RESEARCH ARTICLE

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

Andrea Ariani[1,2] and Paul Gepts[1]*

1 Department of Plant Sciences / MS1, Section of Crop and Ecosystem Sciences, University of California, Davis, CA 95616-8780

2 Current address: BASF Agricultural Solutions Belgium NV, Technologiepark 101, 9052, Ghent, Belgium

* plgepts@ucdavis.edu

Abstract

Landscape genomics integrates population genetics with landscape ecology, allowing the identification of putative molecular determinants involved in environmental adaptation across the natural geographic and ecological range of populations. Wild *Phaseolus vulgaris*, the progenitor of common bean (*P. vulgaris*), has a remarkably extended distribution over 10,000 km 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. Thus, this species provides an opportunity to examine to what extent adaptation of a species can be broadened or, conversely, ecological or geographical distribution can be limited by inherent adaptedness. 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. The current study demonstrates the feasibility of landscape genomics approach for a preliminary identification of specific populations and 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.

Author Summary

The ancestral form of common bean has an unusually large distribution in the Americas, extending over 10,000 km from ~35° N. Lat. to ~35° S. Lat. This wide distribution results from discrete long-range dissemination events to the Andes region from the original environments in Mesoamerica. It also suggests adaptation to new environments that are distinct from those encountered in Mesoamerica. In this research, we identified genes that may be involved in adaptation to climate variables in these new environments using two methods. A first method – outlier detection – was used to identify genome regions that differentiated the wild bean groups in the Andes resulting from discrete dissemination events among themselves and the different groups in Mesoamerica. The second method – genome-wide association – was used to identify candidate genome regions correlated with these same variables across the entire distribution from Mesoamerica to the southern Andes. The two methods identified two sets of candidate genes, several of which were related to the water status of plants, and illustrate how the genetic architecture of adaptation following long-range dissemination. This study provides sets of candidate genes as well as candidate wild bean populations that need to be corroborated for their use in increasing the water use efficiency of domesticated beans.

Introduction

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 (Rippke et al., 2016; Campbell et al., 2016). 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 (Wheeler and von Braun, 2013; Challinor et al., 2014). 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 (Lobell et al., 2008; Field et al., 2012). 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 (Ford-Lloyd et al., 2011). The genetic variability available in domesticated plants is generally low due to the bottleneck effect induced by domestication and subsequent selection during variety improvement (Ford-Lloyd et al., 2011; Zamir 2001; 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 (Ford-Lloyd et al., 2011; Zamire 2001; Gepts 2014; Spillane and Gepts 2001; Brozynska et al., 2016). 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, difficulties in phenotyping a large number of individuals under agricultural conditions, and the existence of linkages between target resistance genes and unfavorable loci subject to linkage drag (Brozynska et al., 2016; Cortés et al., 2013; Zhang et al., 2017). One possible solution for overcoming the

first two 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' (Schoville et al., 2012; Bragg et al., 2015). 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 germplasm accessions and molecular markers that could be readily applicable – pending validation - for breeding improvement of domesticated plants (Anderson et al., 2016).

Several methods have been developed for identifying signatures 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, which identify soft-selection sweeps (Schoville et al., 2012; Wagner and Fortin, 2013). 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} (Wright, 1949; Lewontin and Krakauer, 1973). Although based on the assumption that alleles fixed within sub-populations could confer an evolutionary advantage in the ecological niche occupied (Haldane, 1930; Kimura, 1962), these methods do not take directly into account climatic data and could be biased by complex population structure and/or demography (Narum and Hess, 2011). 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 (Manel et al., 2010). The theory beneath environmental association methods are practically the same as that used in Genome Wide Association Studies (GWAS) (Hirschhorn and Daly, 2005). Both approaches employ mixed model association approaches for correcting the confounding effects that could be introduced by population structure and relatedness in the sample (Lipka et al., 2015).

Common bean (*Phaseolus vulgaris* L.) is an essential staple crop providing most of proteins and micronutrients in the diet of the majority of the population in several developing countries (Gepts et al., 2008). The regular consumption of this crop provides several health benefits, like reducing the risks of heart disease, obesity, and diabetes (Messina, 2014). Its cultivation improves agricultural sustainability thanks to its nitrogen-fixing ability (Rubiales and Mikić, 2015). 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) (Chacón et al., 2007; Koenig and Gepts, 1989; Debouck et al., 1993; Mamidi et al., 2013). Common bean was domesticated independently in Mexico and the Southern Andes, producing locally-adapted varieties and landraces with specific characteristics (Bitocchi et al., 2013; Blair et al., 2012; Gepts et al., 1986, Mamidi et al., 2011, Rossi et al., 2009, Singh et al., 1991). The intermediate gene pool in the Central Andes was not domesticated (Debouck et al., 1993; Kami et al., 1995). This wild group has been recently identified as a cryptic sister species of *P. vulgaris,* named *Phaseolus debouckii* A. Delgado, 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 (Rendón-Anaya et al., 2017a,b).

Wild common bean is an annual vine plant, which is distributed 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 (Gepts, 1998; Porch et al., 2013). 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 (Cortés et al., 2013, Gepts, 1998, Porch et al., 2013). 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 (Porch et al., 2013, Acosta-Gallegos et al., 2007).

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 (Ramirez-Cabral et al., 2016). 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* (Anderson et al., 2016, Westengen et al., 2012, Yoder et al., 2014, Dell'Acqua et al., 2014, Abebe et al., 2015).

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 (Ariani et al., 2018). A similar analysis was performed previously in this species using 148 SNPs located in genes putatively involved in adaptation to biotic or abiotic stresses (Rodriguez et al., 2016). 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.

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) (**S1 Fig**).

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 1B**), while PC1 to PC4 explained cumulatively > 90% of the variance (**S2 Fig**). A plot of PC1 vs. PC2

163 showed some differences in the distribution of the different gene pools of wild common bean in
164 the PC dimensional space. In particular, the majority of genotypes from the Mesoamerican (MW1
165 to MW3) and Intermediate (PhI) gene pools were distributed towards the positive part of PC1,
166 while the Andean group were located in the negative part of this axis (**Fig. 1A**). Given the origin
167 of the genus *Phaseolus* in the Mesoamerican area (with local descendants represented by MW1
168 and MW2), three range expansions characterize this species: 1) PhI, which established wild
169 populations on the western slope of the Andes in Ecuador and northern Peru; 2) AW,
170 encompassing wild populations in the southern Andes; and 3) MW3, a more recent and perhaps
171 ongoing dissemination to Central America and the eastern slope of the northern Andes (Ariani et
172 al., 2018). Inspection of **Fig 1A** and **S3 Table** shows that the distribution of the PhI group, which
173 resulted from the earliest range expansion event, correlates - on bioPC3 - with Isothermality
174 (bio_3), Temperature Seasonality (bio_4), bio_13 (Precipitation of the Wettest Month), and
175 bio_18 (Precipitation of the Warmest Quarter), consistent with a dispersal to an equatorial region.
176 In contrast, the predominant distribution of the southern Andean accessions (AW) in the upper
177 left quadrant of **Fig 1** is consistent with earlier observations that the populations of this gene pool
178 are distributed in cooler and drier locations, as shown by correlations with bio_6 (Minimum
179 Temperature of the Coldest Month), bio_9 (Mean Temperature of the Driest Quarter, bio_11
180 (Mean Temperature of the Coldest Quarter) and bio_1 (Annual Mean Temperature). This
181 dissemination occurred with a concomitant lower potential evapotranspiration (Ariani et al.
182 2018). Dispersal of the MW3 group (**Fig 1**) increased Isothermality (bio_3) and decreased
183 Seasonality (bio_4) and Precipitation Seasonality (bio_15); it also increased Precipitation during
184 the Driest Month (bio_14) and Driest Quarter (bio_17).
185

Genome scan of selection

An analysis of the scree plot of the PCA analysis conducted on SNP data (molecular PCA) showed that a quarter of the variance could be explained by the first principal component, even though molPC2 to molPC5 also explained a considerable amount of variance in the data (**Fig 2A)**. On the other hand, after molPC5, no large increase in the cumulative 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 (Ariani et al., 2018). 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 (**S3 Fig**). 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 groups were mostly localized on the positive part of the molecular PC1 (molPC1) axis, while the AW gene pool was localized towards the negative end of molPC1. Interestingly, molPC1 mostly differentiated MW vs. AW, while molPC2 and molPC3 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 (**S1 Table**). 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).

Of the 70 genes identified by the analysis, 20 of them (28%) were located within a haplotype block (**S1 Table**). When mapping the significant SNPs identified by genome scan analysis to the latest reference genome (v2.1), 62 genes (88%) were confirmed as putatively under selection also in this assembly (**S1 Table**). Interestingly, among the genes not identified in v2.1, three were not present in the annotation file, while one gene (Phvul.001G080400) was renamed Phvul.L006501 and was located to a scaffold instead of chromosome 1.

Among the genes identified, we found several related to drought and/or abscisic acid (ABA) response. Phvul.002G331700, a homolog of the *Arabidopsis* KUP6, is involved in potassium uptake transporter and stomata movement and Phvul.002G143100 is a glycine-rich domain protein (GRP) involved in auxin signaling and stress response. Phvul.004G102800 is a homolog to *Arabidopsis* SLAH3 involved in ABA response; Phvul.008G161000 is a homolog of *Arabidopsis* CAO, a gene related to chlorophyll biosynthesis and ABA signaling; and Phvul.009G050600 is 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 bio_18 variable (Precipitation of Warmest Quarter), 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 bio_3 (Isothermality) with 29 genes, and bio_12 (Annual precipitation) with 11 genes (**S2 Table**).

The associated genes were located in all 11 common-bean reference genome chromosomes, except for chromosome Pv06 where there were no significantly associated SNPs. 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.

