Doctoral School in Agricultural Sciences - XXXIV cycle

Department of Agricultural, Food and Environmental Sciences

Common bean as a model to understand crop evolution

Candidate:

Gaia Cortinovis

Gaia Cortinovis

Supervisor:

Prof. Roberto Papa

Academic year 2020-2021

Table of contents

This Doctoral Thesis, in accordance with the corresponding report authorized by the Thesis Directors, includes five chapters, three of which have been published in journals included in the Science Citation Index and one is currently in submission. The last chapter is a work in progress.

General Introduction

Our agricultural system and hence food security is threatened by a combination of events, such as increasing population, the impacts of climate change, and the need for more sustainable development. Because of their nutritional quality, biological nitrogen fixation capacity, and broad adaptation to several agro-ecological conditions, food legumes are crucial for key agriculture-related societal challenges, such as agrobiodiversity conservation, sustainable agriculture, food security, and human health (Mousavi-Derazmahalleh et al., 2017). Currently, legumes represent the second most agriculturally important crop family on a global scale after cereals (Graham and Vance, 2003). Among legumes, beans, but in particular common bean (*P. vulgaris*), are the most important grain legumes for direct human consumption in the world (Broughton et al., 2003). Moreover, the well-documented history of multiple domestications in *P. vulgaris* and its further adaptation to different environments make it a model system to study crop evolution (Cortinovis et al., 2020). To date, exploitation of the genetic resources in food legumes breeding is limited in comparison to the availability of the materials (Ray et al., 2012). Genetic diversity represents the raw material on which adaptive selection acts, and as such, it has a fundamental role in both evolutionary history and future evolutionary pathways of a species. Analysis of the genetic diversity through population genomics and genotype-phenotype association approaches can be very useful tools to detect genetic variants related to crop adaptation and associated to important agronomic traits. In the last years, numerous sequencing and resequencing efforts have been undertaken in plants and, as a result, reference genome sequences become available for several crops, including the Mesoamerican (Vlasova et al. 2016) and Andean (Schmutz et al. 2014) *P. vulgaris* reference genomes. However, with the assembly of increasing numbers of plant genomes, it is becoming accepted that a single reference does not reflect the complete genetic diversity of a whole species (Springer et al., 2009; Saxena et al., 2014). There is an urgent need for transition from a reference-centric-approach to a pangenomic-reference system that is able to capture the entire genetic diversity present in a species through the direct all-to-all comparisons between all the accessions considered (Eizenga et al., 2020). The main aim of the present study is the exploitation of the evolutionary history of the common bean through the recent progress in genomics and pangenomics, in order to dissect the genetic basis and phenotypic consequences of its parallel domestications and adaptation to new agro environments.

Chapter I

Towards the Development, Maintenance, and Standardized Phenotypic Characterization of Single-Seed-Descent Genetic Resources for Common Bean

Gaia Cortinovis,[1] Markus Oppermann,[2] Kerstin Neumann,[2] Andreas Graner,[2] Tania Gioia,3 Marco Marsella,[4] Saleh Alseekh,[5,6] Alisdair R. Fernie,[5,6] Roberto Papa,[1] Elisa Bellucci,[1,7] and Elena Bitocchi[1,7]

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

[2]Research Group Genebank Documentation, Leibniz Institute of Plant Genetics and Crop Plant Research (IPK) Gatersleben, Seeland, Germany

[3]School of Agricultural, Forestry, Food and Environmental Sciences (SAFE), University of Basilicata, Potenza, Italy

[4]International Treaty on Plant Genetic Resources for Food and Agriculture (FAO), Rome, Italy

[5]Department of Molecular Physiology, Max Planck Institute of Molecular Plant Physiology, Potsdam-Golm, Germany

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

[7]Corresponding authors: e.bellucci@univpm.it, e.bitocchi@univpm.it

DOI: **https://doi.org/10.1002/cpz1.133**

Towards the Development, Maintenance, and Standardized Phenotypic Characterization of Single-Seed-Descent Genetic Resources for Common Bean

Gaia Cortinovis,[1] Markus Oppermann,[2] Kerstin Neumann,[2] Andreas Graner,[2] Tania Gioia,[3] Marco Marsella,[4] Saleh Alseekh,[5,6] Alisdair R. Fernie,[5,6] Roberto Papa,[1] Elisa Bellucci,[1,7] and Elena Bitocchi[1,7]

