Selection and adaptive introgression guided the complex evolutionary history of European common bean

Bellucci E.[1]*, Benazzo A.[2]*, Xu C.[3]*, Bitocchi E.[1]*, Rodriguez M.[4]*, Alseekh S.[5,6]*, Di Vittori V.[1,5]*, Gioia T.[7], Neumann K.[9], Cortinovis G.[1], Frascarelli G.[1], Murube E.[1], Trucchi E.[2], Nanni L.[1], Ariani A.[8], Logozzo G.[7], Shin J.H.[3], Liu C.[10], Jiang L.[5], Ferreira J.J.[11], Campa A.[11], Attene G.[4], Morrell P.L.[10], Bertorelle G.[2], Graner A.[9]*, Gepts P.[8]*, Fernie A.R.[5,6]*, Jackson S.A.[3]*, Papa R.[1]*[#]

[1] Department of Agricultural, Food and Environmental Sciences, Marche Polytechnic University, 60131 Ancona, Italy

[2] Department of Life Sciences and Biotechnology, University of Ferrara, 44100 Ferrara, Italy

[3] Center for Applied Genetic Technologies, University of Georgia, 30602 Athens, GA, USA

[4] Department of Agriculture, University of Sassari, 07100 Sassari, Italy

[5] Max Planck Institute of Molecular Plant Physiology (MPI-MP), 14476 Potsdam-Golm, Germany

[6] Center for Plant Systems Biology, 4000 Plovdiv, Bulgaria

[7] School of Agricultural, Forestry, Food and Environmental Sciences, University of Basilicata, 85100 Potenza, Italy

[8] Department of Plant Sciences, University of California, 95616-8780 Davis, CA, USA

[9] Leibniz Institute of Plant Genetics and Crop Plant Research (IPK), 06466 Seeland, Germany

[10] Department of Agronomy and Plant Genetics, University of Minnesota, St. Paul, MN, USA, 55108-6026

[11] Regional Agrifood Research and Development Service (SERIDA), 33310, Villaviciosa, Asturias, Spain

*These authors contributed equally to this work

[#]**Corresponding author: R. Papa**

Department of Agricultural, Food and Environmental Sciences

Marche Polytechnic University

Via Brecce Bianche 60131 Ancona, Italy

Abstract

Following domestication, humans have disseminated crops over vast geographic regions. After 1492, the common bean was successfully introduced in Europe. Using whole-genome profiling and metabolic fingerprinting combined with phenotypic characterisation, we found that the first common beans seeds introduced in Europe were of Andean origin, after the Francisco Pizarro's expedition to northern Peru in 1529, and we identified the effects of hybridization, selection and recombination in shaping the genomic diversity of the European common bean in parallel with political constrains. There was clear evidence of adaptive introgression into the Mesoamerican-derived European genotypes, with 44 Andean introgressed genomic segments shared by more than 90% of the European accessions and distributed on all chromosomes except for PvChr11. Genomic scans for signatures of selection highlighted the role of genes related to flowering and genes relevant to environmental adaptation, suggesting that introgression was crucial for adapting a tropical crop to temperate regions of Europe.

Introduction

Following the process of domestication, crop species were spread by humans over vast geographic regions, adapting to new and often extreme environments. The Columbian Exchange (*1*) started in 1492 with the transatlantic journey of Christopher Columbus. This large-scale set of reciprocal biological introductions between continents provides a paradigmatic example of the rapid adaptation of crop plants to changing environments. Changes in flowering time and reduced photoperiod sensitivity were selected in parallel in common bean, maize, and potato, to name a few crops that have undergone selection for these adaptive traits (*2*, *3*). Among the crops that originated in the Americas, common bean was rapidly adopted and successfully disseminated across Europe (*4*). It is now possible to identify local varieties of beans in Europe of both Andean and Mesoamerican origins (*5*, *6*, *7*, *8*, *9*, *10*).

The introduction of common bean in Europe from two distinct centres of origin provided the opportunity for widespread inter-gene pool hybridisation and recombination (*9*). Studies of common bean evolution in Europe can exploit the parallel domestications and the strong genetic differentiation between the two common bean gene pools in the Americas. This provides an ideal model to study the role of introgression in the adaptation of common bean to European environments.

Here, we present a whole-genome analysis and metabolic fingerprinting of 218 common bean landraces, integrated with genome-wide association (GWA) to characterise the genetic basis of multiple traits, including flowering time and growth habit under multiple environments with contrasting (i.e., photoperiod) growing conditions. The combined results are used to characterise the effects of selection and inter-gene pool introgression and test the occurrence of adaptive introgression associated with the development and adaptation of common bean in Europe.

The population structure of common bean identifies pervasive admixture in Europe

Nuclear and chloroplast genetic variants were used to reconstruct the ancestry of 218 single seed descent (SSD) purified accessions from the Americas (104) and Europe (114) (Fig. 1a, b, c, d, e). For beans from the Americas, the subdivision into the highly differentiated Andean and Mesoamerican populations was consistent with previous studies (*11*; *12*, *13*) (Fig. 1a). However, among the European

accessions, nuclear variants allowed us to identify several admixed genotypes (Fig. 1a). In a few European accessions (n = 14), the nuclear assignment was inconsistent with the genetic assignment based on the chloroplast genome (Fig. 1a). Indeed, the identification of genotypes in which an Andean chloroplast genome was combined with a Mesoamerican nuclear genome (nine genotypes with ancestry component, or *vice versa* five genotypes with Mesoamerican chloroplast in an Andean nuclear genome), indicates the occurrence of chloroplast capture (*14*) as a result of inter-gene pool hybridisation and subsequent backcrossing. This finding is consistent with molecular phenotyping results (i.e., metabolic fingerprints); indeed, several intermediate phenotypes between Mesoamerican and Andean accessions were observed in European landraces, but these intermediate phenotypes are absent in accessions collected in the Americas (Fig. 2a, b and c). Notably, there was a significant correlation between the admixture coefficients and the PCA1 of the principal component analysis of the metabolic fingerprints for both the American and the European accessions, indicating a tight relationship between the phenotypic and genotypic differences due to the gene pool structure with a reduced difference in Europe due to admixture, particularly in the accessions with a Mesoamerican origin.

Following the nested procedure carried out by Rossi et al. (*15*) we investigated the common bean population structure within each gene pool from the Americas (see Supplementary Note 4.2). Using ADMIXTURE (*16*), we identified two main Mesoamerican groups (**M1**, **M2**) and three main Andean groups (**A1**, **A2**, **A3**) (Fig. 1b, c). In the centres of domestication, there was little evidence of admixture between gene pools. There were only four Andean accessions (two from Peru, two from Chile) with genetic assignment consistent with introgression from the Mesoamerican gene pool (Fig. 1c).

We considered the geographic distribution of the accessions from the Americas that showed low admixture between gene pools (q_i >99%; pure American accessions) and their phenotypic data on growth habits and photoperiod sensitivities (our results and passport information). There was a clear correspondence between these genetic groups in our sample from the Americas and the well-known common bean eco-geographic races described by Singh et al. (*17*): **M1** corresponded to the higher-altitude Durango and Jalisco races, which originated primarily in northern and southern Mexico, respectively; **M2** corresponded to the lower-altitude Mesoamerican race, which is mostly insensitive to

the photoperiod and is distributed in lowland Mexico and Central America; **A1** corresponded to the Nueva Granada race, which is generally insensitive to the photoperiod; **A2** corresponded to the Peru race, which includes entries with vigorous climbing growth habits and sensitivity to the photoperiod; and **A3** corresponded to the Chile race, which has also been identified in archaeological samples from northern Argentina that date from 2,500 to 600 years ago (*18*). The identification of these well-defined ancestral genetic groups in the Americas offers a robust basis to study the inter-gene pool and inter-race introgression that might have promoted adaptation to European environments.

Asymmetric introgression and recombination between gene pools at the basis of European common bean adaptation

Genetic assignment at the chromosome level (ChromoPainter v2.0; *19*) was used to study inter-gene-pool hybridisation and introgression in the evolutionary history of common bean in Europe. The genetic groups identified in American accessions due to the low levels of admixture were used as reference populations for the "chromosome painting" of the European genotypes. We defined as donor (reference/founder) populations the two Mesoamerican (M1, M2) and three Andean groups (A1, A2 A3) identified with ADMIXTURE (*16*) in the Americas (see previous paragraph). On this basis we attributed, for each European genotype, all the SNPs and the chromosomal regions to the various ancestries, taking into account within-accession recombination breakpoints (see Supplementary Note 4). Using this approach, we were also able to detect recombination events between gene pools at the whole-genome level, also in accessions that showed limited introgressions (e.g., 1%). Overall, 71 European accessions were attributed to the Andean gene pool (EU_AND) and 43 were assigned to the Mesoamerican gene pool (EU_MES), in agreement with the results obtained with admixture analysis ($r=0.99$, $p<0.01$) and confirming previous knowledge about the prevalence of Andean genotypes in Europe (*5*, *20*). Globally, the inferred amount of per-accession introgressed material was different between EU_MES and EU_AND accessions (two sided K-S test, $P=3.3\times10^{-3}$), showing a median proportion of 4.7% and 9.2%, respectively. EU_MES accessions had from 0.01% to 44.9% of their genome introgressed from the other genetic pool, with only one EU_MES accession showing less than 1% of their genome introgressed (Figure 3a, Supplementary Note 4, SN4_Fig.24; SN4_Fig.25). These

proportions were similar in the EU_AND samples, ranging from 0% to 42.2% (Figure 3a), although two EU_AND accessions showed no introgression and 10 accessions showed limited introgression from the other genetic pool (i.e., <1%; Figure 3a). The pervasive effect of admixture in European individuals was confirmed by the presence of several accessions showing more than 20% of their genome composed by introgressed material, in both EU_MES (8 of 43 accessions, 18.6%) and EU_AND (11 of 71, 15.5%) groups (Fig. 3a).

Interestingly, the median length of the introgressed genomic segments was higher for the EU_AND than the EU_MES accessions (EU_AND: 217 kb, EU_M: 70 kb; Fig. 3b, c), with more extended introgressed regions into EU_AND particularly on chromosomes PvChr02, PvChr05, PvChr06 and PvChr09 (Figure 3b, c). The EU_AND accessions carried longer Mesoamerican introgressed haplotypes reflecting a more recent introgression of Mesoamerican genomic fragments into the Andean genotypes as compared to the opposite direction. Moreover, several genomic regions that carry haplotypes with a specific Andean ancestry are near fixation in the European accessions group. Many of these regions show clear selection signatures (e.g., position 46 Mb on chromosome Pv01, which carries the *OTU5* locus likely involved in phosphate starvation response; Supplementary Note 4, SN4_Fig. 25) and significant genome-wide association study peaks for flowering time (e.g., position 37.9 Mb on chromosome Pv9; *LHY*), consistently with the hypothesis that natural selection was a crucial step for the adaptation of common bean in Europe.

Our results indicate that the Andean types represent the first population successfully disseminated across Europe. This is shown by the smaller introgression segments of Andean origin and the higher observed frequencies of common bean of Andean origin in Europe. Our data are also consistent with available historical records. Indeed, the first unambiguous evidence of the introduction of common bean in Europe is of Andean cultivars (*21*) probably introduced in Spain by Francisco Pizarro in 1529, after the exploration of Peru. Piero Valeriano Bolzanio received the common bean from Giuliano de Medici (Pope Clement VII, 1523–1534), which had been donated to the same pope by Emperor Charles V's Spanish emissaries from Sicily, where the beans seeds were grown and harvested. The very detailed writing of Piero Valeriano Bolzanio refers to common beans seeds describing in depth several phenotypic traits supporting their Andean origin, as also recently suggested by Myers et al. (*22*).