Of these 49 genes identified by genome-wide association analysis, only 10 (20%) were located within a haplotype block (**S2 Table**). In addition, when mapping significant SNPs identified by association analysis to the latest reference genome (v2.1), 44 genes (88%) were confirmed as putatively associated with environmental variables also in this assembly (**S2 Table**). Four out of five of the missing genes were not present in the v2.1 annotation file.

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 (**S2 Table**). 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. Interestingly, there was no overlap between the genes identified by genome scan and association analysis.

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 (**S4 Fig**). Interestingly, the clustering results closely resembled those obtained in a previous study with more advanced population structure approaches (**S3 Table**) (Ariani et al., 2018).

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 were observed 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 among 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 (Cortés et al., 2013, Gepts 1998, Porch et al., 2013). 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 (Chacón et al., 2007, Koenig and Gepts, 1989, Mamidi et al., 2013, Rendón-Anaya et al., 2017a,b, Bitocchi et al., 2012, Kwak and Gepts, 2009), 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 (Anderson et al., 2016). 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. The concordance between the genes tagged by significant SNPs identified between the two different reference genomes assembly suggests also high concordance between the two versions, with possible minimal differences due to the different sequencing data used for the assembly. In addition, the integration of haploblocks information with the genes identified by our analysis showed that most of these genes (70-80%) are located in regions with low LD in wild common bean. This result suggests that those same genes could be located in regions with relatively high recombination frequencies, thus facilitating possible introgression into the domesticated gene pool.

Genome scan of selection

Molecular PCA analysis clearly separated the three major groups of this species (MW1-MW3, PhI, and AW), as observed in previous research. In particular, the Intermediate gene pool (PhI) was shown again to most diverged group from the Mesoamerican and Andean gene pools, especially along the molPC2 and molPC3 axes (**Fig 2**), further supporting the hypothesis that this gene pool is actually a distinct species of *Phaseolus* (Rendón-Anaya et al., 2017a,b). A genome scan based on molecular 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 several genes involved in plant development and hormone and stress response, suggests that the different populations of this species adapted to their environment by integrating and adjusting to developmental, hormonal and environmental cues. Several genes among those identified could reflect adaptation to abiotic stress. These genes are also of interest for improving stress resilience in common bean, like the KUP6 potassium (K^+) transporter located on chromosome Pv02 (Phvul.002G331700). This gene has been directly linked to drought stress by regulating ABA response and stomata movements in *Arabidopsis* (Osakabe et al., 2013). A homolog of this gene located on chromosome Pv03 (Phvul.003G052900) showed a higher genetic and transcriptional diversity in Mesoamerican domesticated beans than in wild ones (Bellucci et al., 2014; Bitocchi et al., 2017). Due to the possible role of KUP-like genes in response to drought stress and their identification as selected genes in both wild and domesticated populations of common bean, further studies should focus on the evolution and diversity of this gene family in this species.

Another gene identified in the current study and possibly involved in adaptation to drought response in wild common bean is Phvul.004G102800, homolog of SLAH3 of *Arabidopsis*, which

315 was annotated as an S-type anion channel. This type of channels is rapidly regulated by ABA and
316 stimulates stomata closure by inhibiting inward K^+ channels, thus reducing K^+ influx into guard
317 cells (Geiger et al., 2011; Zhang et al., 2016). In addition to being involved in drought stress
318 response, this same gene has been recently identified also as related to salinity stress response in
319 *Arabidopsis* by regulating ion homeostasis between root and shoots (Cubero-Font et al., 2016).
320 The chlorophyll alpha oxygenase (Phvul.008G161000), identified as a gene under selection
321 (and a homolog of *Arabidopsis* CAO), has a primary role in the biosynthesis of chlorophyll b
322 (Espineda et al., 1999). However, *Arabidopsis* mutants for this gene showed a reduction of
323 antioxidant compounds (specifically glutathione) in guard cells and an increased ABA sensitivity
324 in comparison to wild type plants (Jahan et al., 2016), suggesting a possible involvement of this
325 gene in adaptive response to stressful environments.
326 Phvul.002G143100, identified as selected within the different sub-populations of *P. vulgaris,*
327 is annotated as a glycine-rich domain protein, homologous of *Arabidopsis* GRDP2 gene. GRP are
328 a multi-gene superfamily present in several organisms, including plants (Sachetoo-Martins et al.,
329 2000). This gene family has been associated in plants with several developmental processes and
330 in responses to both biotic and abiotic stresses (Mangeon et al., 2010). A recent study focusing on
331 the characterization of the direct *Arabidopsis* homologs of this gene (ATGRDP2) demonstrated
332 that this gene regulates plant growth and flowering by accumulating higher level of indole-3-
333 acetic acid and improves abiotic stress response (Ortega-Amaro et al., 2014). In particular, the
334 over-expression of this gene in transgenic plants increased growth rate and reduced days to
335 flowering. It also increased salt tolerance in comparison to wild-type plants.
336 In addition to the previous genes identified as selected by genome scan analysis and
337 putatively involved in environmental response in plants, Phvul.009G050600 was identified and
338 annotated as an importin β-protein homologous to *Arabidopsis* KPNB1. This gene mediates the

import of proteins and protein complexes between the cytoplasm and the nucleus and is essential in regulating signal transduction pathways in response to environmental and developmental stimuli (Merkle, 2003). In particular, the *Arabidopsis* homolog of Phvul.009G050600 (AtKPNB1) has been directly related to ABA and drought response previously (Luo et al., 2013).

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 bio_12 (Annual Precipitation) and bio_7 (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 (Shabala, 2003, Sharma et al., 2013). In addition, variation in K^+ homeostasis is one of the first responses to several abiotic and biotic stresses in plants, allowing the plants to rapidly respond to stressful conditions (Shabala and Pottosin, 2014) and making the KEA gene identified in the current study an interesting candidate gene for further analysis.

Another gene, significantly associated with bio_14 (Precipitation of Driest Month) is Phvul.010G155000, which is annotated as a phospholipase (PLDα1). This gene is involved in the biosynthesis of phosphatidic acid (PA), which is an important signaling molecule in response to several stresses in plants (Saucedo-García et al., 2015). In particular, PA is involved in the ABA

362 signaling cascade and regulates stomata closure in plants by directly interacting and blocking
363 ABI1, an inhibitor of ABA response in plants (Zhang et al., 2004). This gene regulates stomatal
364 closure and ABA-dependent hydrogen peroxide (H_2O_2) production in *Vicia faba* as well (Qu et
365 al., 2014), making this gene an interesting candidate for improving drought response in common
366 bean.

367 An additional gene, significantly associated with bio_5 (Max Temperature of Warmest
368 Month), is Phvul.010G035200, annotated as a cytokinin response factor homologous of
369 *Arabidopsis* CRF4. Cytokinin is an essential plant hormone involved in growth and
370 developmental processes (Durán-Medina et al., 2017, Kieber and Schaller, 2014), but in recent
371 years it has also been implicated in the response and adaptation to different environmental
372 stresses (Novakova et al., 2007, O'Brien and Benková, 2013). CRF genes are a class of plant
373 transcription factors responsive to cytokinin that integrate hormonal and environmental signals
374 for adapting plant growth and development in response to the environment (Kim, 2016, Rashotte
375 and Goertzen, 2010). The *Arabidopsis* homolog of this gene has been previous related to
376 acclimation to cold temperatures (Zwack et al., 2016). Since this gene has been associated with
377 temperature variables in wild common bean, it could also be involved in adaptation to
378 temperature variation in this species.

379 Another gene of interest, Phvul.008G161700, is significantly associated with bio_3
380 (Isothermality) and is annotated as a thioredoxin protein. These proteins are involved in the
381 regulation of oxidative stress response and in scavenging reactive oxygen species (ROS) in plants
382 (Gelhaye et al., 2005). Other than being simple byproducts of cellular metabolism, ROS
383 molecules has been recognized as important signaling molecules that regulate the response to
384 several environmental stresses in plants (D'Autréaux and Toledano, 2007, Sewelam et al., 2016).
385 Due to their ability to control the redox state of the cell, thioredoxin represents a key component

386 of the ROS signal transduction pathways in plants and in the response to environmental stress
387 (Sevilla et al., 2015). Thus, this gene could constitute another interesting candidate gene for
388 improving stress resistance in domesticated common bean.
389

390 ## Comparison of genes identified by genome scan and GWAS

391 Even though the genes identified by outlier-detection methods (hard-selection sweeps) and
392 association methods (soft-selection sweeps) are involved in similar processes, there was no
393 overlap between the candidate genes identified by the two approaches in this study. This could be
394 the direct result of the different assumptions underlying these methods. Indeed, genome scan
395 analysis identify genes that shows drastic variations of allele frequencies between natural
396 subpopulations (Schoville et al., 2012, Wagner and Fortin, 2013). This approach is independent
397 from bio-climatic variables, thus the SNPs identified as under selection by this analysis could be
398 the results of selective mechanisms not considered by association analysis, like soil composition,
399 pathogen pressure and/or competition with other plants. On the other hand, association analyses
400 identify SNPs showing slight variations in allele frequencies across environmental gradients that
401 can increase environmental adaptation in natural populations (Schoville et al., 2012, Wagner and
402 Fortin, 2013). This selection process usually acts on natural standing variations and favor the
403 presence of multiple alleles and haplotypes, instead of allele fixation within populations
404 (Hermisson and Pennings, 2005).
405

406 # Epilogue

407 In conclusion, landscape genomic analysis of wild common bean genotypes allowed us to
408 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 (Dell'Acqua et al., 2014, Pyhäjärvi et al., 2013) 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.