[1]Department of Agricultural, Food and Environmental Sciences, Polytechnic University of Marche, Ancona, Italy
[2]Research Group Genebank Documentation, Leibniz Institute of Plant Genetics and Crop Plant Research (IPK) Gatersleben, Seeland, Germany
[3]School of Agricultural, Forestry, Food and Environmental Sciences (SAFE), University of Basilicata, Potenza, Italy
[4]International Treaty on Plant Genetic Resources for Food and Agriculture (FAO), Rome, Italy
[5]Department of Molecular Physiology, Max Planck Institute of Molecular Plant Physiology, Potsdam-Golm, Germany
[6]Center for Plant Systems Biology, Plovdiv, Bulgaria
[7]Corresponding authors: *e.bellucci@univpm.it*, *e.bitocchi@univpm.it*

The optimal use of legume genetic resources represents a key prerequisite for coping with current agriculture-related societal challenges, including conservation of agrobiodiversity, agricultural sustainability, food security, and human health. Among legumes, the common bean (*Phaseolus vulgaris*) is the most economically important for human consumption, and its evolutionary trajectories as a species have been crucial to determining the structure and level of its present and available genetic diversity. Genomic advances are considerably enhancing the characterization and assessment of important genetic variants. For this purpose, the development and availability of, and access to, well-described and efficiently managed genetic resource collections that comprise pure lines derived by single-seed-descent cycles will be paramount for the use of the reservoir of common bean variability and for the advanced breeding of legume crops. This is one of the main aims of the new and challenging European project INCREASE, which is the implementation of Intelligent Collections with appropriate standardized protocols that must be characterized, maintained, and made available, along with the related data, to users such as breeders and researchers.
© 2021 The Authors. Current Protocols published by Wiley Periodicals LLC.

Basic Protocol 1: Characterizing common bean seeds for seed trait descriptors
Basic Protocol 2: Bean seed imaging
Basic Protocol 3: Characterizing bean lines for plant trait descriptors specific for common bean Primary Seed Increase

Keywords: common bean • genetic diversity • genetic resources • intelligent collections • single-seed-descent line • standardized phenotyping protocols

CURRENT PROTOCOLS
A Wiley Brand

Current Protocols e133, Volume 1
Published in Wiley Online Library (wileyonlinelibrary.com).
doi: 10.1002/cpz1.133
© 2021 The Authors. Current Protocols published by Wiley Periodicals LLC. This is an open access article under the terms of the Creative Commons Attribution License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited.

How to cite this article:
Cortinovis, G., Oppermann, M., Neumann, K., Graner, A., Gioia, T., Marsella, M., Alseekh, S., Fernie, A. R., Papa, R., Bellucci, E., & Bitocchi, E. (2021). Towards the development, maintenance, and standardized phenotypic characterization of single-seed-descent genetic resources for common bean. *Current Protocols*, *1,* e133. doi: 10.1002/cpz1.133

INTRODUCTION

Origin, Domestication and Diffusion, and Evolution Out of the Centers of Origin

The Leguminosae family consists of about 770 genera and over 19,500 species (Azani et al., 2017; Lewis et al., 2013). It currently represents the second most agriculturally important crop family globally, after the cereals (Graham & Vance, 2003). Due to their nutritional quality, biological nitrogen fixation capacity, and broad adaptation to various agro-ecological conditions, legumes have a crucial role in helping to overcome key agriculture-related societal challenges, such as the mitigation of and adaptation to climate change, agrobiodiversity conservation, agricultural sustainability, food security, and human health. Among the legumes, beans (*Phaseolus* ssp.) are the most important for direct human consumption throughout the world, particularly the common bean (*Phaseolus vulgaris*; Broughton et al., 2003). Here, we provide a brief overview of the evolutionary history of *P. vulgaris* and the ongoing efforts for innovative and sustainable conservation and management of its broad genetic diversity.

P. vulgaris originated in Mexico (Bitocchi et al., 2012) and later, through different migration events, became widespread across the highlands of Latin America and into northwestern Argentina (Toro et al., 1990). The common bean is characterized by three ecogeographic gene pools: Mesoamerica and the Andes, the two major gene pools, which include both wild and domesticated forms; and northern Peru–Ecuador, with a relatively narrow distribution (i.e., western slopes of the Andes) that includes only wild forms. The Mesoamerican origin of the common bean was confirmed relatively recently (Ariani, Mier y Teran, & Gepts, 2017; Desiderio et al., 2013; Rendón-Anaya et al., 2017; Schmutz et al., 2014). However, debate continues on the timing of its dispersal to South America and the evolutionary consequences (see Cortinovis, Frascarelli, et al., 2020, for a review).