Valeriano documented his efforts, along with a network of collaborators in the northeast of Italy, Slovenia and Dalmatia, to grow and reproduce beans starting in 1532 (*21*). Thus, historical information and timelines support our results suggesting an early introduction of the Andean gene pool in Europe. It may also explain the high frequency of Nueva Granada (A1) Andean ancestries (Fig. 1e) in Sicily, the south and the northeast of Italy, in Slovenia and Croatia as they could have been among the firsts European areas of common bean cultivation with the early introduced Andean genotypes likely from the Nueva Granada race.

Adaptive differences among common beans in the New World could also have played a role in distributions in Europe. For example, Mesoamerican genotypes from M1 (Durango-Jalisco race) can be highly sensitive to photoperiod. Thus, it is possible that the Mesoamerican genotypes were not well adapted to many European environments, which would have limited their dissemination, particularly in central and northern Europe (Fig. 1e). In contrast, southern Spain, southern Italy, Sicily, North Africa, Madeira Island, and the Canary Islands are characterised by mild winters. In these environments, genotypes sensitive to photoperiod, late-flowering, or adapted to warmer conditions might have easily completed the crop cycle. As also reported by Gepts and Bliss (*5*) and Bellucci et al. (*23*), we also found that Mesoamerican genotypes are more frequent in specific European regions, particularly in southeastern Europe (Fig. 1e), which also suggests that the history of their introduction may have contributed to their actual distribution. As with the role of Charles V and Pope Clement VII in the early dissemination of the Andean beans, the political subdivision of Europe and the Mediterranean basin in the 16th century likely has impacted the dissemination of the Mesoamerican gene pool. The Ottoman Empire dominated the southern shores of the Mediterranean, the Nile Basin, the Red Sea into eastern Africa, and southeastern Europe, spanning the area from modern-day Greece to Austria. The prevalence of Mesoamerican genotypes in eastern Africa and China (*23*, *24*) might result from their initial introduction into Africa from Spain during the Ottoman Empire, which extended its rule in northeastern Africa controlling the exchange with China through the Silk Road. An important role of political/cultural factors associated with the dissemination of bean genotypes in Europe is also suggested by the lack of significant spatial and ecological patterns between genetic, geographic, ecological distances. Indeed, the routes of dissemination based on cultural and political factors are often

independent from geographic and environmental distances thus less prone to determine correlations between genetic distances and geographical or environmental differences (see Supplementary Note 4). We used the geographic distribution of the five ancestry components (A1, A2, A3 M1, and M2) in Europe, as inferred from the ChromoPainter analysis, in an association analysis with biogeographical variables. Interesting, ancestry components of race Chile (A3) is negatively correlated with latitude (Supplementary Note 4, SN4_Tab.11; r=-0.35, p=0.0001) and it is never observed above the 47th parallel (Fig. 1e). Moreover, the A3 component is associated with warmer climates, particularly the maximum temperature in September (Supplementary Note 4, SN4_Tab.11; r=0.29, P <0.002). Also, the Chile race (genetic group A3) is more sensitive to the photoperiod than the Nueva Granada race (A1) (Fig. 2d), which highlights the importance of this trait in the dissemination of common bean in Europe (Fig. 1e, f). However, compared to the Mesoamerican introductions, the race Chile genotypes were more uniformly distributed in Europe across different longitudes (Fig. 1e). This is also congruent with the hypothesis of earlier introduction of Andean as compared to Mesoamerican genotypes. Considering the other genetic groups, only a few weak associations with environmental variables were detected (see Supplementary Note 4).

Analysis of the genetic diversity in the European common bean

Due to evidence of widespread admixture in Europe, we developed a masked dataset of European accessions where all of the introgressed alleles or those with an ambiguous assignment were filtered out (see Supplementary Note 4). By this masking approach, we studied the nucleotide diversity using the frequencies of two reconstructed non-admixed populations of Andean and Mesoamerican origins. From each European genotype, all the Andean SNPs were separated from the Mesoamerican SNPs and included in the two masked datasets. Using both the unmasked and masked datasets, common bean from the Americas showed moderately higher nucleotide diversity than European ones (see Supplementary Note 4.3), which appears to be due to the introduction bottleneck in Europe (Fig. 4a, b). Differently, a contrasting pattern was seen when the American and European genetic diversities were compared within their respective gene pools (AM_AND *vs* EU_AND; AM_MES *vs* EU_MES). Due to admixture, the European diversity was always higher than that in the Americas with the unmasked

dataset, while the opposite was found using the masked dataset (Fig. 4b). In other words, particularly the Andean common bean from Europe shows a higher diversity than those from America, because of admixed ancestry with the Mesoamerican gene pool. This comparison of the estimated levels of genetic diversity in Europe confirms the key role of inter-gene pool hybridisation and recombination in shaping the diversity of the European common bean. Interestingly, as compared to the Mesoamerican gene pool, the Andean gene pool showed higher contrast between unmasked and masked datasets, as the diversity of the Andean germplasm in the centre of origin still reflects the bottleneck that occurred in the Andean wild populations during the expansion into South America before domestication (*25*) that was reflected in the domesticated pool (*26*). Indeed, here in non-coding regions we have detected reduced diversity (θ_π/bp) of approximately 70% in the Andes compared to Mesoamerica in the American accessions.

Considering the various genetic groups, there was significantly higher diversity in the Durango-Jalisco race (i.e., M1) compared to race Mesoamerica (M2) and in race Peru (A2) compared to the Nueva Granada race (A1). Moreover, a very low level of diversity was found in race Chile (A3). This, along with the NJ tree shown in Fig. 1d, indicates that race Peru in the Andes and race Durango-Jalisco in Mesoamerica were likely the first domesticated populations from which the other races were derived by secondary domestication associated with the loss of photoperiod sensitivity (Fig. 1f). Indeed, earliness and loss or reduction of photoperiod sensitivity were major traits under selection during common bean expansion in Europe. Considering the Andean gene pool, this was connected to the domestication pattern in the centre of origin. The earlier domesticated genotypes that are sensitive to the photoperiod were less successfully disseminated in Europe. Indeed, the relationship between the American and European genetic groups of Andean origin (as defined using the ChromoPainter approach; [*19*]; see Supplementary Note 4), coupled with the phenotypic data for flowering (Fig. 2d), show that genetic group A2 (Peru race) that was more sensitive to the length of the photoperiod was not successfully introduced into Europe due to the lack of adaptation (we observed a single exception of a highly admixed accessions; qA2=43.6%).

In contrast, the remaining two Andean genetic groups (A1 and partially A3) became widespread in Europe. An opposite scenario was seen for the Mesoamerican gene pool, especially for the Jalisco-Durango genotypes, where introgression appears to have been the critical element determining this

genetic group's dissemination in Europe (Fig. 1f, 2d, e). The Durango-Jalisco race (M1) showed very high levels of admixture in the European material due to introgression from the Mesoamerica race (M2) and the Andean populations (A1 and A3) (Fig. 2e) that likely contributed to the reduced sensitivity to photoperiod compared to the American counterpart (AM_M1) (Figure 2d) and to its spread over Europe (Figure 1e, f).

For the Andean genotypes, the parallel pattern of diversity associated with photoperiod sensitivity (Fig. 2d) suggests the occurrence of at least two steps of domestication: (i) primary domestication, as the domestication of the photoperiod sensitive populations (Peru race); and (ii) secondary domestication, which was characterised by reduced sensitivity to the photoperiod (Chile and particularly Nueva Granada races). This indicates that secondary domestication (*27*) was a crucial precondition for the successful dissemination of the Andean common bean in Europe (Fig 1f). For the Mesoamerican genotypes, an open question is where and when the introgression from the Andean gene pool occurred. We suggest that this is likely to have happened along the southern Mediterranean shore and in northern Africa, where the warmer climate might have favoured the Mesoamerican genotypes.

The average linkage disequilibrium (LD) decay in accessions from Europe and the Americas (Fig. 5a) is consistent with the historical differences between the gene pools and the effects of high inter-gene-pool hybridisation and introgression at the whole-genome scale in Europe. Admixture in Europe increased the molecular diversity (i.e., effective population size). It also generated new genome-wide admixture LD due to new combinations of alternative alleles in each gene pool. Thus, inter-gene-pool hybridisation followed by recombination reduced the LD at a long distance but as expected, has limited effect on LD decay at short distances as regions are directly inherited from the source populations (*28*). When we compared the accessions from the Americas and Europe, the LD decay was much faster over short distances (<1.5-2.0 Mb) in genotypes from the Americas. In contrast, over greater distances (>3 Mb), there was faster decay of LD in European populations (Fig. 5a). This reflects higher historical rates of recombination in the American sample over short distances and the effect of recombination due to the inter-gene-pool introgression in Europe over long distances. A similar pattern was seen when the Mesoamerican and Andean gene pools were analysed separately (Supplementary Note 7, SN7_Fig.45). However, the Andean accessions were characterised by higher baseline LD

levels. Indeed, the AM_M and AM_A populations reached r^2 of 0.2 at 500 kb and 1 Mb, respectively, while r^2 of 0.3 was reached at 250 kb and 1.5 Mb for the EU_MES and EU_AND samples, respectively.

Synonymous and missense mutations

The ratios between missense and loss-of-function mutations over synonymous mutations were used to reveal the patterns of genetic load across gene pools and continents. We observed a clear pattern in the genetic load that reflects the differences between the Andean and Mesoamerican origins, with the Andean accessions showing a higher genetic load due to the bottleneck before domestication (Fig. 4c). Interestingly, EU_AND, and to a lesser extent EU_MES, showed reduced genetic load when the loss-of-function pattern was considered (Fig. 4c). This suggests that the relatively short period of inter-gene-pool hybridisation, followed by selfing and recombination, promoted the purging of deleterious alleles accumulated in the European Andean pool. The role of hybridisation and subsequent recombination was also supported by the pattern of long-range LD in Europe compared to America (Fig. 5a). The pattern for private alleles (i.e., not identified in other gene pools or populations) in American and European accessions for low-frequency mutations (<5%) revealed a higher frequency of non-synonymous over synonymous mutations in Europe (i.e., a ratio of 1.44) (Supplementary Note 4, SN4_Fig.40). This might have resulted from the pattern of crop dissemination; indeed, this was probably characterised by the exchange of small quantities of seeds and several sequential bottlenecks, followed by rapid population growth at the single farm level. This might have resulted in the fixation of most mutations due to the small population size (i.e., founder effect). In this demographic context, most mutations could be fixed rapidly at the local level (within the population grown by a single farmer). However, there is the possibility that the purging of deleterious mutations, due to hybridisation following seed exchange among farmers and co-occurrence of different varieties in the same farmer fields (*20*), facilitated the combined effects of natural and human selection against deleterious recessive alleles and the capture of valuable variants.