Our long-term objective is to identify both populations (Ariani et al., 2018) and genes (this study) that have been putatively under selection by the abiotic stresses of temperature, rainfall, and the related variable, potential evapotranspiration. Identification of these potential sources of genetic tolerance are only a first step towards the development of more stress-resilient beans. The next step is to corroborate the effectiveness of these wild populations and these candidate genes as a source of stress-tolerance through indirect selection for these genes in selected populations resulting from the cross between candidate populations and domesticated testers (Acosta-Gallegos et al., 2007). In a recent paper, Cortés and Blair (2018) identified 115 SNPs tagging 77 annotated genes potentially selected by drought tolerance among a set of wild bean populations, using correlations between SNPs and an average yearly drought index. They argued that drought tolerance and performance under well-watered conditions were mutually incompatible. The testcrosses just mentioned between domesticated testers and wild populations that have been subjected to different drought stress conditions, as identified in this study and Ariani et al., (2018), will allow us to examine this hypothesis, which has considerable implications for breeding for stress tolerance.

MATERIALS AND METHODS

Plant material and genotypic data

A panel of 246 wild *P. vulgaris* accessions, previously genotyped with a Genotyping-By-Sequencing (GBS) protocol using the *Cvi*AII restriction enzyme (Ariani et al., 2018), 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 **S4 Table**, while genotyping data in VCF format are available as a Dash dataset **(**https://doi.org/10.25338/B8DW39). The seeds were provided by the Genetic Resources Unite at 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

(Wehrens 2011). We then selected the first two variables with the highest positive and negative loading in the first four principal components (PC1 to PC4) (**S3 Table**). Since some of the selected bio-climatic variables showed a high correlation (**S1 Table**), 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 (Luu et al., 2017), an algorithm able to detect population structure and outlier loci by performing a PCA analysis on SNP genotypic 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 (Frichot et al., 2013) implemented in the LEA R package (Frichot and François, 2015). This method was developed specifically for identifying signature of environmental selection in

genomic data and can efficiently correct for population history and isolation-by-distance (IBD). 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 (Devlin and Roeder, 1999). Plots of the inflation factor for different values of K (**S5 Fig**) 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) (Schmutz et al., 2014), was evaluated using the GenomicRanges/rtracklayer packages or R (Lawrence et al., 2009, 2013). 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 considered the

presence of possible regulatory regions immediately adjacent to gene sequences (Li et al., 2012). To understanding if the genes identified by significant SNPs were in regions with high linkage disequilibrium (LD), we identified haploblocks from the complete set of SNPs data using the PLINK program (Purcell et al., 2007) with default parameters. For downstream analysis, we considered only blocks longer than 100 bp. We then integrated this information with the genes identified as putatively selected by genome scan or association analysis, to determine if these candidate genes were located in haploblock regions. This analysis identified 1338 haplotype blocks evenly distributed across the 11 chromosomes (Dash dataset: https://doi.org/10.25338/B8DW39).

Comparison with latest genome reference

A new genome reference for *P. vulgaris* (v2.1) has been released on Phytozome although it has yet to be peer-reviewed. We compared the genes and the SNPs identified by our analysis between the old (v1.0) and the newest (v2.1) genome version. To compare the results between the two genomes, we mapped the significant SNPs, identified by genome scan and association analysis in the v1.0 genome reference, onto the v2.1 version. For this analysis, we extracted the 100 bp upstream and downstream of a significant SNPs (200 bp window) in the v1.0 version and mapped them to the v2.1 reference genome using nucleotide BLAST (Camacho et al., 2009). For each SNP and relative flanking region, we then selected the best hit in the v2.1 genome and identified the genes annotated in the new reference located within 5 Kb of the hit (as described in the 'Identification of putatively selected genes' section).

Candidate genes evaluation across genetic groups

For clustering individuals based on genetic groups and visualizing allele frequency variations across clusters, we applied a K-means clustering approach using the first 5 PCs obtained from pcadapt analysis. We selected K=5 as the best number of clusters, based on the scree plot of the eigenvalues obtained with pcadapt. The clustering analysis was performed using the python scikit-learn library (Pedregosa et al., 2011). For each genetic cluster, we calculated allele frequencies for SNPs tagging candidate genes using VCFtools (Danecek et al., 2011) and plotted them on genetic maps using R.

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.

566 **Figure legends**

567 **Fig 1** Bio-climatic data analysis. **(A)** Loading plot of the PCA analysis. **(B)** Principal Component
568 Analysis (PCA) of the bio-climatic data. Groups are colored according to the K-mean clustering
569 analysis conducted in this study, which gave results very similar to the STRUCTURE analysis
570 conducted by Ariani et al. (2018): MW1, MW2, and MW3: Mesoamerican wild gene pools; AW:
571 Andean wild gene pool; PhI: Intermediate wild gene pool.

572 **Fig 2** Principal Component Analysis on SNP data. **(A)** Screeplot of the PCA explained variance.
573 **(B)** PCA plot based on molecular data of the different genotyped analyzed in the current study.
574 Groups are colored according to the K-mean clustering analysis conducted in this study, which
575 gave results very similar to the STRUCTURE analysis conducted by Ariani et al. (2018): MW1,
576 MW2, and MW3: Mesoamerican wild gene pools; AW: Andean wild gene pool; PhI:
577 Intermediate wild gene pool.

578 **Fig 3** Three-dimensional plot of the PCA analysis on molecular data. Points are colored as in **Fig**
579 **2B**.

580 **Fig 4** Manhattan plot of the genome scan data with 5 sub-populations (K). The blue dashed line
581 represents the significance threshold (Bonferroni p-value ≤ 0.001).

582 **Fig 5** Chromosome ideogram of the genes identified as associated with the bio-climatic variables
583 analyzed. Only chromosomes with significantly associated variants are shown. Each circle
584 represents a different bio-climatic variable. When available, gene annotations are shown. The
585 centromeric regions shown are based on the results from Sevilla et al. (2015).

586 **Fig 6** Allele frequency distribution across different genetic groups for candidate genes identified
587 by genome scan analysis. *P. vulgaris v1.0* genes annotation and ID: **(A)** Potassium uptake
588 transporter (Phvul.002G331700); **(B)** Glycine-rich domain protein (Phvul.002G143100); **(C)**
589 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 because SNP data were completely missing. REF: Reference allele, in red, ALT: Alternative allele, according to the *P. vulgaris* v1.0 gene version, in blue (Sevilla et al., 2015).

Fig 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 **Fig 6.**

Supporting information

Tables

S1 Table. Candidate genes identified by genome scans.

S2 Table. Candidate genes identified by genome-wide association analysis.

S3 Table. Eigenvalues of the different bioclimatic variables along the first four principal components.

S4 Table. List of the final wild *Phaseolus vulgaris* analyzed in this study. Accession ID, country of origin, geographical coordinates of collection, and gene pool information are shown (from Ariani et al. 2018).

Figures

S1 Fig. 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.

614 **S2 Fig.** Cumulative variance explained by the different PCs when performing a PCA on bio-

615 climatic variables.

616 **S3 Fig.** P-values distribution for genome scans with 2 (A), 3 (B), 4 (C) or 5 (D) sub-populations.

617 **S4 Fig.** Plot of geographic distribution of the wild *P. vulgaris* analyzed in the current studies.

618 Genotypes are colored based on the different clusters identified by K-means clustering **(Fig 1B,**

619 **Fig. 2B, S4 Table**).

620 **S5 Fig.** Plots of the inflation factor for different values of K across the climatic variables

621 selected for association study.

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. Nature Climate Change. 2016;6: 605–609. doi:10.1038/nclimate2947 W

2. Campbell BM, Vermeulen SJ, Aggarwal PK, Corner-Dolloff C, Girvetz E, Loboguerrero AM, et al., Reducing risks to food security from climate change. Global Food Security. 2016;11: 34–43. doi:10.1016/j.gfs.2016.06.002

3. Wheeler T, Braun J von. Climate Change Impacts on Global Food Security. Science. 2013;341: 508–513. doi:10.1126/science.1239402

4. Challinor AJ, Watson J, Lobell DB, Howden SM, Smith DR, Chhetri N. A meta-analysis of crop yield under climate change and adaptation. Nature Climate Change. 2014;4: 287–291. doi:10.1038/nclimate2153

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: 607–610. doi:10.1126/science.1152339

6. Field CB, Barros V, Stocker TF, Dahe Q, editors. Managing the Risks of Extreme Events and Disasters to Advance Climate Change Adaptation: Special Report of the Intergovernmental Panel on Climate Change (Internet). Cambridge: Cambridge University Press; 2012. doi:10.1017/CBO9781139177245

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: 559–565. doi:10.1525/bio.2011.61.7.10

8. Zamir D. Improving plant breeding with exotic genetic libraries. Nat Rev Genet. 2001;2: 983–989. doi:10.1038/35103589

9. Gepts P. The contribution of genetic and genomic approaches to plant domestication studies. Curr Opin Plant Biol. 2014;18: 51–59. doi:10.1016/j.pbi.2014.02.001

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

11. Brozynska M, Furtado A, Henry RJ. Genomics of crop wild relatives: expanding the gene pool for crop improvement. Plant Biotechnol J. 2016;14: 1070–1085. doi:10.1111/pbi.12454

12. Cortés AJ, Monserrate FA, Ramírez-Villegas J, Madriñán S, Blair MW. Drought tolerance in wild plant populations: the case of common beans (Phaseolus vulgaris L.). PLoS ONE. 2013;8: e62898. doi:10.1371/journal.pone.0062898

13. Zhang H, Mittal N, Leamy LJ, Barazani O, Song B-H. Back into the wild-Apply untapped genetic diversity of wild relatives for crop improvement. Evol Appl. 2017;10: 5–24. doi:10.1111/eva.12434

14. Schoville SD, Bonin A, François O, Lobreaux S, Melodelima C, Manel S. Adaptive Genetic Variation on the Landscape: Methods and Cases. Annual Review of Ecology, Evolution, and Systematics. 2012;43: 23–43. doi:10.1146/annurev-ecolsys-110411-160248

15. Bragg JG, Supple MA, Andrew RL, Borevitz JO. Genomic variation across landscapes: insights and applications. New Phytol. 2015;207: 953–967. doi:10.1111/nph.13410

16. Anderson JE, Kono TJY, Stupar RM, Kantar MB, Morrell PL. Environmental Association Analyses Identify Candidates for Abiotic Stress Tolerance in Glycine soja, the Wild Progenitor of Cultivated Soybeans. G3 (Bethesda). 2016;6: 835–843. doi:10.1534/g3.116.026914

17. Wagner HH, Fortin M-J. A conceptual framework for the spatial analysis of landscape genetic data. Conserv Genet. 2013;14: 253–261. doi:10.1007/s10592-012-0391-5

18. Wright S. The Genetical Structure of Populations. Annals of Eugenics. 1949;15: 323–354. doi:10.1111/j.1469-1809.1949.tb02451.x

19. Lewontin RC, Krakauer J. Distribution of gene frequency as a test of the theory of the selective neutrality of polymorphisms. Genetics. 1973;74: 175–195.

