under threat? BioScience. 2011;61:559–65.

8. Zamir D. Improving plant breeding with exotic genetic libraries. Nat Rev

613 Genet. 2001;2:983–9.

614 9. Gepts P. The contribution of genetic and genomic approaches to plant
615 domestication studies. Curr Opin Plant Biol. 2014;18:51–9.

616 10. Spillane C, Gepts P. Evolutionary and genetic perspectives on the dynamics
617 of crop genepools. (2001); Available from:
618 http://agris.fao.org/agris-search/search.do?recordID=XF2003411459

619 11. Brozynska M, Furtado A, Henry RJ. Genomics of crop wild relatives:
620 expanding the gene pool for crop improvement. Plant Biotechnol J.
621 2016;14:1070–85.

622 12. Cortés AJ, Monserrate FA, Ramírez-Villegas J, Madriñán S, Blair MW. Drought
623 tolerance in wild plant populations: the case of common beans (*Phaseolus*
624 *vulgaris* L.). PLoS ONE. 2013;8:e62898.

625 13. Zhang H, Mittal N, Leamy LJ, Barazani O, Song B. Back into the wild—Apply
626 untapped genetic diversity of wild relatives for crop improvement. Evol Appl.
627 2016;10:5–24.

628 14. Schoville SD, Bonin A, François O, Lobreaux S, Melodelima C, Manel S.
629 Adaptive genetic variation on the landscape: methods and cases. Annu Rev
630 Ecol Evol Syst. 2012;43:23–43.

631 15. Bragg JG, Supple MA, Andrew RL, Borevitz JO. Genomic variation across
632 landscapes: insights and applications. New Phytol. 2015;207:953–67.

633 16. Anderson JE, Kono TJY, Stupar RM, Kantar MB, Morrell PL. Environmental
634 association analyses identify candidates for abiotic stress tolerance in *Glycine*

635 *soja*, the wild progenitor of cultivated soybeans. G3 GenesGenomesGenetics.
636 2016;6:835–43.

637 17. Wagner HH, Fortin M-J. A conceptual framework for the spatial analysis of
638 landscape genetic data. Conserv Genet. 2012;14:253–61.

639 18. Wright S. The genetical structure of populations. Ann Eugen. 1949;15:323–
640 54.

641 19. Lewontin RC, Krakauer J. Distribution of gene frequency as a test of the
642 theory of the selective neutrality of polymorphisms. Genetics. 1973;74:175–95.

643 20. Haldane JBS. A mathematical theory of natural and artificial selection. (Part
644 VI, Isolation.). Math Proc Camb Philos Soc. 1930;26:220–30.

645 21. Kimura M. On the probability of fixation of mutant genes in a population.
646 Genetics. 1962;47:713–9.

647 22. Narum SR, Hess JE. Comparison of F(ST) outlier tests for SNP loci under
648 selection. Mol Ecol Resour. 2011;11 Suppl 1:184–94.

649 23. Manel S, Joost S, Epperson BK, Holderegger R, Storfer A, Rosenberg MS, et
650 al. Perspectives on the use of landscape genetics to detect genetic adaptive
651 variation in the field. Mol Ecol. 2010;19:3760–72.

652 24. Hirschhorn JN, Daly MJ. Genome-wide association studies for common
653 diseases and complex traits. Nat Rev Genet. 2005;6:95–108.

654 25. Lipka AE, Kandianis CB, Hudson ME, Yu J, Drnevich J, Bradbury PJ, et al. From
655 association to prediction: statistical methods for the dissection and selection of

656 complex traits in plants. Curr Opin Plant Biol. 2015;24:110–8.

657 26. Gepts P, Aragão FJL, Barros E de, Blair MW, Brondani R, Broughton W, et al.
658 genomics of *Phaseolus* beans, a major source of dietary protein and
659 micronutrients in the Tropics. In: Moore PH, Ming R, editors. Genomics Trop Crop
660 Plants. Springer New York; 2008 p. 113–43.