Selection and adaptive introgression

We consider putatively '*adaptive introgression genomic regions*' (AIGR) to be those simultaneously meeting the following requirements: 1) an 'excess of introgression' based on Chromopainter results (Supplementary Note 6.), 2) a signature of selection detected using the hapFLK method (that analyse multiple populations jointly considering their hierarchical structure; *29*), and 3) an outlier F_{ST} value between Europe and America, suggesting a different pattern of diversity between America and Europe (see Supplementary Notes 6 and 8). Adaptive introgression appears to be particularly important for the evolution of the European genotypes of Mesoamerican origin (EU_MES). We identified 44 Andean regions with an excess of introgression (of which 22 AIGR) shared by >90% of the European genotypes, spanning all chromosomes except for chromosome PvChr11 and ranging from 5.016 kb to 118.424 kb in length. (Supplementary Note 6). An Andean allele frequency of 96% was also detected along a genomic segment of PvChr01, near the gene *Phvul.001G203400*, which is orthologous to *OVARIAN TUMOR DOMAIN-CONTAINING DEUBIQUITINATING ENZYME 5* (*OTU5*) (Supplementary Note 8, SN8_ Supplemental Dataset 6; row 16), which might be involved in phosphate starvation response, according to the ortholoous function in *Arabidopsis thaliana*, where it recalibrates and maintains cellular inorganic phosphate homeostasis (*30*, *31*). Moreover, adaptive introgression with a strong signature of selection has been identified for many *P. vulgaris* flowering-related genes (see Supplementary Note 8.3) orthologous to those involved in the four major *A. thaliana* flowering pathways for which the flowering role has been established and described (Fig. 6). Significant examples here are seen for *Phvul.009G259400* and *Phvul.009G259650* (Supplementary Note 8, SN8_ Supplemental Dataset 6; rows 90, 92), which are orthologues of the *LATE ELONGATED HYPOCOTYL* (*LHY*) gene of *A. thaliana*, with both located within the same introgressed region of chromosome PvChr09, and characterised by an Andean allele frequency of 96% in the European genotypes. Notably, *LHY* encodes a transcription factor that is a pivotal oscillator in the morning stage of the circadian clock and is interconnected with *CIRCADIAN CLOCK ASSOCIATED 1* (*CCA1*) indirect suppression of the middle, evening, and night complex genes (*32*) (Fig. 6). In the EU_MES population, these two *LHY* orthologues show private and significant inter-chromosomal LD with *Phvul.011G050600* (Supplementary Note 8, SN8_ Supplemental Dataset 6; row 97) (Fig. 6), which is orthologous to the *A. thaliana VERNALISATION 1* (*VRN1*) and *RELATED TO VERNALISATION1 1* (*RTV1*) genes (Fig. 6).

In *A. thaliana*, *VRN1* and *RTV1* are essential for activation of the floral integrator genes after exposure to long-term cold temperatures (*33*). The identification of the inter-chromosomal LD between these flowering genes, which are private (i.e., not identified in other gene pools) to the EU_MES accessions, may suggest the effect of epistatic selection. An analogous example is seen for *Phvul.001G204600* (Supplementary Note 8, SN8_ Supplemental Dataset 6; row 29) and *Phvul.001G204700* (Supplementary Note 8, SN8_ Supplemental Dataset 6; row 30), which are orthologous to *LUMINIDEPENDENS* (*LD*) and *NUTCRACKER* (*NUC*), respectively. Both *Phvul.001G204600* and *Phvul.001G204700* are located in a region of PvChr01 and are in private inter-chromosomal LD with *Phvul.003G137100* (Supplementary Note 8, SN8_ Supplemental Dataset 6; row 38) on PvChr03 (Fig. 6), which is orthologous to *GATA, NITRATE-INDUCIBLE, CARBONMETABOLISM INVOLVED* (*GNC*), and *CYTOKININ-RESPONSIVE GATA FACTOR 1* (*CGA1*). All of these genes are functionally involved in the flowering process. *LD* is one of the eight genes identified so far in the autonomous pathway of *A. thaliana* that acts as a repressor of *FLOWERING LOCUS C* (*FLC*) and consequently promotes the transition from vegetative to flowering stages. *NUC* encodes a transcription factor that positively regulates photoperiodic flowering by modulation of sugar transport and metabolism via the *FLOWERING LOCUS T* (*FT*) gene (*34*; *35*). The paralogous *GNC* and *CGA1* genes act in a redundant way to promote greening downstream from the gibberellic acid signalling network (*36*).

We found that genes identified in adaptive introgression regions (AIGR) represent ~17% of genes identified by the selection scan with HapFLK (see Supplementary Notes 6 and 8) and show enrichment in seven Gene Ontology categories including GO:0048523, negative regulation of cellular processes; GO:0010228, vegetative to the reproductive phase transition of the meristem; GO:0042445, hormone metabolic processes; GO:0009657, plastid organisation; GO:0042440, pigment metabolic processes; GO:0009733, response to auxin; and GO:0070647, protein modification by small protein conjugation or removal (Supplementary Note 8.2). Enrichment analysis confirmed that the primary trait under selection for adaptive introgression is flowering time. Still, it also highlighted the important role of genes associated with the adaptation to abiotic and biotic stresses.

Conclusions

Here, we show that adaptive introgression was crucial for the successful dissemination and adaptation of common bean in Europe. We use a combination of genome resequencing, molecular (metabolomics) and classical phenotyping, and data analysis approaches, such as chromosomal-level genetic assignment and environmental association. Our data indicate that the Andean gene pool was the first to be successfully introduced in Europe most likely from Francisco Pizarro's expedition to northern Peru in 1529. Most of the genetic background of the European common bean of Andean origin has been determined by the secondarily domesticated Nueva Granada (A1) and Chile (A3) races. In contrast, the Peru race, which is more sensitive to the photoperiod, contributed little to the European common bean germplasm. Indeed, the secondary domestication of these Andean races that was related to the latitudinal expansion of the cultivation areas from the Andean centres of origin was the key element that guaranteed the successful dissemination in the Old World of the Andean common bean. In contrast, the key element for dissemination of the Mesoamerican gene pool in Europe was the adaptive introgression of flowering time genes from the Andean genotypes. Indeed, our genomic analysis indicated that Andean types were rapidly disseminated, while Mesoamerican genotypes were eventually disseminated in Europe only following introgression from the Andean types. As expected, selection strongly influenced common bean orthologues of the major flowering pathways described for *A. thaliana* and environmental adaptative traits, such as the orthologues of *OTU5*, which is involved in the inorganic phosphate starvation response. Finally, we suggest that the pattern of dissemination of common bean was greatly affected by political factors and constraints present in the XVI century as the interaction between the political and religious power in western Europe and the subdivision of the European continent into Islamic and Christian countries.

Acknowledgements

This manuscript is dedicated to our former collaborator Monica Rossi who passed away at the age of 44 in 2019.

This work was carried out within the BEAN ADAPT project, founded through the ERA-CAPS Programme, 2014 Call, Expanding the European Research Area in Molecular Plant Sciences

S.A. and A.R.F additionally acknowledge the funding of the European Commission for the PlantaSyst project (SGA-CSA no. 664621 and no. 739582 under FPA no. 664620)

Author contributions

A.G., P.G., A.R.F., S.A.J. and R.P. conceived and managed the project; E.Be. and R.P. wrote the article; E.Be., A.B., C.X., E.Bi., M.R., S.A., V.D.V., K.N., G.C., G.F., P.L.M., A.G., P.G., A.R.F., S.A.J. and R.P. contributed to the writing and drafting; E.Be., A.B., C.X., E.Bi., M.R., S.A., V.D.V., T.G., K.N., G.C., G.F., L.N., J.J.F., A.C., G.A., P.L.M., G.B., A.G., P.G., A.R.F., S.A.J. and R.P. contributed to the editing of the article; E.Be., A.B., C.X., E.Bi., M.R., S.A., V.D.V., T.G., K.N., G.C., G.F., E.M., E.T., L.N., A.A., C.L., J.J.F., A.C., P.L.M., GB, A.G., A.R.F., S.A.J. and R.P. contribute to write and organize the supplementary material; E.Be., E.Bi., S.A., T.G., K.N., L.N., G.L., L.J., J.J.F. and A.C. performed DNA extraction, field and greenhouse experiments; C.X., J.H.S. and S.A.J. conducted sequencing and primary bioinformatic analysis; S.A. and A.R.F. conducted metabolomics analysis; A.B., E.Bi., M.R., G.F., E.T., C.L., G.B. and R.P. contributed to data analysis; E.Be., A.B., C.X., E.Bi., M.R., S.A., V.D.V., G.B., A.R.F., S.A.J. and R.P. contributed to coordinate data analysis and data integration. All authors read and approved the article.

Data availability

The raw sequence reads generated and analyzed during the current study are available in the Sequence Read Archive (SRA) of the National Center of Biotechnology Information (NCBI) with the following BioProject number PRJNA573595.

References

1. Crosby AW (1972) The Columbian Exchange; Biological and Cultural Consequences of 1492. Westport, Conn.: Greenwood Pub.Co.

2. Cortinovis G, Di Vittori V, Bellucci E, Bitocchi E, Papa R (2020) Adaptation to novel environments during crop diversification. Curr. Opin. Plant Biol., 56:218–222, doi: 10.1016/j.pbi.2019.12.011

3. Brandenburg J-T, Mary-Huard T, Rigaill G, Hearne SJ, Corti H, Joets J, et al. (2017) Independent introductions and admixtures have contributed to adaptation of European maize and its American counterparts. PLoS Genet 13(3): e1006666. https://doi.org/10.1371/journal. pgen.1006666

4. Gepts P (2002) A Comparison between Crop Domestication, Classical Plant Breeding, and Genetic Engineering. Crop Sci., 42: 1780-1790. https://doi.org/10.2135/cropsci2002.1780

5. Gepts P and Bliss FA (1988) Dissemination pathways of common bean (*Phaseolus vulgaris*, Fabaceae), deduced from phaseolin elettrophoretic variability. II Europe and Africa. Econ. Bot. 42, 86-104. https://doi.org/10.1007/BF02859038

6. Santalla M, Rodiño AP, De Ron AM (2002) Allozyme evidence supporting southwester Europe as a secondary center of genetic diversity for common bean. Theor Appl Genet 104:934–944

7. Sicard D, Nanni L, Porfiri O, Bulfon D and PapanR (2005) Genetic diversity of *Phaseolus vulgaris* L. and *P. coccineus* L. landraces in central Italy. Plant Breed., 124: 464-472. https://doi.org/10.1111/j.1439-0523.2005.01137.x

8. Zeven AC, Waninge J, van Hintum T, Singh SP (1999) Phenotypic variation in a core collection of common bean (*Phaseolus vulgaris* L.) in the Netherlands. Euphytica 109:93-106 doi: 10.1023/A:1003665408567

9. Angioi SA, Rau D, Attene G, Nanni L, Bellucci E, Logozzo G, Negri V, Spagnoletti Zeuli PL, Papa R (2010) Beans in Europe: origin and structure of the European landraces of *Phaseolus vulgaris* L. Theor Appl Genet. 121:829-843, doi: 10.1007/ s00122-010-1353-2

10. Angioi SA, Rau D, Nanni L, Bellucci E, Papa R, Attene G (2011) The genetic make-up of the European landraces of the common bean. Plant Genet Resour. 9:197-201 doi: doi:10.1017/S1479262111000190

11. Ariani A, Berny Mier y Teran JC, Gepts P (2018) Spatial and temporal scales of range expansion in wild *Phaseolus vulgaris*. Mol. Biol. Evol. 35(1): 119-131 doi: 10.1093/molbev/msx273

12. Bitocchi E, Rau D, Bellucci E, Rodriguez M, Murgia ML, Gioia T, Santo D, Nanni L, Attene G, Papa R (2017) Beans (*Phaseolus* ssp.) as a model for understanding crop evolution. Front Plant Sci. 8:722, doi: 10.3389/fpls.2017.00722

13. Schmutz J, McClean P, Mamidi S, et al. (2014) A reference genome for common bean and genome-wide analysis of dual domestications. Nat Genet 46, 707–713. https://doi.org/10.1038/ng.3008

14. Tsitrone A, Kirkpatrick M, Levin DA (2003) A model for chloroplast capture. Evolution 57(8):1776-82. PMID: 14503619 doi: 10.1111/j.0014-3820.2003.tb00585.x

15. Rossi M, Bitocchi E, Bellucci E, Nanni L, Rau D, Attene G, Papa R (2009) Linkage disequilibrium and population structure in wild and domesticated populations of *Phaseolus vulgaris* L. Evol. Appl. 2, 504–522, doi: 0.1111/j.1752-4571.2009.00082.x

16. Alexander DH, Novembre J, Lange K (2009) Fast model-based estimation of ancestry in unrelated individuals. Genome res., 19(9):1655-64

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

18. Trucchi E, Benazzo A, Lari M, et al. (2021) Ancient genomes reveal early Andean farmers selected common beans while preserving diversity. Nat. Plants 7, 123–128, doi: 10.1038/s41477-021-00848-7

19. Lawson DJ, Hellenthal G, Myers S, Falush D (2012) Inference of population structure using dense haplotype data. PLoS Genet. 8(1): e1002453, doi:10.1371/journal.pgen.1002453

20. Zeven AC (1997) The introduction of the common bean (*Phaseolus vulgaris* L.) into western Europe and the phenotypic variation of dry beans collected in The Netherlands in 1946. Euphytica 94, 319–328, doi: 10.1023/A:1002940220241

21. Perale M (2001) Milacis Cultus Aperire Paramus. “De milacis cultura” di Pierio Valeriano, il Primo Testo Europeo Dedicato al Fagiolo. Belluno: Momenti AiCS

22. Myers JR, Formiga AK and Janick J (2022) Iconography of beans and related legumes Following the Columbian Exchange. Front. Plant Sci. 13:851029. doi: 10.3389/fpls.2022.851029

23. Bellucci E, Bitocchi E, Rau D, Rodriguez M, Biagetti E, Giardini A, et al. (2014) Genomics of origin, domestication and evolution of *Phaseolus vulgaris*. Genomics of Plant Genetic Resources, eds R. Tuberosa, A. Graner, and E. Frison (Berlin: Springer), 483–507. doi: 10.1007/978-94-007-7572-5_20.