20. Haldane JBS. A mathematical theory of natural and artificial selection. (Part VI, Isolation.). Mathematical Proceedings of the Cambridge Philosophical Society. 1930;26: 220–230. doi:10.1017/S0305004100015450

21. Kimura M. On the probability of fixation of mutant genes in a population. Genetics. 1962;47: 713–719.

22. Narum SR, Hess JE. Comparison of F(ST) outlier tests for SNP loci under selection. Mol Ecol Resour. 2011;11 Suppl 1: 184–194. doi:10.1111/j.1755-0998.2011.02987.x

23. Manel S, Joost S, Epperson BK, Holderegger R, Storfer A, Rosenberg MS, et al., Perspectives on the use of landscape genetics to detect genetic adaptive variation in the field. Mol Ecol. 2010;19: 3760–3772. doi:10.1111/j.1365-294X.2010.04717.x

24. Hirschhorn JN, Daly MJ. Genome-wide association studies for common diseases and complex traits. Nat Rev Genet. 2005;6: 95–108. doi:10.1038/nrg1521

25. Lipka AE, Kandianis CB, Hudson ME, Yu J, Drnevich J, Bradbury PJ, et al., From association to prediction: statistical methods for the dissection and selection of complex traits in plants. Current Opinion in Plant Biology. 2015;24: 110–118. doi:10.1016/j.pbi.2015.02.010

26. Gepts P, Aragão FJL, Barros E de, Blair MW, Brondani R, Broughton W, et al., Genomics of Phaseolus Beans, a Major Source of Dietary Protein and Micronutrients in the Tropics. In: Moore PH, Ming R, editors. Genomics of Tropical Crop Plants. New York, NY: Springer New York; 2008. pp. 113–143. doi:10.1007/978-0-387-71219-2_5

27. Messina V. Nutritional and health benefits of dried beans. Am J Clin Nutr. 2014;100 Suppl 1: 437S–42S. doi:10.3945/ajcn.113.071472

28. Rubiales D, Mikic A. Introduction: Legumes in Sustainable Agriculture. Critical Reviews in Plant Sciences. 2015;34: 2–3. doi:10.1080/07352689.2014.897896

29. Chacón S. MI, Pickersgill B, Debouck DG, Arias JS. Phylogeographic analysis of the chloroplast DNA variation in wild common bean (Phaseolus vulgaris L.) in the Americas. Plant Syst Evol. 2007;266: 175–195. doi:10.1007/s00606-007-0536-z

30. Koenig R, Gepts P. Allozyme diversity in wild Phaseolus vulgaris: further evidence for two major centers of genetic diversity. Theor Appl Genet. 1989;78: 809–817. doi:10.1007/BF00266663

31. Mamidi S, Rossi M, Moghaddam SM, Annam D, Lee R, Papa R, et al., Demographic factors shaped diversity in the two gene pools of wild common bean Phaseolus vulgaris L. Heredity (Edinb). 2013;110: 267–276. doi:10.1038/hdy.2012.82

32. Bitocchi E, Bellucci E, Giardini A, Rau D, Rodriguez M, Biagetti E, et al., Molecular analysis of the parallel domestication of the common bean (Phaseolus vulgaris) in Mesoamerica and the Andes. New Phytol. 2013;197: 300–313. doi:10.1111/j.1469-8137.2012.04377.x

33. Blair MW, Soler A, Cortés AJ. Diversification and population structure in common beans (Phaseolus vulgaris L.). PLoS ONE. 2012;7: e49488. doi:10.1371/journal.pone.0049488

34. Chacón S MI, Pickersgill B, Debouck DG. Domestication patterns in common bean (Phaseolus vulgaris L.) and the origin of the Mesoamerican and Andean cultivated races. Theor Appl Genet. 2005;110: 432–444. doi:10.1007/s00122-004-1842-2

35. Gepts P, Osborn TC, Rashka K, Bliss FA. Phaseolin-protein Variability in Wild Forms and Landraces of the Common Bean(Phaseolus vulgaris): Evidence for Multiple Centers of Domestication. Econ Bot. 1986;40: 451–468. doi:10.1007/BF02859659

36. Mamidi S. Investigation of the domestication of common bean (Phaseolus vulgaris) using multilocus sequence data. Functional plant biology. 2011;v. 38: 953–967. doi:10.1071/FP11124

37. Rossi M, Bitocchi E, Bellucci E, Nanni L, Rau D, Attene G, et al., Linkage disequilibrium and population structure in wild and domesticated populations of Phaseolus vulgaris L. Evol Appl. 2009;2: 504–522. doi:10.1111/j.1752-4571.2009.00082.x

38. Singh SP, Gepts P, Debouck DG. Races of common bean (Phaseolus vulgaris, Fabaceae). Econ Bot. 1991;45: 379–396. doi:10.1007/BF02887079

39. Debouck DG, Toro O, Paredes OM, Johnson WC, Gepts P. Genetic diversity and ecological distribution of Phaseolus vulgaris (Fabaceae) in northwestern South America. Econ Bot. 1993;47: 408–423. doi:10.1007/BF02907356

40. Kami J, Velásquez VB, Debouck DG, Gepts P. Identification of presumed ancestral DNA sequences of phaseolin in Phaseolus vulgaris. PNAS. 1995;92: 1101–1104. doi:10.1073/pnas.92.4.1101

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: 259–266. doi:10.11646/phytotaxa.313.3.3

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. doi:10.1186/s13059-017-1190-6

43. Gepts P. Origin and Evolution of Common Bean: Past Events and Recent Trends. HortScience. 1998;33: 1124–1130.

44. Porch T, Beaver J, Debouck D, Jackson S, Kelly J, Dempewolf H, et al., Use of Wild Relatives and Closely Related Species to Adapt Common Bean to Climate Change. Agronomy. 2013;3: 433–461. doi:10.3390/agronomy3020433

45. Acosta-Gallegos JA, Kelly JD, Gepts P. Prebreeding in Common Bean and Use of Genetic Diversity from Wild Germplasm. Crop Science. 2007;47: S-44-S-59. doi:10.2135/cropsci2007.04.0008IPBS

46. Ramirez-Cabral NYZ, Kumar L, Taylor S. Crop niche modeling projects major shifts in common bean growing areas. Agricultural and Forest Meteorology. 2016;218–219: 102–113. doi:10.1016/j.agrformet.2015.12.002

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. doi:10.1371/journal.pone.0047832

48. Yoder JB, Stanton-Geddes J, Zhou P, Briskine R, Young ND, Tiffin P. Genomic signature of adaptation to climate in Medicago truncatula. Genetics. 2014;196: 1263–1275. doi:10.1534/genetics.113.159319

49. Dell'Acqua M, Zuccolo A, Tuna M, Gianfranceschi L, Pè ME. Targeting environmental adaptation in the monocot model Brachypodium distachyon: a multi-faceted approach. BMC Genomics. 2014;15. doi:10.1186/1471-2164-15-801

761 50. Abebe TD, Naz AA, Léon J. Landscape genomics reveal signatures of local adaptation in
762 barley (Hordeum vulgare L.). Front Plant Sci. 2015;6: 813. doi:10.3389/fpls.2015.00813

763 51. Ariani A, Berny Mier Y Teran JC, Gepts P. Spatial and temporal scales of range expansion
764 in wild *Phaseolus vulgaris*. Mol Biol Evol. 2018;35: 119–131. doi:10.1093/molbev/msx273

765 52. Rodriguez M, Rau D, Bitocchi E, Bellucci E, Biagetti E, Carboni A, et al., Landscape
766 genetics, adaptive diversity and population structure in *Phaseolus vulgaris*. New Phytol.
767 2016;209: 1781–1794. doi:10.1111/nph.13713

768 53. Wehrens R. Chemometrics with R: Multivariate data analysis in the natural sciences and life
769 sciences (internet). Berlin Heidelberg: Springer-Verlag; 2011. Available:
770 //www.springer.com/us/book/9783642178405

771 54. Luu K, Bazin E, Blum MGB. pcadapt: an R package to perform genome scans for selection
772 based on principal component analysis. Mol Ecol Resour. 2017;17: 67–77.
773 doi:10.1111/1755-0998.12592

774 55. Frichot E, Schoville SD, Bouchard G, François O. Testing for associations between loci and
775 environmental gradients using latent factor mixed models. Mol Biol Evol. 2013;30: 1687–
776 1699. doi:10.1093/molbev/mst063

777 56. Frichot E, François O. LEA: An R package for landscape and ecological association studies.
778 Methods in Ecology and Evolution. 2015;6: 925–929. doi:10.1111/2041-210X.12382

779 57. Devlin B, Roeder K. Genomic control for association studies. Biometrics. 1999;55: 997–
780 1004.