661 27. Messina V. Nutritional and health benefits of dried beans. Am J Clin Nutr.
662 2014;100 Suppl 1:437S–42S.

663 28. Rubiales D, Mikic A. Introduction: legumes in sustainable agriculture. Crit
664 Rev Plant Sci. 2015;34:2–3.

665 29. Chacón SMI, Pickersgill B, Debouck DG, Arias JS. Phylogeographic analysis
666 of the chloroplast DNA variation in wild common bean (*Phaseolus vulgaris* L.) in
667 the Americas. Plant Syst Evol. 2007;266:175–95.

668 30. Koenig R, Gepts P. Allozyme diversity in wild *Phaseolus vulgaris*: further
669 evidence for two major centers of genetic diversity. Theor Appl Genet.
670 1989;78:809–17.

671 31. Mamidi S, Rossi M, Moghaddam SM, Annam D, Lee R, Papa R, et al.
672 Demographic factors shaped diversity in the two gene pools of wild common
673 bean *Phaseolus vulgaris* L. Heredity. 2013;110:267–76.

674 32. Bitocchi E, Bellucci E, Giardini A, Rau D, Rodriguez M, Biagetti E, et al.
675 Molecular analysis of the parallel domestication of the common bean
676 (*Phaseolus vulgaris*) in Mesoamerica and the Andes. New Phytol.
677 2013;197:300–13.

678 33. Blair MW, Soler A, Cortés AJ. Diversification and population structure in
679 common beans (*Phaseolus vulgaris* L.). PLOS ONE. 2012;7:e49488.

680 34. Chacón SMI, Pickersgill B, Debouck DG. Domestication patterns in common
681 bean (*Phaseolus vulgaris* L.) and the origin of the Mesoamerican and Andean
682 cultivated races. Theor Appl Genet. 2005;110:432–44.

683 35. Gepts P, Osborn TC, Rashka K, Bliss FA. Phaseolin-protein variability in wild
684 forms and landraces of the common bean (*Phaseolus vulgaris*): evidence for
685 multiple centers of domestication. Econ Bot. 1986;40:451–68.

686 36. Mamidi S, Rossi M, Annam D, Moghaddam S, Lee R, Papa R, et al.
687 Investigation of the domestication of common bean (*Phaseolus vulgaris*) using
688 multilocus sequence data. Funct Plant Biol. 2011;38:953–67.

689 37. Rossi M, Bitocchi E, Bellucci E, Nanni L, Rau D, Attene G, et al. Linkage
690 disequilibrium and population structure in wild and domesticated populations of
691 *Phaseolus vulgaris* L. Evol Appl. 2009;2:504–22.

692 38. Singh SP, Gepts P, Debouck DG. Races of common bean (*Phaseolus*
693 *vulgaris*, Fabaceae). Econ Bot. 1991;45:379–96.

694 39. Debouck DG, Toro O, Paredes OM, Johnson WC, Gepts P. Genetic diversity
695 and ecological distribution of *Phaseolus vulgaris* (fabaceae) in Northwestern
696 South America. Econ Bot. 1993;47:408–23.

697 40. Kami J, Velásquez VB, Debouck DG, Gepts P. Identification of presumed
698 ancestral DNA sequences of phaseolin in *Phaseolus vulgaris*. Proc Natl Acad Sci.
699 1995;92:1101–4.

41. Rendón-Anaya M, Herrera-Estrella A, Gepts P, Delgado-Salinas A. A new species of *Phaseolus* (Leguminosae, Papilionoideae) sister to *Phaseolus vulgaris*, the common bean. Phytotaxa. 2017;313:259–66.

42. Rendón-Anaya M, Montero-Vargas JM, Saburido-Álvarez S, Vlasova A, Capella-Gutierrez S, Ordaz-Ortiz JJ, et al. Genomic history of the origin and domestication of common bean unveils its closest sister species. Genome Biol. 2017;18:60.