24. Wu J, Wang L, Fu J, et al. (2020) Resequencing of 683 common bean genotypes identifies yield component trait associations across a north–south cline. Nat Genet 52, 118–125. https://doi.org/10.1038/s41588-019-0546-0

25. Bitocchi E, Nanni L, Bellucci E, Rossi M, Giardini A, Spagnoletti Zeuli P, Logozzo G, Stougaard J, McClean P, Attene G et al. (2012) Mesoamerican origin of the common bean (*Phaseolus vulgaris* L.) is revealed by sequence data. PNAS, USA 109: E788–E796

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

27. Meyer RS and Purugganan MD (2013) Evolution of crop species: genetics of domestication and diversification. Nat Rev Genet. 14(12):840-52. doi: 10.1038/nrg3605. PMID: 24240513

28. Chakraborty R and Weiss KM (1988) Admixture as a tool for finding linked genes and detecting that difference from allelic association between loci. PNAS, USA 85 (23) 9119-9123; doi: 10.1073/pnas.85.23.9119

29. Fariello MI, Boitard S, Naya H, SanCristobal M, Servin B (2013) Detecting signatures of selection through haplotype differentiation among hierarchically structured populations. Genetics, 193(3):929-41

30. Yen MR, Suen DF, Hsu FM, Tsai YH, Fu H, Schmidt W, Chen PY (2017) Deubiquitinating enzyme OTU5 contributes to DNA methylation patterns and is critical for phosphate nutrition signals. Plant Physiol. 175(4):1826-1838, doi: 10.1104/pp.17.01188

31. Suen DF, Tsai YH, Cheng YT, Radjacommare R, Ahirwar RN, Fu H, Schmidt W (2018) The deubiquitinase OTU5 regulates root responses to phosphate starvation. Plant Physiol. 176(3):2441-2455, doi: 10.1104/pp.17.01525

32. Adams S, Manfield I, Stockley P, Carré IA (2015) Revised morning loops of the *Arabidopsis* circadian clock based on analyses of direct regulatory interactions. Plos One 10(12): e0143943, doi: 10.1371/journal.pone.0143943

33. Heo JB, Sung S, Assmann SM (2012) Ca2+-dependent GTPase, extra-large G protein 2 (XLG2), promotes activation of DNA-binding protein related to vernalization 1 (RTV1), leading to activation of floral integrator genes and early flowering in *Arabidopsis*. J. Biol. Chem. 287(11):8242-8253, doi: 10.1074/jbc.M111.317412

34. King RW, Hisamatsu T, Goldschmidt EE, Blundell C (2008) The nature of floral signals in *Arabidopsis*. I. Photosynthesis and a far-red photoresponse independently regulate flowering by increasing expression of FLOWERING LOCUS T (FT). J. Exp. Bot. 59(14):3811–3820, doi: 10.1093/jxb/ern231

35. Seo PJ, Ryu J, Kang SK, Park CM (2011) Modulation of sugar metabolism by an INDETERMINATE DOMAIN transcription factor contributes to photoperiodic flowering in *Arabidopsis*. Plant J. 65(3):418-429, doi: 10.1111/j.1365-313X.2010.04432.x

36. Richter R, Bastakis E, Schwechheimer C (2013) Cross-repressive interactions between SOC1 and the GATAs GNC and GNL/CGA1 in the control of greening, cold tolerance, and flowering time in *Arabidopsis*. Plant Physiol. 162(4):1992-2004; doi: 10.1104/pp.113.219238

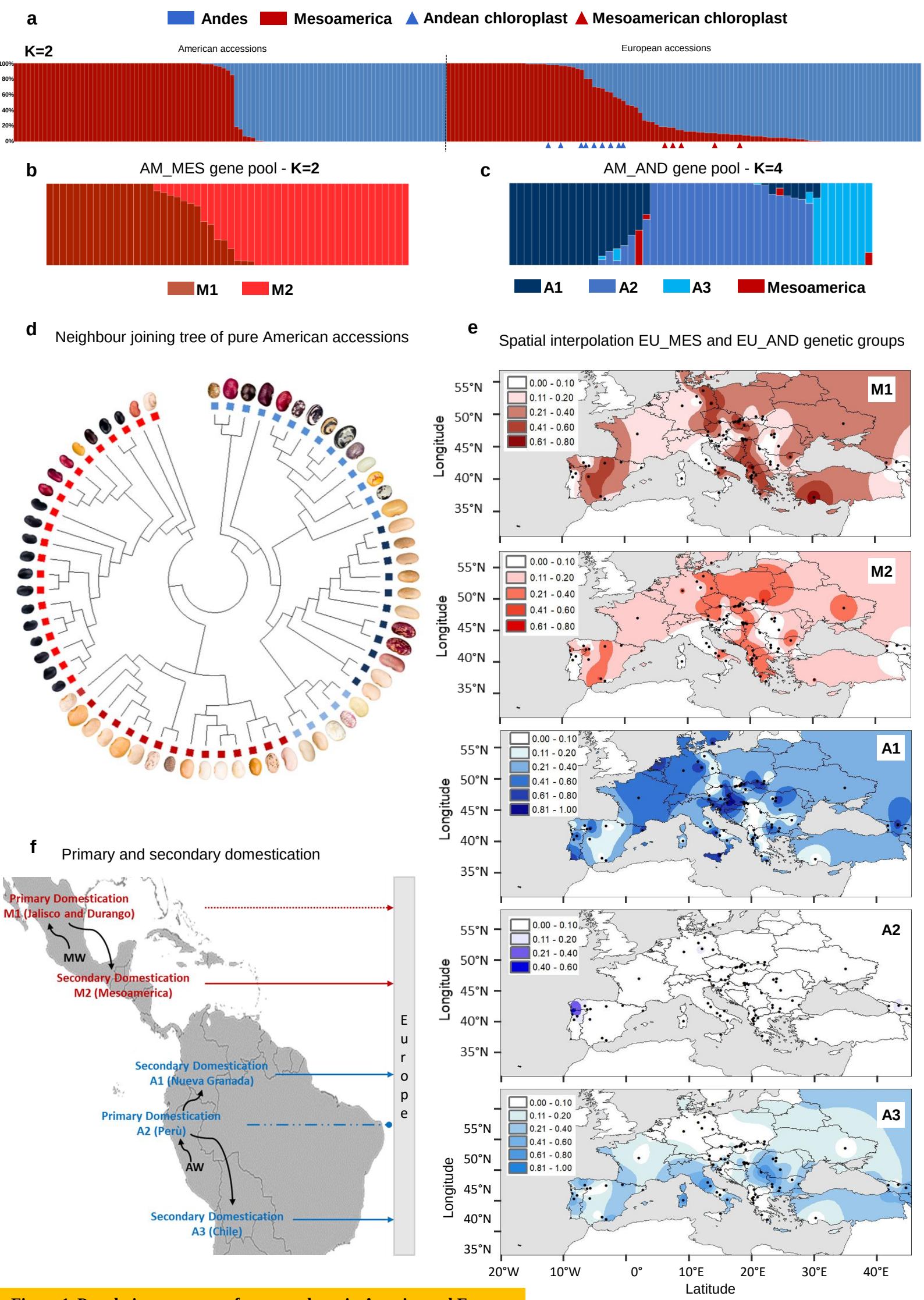
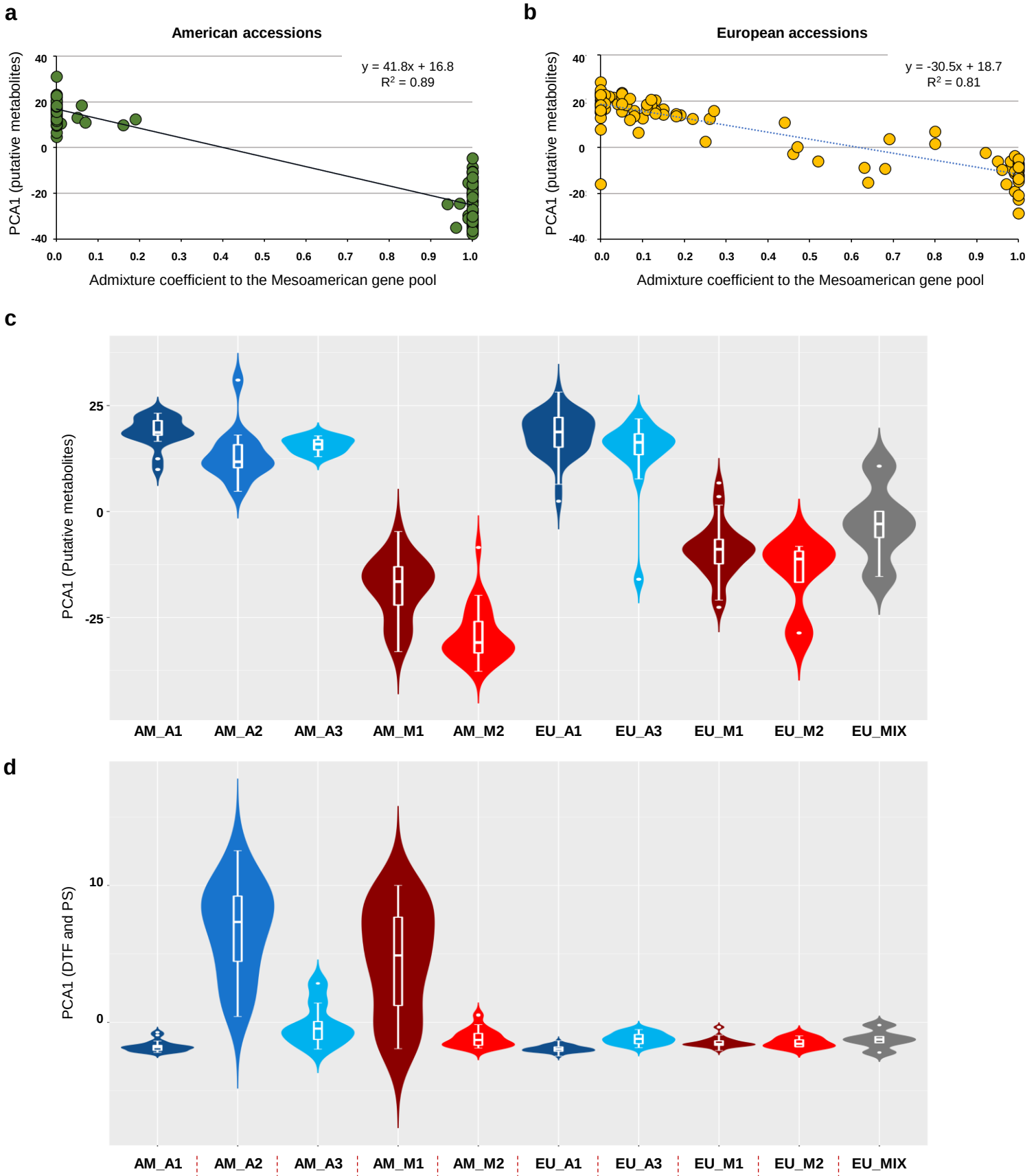


Figure 1. Population structure of common bean in America and Europe.