781 58. Schmutz J, McClean PE, Mamidi S, Wu GA, Cannon SB, Grimwood J, et al., A reference
782 genome for common bean and genome-wide analysis of dual domestications. Nat Genet.
783 2014;46: 707–713. doi:10.1038/ng.3008

784 59. Lawrence M, Gentleman R, Carey V. rtracklayer: an R package for interfacing with genome
785 browsers. Bioinformatics. 2009;25: 1841–1842. doi:10.1093/bioinformatics/btp328

786 60. Lawrence M, Huber W, Pagès H, Aboyoun P, Carlson M, Gentleman R, et al., Software for
787 computing and annotating genomic ranges. PLoS Comput Biol. 2013;9: e1003118.
788 doi:10.1371/journal.pcbi.1003118

789 61. Li X, Zhu C, Yeh C-T, Wu W, Takacs EM, Petsch KA, et al., Genic and nongenic
790 contributions to natural variation of quantitative traits in maize. Genome Res. 2012;22:
791 2436–2444. doi:10.1101/gr.140277.112

792 62. Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira MAR, Bender D, et al., PLINK: A
793 Tool Set for Whole-Genome Association and Population-Based Linkage Analyses. Am J
794 Hum Genet. 2007;81: 559–575.

795 63. Camacho C, Coulouris G, Avagyan V, Ma N, Papadopoulos J, Bealer K, et al., BLAST+:
796 architecture and applications. BMC Bioinformatics. 2009;10: 421. doi:10.1186/1471-2105-
797 10-421

798 64. Pedregosa F, Varoquaux G, Gramfort A, Michel V, Thirion B, Grisel O, et al., Scikit-learn:
799 Machine Learning in Python. Journal of Machine Learning Research. 2011;12: 2825−2830.

800 65. Danecek P, Auton A, Abecasis G, Albers CA, Banks E, DePristo MA, et al., The variant
801 call format and VCFtools. Bioinformatics. 2011;27: 2156–2158.
802 doi:10.1093/bioinformatics/btr330

803 66. Bitocchi E, Nanni L, Bellucci E, Rossi M, Giardini A, Zeuli PS, et al., Mesoamerican origin
804 of the common bean (Phaseolus vulgaris L.) is revealed by sequence data. Proc Natl Acad
805 Sci USA. 2012;109: E788-796. doi:10.1073/pnas.1108973109

806 67. Kwak M, Gepts P. Structure of genetic diversity in the two major gene pools of common
807 bean (Phaseolus vulgaris L., Fabaceae). Theor Appl Genet. 2009;118: 979–992.
808 doi:10.1007/s00122-008-0955-4

809 68. Osakabe Y, Arinaga N, Umezawa T, Katsura S, Nagamachi K, Tanaka H, et al., Osmotic
810 stress responses and plant growth controlled by potassium transporters in Arabidopsis. Plant
811 Cell. 2013;25: 609–624. doi:10.1105/tpc.112.105700

812 69. Bellucci E, Bitocchi E, Ferrarini A, Benazzo A, Biagetti E, Klie S, et al., Decreased
813 Nucleotide and Expression Diversity and Modified Coexpression Patterns Characterize
814 Domestication in the Common Bean[W][OPEN). Plant Cell. 2014;26: 1901–1912.
815 doi:10.1105/tpc.114.124040

816 70. Bitocchi E, Rau D, Bellucci E, Rodriguez M, Murgia ML, Gioia T, et al., Beans (Phaseolus
817 ssp.) as a Model for Understanding Crop Evolution. Front Plant Sci. 2017;8: 722.
818 doi:10.3389/fpls.2017.00722

819 71. Geiger D, Maierhofer T, Al-Rasheid KAS, Scherzer S, Mumm P, Liese A, et al., Stomatal
820 closure by fast abscisic acid signaling is mediated by the guard cell anion channel SLAH3
821 and the receptor RCAR1. Sci Signal. 2011;4: ra32. doi:10.1126/scisignal.2001346

822 72. Zhang A, Ren H-M, Tan Y-Q, Qi G-N, Yao F-Y, Wu G-L, et al., S-type Anion Channels
823 SLAC1 and SLAH3 Function as Essential Negative Regulators of Inward K+ Channels and
824 Stomatal Opening in Arabidopsis. Plant Cell. 2016;28: 949–955. doi:10.1105/tpc.16.01050

825 73. Cubero-Font P, Maierhofer T, Jaslan J, Rosales MA, Espartero J, Díaz-Rueda P, et al.,
826 Silent S-Type Anion Channel Subunit SLAH1 Gates SLAH3 Open for Chloride Root-to-
827 Shoot Translocation. Curr Biol. 2016;26: 2213–2220. doi:10.1016/j.cub.2016.06.045

828 74. Espineda CE, Linford AS, Devine D, Brusslan JA. The AtCAO gene, encoding chlorophyll
829 a oxygenase, is required for chlorophyll b synthesis in Arabidopsis thaliana. Proc Natl Acad
830 Sci USA. 1999;96: 10507–10511.

831 75. Jahan MS, Nozulaidi M, Khairi M, Mat N. Light-harvesting complexes in photosystem II
832 regulate glutathione-induced sensitivity of Arabidopsis guard cells to abscisic acid. J Plant
833 Physiol. 2016;195: 1–8. doi:10.1016/j.jplph.2016.03.002

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

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

838 78. Ortega-Amaro MA, Rodríguez-Hernández AA, Rodríguez-Kessler M, Hernández-Lucero E,
839 Rosales-Mendoza S, Ibáñez-Salazar A, et al., Overexpression of AtGRDP2, a novel glycine-
840 rich domain protein, accelerates plant growth and improves stress tolerance. Front Plant Sci.
841 2014;5: 782. doi:10.3389/fpls.2014.00782

842 79. Merkle T. Nucleo-cytoplasmic partitioning of proteins in plants: implications for the
843 regulation of environmental and developmental signalling. Curr Genet. 2003;44: 231–260.
844 doi:10.1007/s00294-003-0444-x

845 80. Luo Y, Wang Z, Ji H, Fang H, Wang S, Tian L, et al., An Arabidopsis homolog of importin
846 β1 is required for ABA response and drought tolerance. Plant J. 2013;75: 377–389.
847 doi:10.1111/tpj.12207

848 81. Shabala S. Regulation of Potassium Transport in Leaves: from Molecular to Tissue Level.
849 Ann Bot. 2003;92: 627–634. doi:10.1093/aob/mcg191

850 82. Sharma T, Dreyer I, Riedelsberger J. The role of K+ channels in uptake and redistribution of
851 potassium in the model plant Arabidopsis thaliana. Front Plant Sci. 2013;4.
852 doi:10.3389/fpls.2013.00224

853 83. Shabala S, Pottosin I. Regulation of potassium transport in plants under hostile conditions:
854 implications for abiotic and biotic stress tolerance. Physiol Plant. 2014;151: 257–279.
855 doi:10.1111/ppl.12165

856 84. Saucedo-García M, Gavilanes-Ruíz M, Arce-Cervantes O. Long-chain bases, phosphatidic
857 acid, MAPKs, and reactive oxygen species as nodal signal transducers in stress responses in
858 Arabidopsis. Front Plant Sci. 2015;6: 55. doi:10.3389/fpls.2015.00055

859 85. Zhang W, Qin C, Zhao J, Wang X. Phospholipase Dα1-derived phosphatidic acid interacts
860 with ABI1 phosphatase 2C and regulates abscisic acid signaling. Proc Natl Acad Sci U S A.
861 2004;101: 9508–9513. doi:10.1073/pnas.0402112101

862 86. Qu Y, An Z, Zhuang B, Jing W, Zhang Q, Zhang W. Copper amine oxidase and
863 phospholipase D act independently in abscisic acid (ABA)-induced stomatal closure in
864 Vicia faba and Arabidopsis. J Plant Res. 2014;127: 533–544. doi:10.1007/s10265-014-
865 0633-3

87. Durán-Medina Y, Díaz-Ramírez D, Marsch-Martínez N. Cytokinins on the Move. Front Plant Sci. 2017;8: 146. doi:10.3389/fpls.2017.00146

88. Kieber JJ, Schaller GE. Cytokinins. Arabidopsis Book. 2014;12: e0168. doi:10.1199/tab.0168

89. Novakova M, Dobrev P, Motyka V, Gaudinova A, Malbeck J, Pospisilova J, et al., Cytokinin Function in Drought Stress Response and Subsequent Recovery. In: Xu Z, Li J, Xue Y, Yang W, editors. Biotechnology and Sustainable Agriculture 2006 and Beyond. Springer Netherlands; 2007. pp. 171–174.

90. O'Brien JA, Benková E. Cytokinin cross-talking during biotic and abiotic stress responses. Front Plant Sci. 2013;4: 451. doi:10.3389/fpls.2013.00451

91. Kim J. CYTOKININ RESPONSE FACTORs Gating Environmental Signals and Hormones. Trends Plant Sci. 2016;21: 993–996. doi:10.1016/j.tplants.2016.10.004

92. Rashotte AM, Goertzen LR. The CRF domain defines cytokinin response factor proteins in plants. BMC Plant Biol. 2010;10: 74. doi:10.1186/1471-2229-10-74

93. Zwack PJ, Compton MA, Adams CI, Rashotte AM. Cytokinin response factor 4 (CRF4) is induced by cold and involved in freezing tolerance. Plant Cell Rep. 2016;35: 573–584. doi:10.1007/s00299-015-1904-8

94. Gelhaye E, Rouhier N, Navrot N, Jacquot JP. The plant thioredoxin system. Cell Mol Life Sci. 2005;62: 24–35. doi:10.1007/s00018-004-4296-4

95. D'Autréaux B, Toledano MB. ROS as signalling molecules: mechanisms that generate specificity in ROS homeostasis. Nat Rev Mol Cell Biol. 2007;8: 813–824. doi:10.1038/nrm2256

96. Sewelam N, Kazan K, Schenk PM. Global Plant Stress Signaling: Reactive Oxygen Species at the Cross-Road. Front Plant Sci. 2016;7. doi:10.3389/fpls.2016.00187

97. Sevilla F, Camejo D, Ortiz-Espín A, Calderón A, Lázaro JJ, Jiménez A. The thioredoxin/peroxiredoxin/sulfiredoxin system: current overview on its redox function in plants and regulation by reactive oxygen and nitrogen species. J Exp Bot. 2015;66: 2945–2955. doi:10.1093/jxb/erv146