The wide geographic extent where wild forms of *P. vulgaris* grow implies that they are characterized by adaptation to different environments, as distinct from those of the Mesoamerican population. In this regard, Rodriguez et al. (2016), and more recently Ariani and Gepts (2019), carried out landscape genomics analyses based on wide samples of wild *P. vulgaris* genotypes and high-throughput genomic data to identify several genomic regions that show signatures of selection for adaptation.

Common bean also underwent two parallel and independent domestication events, one in Mesoamerica and the other in the Andes, which gave rise to the current two major domesticated gene pools. At the genomic level, domestication caused a reduction in the genetic diversity of the domesticated germplasm (for review, see Bitocchi, Rau, Bellucci et al., 2017, and Cortinovis, Frascarelli, et al., 2020) due to demographic factors that affected the entire genome, and to natural and artificial selection at target loci. Domestication had a major impact not only by reducing the diversity of domesticated forms compared to the wild population at the nucleotide level but also reducing the diversity of gene expression,

as was seen through a scan of transcriptome diversity performed by Bellucci, Bitocchi, Ferrarini, et al. (2014).

In addition, some observations have identified an increase in functional diversity at target loci that harbor genes involved in environmental adaptation, including adaptation to both biotic and abiotic factors (Bellucci, Bitocchi, Ferrarini et al., 2014; Bitocchi, Rau, Benazzo, et al., 2017).

The next step in the evolution of the common bean was its spread out of the Americas (Cortinovis, Di Vittori, et al., 2020). This was a very complex process involving several introductions from the American continent, accompanied by several exchanges between different continents and countries, due to intensive commercial interactions (for reviews, see Bellucci, Bitocchi, Rau, Rodriguez, et al., 2014; Bitocchi, Rau, Bellucci, et al., 2017; Cortinovis, Di Vittori, et al., 2020). Particularly interesting, in terms of genetic variability and adaptation, is the breakdown of the spatial geographical barriers between the Mesoamerican and Andean genotypes that characterized the evolution of the common bean out of the New World. In particular, in Europe, this process favored hybridization and introgression between gene pools, increasing the possibility that novel genotypes and phenotypes would arise (Angioi et al., 2010; Gioia et al., 2013).

Worldwide Germplasm Collections

Nikolai Ivanovich Vavilov was one of the first pioneers to recognize the exceptional importance and potential value of collecting and conserving the wide genetic diversity of a crop and its wild relatives (Vavilov, 1920, 1922), which still remain largely unexploited. He highlighted the crucial role of wild relatives of crop plants as sources of genes for the exploitation of natural and artificial introgression. This specific point was formalized by Harlan and de Wet (1971) through the introduction of the "gene-pool concept," which was very useful for investigating how genes can be transferred between species. Each crop is characterized by a pool of genetic diversity that can potentially be available for use in breeding. This pool can be classified on the basis of the degree of crossing ability between the crop itself and its wild relatives. The primary gene pool (GP-1) corresponds to the biological species, which includes the crop and individuals that have no barriers to reproduction. The secondary gene pool (GP-2) includes less closely related species, for which hybridization with the crop is possible, but difficult. Within the tertiary gene pool (GP-3), crosses are feasible through advanced techniques, such as protoplast fusion, embryo rescue, or genetic engineering, whereas the quaternary gene pool (GP-4) is characterized by a lack of success in obtaining fertile hybrids by any means. Based on this concept, Figure 1 illustrates the species related to *P. vulgaris* and their degrees of relationship with *P. vulgaris* in terms of crossing ability.

Considering the current challenges posed by climate, agriculture, and food production, the conservation of plant genetic resources (PGRs) is now becoming imperative, along with their characterization and use (McCouch et al., 2020; Mousavi-Derazmahalleh et al., 2019). Generally, the major proportion of the genetic diversity of a crop is contributed by its wild relatives, as they have not experienced domestication and the resulting reduction of diversity (Diamond, 2002; Gepts, 2010; Glémin & Bataillon, 2009). Landraces are also important repositories of the genetic diversity of a crop, as these represent local varieties that have evolved through natural and artificial selection over millennia adapting to specific and diversified agro-environmental conditions, without undergoing genetic bottlenecks due to modern breeding techniques (Zeven, 1998) or rapid adaptation to specific and diversified agro-environmental conditions (Bellucci et al., 2013; Bitocchi et al., 2009, 2015; Dwivedi et al., 2016; Mir, Sharma, & Mahajan, 2020; Zhu et al., 2000). Thus, landraces and wild relatives harbor functional and adaptive genetic variation that needs to be more easily managed and used.