Figure 1. Population structure of common bean in America and Europe.

a, Admixture analysis (K=2) showing the inferred ancestry in the American (AM; left) and European (EU; right) accessions, with identification of two main gene pools, and several intermediates and admixed genotypes in Europe. **b**, Admixture plots for the AM Mesoamerican accessions (K=2) grouped by geographic origin (i.e., latitude and state), which identifies two main subgroups (M1, M2). **c**, Admixture plots for the AM Andean accessions (K=4) grouped by geographic origin (i.e., latitude, states), which identifies three Andean genetic subgroups (A1, A2, A3). **d**, Neighbour joining tree and seed pictures of the 66 pure American accessions. **e**, Spatial interpolation of the geographic distributions of the EU Mesoamerican (M1, M2) and EU Andean (A1, A2, A3) ancestry components in Europe, as inferred by ChromoPainter analysis. **f**, Primary and secondary domestications of Mesoamerican and Andean genetic groups/races in America. Loss of photoperiod sensitivity during the secondary domestication was a key factor for the introduction of the Andean gene pool (genetic groups A1 and A3; Races Nueva Granada and Chile, respectively) and the Mesoamerican one (genetic group M2; Race Mesoamerica) in Europe (solid arrow). M1 genetic group (Race Durango/Jalisco) was successfully introduced in Europe after introgression from other genetic groups characterized by absent or reduced photoperiod sensitivity (dashed arrow). A2 genetic group (Race Perù) was not introduced in Europe due to its high sensitivity to the photoperiod (discontinuous and truncated line).



	AM_A1	AM_A2	AM_A3	AM_M1	AM_M2	EU_A1	EU_A3	EU_M1	EU_M2	EU_MIX
N.	20	18	8	22	31	40	18	33	4	5
Flow_diff North_EU (N. & %)	*3* 15%	*13* 72%	*4* 50%	*13* 59%	*8* 26%	*2* 5%	*7* 39%	*2* 6%	*1* 25%	*2* 40%
Flow_diff South_EU (N. & %)	*0* 0%	*7* 56%	*0* 0%	*7* 32%	*0* 0%	*0* 0%	*0* 0%	*0* 0%	*0* 0%	*0* 0%

Figure 2. Phenotyping of the genetic structure.

Figure 2. Phenotyping of the genetic structure.

a, b, Molecular phenotypes (PCA1 from 1493 putative secondary metabolites, showing H^2 >0.65 on the entire dataset): **a**) 94 American; and **b)** 96 European accessions; confirm the subdivision into the two main groups seen using the admixture coefficient (derived from nuclear genomic data; K=2). Intermediate phenotypes and genotypes are seen in Europe. **c,** Violin plots showing the distribution for the PCA1 values related to secondary metabolites showing high heritability (H^2 >0.65), by genetic subgroups in the American and European accessions. PCA1 was used as a representative molecular phenotype, and it explains 25.7% of the total variance for these traits. **d**, Violin plots showing the PCA1 values related to the days to flowering (DTF) and photoperiod sensitivity (PS) by genetic subgroups in the American and European accessions. PCA1 was used as a representative phenotypic trait for DTF and photoperiod sensitivity, and it explains 68.8% of the total variance for these traits. **e**, Proportions of the genetic memberships (i.e., P(AM_A1), P(AM_A2), P(AM_A3), P(AM_M1), P(AM_M2), P(SAND), and P(SMES)) inferred from the donor accessions and composing the American and European accessions (grouped as mainly AM_A1, AM_A2, AM_A3_AM_M1, AM_M2, EU_A1, EU_A3, EU_M1, EU_M2, and EU_MIX) are shown in the pie charts below the corresponding groups and flowering data (n. of individuals and %) in Northen and in Southern Europe, related to the corresponding groups.

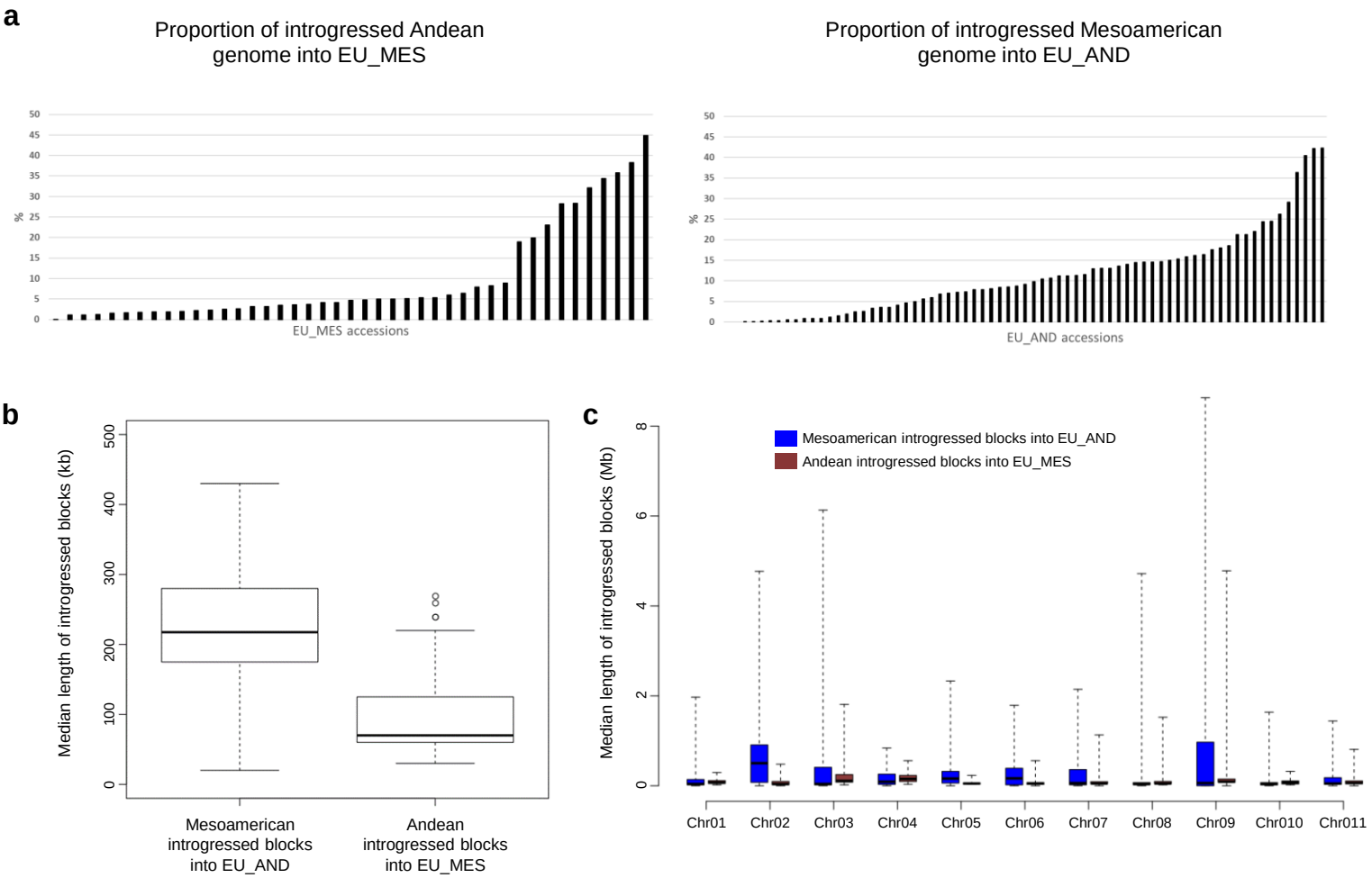


Figure 3. Mapping the introgression in the European common bean using Chromopainter.

Figure 3. Mapping the introgression in the European common bean using Chromopainter.

a, Proportion of introgressed genome in the Mesoamerican (EU_MES; n=43) and Andean (EU_AND; n=71) groups. **b**, **c**, Boxplots showing the median length of the introgressed blocks identified in each of the EU_AND and EU_MES accessions across all of the chromosomes (**b**) and the median length of the introgressed blocks identified in each of the EU_AND and EU_MES individuals by chromosome (**c**).

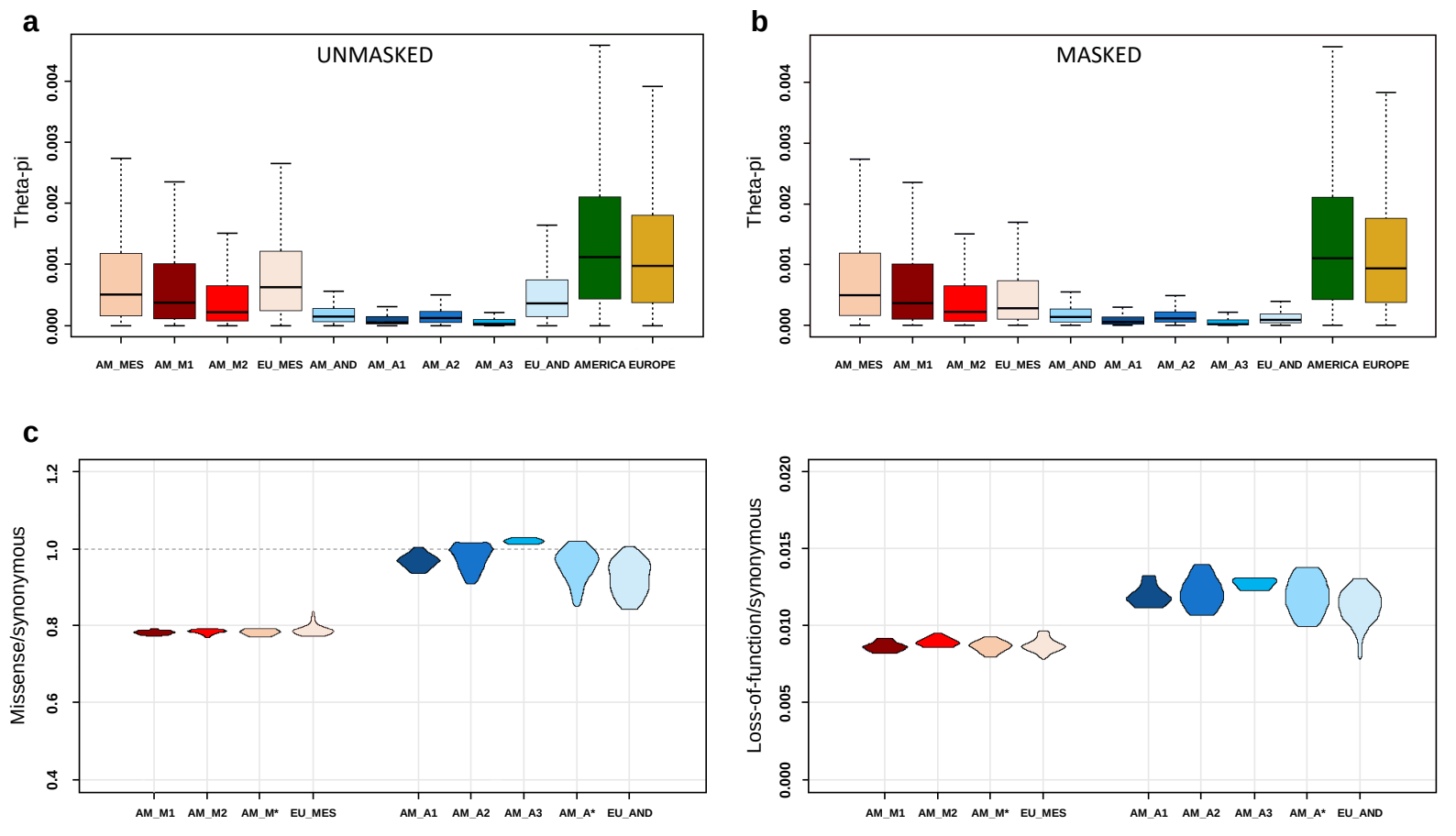


Figure 4. Boxplots of θπ averaged over 100-kb non-overlapping sliding windows, and genetic load.