98. Hermisson J, Pennings PS. Soft sweeps: molecular population genetics of adaptation from standing genetic variation. Genetics. 2005;169: 2335–2352. doi:10.1534/genetics.104.036947

99. Pyhäjärvi T, Hufford MB, Mezmouk S, Ross-Ibarra J. Complex patterns of local adaptation in teosinte. Genome Biol Evol. 2013;5: 1594–1609. doi:10.1093/gbe/evt109

899 100. Cortés AJ, Blair MW. Genotyping by Sequencing and Genome–Environment Associations
900 in Wild Common Bean Predict Widespread Divergent Adaptation to Drought. Front Plant
901 Sci. 2018;9. doi:10.3389/fpls.2018.00128

902

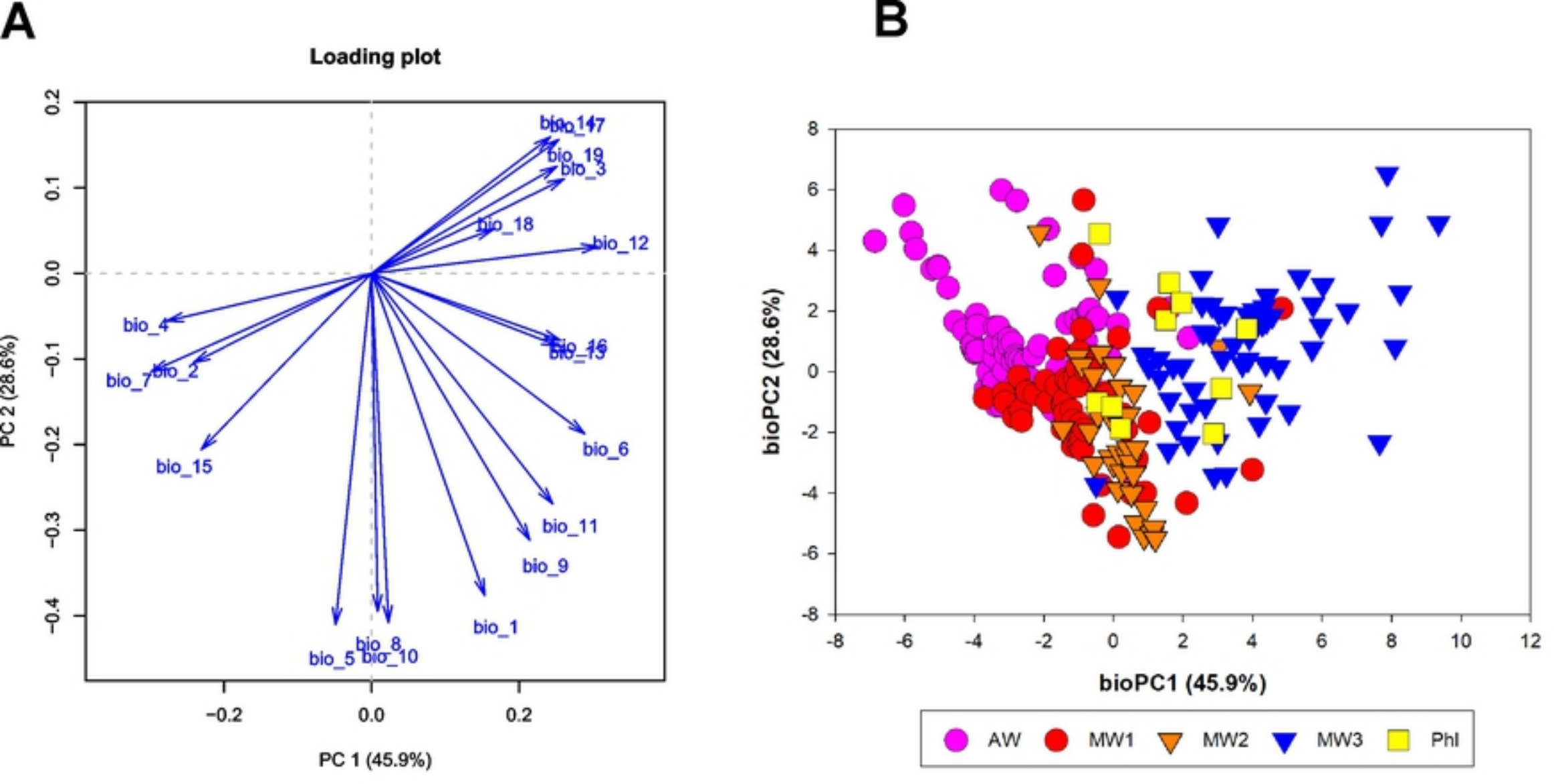
A
Loading plot
bio_17
bio_19
bio_3
bio_18
bio_12
bio_4
bio_16
bio_13
bio_7
bio_2
bio_15
bio_6
bio_11
bio_9
bio_1
bio_8
bio_5
bio_10
PC 2 (28.6%)
PC 1 (45.9%)
0.2
0.1
0.0
−0.1
−0.2
−0.3
−0.4
−0.2
0.0
0.2
B
bioPC2 (28.6%)
bioPC1 (45.9%)
8
6
4
2
0
-2
-4
-6
-8
-8
-6
-4
-2
0
2
4
6
8
10
12
AW
MW1
MW2
MW3
Phl

Figure 1

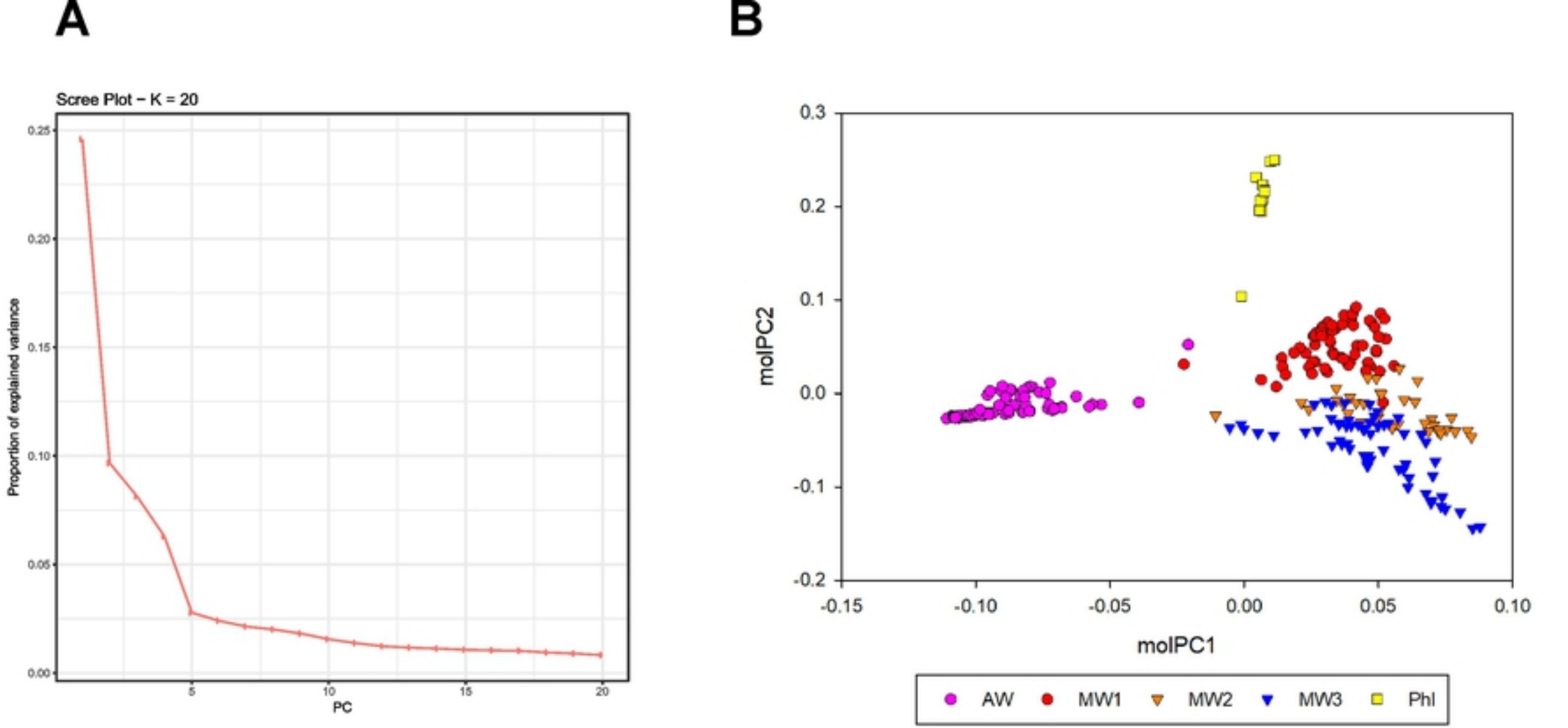
A
Scree Plot − K = 20
Proportion of explained variance
0.25
0.20
0.15
0.10
0.05
0.00
5
10
15
20
PC
B
molPC2
0.3
0.2
0.1
0.0
-0.1
-0.2
-0.15
-0.10
-0.05
0.00
0.05
0.10
molPC1
AW
MW1
MW2
MW3
Phl