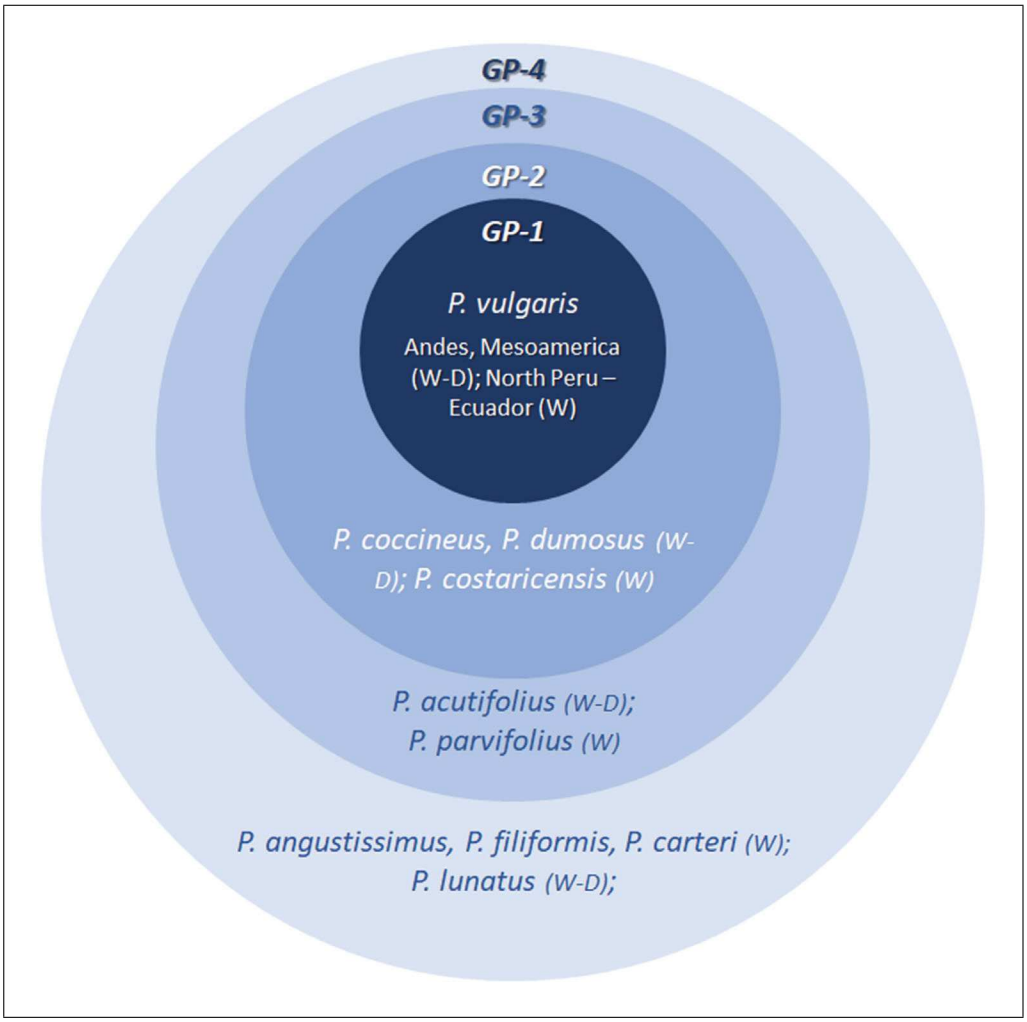


Figure 1 Primary (GP-1), secondary (GP-2), tertiary (GP-3), and quaternary (GP-4) gene pools of *P. vulgaris*. D, domesticated; W, wild.