Figure 4. Boxplots of θπ averaged over 100-kb non-overlapping sliding windows, and genetic load.

a, Genetic diversity computed using whole chromosomes and the unmasked dataset. **b**, Genetic diversity computed after the admixture masking process using whole chromosomes and linkage disequilibrium decay according to the physical distance. **c**, Genome-wide measure of genetic load in the American and European accessions; the ratios are shown for missense (left) and loss-of-function (right) over synonymous mutations in the different groups. AM_M* and AM_A* are the admixed American accessions (not pure American individuals).

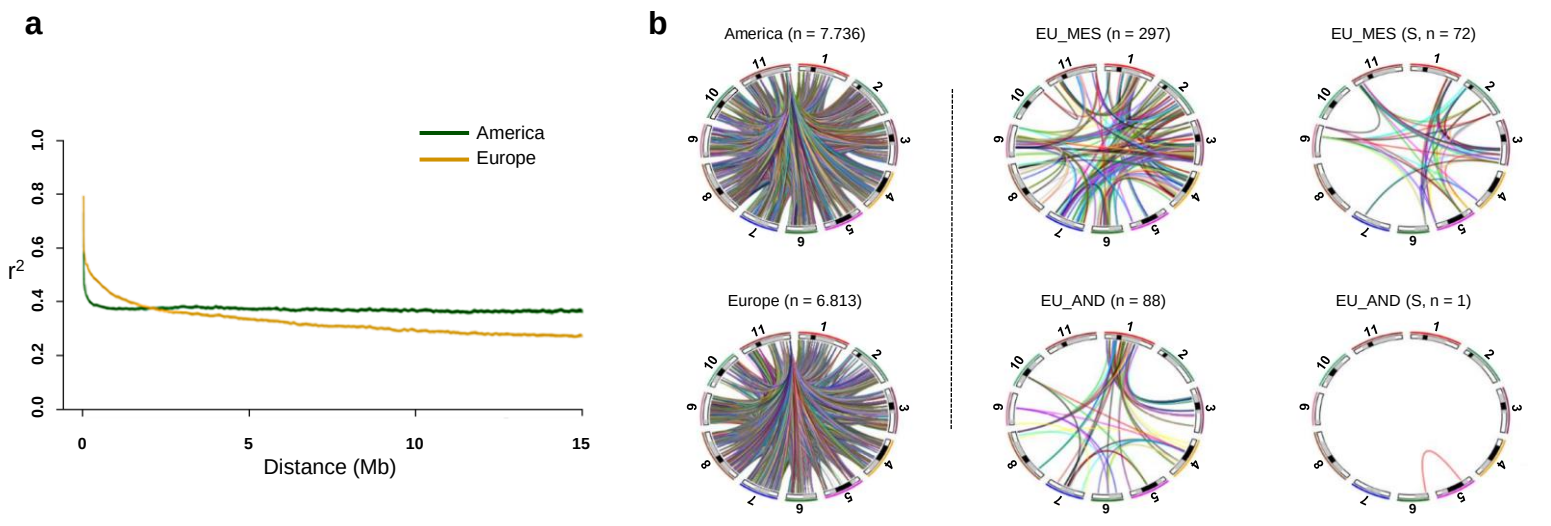


Figure 5. Linkage disequilibrium decay and inter-chromosomal linkage disequilibrium.

Figure 5. Linkage disequilibrium decay and inter-chromosomal linkage disequilibrium.

a, Linkage disequilibrium (LD) decay comparing the American and European accessions. **b**, Private inter-chromosomal linkage disequilibrium in American and European accessions (left), in the Mesoamerican and Andean European accessions (middle), and considering genomic regions under selection (S) in the Mesoamerican and Andean European accessions (right).

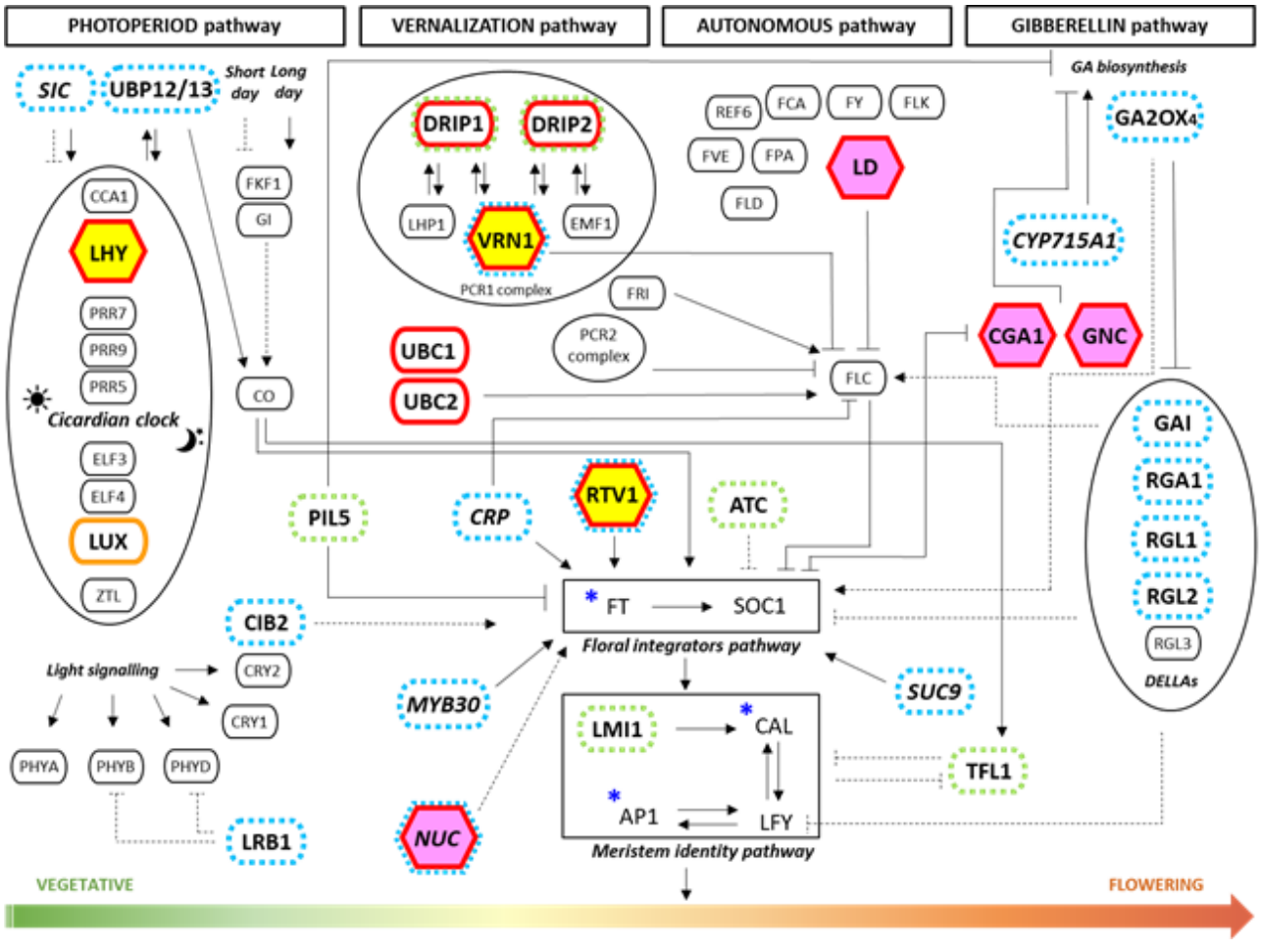


Figure 6. Candidate genes for adaptation.

Figure 6. Candidate genes for adaptation.

Schematic representation of the regulatory networks underlying the four major flowering pathways in *Arabidopsis thaliana*. The genes involved in the photoperiod, vernalization, autonomous and gibberellin pathways, that lead to the transition from vegetative to the flowering stage, are shown below to the corresponding pathway. Additional genes belonging to secondary pathways and that interact with the main regulatory flowering networks are reported in italic style. The orthologous to these flowering genes in *A. thaliana* were identified in *P. vulgaris* by using the OrthoFinder algorithm (see *Supplementary Note 8.1*). The genes with signature of selection and adaptive introgression, and those located in GWAS peaks for days to flowering and growth habit in common bean are highlighted as follow: **yellow hexagons**, the orthologous genes in common bean of LHY (Phvul.009G259400, Phvul.009G259650) and VRN1 and RTV1 (Phvul.011G050600) showing private inter-chromosomal linkage disequilibrium in the EU_M pool (see *Supplementary Note 7*; Inter-chromosomal linkage disequilibrium); **pink hexagons**, the orthologous genes in common bean of LD (Phvul.001G204600), NUC (Phvul.001G204700), CGA1 and GNC (Phvul.003G137100) showing private inter-chromosomal linkage disequilibrium in the EU_M pool (see *Supplementary Note 7*; Inter-chromosomal linkage disequilibrium); **red outlines**, at least one orthologous gene in common bean showing signature of selection, introgression and with a significant differentiation (Fst index) between American and European accessions ($p<0.05$); **orange outlines**, at least one orthologous gene in common bean showing signature of selection with no significant Fst ($p<0.05$); **blue asterisks,** at least one orthologous gene in common bean showing signature of introgression; **dashed blue outlines**, at least one orthologous gene in common bean is located within 50 kb centered on a significant GWAS peak for the days to flowering (DTF); **dashed green outlines**, at least one orthologous gene in common bean is located within 50 kb centered on a significant GWAS peak for the growth habit (GH); **arrows**, positive regulation of gene expression; **truncated arrows**, repression of gene expression; **solid lines**, direct interactions; **dashed lines**, indirect interactions.

The candidate genes for adaptation or post-domestication process of the common bean in European environments, orthologous to those involved in flowering-related pathways, are reported between brackets: UBP12/13 (Phvul.007G234000); LHY (Phvul.009G259400, Phvul.009G259650); LUX (Phvul.011G062100); PIL5 (Phvul.001G168700); CIB2 (Phvul.008G133600); LRB1 (Phvul.006G109600); DRIP1/2 (Phvul.001G157400, Phvul.007G177500); VRN1, RTV1 (Phvul.011G050600); UBC1/2 (Phvul.003G191900); LD (Phvul.001G204600); TFL1, ATC (Phvul.001G189200); $GA2OX_4$ (Phvul.006G120700); CGA1, GNC (Phvul.003G137100); GAI, RGA1, RGL1, RGL2 (Phvul.001G230500); LMI1 (Phvul.001G184800, Phvul.001G184900); SIC (Phvul.008G182500); CRP (Phvul.008G142400); MYB30 (Phvul.008G041500); NUC (Phvul.001G154800, Phvul.001G204700, Phvul.011G074100); SUC9 (Phvul.004G085100, Phvul.004G085400, Phvul.004G085594); CYP715A1 (Phvul.007G071500).

Refer to **SN8_ Supplemental Dataset 5** for detailed information on each candidate gene (i.e., selection, top selection, Fst, introgression and GWAS data).

Methods

Plant Materials

Original seeds of 218 accessions of common bean (*P. vulgaris*) were collected from International Gene Banks or from individual Institutions/Organizations collections.