Figure 2

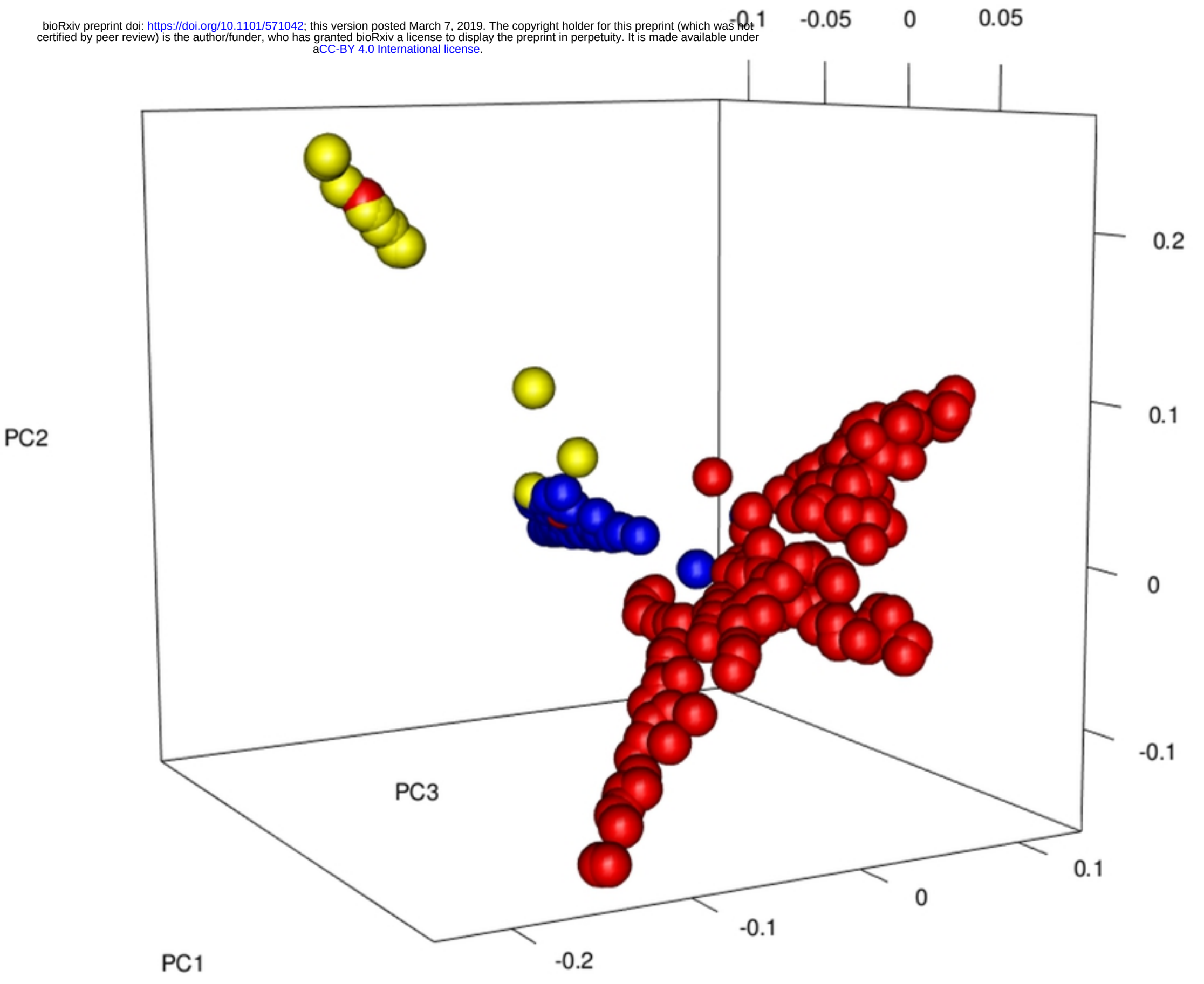

Figure 3

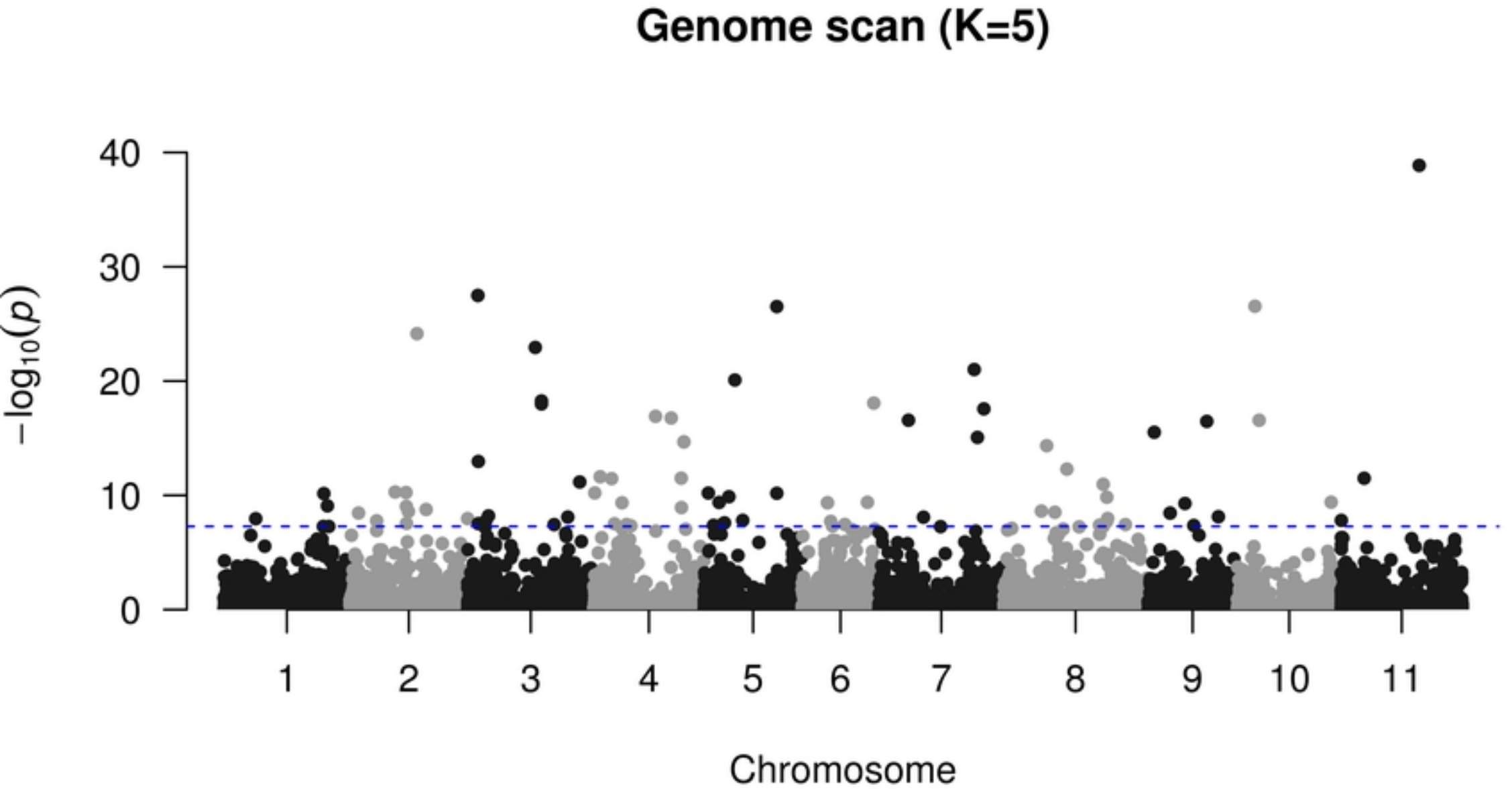
Genome scan (K=5)
40
30
20
10
0
$-\log_{10}(p)$
1
2
3
4
5
6
7
8
9
10
11
Chromosome

Figure 4

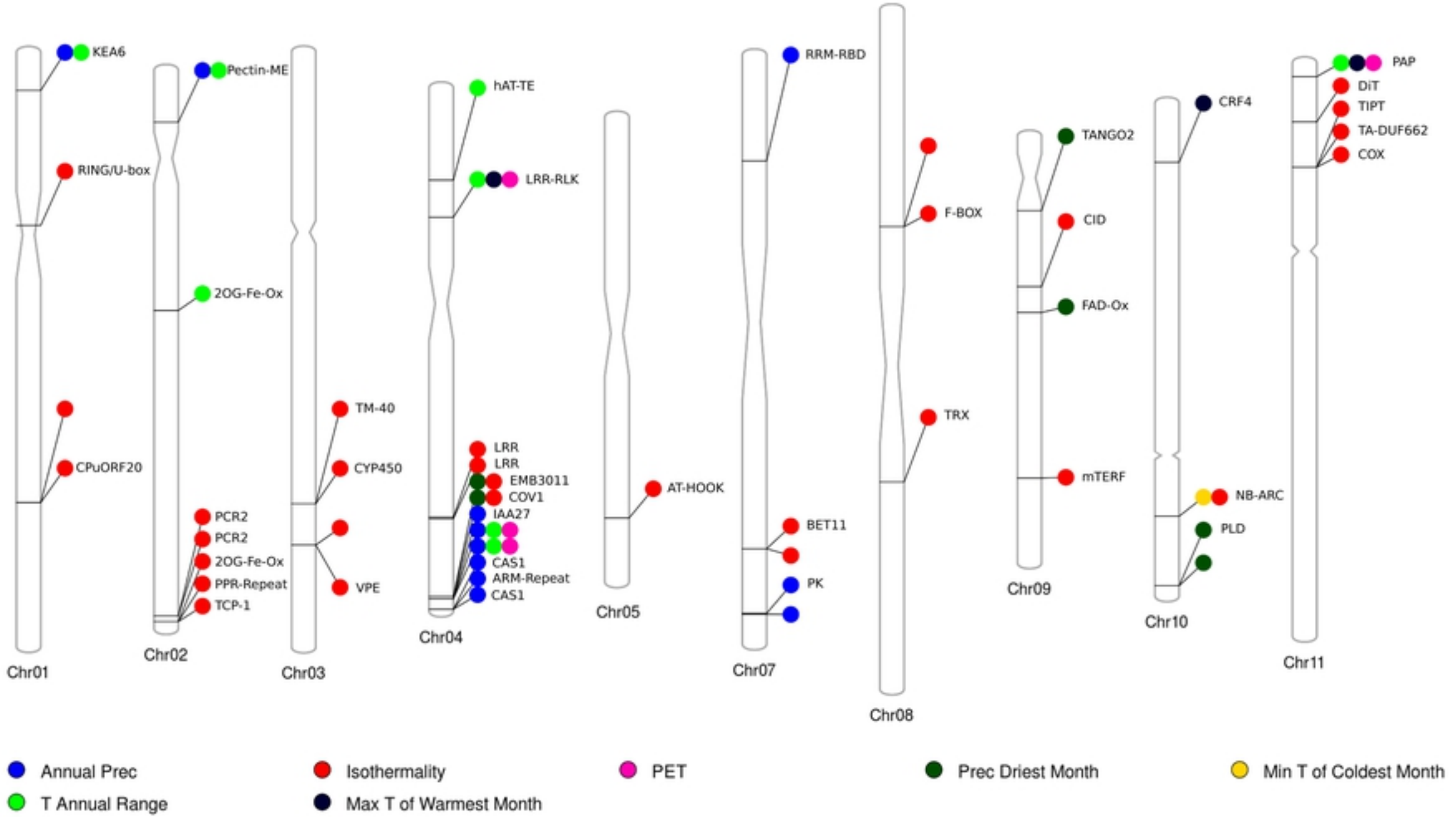
KEA6
RING/U-box
CPuORF20
Chr01
Pectin-ME
2OG-Fe-Ox
PCR2
PCR2
2OG-Fe-Ox
PPR-Repeat
TCP-1
Chr02
TM-40
CYP450
VPE
Chr03
hAT-TE
LRR-RLK
LRR
LRR
EMB3011
COV1
IAA27
CAS1
ARM-Repeat
CAS1
Chr04
AT-HOOK
Chr05
RRM-RBD
BET11
PK
Chr07
F-BOX
TRX
Chr08
TANGO2
CID
FAD-Ox
mTERF
Chr09
CRF4
NB-ARC
PLD
Chr10
PAP
DiT
TIPT
TA-DUF662
COX
Chr11
Annual Prec
T Annual Range
Isothermality
Max T of Warmest Month
PET
Prec Driest Month
Min T of Coldest Month