Thus, as the first step, PGRs and their wide diversity need to be maintained. Germplasm banks can guarantee the conservation ex situ of such biodiversity. Genesys PGR is a free online global portal that allows the exploration of plant species diversity through a single website (accessible at *www.genesys-pgr.org*). The data published on Genesys follow the standards for Multi-Crop Passport Descriptors (as defined by the United Nations Food and Agriculture Organization [FAO] and Bioversity). For *P. vulgaris*, Genesys contains over 135,500 accessions (as of October 19, 2020), which are stored in several holding institutions worldwide. Most of these accessions are landraces (~71,000), followed by improved cultivars (~20,000), and wild (~2000) and breeding/research materials (~2600). Figure 2 shows the geographic distribution of the wild and landrace accessions of common bean for which information on geographic coordinates is available in Genesys. The International Centre for Tropical Agriculture (CIAT) in Colombia holds the largest *P. vulgaris* collection, with ~32,000 accessions, followed by the United States Department of Agriculture Agricultural Research Service (USDA-ARS), with nearly 14,000 accessions, and the Leibniz Institute of Plant Genetics and Crop Plant Research (IPK) in Germany, with ~8400 accessions. Numerous common bean accessions (~27,700 overall) are also held by a Brazilian gene bank (the Empresa Brasileira de Pesquisa Agropecuária, or EMBRAPA). Recursos Genéticos e Biotecnologia CENARGEN, Brasília, DF, and Arroz e Feijão, CNPAF, Goiânia, GO). Furthermore, the Russian collection of common beans at the N.I. Vavilov Institute of Plant Genetic Resources (VIR) is one of the oldest in the world. The VIR collection now numbers ~6400 accessions (according to information in

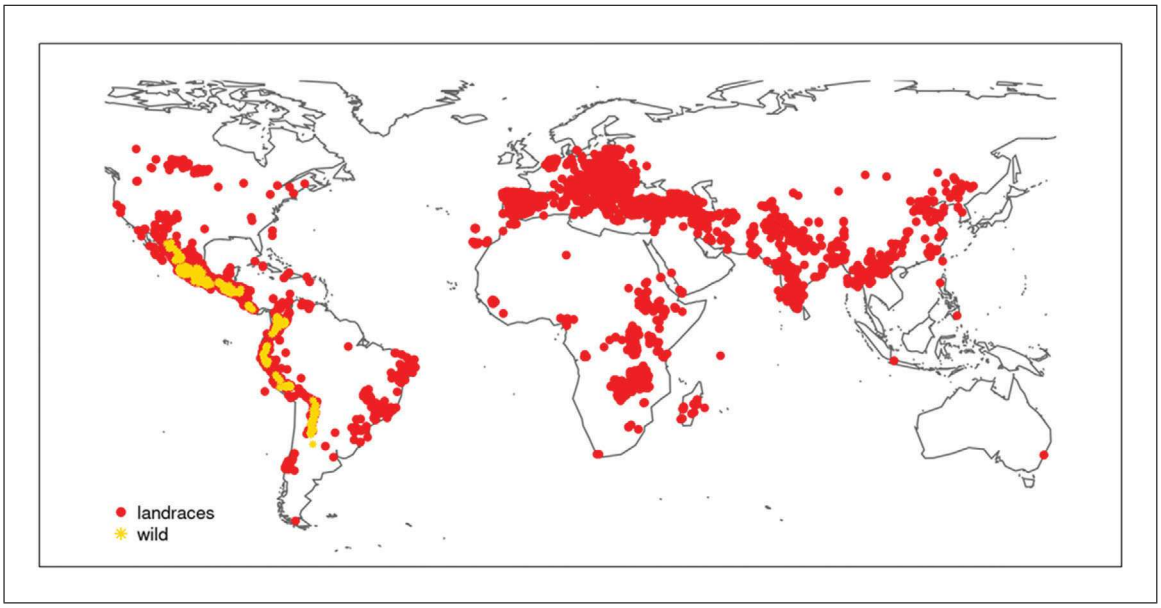


Figure 2 Worldwide geographic distribution of wild and landrace accessions of common bean maintained in gene banks (coordinates from Genesys, *www.genesys-pgr.org*).

Genesys) and is the result of the collective efforts of several generations of scientists and explorers, beginning in the early twentieth century.

P. vulgaris germplasm collections stored in gene banks all around the world are essential for national and international attempts to ensure the safety (i.e., conservation and maintenance) and use of such PGRs.