We produce Single Seed Descent (SSD) for 199 lines through at least three cycles of self-fertilization. For the remaining 19 accessions, one seed per accession was sampled directly from bank original seeds provided by the donor.

Experimental design and Phenotyping

Plants were grown across ten different environments both in fields and greenhouses, applying Long Day (7), Short Day (2) and intermediate photoperiod conditions. During the summers of 2016 and 2017 four field trials were carried out in Italy (Villa d'Agri, Marsicovetere, Potenza) and in Germany (Gatersleben IPK) (Supplementary Note 2.1). Six additional greenhouse experiments were performed with controlled condition in Golm (Potsdam, Germany), Potenza (Italy) and Villaviciosa (Spain) during 2016, 2017 and 2018 (Supplementary Note 2.2).

Classical phenotyping was carried out on the 199 SSD lines, focusing on two main traits, i) day to flowering (DTF) as the number of days from sowing until 50% of plants showed at least one open flower and ii) grow habit (GH) recorded as determinacy *versus* indeterminacy on a single plant basis. Photoperiod sensitivity (PS) was calculated as the ratio between DTF in long day experiments and DTF in a short-day experiment.

Descriptive statistics of the different phenotypic traits were calculated by using R (https://cran.r-project.org/) or JMP 7.0.0 (*37*). The restricted maximum likelihood (REML) model implemented in JMP 7.00 was used to calculate the least square means (LSM) and the Best Linear Unbiased Predictors (BLUPs) of each genotype. REML model was also used to calculate the broad sense heritability (h^2B) for each quantitative trait, by assuming genotypes and environments as random effects. Distribution of DTF in each environment and Pearson's pairwise correlation among environments were calculated using *corrplo*t and *PerformanceAnalytics* R packages (*38*, *39*).

Molecular phenotyping of the 199 accessions was performed on the first trifoliate fully expanded leaves harvested from the long-day conditions experiment with three biological replicates. Secondary metabolites were measured according to Perez de Souza et al. (*40*). For non-targeted metabolomics, chromatograms were processed, and peak detection and integration were performed using REFINER MS® 10.0 (GeneData, http://www.genedata.com). To explore the molecular phenotypic diversity, we performed a non-targeted metabolic fingerprinting analysis using the high-throughput LC-MS analysis. Mass signals which were not detected in ≥ 50% of the samples, and/or having ≤ 1000 peak intensity were excluded. H2B was analyzed following the same approach adopted before, setting genotype and Continent with random effect. The heritability was calculated based on 190 accessions (94 from Americas and 96 from Europe) having more than one replicate.

Sequencing, variant calling and annotation

Genomic DNA of the 199 SSD lines was extracted from frozen young leaves of plants grown in greenhouse and directly from seeds for the remaining 19 accessions. DNA was extracted using the Qiagen DNeasy Plant Mini Kit (#69106 Qiagen, Hilden, Germany) and it was sheared with Covaris E220 to fragment sizes of approximate 550 bp. PCR-free libraries were constructed according to manufacturer's instructions (KAPA HyperPrep Kit PCR-free). Paired-end sequencing libraries were sequenced on Illumina HiSeq2500 or HiSeq4000 sequencers and labeled with different barcodes.

Sequencing data were aligned to the common bean reference genome (V2.0) (*13*) using BWA-mem (V0.7.15) (*41*). The unmapped reads were mapped to the *P. vulgaris* chloroplast genome (NCBI Entry: NC_009259). In both cases, SNP calling was performed using SAMtools (*42*) and GATK (V3.6) (*43*; *44*). In SAMtools, duplicated reads were removed with "rmdup", and SNPs were discovered with "mpileup" for filtered high quality alignments (-q 10) and bases (-Q 20) and then genotyped with BCFtools (*45*). With GATK, duplicated reads were sorted and filtered with Picard (V2.4.1) (http://broadinstitute.github.io/picard) for sequencing duplicates. Variants were then called following the best practices reported in the Genome Analysis ToolKit (GATK) and pre-filtered using the recommended parameters for hard filtering. Chromosomal and overlapping SNPs reported by both methods were retained and the genotypes produced by GATK were selected. We applied an additional filtering with VCFtools (*46*) (--minDP 3 --max-missing 0.5 --maf 0.05) and

we excluded SNP sites whose proportions of heterozygous genotypes were higher than 0.01. SNPs were annotated with SnpEff (V4.3s) (*47*).

Population structure analysis

A population structure analysis was conducted on the SAMtools/GATK overlapping SNP callset for which a further filtering was performed. Only the genomic positions having a QUAL >=30 and a global depth of coverage between 1/3 and 4 times the mean value, were retained and individual genotypes called using two reads or less were marked as missing data. Imputation and phasing were performed with Beagle (*48*). ADMIXTURE v1.3 (*16*) was used for the population structure analysis. The unphased variants were filtered taking one SNP every 250kb using VCFtools. ADMIXTURE was run varying K from 1 to 20 and performing 20 replicates. The analysis was performed independently over the whole sample of American and European (n=218) accessions or using the American (AM, n=104) accessions only. We applied the approach already used in Rossi et al. (*15*); Bitocchi et al. (*25*, *26*) to deal with possible cryptic population structure within pool.

Population structure determined by chloroplast data was inferred using the Bayesian Analysis of Population Structure (BAPS) software, version 5.3 (*49*, *50*). A mixture analysis was performed to determine the most probable K according to the data. The `clustering with linked loci' analysis was chosen to account for the linkage between sites. Ten repetitions of the algorithm for each K (i.e., from 2 to 20) were applied. The relationships among the genotypes were investigated based on Neighbor-Joining (NJ) method implemented in MEGA X (*51*) using a bootstrap value of 1,000. Gaps and missing data were excluded.

Chromosome painting

The "chromosome painting" approach implemented in ChromoPainter v2.0 (*19*) was applied to the phased variants. The effective population size (Ne) and mutation rates (Mu) were estimated individually for each accession using 10 iterations of the Expectation-Maximization (EM) algorithm implemented in ChromoPainter. The estimated parameters were fixed in a new round of analysis producing the final chromosome painting of the "recipient" haplotypes. Donor individuals were chosen according to their ancestry

proportion inferred by admixture, as follow: i) Mesoamerican individuals showing a q value > 0.99 in the admixture run with K=3 using all American accessions, and ii) Andean individuals constantly having a q value > 0.99 from K=2 to K=4 in the admixture run restricted to Andean accessions. Donors were subdivided into the five groups inferred by ADMIXTURE (AM_M1, AM_M2, AM_A1, AM_A2 and AM_A3) and used to estimate their contribution to the ancestry of each SNP of the recipient individuals. Individual SNP probabilities were then combined in 10Kb not-overlapping sliding windows along chromosomes and each window in each recipient haplotype was assigned to one of the five donor groups if a probability >=0.8 was observed (Supplementary Note 4.2). The total proportion of genetic material coming from the seven groups or "unknown" (genotypes assigned to none of the groups) was computed for each recipient individual and for each chromosome (both pairs). The final assignment of each recipient accession to the gene pools was done according to i) the total proportion of windows attributed to Mesoamerica or Andes, and ii) the number of chromosomes assigned to the two gene pools following the majority rule criterion.

The attribution of each genomic window to the seven groups was also used to estimate the length of the introgressed blocks within each European accession. Each haplotype of the EU_AND accessions was traversed merging consecutive windows, attributed to any of the Mesoamerican clusters. Bedtools (*52*) was used to join windows within a maximum distance between elements of 50Kb to deal with artificially broken introgressed blocks. The length of each Mesoamerican block in each EU_AND individual was recorded for each chromosome and was then filtered removing blocks composed by single windows (10kb). The final within-individual distribution of lengths was characterized by the median, due to its non-normality.

For the spatial analyses, the ecological data (about 1-km2 resolution) were downloaded from WorldClim data (http://www.worldclim.org, *53*) for a total of 19 bioclimatic variables and 24 monthly variables (**SN4_ Supplemental Dataset 5**). The vegan R package (*54*) was used to calculate the geographical and ecological distances, the Mantel statistics, and the spatial autocorrelation. At first, the Mantel statistics was tested by 103 permutations and the autocorrelogram was calculated among 10 distance classes of nearly 540 km each, calculating the significance of the correlation per each class by 9999 permutations. Subsequently, we performed an environmental association analysis with a multivariate correlation analysis between the Pvalues (proportion of the genetic membership to the five genetic group M1, M2, A1, A2 and A3) assigned to each European accession and the ecological variables registered at the collection site.

Genetic diversity

The genetic diversity within groups of accessions, defined according to their geographic origin and gene pool, was quantified using the theta estimator (θπ, *55*). The "--site-pi" VCFtools flag was used to obtain a per-SNP estimate that was subsequently filtered, according to the genome annotation, including only positions located: i) in callable regions ii) in coding regions iii) in neutral regions (Supplementary Note 4.3). The per-site θπ estimate was then summed up and divided by the size of each specific region to calculate a global estimate over it. A raw estimate of θπ along chromosomes, averaged over 100kb not overlapping windows, was also computed to highlight chromosomal regions having different levels of genetic diversity. To evaluate the stability of the θπ estimate at different missing data levels, a masked dataset was obtained filtering alleles identified to be introgressed by the ChromoPainter analysis or with an ambiguous assignment, within European accessions. The "--site-pi" and "--missing-site" commands in VCFtools were used to obtain a per-site θπ estimate and the proportion of missing data for each position, respectively. The global within group θπ was computed for the callable, the coding and the neutral genomic partitions, excluding regions with an average (over SNP) minimum mean proportion of not-masked individuals (PIND) from 0 to 100%.

To detect patterns of private alleles, missense and synonymous variants were screened in American and European accessions (Supplementary Note 4.3). Variants that were private of the European or the American group were retained and divided in low (below 5%) and medium-high (above 5%) within-sample frequency. The genomic coordinates related to private alleles segregating at different frequencies in the American and European groups of accessions were intersected with the gene annotation, and the burden of missense and synonymous mutations was recorded for each gene element.

The magnitude of the genomic differentiation between and within America and Europe was evaluated using the Weir & Cockerham estimator of F_{ST} (*56*). We estimated the baseline differentiation between and within the two continents. In addition, the F_{ST} was then computed in 10kb not-overlapping sliding windows between each pair of groups using VCFtools. The mean and the interquartile range (IQR) of the windows-based distribution were used as point estimate of the differentiation between groups and to evaluate its dispersion.

Phenotyping of the genetic structure

The Analysis of Variance (ANOVA) between the genetic groups was performed using the first principal component related to the DTF and photoperiod sensitivity (Supplementary Note 2.4) as representative phenotypic trait. Following the same approach, we performed an ANOVA between the genetic subgroups. The PC1, that was obtained from the secondary metabolites having a high hereditability (H2>0.65) (Supplementary Note 2.3), was used as a phenotype for comparison between the genetic subgroups.

Tagging the signatures of adaptation in Europe

The occurrence of "excess of introgression and selection" was investigated in Europe. To detect deviations from the frequencies expected in absence of demographic and selection forces, the ChromoPainter output was parsed tracing the assignment of each SNP to the corresponding Mesoamerican or Andean groups. For each SNP, we computed the proportion of haplotypes assigned to the Mesoamerican or Andean groups.