Figure 5

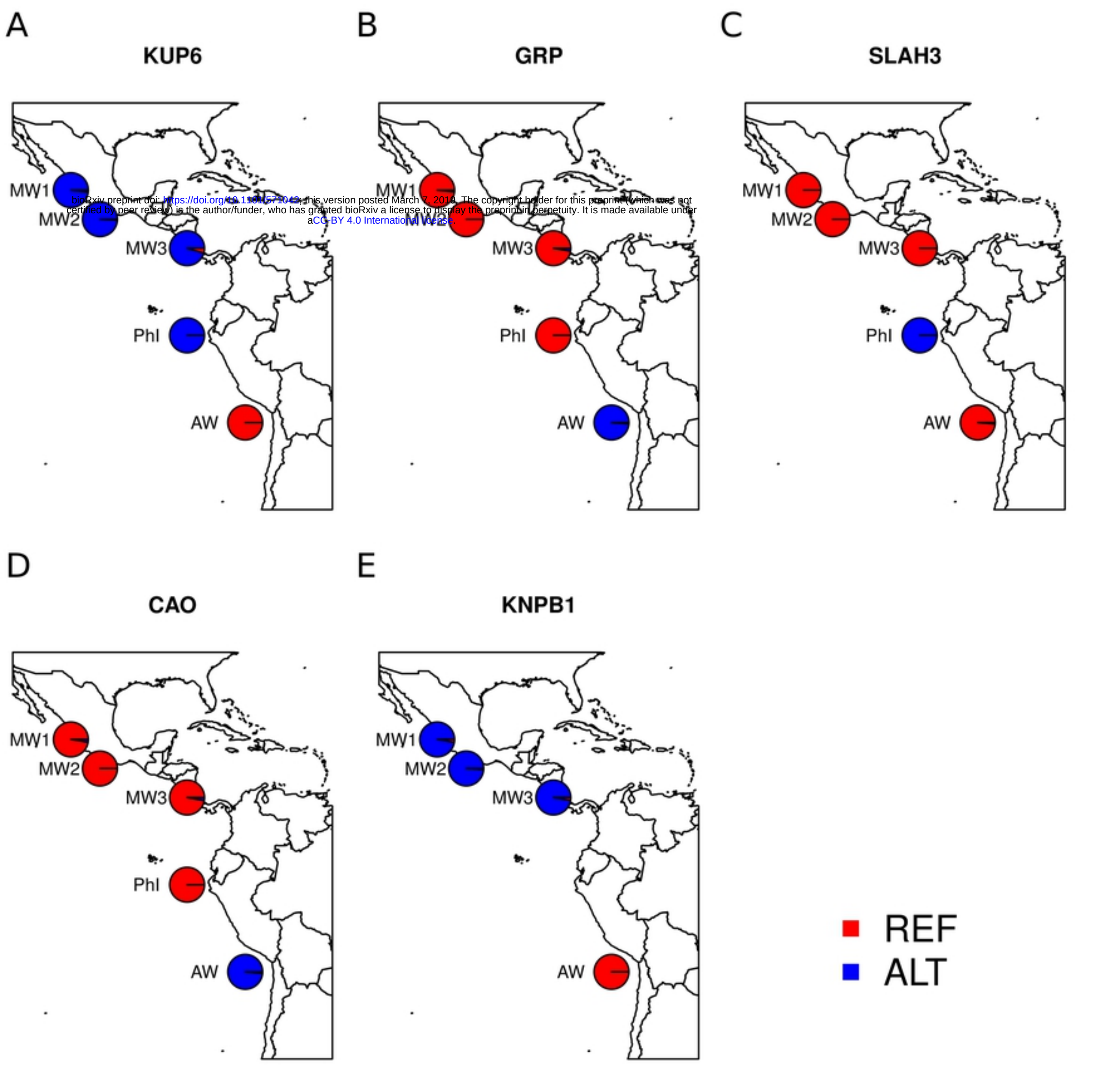

Figure 6

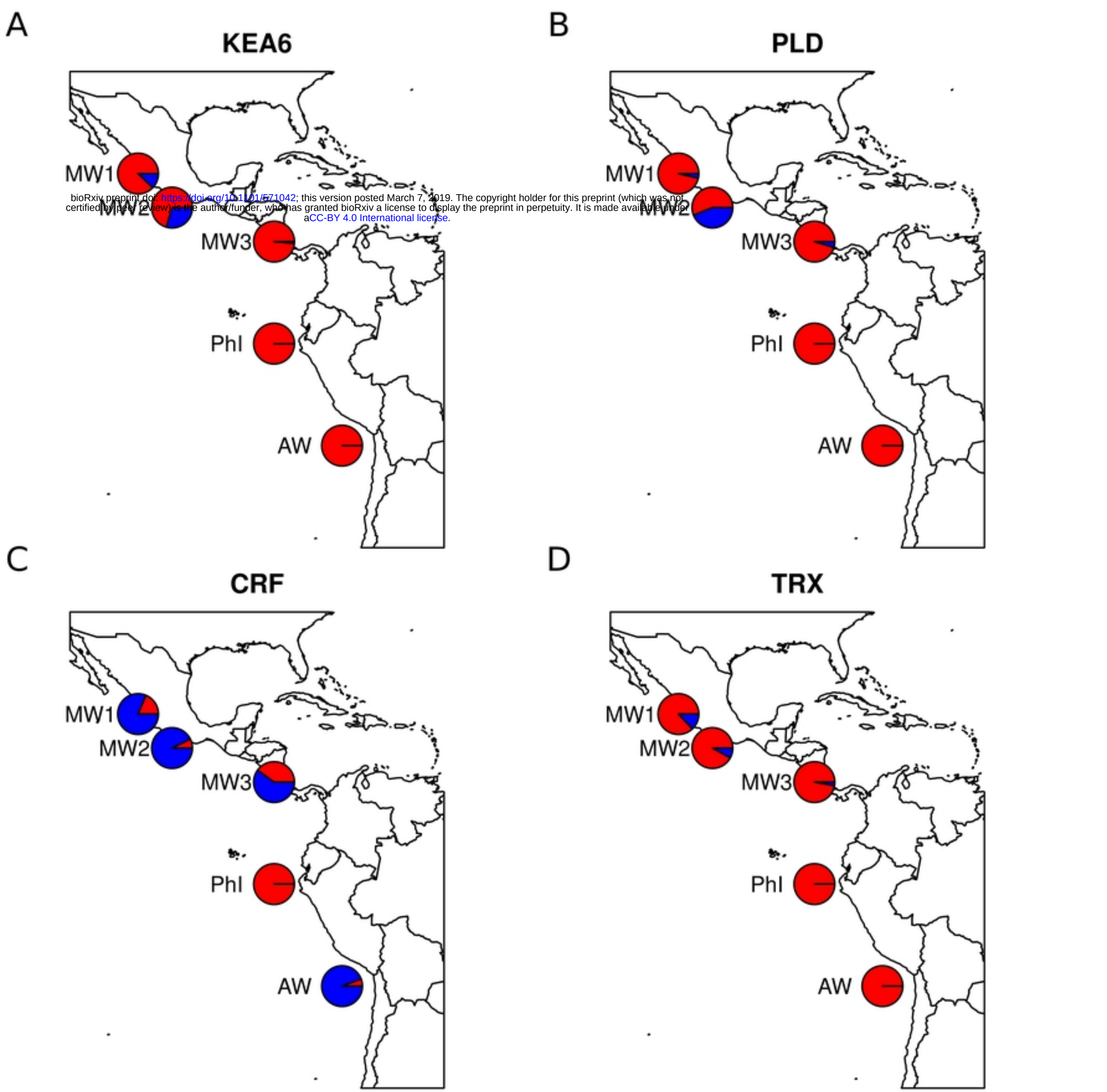

Figure 7