43. Gepts P. Origin and evolution of common bean: past events and recent trends. HortScience. 1998;33:1124–30.

44. Porch TG, Beaver JS, Debouck DG, Jackson SA, Kelly JD, Dempewolf H. Use of wild relatives and closely related species to adapt common bean to climate change. Agronomy. 2013;3:433–61.

45. Acosta-Gallegos JA, Kelly JD, Gepts P. Prebreeding in common bean and use of genetic diversity from wild germplasm. Crop Sci. 2007;47:S-44.

46. Ramirez-Cabral NYZ, Kumar L, Taylor S. Crop niche modeling projects major shifts in common bean growing areas. Agric For Meteorol. 2016;218–219:102–13.

47. Westengen OT, Berg PR, Kent MP, Brysting AK. Spatial structure and climatic adaptation in African maize revealed by surveying SNP diversity in relation to global breeding and landrace panels. PloS One 2012;7:e47832.

48. Yoder JB, Stanton-Geddes J, Zhou P, Briskine R, Young ND, Tiffin P. Genomic signature of adaptation to climate in *Medicago truncatula*. Genetics

722 2014;196:1263–75.

723 49. Dell’Acqua M, Zuccolo A, Tuna M, Gianfranceschi L, Pè ME. Targeting
724 environmental adaptation in the monocot model Brachypodium distachyon: a
725 multi-faceted approach. BMC Genomics 2014;15:801.

726 50. Abebe TD, Naz AA, Léon J. Landscape genomics reveal signatures of local
727 adaptation in barley (*Hordeum vulgare* L.). Front Plant Sci. 2015;6.

728 51. Ariani A, Berny Mier Y Teran J, Gepts P. Spatial and temporal scales of range
729 expansion in wild *Phaseolus vulgaris*. Mol Biol Evol. 2017;35:119-131.

730 52. Rodriguez M, Rau D, Bitocchi E, Bellucci E, Biagetti E, Carboni A, et al.
731 Landscape genetics, adaptive diversity and population structure in *Phaseolus*
732 *vulgaris*. New Phytol. 2016;209:1781–94.

733 53. Wehrens R. Chemometrics with R. Berlin, Heidelberg: Springer Berlin
734 Heidelberg; 2011

735 54. Luu K, Bazin E, Blum MGB. pcadapt: an R package to perform genome
736 scans for selection based on principal component analysis. bioRxiv.
737 2016;056135.

738 55. Frichot E, Schoville SD, Bouchard G, François O. Testing for associations
739 between loci and environmental gradients using latent factor mixed models.
740 Mol Biol Evol. 2013;30:1687–99.

741 56. Frichot E, François O. LEA: An R package for landscape and ecological
742 association studies. Methods Ecol Evol. 2015;6:925–9.

743 57. Devlin B, Roeder K. Genomic control for association studies. Biometrics.
744 1999;55:997–1004.

745 58. Lawrence M, Gentleman R, Carey V. rtracklayer: an R package for
746 interfacing with genome browsers. Bioinformatics. 2009;25:1841–2.

747 59. Lawrence M, Huber W, Pagès H, Aboyoun P, Carlson M, Gentleman R, et al.
748 Software for computing and annotating genomic ranges. PLoS Comput Biol.
749 2013;9:e1003118.

750 60. Li X, Zhu C, Yeh C-T, Wu W, Takacs EM, Petsch KA, et al. Genic and nongenic
751 contributions to natural variation of quantitative traits in maize. Genome Res.
752 2012;22:2436–44.

753 61. Pedregosa F, Varoquaux G, Gramfort A, Michel V, Thirion B, Grisel O, et al.
754 Scikit-learn: machine learning in Python. J Mach Learn Res.
755 2011;12:2825−2830.

756 62. Danecek P, Auton A, Abecasis G, Albers CA, Banks E, DePristo MA, et al. The
757 variant call format and VCFtools. Bioinforma Oxf Engl. 2011;27:2156–8.

758 63. Bitocchi E, Nanni L, Bellucci E, Rossi M, Giardini A, Zeuli PS, et al.
759 Mesoamerican origin of the common bean (*Phaseolus vulgaris* L.) is revealed by
760 sequence data. Proc Natl Acad Sci. 2012;109:E788–96.