We extracted the genomic coordinates of SNPs showing an unexpected proportion of Andean alleles (threshold: EU_A, 71*2=142 haplotype, plus the 50% of the Mesoamerican ones EU_M, 43*2*0.5=43, Fobs>=0.811). The putative SNPs targets of Mesoamerican introgression events were identified according to the same rationale (threshold: EU_M, 43*2=86 plus the 50% of Andean ones EU_A, 71*2*0.5=71, Fobs=0.688). The Bedtools "slop -b 2500" and "merge -d 10000" functions were used to pass from SNP point coordinates to 5Kb regions and then merge it in larger genomic blocks if the relative distance between each of them was lower than 10kb. Only genomic regions supported by at least three SNPs were retained (Supplementary Note 6.1)

The hapFLK (*29*) method was used to identify selection signatures. The local genomic differentiation along chromosomes, as measured by haplotypic FST, was compared to the expectation given by the inferred genomic relationships between groups, considering the genetic drift within groups. Accessions were subdivided in the AM_A (n=30), AM_M (n=36), EU_A (n=71) and EU_M (n=43) groups and VCFtools was used to sample a single SNP every 250kb. This set of SNPs was used to estimate a neighbor joining tree and a

kinship matrix according to the Reynolds' genetic distance matrix between the four groups of accessions, constituting a genome wide estimate of population structure. The hapFLK statistics was then computed on each chromosome independently over the complete SNP dataset and averaged over 20 expectation maximization cycles to fit the LD model. A first analysis was performed fixing the number of haplotype clusters to 5, according to the admixture analysis. A second run was conducted selecting the appropriate number of haplotype clusters based on the fastPHASE (*57*) cross-validation procedure, implemented in the *imputeqc* R package (https://github.com/inzilico/imputeqc). VCFtools was used to extract a subset of SNPs spaced at least 100kb for each chromosome and five independent copies of such SNP set were generated, randomly masking the 10% of the variants. The fastPHASE v1.4.8 software was used for imputing the missing genotypes in each dataset, setting the number of haplotype clusters K to 5, 10, 20, 30, 40 and 50. The EstimateQuality function was used to compute the proportion of wrongly imputed genotypes (Wp) for each combination, and the K value, minimizing the mean Wp proportion across the five SNP set replicates, was selected as the most supported number of haplotype clusters. The analysis was replicated using all or only American accessions. The "scaling_chi2_hapflk.py" script was used to scale hapFLK values and compute the corresponding p-values. The significant SNPs, showing a pvalue < 10-3 (fdr < 0.05), were extracted and bedtools was used to create a region of 10kb centered on each significant SNP and to merge overlapping regions within a maximum distance of 5 Kb. The two set of regions were merged forming the extended set, constituted by the union of the two sets, and the restricted set, containing only regions supported by both runs. To pinpoint putative regions under selection in Europe, the "Extended" and the "Restricted" set were intersected with the F_{ST} windowed analysis, and only regions containing at least one F_{ST} window located in the top 5% or top 1% were retained.

Linkage disequilibrium

The relationship between linkage disequilibrium (LD) and physical distance along chromosomes was evaluated in America and Europe, and successively within the American subgroups. The PopLDdecay (*58*) tool was used to compute r^2 correlation between allele frequencies at pairs of SNPs along the chromosomes, setting a minimum minor allele frequency of 0.1 and a maximum distance between SNPs of 5 Mbp.

The level of inter-chromosomal linkage disequilibrium was also evaluated. VCFtools was used to sample one SNP every 10kb and compute the r^2 correlation index between pairs of markers located on different chromosomes. The analysis was performed independently over the American subgroups, using only SNPs that were segregating within each group of accessions with a minor allele frequency higher than 0.05, and only pairs of SNPs showing an r^2 value >= 0.8 were retained. Multiple pairs of SNPs pointing to the same chromosomal regions were merged if located within 100kb from each other and only pairs of regions spanning at least 500kb at each side were retained. The whole analysis was also repeated including only SNPs falling in the putative regions under selection, decreasing the minimum width of retained regions from 500kb to 50kb. Link plot showing regions in high linkage disequilibrium were produced using the Rcircos (*59*) package.

Genome-wide association study (GWAS)

GWAS was performed for the growth habit, flowering time and photoperiod sensitivity data (Supplementary Note 2.4). A single-locus mixed linear model (MLM), implemented in the MVP R package (*60*; https://github.com/XiaoleiLiuBio/MVP), was run at first. The Bonferroni correction at $\alpha = 0.01$ was set up as the significance threshold for each trait. The analysis was then conducted using the multi-locus stepwise linear mixed-model (MLMM; [*61*]; https://github.com/Gregor-Mendel-Institute/MultLocMixMod) that, using a step-wise approach, includes the most significant SNPs as cofactors to the mixed-model. The mBonf criterion was used to identify the optimal results and the $\alpha = 0.05$ Bonferroni-corrected threshold was used.

Investigation on function of candidate genes for adaptation

Orthologous to common bean genes were identified with Orthofinder using legume species and *A. thaliana* (see *62*). The putative function of not well characterized genes was predicted based on the orthologous relationship coupled with a literature screening of functionally characterized genes. The orthologue and known genes involved in flowering time, photoperiod and growth habit were selected and checked if located within the GWAS results.

Genes within a 100 kb interval including 50 kb up and down-stream of each significant SNP associated to DTF, and GH, and genes located within selection scan and introgression scan regions were subjected to GO term enrichment analysis including biological process, cellular component, and molecular function using the enrichment analysis available on the Metascape tool (*63*; http://metascape.org).

References

37. JMP®, Version 7.0.0. SAS Institute Inc., Cary, NC, 1989–2012

38. Wei T, Simko V (2017) R package "corrplot": Visualization of a Correlation Matrix (Version 0.84). Available from https://github.com/taiyun/corrplot

39. Peterson BG, Peter CP, Boudt K, Bennett R, Ulrich J, Zivot E, Cornilly D, Hung E, Lestel M, Balkissoon K, Wuertz D, Christidis AA, Martin RD, Zhou Z, Shea JM (2020) PerformanceAnalytics: Econometric Tools for Performance and Risk Analysis (Version 2.0.4). Available from https://github.com/braverock/PerformanceAnalytics

40. Perez de Souza L, Scossa F, Proost S, Bitocchi E, Papa R, Tohge T and Fernie AR (2019) Multi-tissue integration of transcriptomic and specialized metabolite profiling provides tools for assessing the common bean (*Phaseolus vulgaris*) metabolome. Plant J, 97: 1132-1153. https://doi.org/10.1111/tpj.14178

41. Li H (2013) Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. arXiv:1303.3997 [q-bio.GN].

42. Li H (2011) A statistical framework for SNP calling, mutation discovery, association mapping and population genetical parameter estimation from sequencing data. Bioinformatics 27:2987-2993.

43. McKenna A, Hanna M, Banks E, Sivachenko A, Cibulskis K, Kernytsky A, Garimella K, Altshuler D, Gabriel S, Daly M, DePristo MA (2010) The Genome Analysis Toolkit: A MapReduce framework for analyzing next-generation DNA sequencing data. Genome Res. 20:1297-1303.

44. Van der Auwera GA, Carneiro MO, Hartl C, Poplin R, Del Angel G, Levy-Moonshine A, Jordan T, Shakir K, Roazen D, Thibault J, Shakir K, Roazen D, Thibault J, Banks E, Garimella KV, Altshuler D, Gabriel S, DePristo MA (2013) From FastQ data to high confidence variant calls: the Genome Analysis Toolkit best practices pipeline. Curr Protoc Bioinformatics 43:11 10 11-33.

45. Danecek P, Bonfield JK, Liddle J, Marshall J, Ohan V, Pollard MO, Whitwham A, Keane T, McCarthy SA, Davies RM, Li H (2021) GigaScience, Volume 10, Issue 2, giab008, https://doi.org/10.1093/gigascience/giab008

46. Danecek P, Auton A, Abecasis G, Albers CA, Banks E, DePristo MA, Handsaker RE, Lunter G, Marth GT, Sherry ST, McVean G (2011) The variant call format and VCFtools. Bioinformatics. 2011, 27(15):2156-8.

47. Cingolani P, Platts A, Wang LL, Coon M, Nguyen T, Wang L, Land SJ, Lu XY, Ruden DM (2012) A program for annotating and predicting the effects of single nucleotide polymorphisms, SnpEff: SNPs in the genome of Drosophila melanogaster strain w(1118); iso-2; iso-3. Fly 6:80-92.

48. Browning S R and Browning B L (2007) Rapid and accurate haplotype phasing and missing data inference for whole genome association studies by use of localized haplotype clustering. Am J Hum Genet 81:1084-1097. doi:10.1086/521987

49. Corander J, Waldmann P, Sillanpää MJ (2003) Bayesian analysis of genetic differentiation between populations. Genetics 163(1):367-74. doi: 10.1093/genetics/163.1.367. PMID: 12586722; PMCID: PMC1462429.

50. Corander J, Marttinen P, Sirén J, Tang J (2008) Enhanced Bayesian modelling in BAPS software for learning genetic structures of populations. BMC Bioinformatics 9, 539 https://doi.org/10.1186/1471-2105-9-539

51. Kumar S, Stecher G, Li M, Knyaz C, Tamura K (2018) MEGA X: Molecular Evolutionary Genetics Analysis across Computing Platforms. Mol. Biol. Evol. 35(6):1547–49. doi: 10.1093/MOLBEV/MSY096.

52. Quinlan AR, Hall IM (2010) BEDTools: a flexible suite of utilities for comparing genomic features, Bioinformatics, Volume 26, Issue 6, 15, 841–842, https://doi.org/10.1093/bioinformatics/btq033

53. Hijmans RJ, Cameron SE, Parra JL, Jones PG, Jarvis A (2005) Very high resolution interpolated climate surfaces for global land areas. Int. J. Climatol., 25, 1965–1978.

54. Oksanen J, Blanchet FG, Friendly M, Kindt R, Legendre P, McGlinn D, Minchin PR, O'Hara RB, Simpson GL, Solymos P, Stevens MHH, Szoecs E, Wagner H (2020) vegan: Community Ecology Package. R package version 2.5-7. https://CRAN.R-project.org/package=vegan

55. Tajima F (1983) Evolutionary relationship of DNA sequences in finite populations. Genetics, 105(2):437-60.

56. Weir BS, Cockerham CC (1984) Estimating F-statistics for the analysis of population structure. evolution, 1358-1370.

57. Scheet P, Stephens M (2006) A fast and flexible statistical model for large-scale population genotype data: applications to inferring missing genotypes and haplotypic phase. Am. J. Hum. Genet., 78(4):629-44.

58. Zhang C, Dong SS, Xu JY, He WM, Yang TL (2019) PopLDdecay: a fast and effective tool for linkage disequilibrium decay analysis based on variant call format files. Bioinformatics, Volume 35, Issue 10, 1786–1788, https://doi.org/10.1093/bioinformatics/bty875

59. Zhang H, Meltzer P, Davis S (2013) RCircos: an R package for Circos 2D track plots BMC bioinformatics, 14(1):244.

60. Zhou X, Stephens M (2012) Genome-wide efficient mixed-model analysis for association studies. Nat. genet., 44(7), 821.

61. Segura V, Vilhjálmsson BJ, Platt A, Korte A, Seren Ü, Long Q, Nordborg M (2012) An efficient multi-locus mixed-model approach for genome-wide association studies in structured populations. Nat. genet., 44(7), 825.

62. Di Vittori V, Bitocchi E, Rodriguez M, Alseekh S, Bellucci E, Nanni L, Gioia T, Marzario S, Logozzo G, Rossato M, De Quattro C, Murgia ML, Ferreira JJ, Campa A, Xu C, Fiorani F, Sampathkumar A, Fröhlich A, Attene G, Delledonne M, Usadel B, Fernie AR, Rau D, Papa R (2021) Pod indehiscence in common bean is associated with the fine regulation of PvMYB26. J. Exp. Bot., Volume 72, Issue 5, 27 February 2021, Pages 1617–1633, https://doi.org/10.1093/jxb/eraa553

63. Zhou Y, Zhou B, Pache L, Chang M, Khodabakhshi AH, Tanasichuk O, Benner C, Chanda SK (2019) Metascape provides a biologist-oriented resource for the analysis of systems-level datasets. Nat. Commun. 10, 1523. https://doi.org/10.1038/s41467-019-09234-6