Genomics of Genetic Resources

The genomics era has led to a rapid increase in available sequence data, which has thus provided a more detailed picture of the genetic diversity and structure of crop germplasm, as well as enabling the identification of genetic variants that are the bases of important heritable target traits (Luikart et al., 2018). Genome-wide fingerprinting of single-seed descent (SSD) gene-bank accessions allowed the characterization of a complete barley collection (Milner et al., 2019) and resulted in further identification of potential duplicated material within (Milner et al., 2019) and across (Singh et al., 2019) collections. Moreover, the accurate description of the genetic topology of a gene pool, as revealed by principal-component analysis or other population genomics approaches, allows the identification of hitherto under-represented parts of the overall genetic diversity, and can suggest strategies for completing a collection. Furthermore, once a collection is fingerprinted, existing historic phenotype data obtained during seed multiplication can be used. In the case of barley, the statistical analysis of historic data of key agronomic traits resulted in high heritability estimates for the collection (González et al., 2018), so that users can make informed selections of lines of interest.

For the common bean, the current availability of high-throughput sequencing platforms has allowed the release of high-quality reference genomes of the Andean G19833 (Schmutz et al., 2014) and Mesoamerican BAT93 (Vlasova et al., 2016) genotypes. A further high-quality common bean reference genome of the race Durango pinto UI111 genotype was also released recently (*P. vulgaris* UI111 v1.1, DOE-JGI and USDA-NIFA, *http://phytozome.jgi.doe.gov/*). Genetic characterization through high-throughput sequencing of common bean collections and segregant populations is increasing in the literature. One major study that aims to characterize a very wide sample of common bean genetic resources is the BEAN_ADAPT project, which is funded through the second

European Research Area Network for Coordinating Action in Plant Sciences (ERA-CAPS) joint funding call. As PGRs are often heterogeneous, working with them has required the development of SSD seeds to obtain purified and genetically homogeneous material that is suitable for genetic studies. Therefore, BEAN_ADAPT is based on the characterization of nested core collections, developed by SSD, that have been genotyped at different levels on the basis of their size. The Pv_ALL collection, a large collection of ∼10,000 accessions (mostly European and from the American centers of origin), and a subsample of 500 lines (Pv_core1) have both been characterized through genotyping-by-sequencing methods, while a further subsample of 220 accessions have been genotyped by whole-genome sequencing. The genetic data are now being used in combination with the different phenotypic evaluations to identify the genetic basis of phenotypic variants related to adaptation to new environments, through the study of the introduction and wide spread of the common bean from the Americas into Europe. Moreover, the overall design of the project and the data obtained can be also used to predict the phenotypes of the whole collection based on the genotype data, resulting in a very useful tool for exploiting the huge number of PGRs that are currently maintained in gene banks and that remain widely underexploited.

Several genomic studies have been performed based on two particular panels: the Middle American Diversity Panel, which was developed as part of the USDA-funded Common Bean Coordinated Agricultural Project (BeanCAP) and consists of 280 modern bean genotypes that are mostly from the race Mesoamerica, but also from Durango and Jalisco (Moghaddam et al., 2016); and the Andean Diversity Panel, which consists of 396 accessions, the majority of which belong to the Andean gene pool (349), while 21 are Mesoamerican, and 26 are from admixtures between the Mesoamerican and Andean gene pools (Chichy, Porch, et al., 2015). Moghaddam et al. (2016) genotyped the Middle American Diversity Panel with two Illumina iSelect 6K gene chip sets (BARCBEAN6K_1 and BARCBEAN6K_2; Song et al., 2015) and two low-pass sequencing protocols (Elshire et al., 2011; Schröder et al., 2016). Overall, the panel was genotyped at 217,486 mapped single-nucleotide polymorphisms (SNPs). These data were combined with phenotype evaluations for classical agronomic traits (carried out in four different American locations), which were used to perform genome-wide association studies (GWAS). This identified new and known genomic regions that affect various agronomic traits, such as days to flowering, days to maturity, growth habit, lodging, canopy height, and seed weight (Moghaddam et al., 2016).

The Andean Diversity Panel was developed and genotyped at 5398 SNPs using the Illumina BARCBean6K_3 SNP BeadChip (Cichy, Porch, et al., 2015). The SNP data were used to genetically characterize the panel by investigating the level and structure of its genetic diversity, and they were also coupled with phenotypic data (i.e., plant determinacy) to identify marker-trait associations. These panels (or subsamples of them, or as part of larger collections) have been used in further GWAS studies (Cichy, Wiesinger, & Mendoza, 2015; Kamfwa, Cichy, & Kelly, 2015a, 2015b; McClean et al., 2017, 2018; Moghaddam et al., 2017; Oladzad, Zitnick-Anderson, et al., 2019; Parker et al., 2020; Soltani et al., 2017; Tock et al., 2017; Zuiderveen, Padder, Kamfwa, Song, & Kelly, 2016), demonstrating the potential of such materials for pre-breeding and breeding studies. Recently, Almeida et al. (2020) also efficiently used SNPs to characterize the diversity of an association panel of Carioca strains from Brazil.