761 64. Kwak M, Gepts P. Structure of genetic diversity in the two major gene pools
762 of common bean (*Phaseolus vulgaris* L., Fabaceae). Theor Appl Genet.
763 2009;118:979–92.

764 65. Osakabe Y, Arinaga N, Umezawa T, Katsura S, Nagamachi K, Tanaka H, et

765 al. Osmotic stress responses and plant growth controlled by potassium
766 transporters in Arabidopsis. Plant Cell 2013;25:609–24.

767 66. Bellucci E, Bitocchi E, Ferrarini A, Benazzo A, Biagetti E, Klie S, et al.
768 Decreased nucleotide and expression diversity and modified coexpression
769 patterns characterize domestication in the common bean. Plant Cell Online.
770 2014;26:1901–12.

771 67. Bitocchi E, Rau D, Bellucci E, Rodriguez M, Murgia ML, Gioia T, et al. Beans
772 (*Phaseolus* ssp.) as a model for understanding crop evolution. Front Plant Sci.
773 2017;8.

774 68. Geiger D, Maierhofer T, AL-Rasheid KAS, Scherzer S, Mumm P, Liese A, et al.
775 Stomatal closure by fast abscisic acid signaling is mediated by the guard cell
776 anion channel SLAH3 and the receptor RCAR1. Sci Signal. 2011;4:ra32-ra32.

777 69. Zhang A, Ren H-M, Tan Y-Q, Qi G-N, Yao F-Y, Wu G-L, et al. S-type anion
778 channels SLAC1 and SLAH3 Function as essential negative regulators of inward
779 k+ channels and stomatal opening in Arabidopsis. Plant Cell Online
780 2016;28:949–65.

781 70. Cubero-Font P, Maierhofer T, Jaslan J, Rosales MA, Espartero J, Díaz-Rueda P,
782 et al. Silent S-type anion channel subunit SLAH1 gates SLAH3 open for chloride
783 root-to-shoot translocation. Curr Biol. 2016;26:2213–20.

784 71. Espineda CE, Linford AS, Devine D, Brusslan JA. The AtCAO gene, encoding
785 chlorophyll a oxygenase, is required for chlorophyll b synthesis in *Arabidopsis*
786 *thaliana*. Proc Natl Acad Sci. 1999;96:10507–11.

787 72. Jahan MS, Nozulaidi M, Khairi M, Mat N. Light-harvesting complexes in
788 photosystem II regulate glutathione-induced sensitivity of Arabidopsis guard
789 cells to abscisic acid. J Plant Physiol. 2016;195:1–8.

790 73. Sachetto-Martins G, Franco LO, de Oliveira DE. Plant glycine-rich proteins: a
791 family or just proteins with a common motif? Biochim Biophys Acta
792 2000;1492:1–14.

793 74. Mangeon A, Junqueira RM, Sachetto-Martins G. Functional diversity of the
794 plant glycine-rich proteins superfamily. Plant Signal Behav. 2010;5:99–104.

795 75. Ortega-Amaro MA, Rodríguez-Hernández AA, Rodríguez-Kessler M,
796 Hernández-Lucero E, Rosales-Mendoza S, Ibáñez-Salazar A, et al.
797 Overexpression of AtGRDP2, a novel glycine-rich domain protein, accelerates
798 plant growth and improves stress tolerance. Front Plant Sci. 2015;5.

799 76. Merkle T. Nucleo-cytoplasmic partitioning of proteins in plants: implications
800 for the regulation of environmental and developmental signalling. Curr Genet.
801 2003;44:231–60.

802 77. Luo Y, Wang Z, Ji H, Fang H, Wang S, Tian L, et al. An Arabidopsis homolog
803 of importin β1 is required for ABA response and drought tolerance. Plant J.
804 2013;75:377–89.

805 78. Shabala S. Regulation of Potassium Transport in Leaves: from Molecular to
806 Tissue Level. Ann Bot. 2003;92:627–34.

807 79. Sharma T, Dreyer I, Riedelsberger J. The role of K+ channels in uptake and
808 redistribution of potassium in the model plant Arabidopsis thaliana. Front Plant

809 Sci. 2013;4.

810 80. Shabala S, Pottosin I. Regulation of potassium transport in plants under
811 hostile conditions: implications for abiotic and biotic stress tolerance. Physiol
812 Plant. 2014;151:257–79.

813 81. Saucedo-García M, Gavilanes-Ruíz M, Arce-Cervantes O. Long-chain bases,
814 phosphatidic acid, MAPKs, and reactive oxygen species as nodal signal
815 transducers in stress responses in Arabidopsis. Front Plant Sci. 2015;6.

816 82. Zhang W, Qin C, Zhao J, Wang X. Phospholipase Dα1-derived phosphatidic
817 acid interacts with ABI1 phosphatase 2C and regulates abscisic acid signaling.
818 Proc Natl Acad Sci 2004;101:9508–13.

819 83. Qu Y, An Z, Zhuang B, Jing W, Zhang Q, Zhang W. Copper amine oxidase
820 and phospholipase D act independently in abscisic acid (ABA)-induced stomatal
821 closure in Vicia faba and Arabidopsis. J Plant Res. 2014;127:533–44.

822 84. Durán-Medina Y, Díaz-Ramírez D, Marsch-Martínez N. Cytokinins on the
823 Move. Front Plant Sci. 2017;8.

824 85. Kieber JJ. Cytokinins. Arab Book Am Soc Plant Biol. 2002;1.

825 86. Novakova M, Dobrev P, Motyka V, Gaudinova A, Malbeck J, Pospisilova J, et
826 al. Cytokinin function in drought stress response and subsequent recovery.
827 Biotechnol Sustain Agric 2006 Beyond. Springer, Dordrecht; 2007 p. 171–4.

828 87. O'Brien JA, Benková E. Cytokinin cross-talking during biotic and abiotic
829 stress responses. Front Plant Sci. 2013;4.

830 88. Kim J. CYTOKININ RESPONSE FACTORs gating environmental signals and
831 hormones. Trends Plant Sci. 2016;21:993–6.

832 89. Rashotte AM, Goertzen LR. The CRF domain defines Cytokinin Response
833 Factor proteins in plants. BMC Plant Biol. 2010;10:74.

834 90. Zwack PJ, Compton MA, Adams CI, Rashotte AM. Cytokinin response factor 4
835 (CRF4) is induced by cold and involved in freezing tolerance. Plant Cell Rep.
836 2016;35:573–84.

837 91. Gelhaye E, Rouhier N, Navrot N, Jacquot JP. The plant thioredoxin system.
838 Cell Mol Life Sci 2005;62:24–35.

839 92. D'Autréaux B, Toledano MB. ROS as signalling molecules: mechanisms that
840 generate specificity in ROS homeostasis. Nat Rev Mol Cell Biol. 2007;8:813–24.

841 93. Sewelam N, Kazan K, Schenk PM. Global plant stress signaling: reactive
842 oxygen species at the cross-road. Front Plant Sci. 2016;7.

843 94. Sevilla F, Camejo D, Ortiz-Espín A, Calderón A, Lázaro JJ, Jiménez A. The
844 thioredoxin/peroxiredoxin/sulfiredoxin system: current overview on its redox
845 function in plants and regulation by reactive oxygen and nitrogen species. J Exp
846 Bot. 2015;66:2945–55.

847 95. Hermisson J, Pennings PS. Soft sweeps: molecular population genetics of
848 adaptation from standing genetic variation. Genetics 2005;169:2335–52.

849 96. Pyhäjärvi T, Hufford MB, Mezmouk S, Ross-Ibarra J. Complex patterns of
850 local adaptation in teosinte. Genome Biol Evol. 2013;5:1594–609.

851 97. Schmutz J, McClean PE, Mamidi S, Wu GA, Cannon SB, Grimwood J, et al. A
852 reference genome for common bean and genome-wide analysis of dual
853 domestications. Nat Genet. 2014;46:707–13.

## Figure legends

**Figure 1** Bio-climatic data analysis. **(A)** Loading plot of the PCA analysis. **(B)** Principal Component Analysis (PCA) of the bio-climatic data. Groups are colored according to the K-mean clustering analysis conducted in this study, which gave results very similar to the STRUCTURE analysis conducted by Ariani et al. (2016): MW1, MW2, and MW3: Mesoamerican wild gene pools; AW: Andean wild gene pool; PhI: Intermediate wild gene pool.

**Figure 2** PCA analysis on molecular data. **(A)** Screeplot of the PCA explained variance. **(B)** PCA plot based on molecular data of the different genotyped analyzed in the current study. MW: Mesoamerican Wild; AW: Andean Wild; PhI: Intermediate gene pool.

**Figure 3** Three-dimensional plot of the PCA analysis on molecular data. Points are colored as in Figure 2A.

**Figure 4** Manhattan plot of the genome scan data with 5 sub-populations (K). The blue dashed line represents the significance threshold (Bonferroni p-value $\leq 0.001$).

**Figure 5** Chromosome ideogram of the genes identified as associated with the bio-climatic variables analyzed. Only chromosomes with significantly associated variants are shown. Each circle represents a different bio-climatic variable. When available, gene annotations are shown. The centromeric regions shown are based on the results from [97].

**Figure 6** Allele frequency distribution across different genetic groups for candidate genes identified by genome scan analysis. *P. vulgaris v1.0* genes annotation and ID: **(A)** Potassium uptake transporter (Phvul.002G331700); **(B)** Glycine-rich domain protein (Phvul.002G143100); **(C)** ABA response

(Phvul.004G102800); **(D)** Chlorophyll biosynthesis and ABA signaling (Phvul.008G161000); **(E)** ABA and drought response (Phvul.009G050600). For panel **(E)** the PhI group was removed due to complete missing data on the SNPs. REF: Reference allele, ALT: Alternative allele, according to the *P. vulgaris* v1.0 gene version.

**Figure 7** Allele frequency distribution across different genetic groups for candidate genes identified by association analysis. *P. vulgaris v1.0* genes annotation and ID: **(A)** Potassium efflux antiporter (Phvul.001G034400); **(B)** Phospholipase D α 1 (Phvul.010G155000); **(C)** Cytokinin responsive factor (Phvul.010G035200); **(D)** Thioredoxin (Phvul.008G161700). Reference and Alternative alleles are colored as in Figure 6.

**Additional Files**

**Additional File 1 (Additional_File_1.csv):** List of the *P. vulgaris* accessions sequenced in the current study, with gene pool information, geographic coordinates, gene pools and genetic group information.

**Additional File 2 (Additional_File_2.vcf):** Genotypic data of the *P. vulgaris* accessions analyzed in the current study

**Additional File 3 (Additional_File_3.csv):** Eigenvalues of the different bioclimatic variables along the first four principal components.

**Additional File 4 (Additional_File_4.pdf):** Correlation graphs between bio-climatic variables for the different *P. vulgaris* accessions analyzed. Correlation coefficients are rendered using circles (upper-right part) or by showing directly the value (lower-left part). Color are based on color-bar in the right side of the graph.

**Additional File 5 (Additional_File_5.pdf):** Plots of the inflation factor for different values of K across the climatic variables selected for association study

**Additional File 6 (Additional_File_6.pdf):** Cumulative variance explained by the different PCs when performing a PCA on bio-climatic variables

**Additional File 7 (Additional_File_7.pdf):** P-values distibution for genome scans with 2 (A), 3 (B), 4 (C) or 5 (D) sub-populations.

**Additional File 8 (Additional_File_8.pdf):** Plot of geographic distribution of the wild *P. vulgaris* analyzed in the current studies. Genotypes are colored based on the different clusters identified by K-means clustering (Additional File 1).