Recently, Oladzad, Porch, et al. (2019) reported the development of the moderately sized Bean Abiotic Stress Evaluation panels for the U.S. Agency for International Development (USAID) Climate Resilience Bean (CRIB) project, which aims to investigate the genetics and physiological mechanisms of the responses of dry beans cultivated under abiotic stress. These panels are of modest size (∼120 lines) and are managed by research

groups with limited resources; however, at the same time, they are large enough to define the genetic variance at the basis of phenotypic variability. Oladzad, Porch, et al. (2019) carried out genotyping by sequencing for the Bean Abiotic Stress Evaluation panels, and they analyzed all of the data to identify markers associated with production traits under both heat- and drought-stress environments.

Collections and/or panels have also been developed with a focus on snap bean genetic resources. Within the BeanCAP project, a snap bean diversity panel was developed that consisted of 149 cultivars from North America and Europe (Kleintop, Myers, Echeverria, Thompson, & Brick, 2016). Wallace, Arkwazee, Vining, and Myers (2018) genotyped the BeanCAP Snap Bean Panel along with a further 55 Chinese and 19 Spanish snap bean genotypes and 24 heirloom beans at 5398 SNPs. This study provided a deep investigation of the genetic diversity and structure of this collection, which represents a very useful tool for further GWAS. Myers et al. (2019) characterized this panel for total phenolic contents and genotyped them at 10,073 SNPs, and they used GWAS for identification of 11 quantitative trait nucleotides associated with this trait.

Another interesting example of the use of whole-genome scan analysis to characterize *P. vulgaris* genetic resources was conducted by Wu et al. (2020) and focused on the variability of Chinese common bean germplasm compared to worldwide accessions, and on the relevant contributions of landrace material. They applied whole-genome resequencing to a collection of 683 accessions, which comprised 529 landraces and 154 breeding lines that were representative of both the Mesoamerican and Andean gene pools. The plant material was grown over 3 years and in four locations in China that are characterized by different agro-ecological conditions, with the aim of investigating trait associations related to *P. vulgaris* yield across the north-south geographic clines. Wu et al. (2020) detected a total of 4,811,097 SNPs and several genetic differences in terms of structural variations affecting DNA segments of >1 kbp, such as insertions, deletions, copy number variations, and presence/absence variations. By using GWAS, these researchers defined several marker-trait associations that were dispersed across the entire genome for all of the traits considered, even though they were found in different proportions depending on the inheritance level and the complexity of the trait itself.

Segregant populations also represent valuable PGRs. As an example of the use of wild individuals to introgress new variability into elite germplasm, for instance for the common bean, Murgia et al. (2017) phenotypically characterized a very interesting introgression line (IL) population. This population was developed from an initial cross between the domesticated Andean variety Midas and the G12873 line, a wild Mesoamerican accession (Koinage et al., 1996). Several cycles of selfing and subsequent backcrosses with the recurrent parent Midas allowed the development of a set of 70 ILs from BC_3/F_4:F_5 families and 217 ILs from BC_3/F_6:F_7 families (Murgia et al., 2017; Rau et al., 2019). Phenotypic characterization for pod shattering (Murgia et al., 2017), combined with genetic characterization through genotyping by sequencing at 14,196 SNPs, allowed Rau et al. (2019) to investigate the genetic architecture of the shattering trait in common bean. They identified a locus in the distal part of chromosome Pv05 as the primary locus responsible for the pod indehiscence phenotype, along with numerous other secondary quantitative trait loci that contributed to the modulation of this phenotype. Recently, Di Vittori et al. (2020) continued the development of this IL population: from the Murgia et al. (2017) IL population, they selected six highly shattering ILs and used these as donor parents for high pod shattering for further backcrosses (BC_4) with Midas, to provide an IL population of 1197 BC_4/F_4 individuals that was then genetically (19,420 SNPs) and phenotypically evaluated for pod shattering. Along with gene expression and parallel histological analysis of dehiscent and indehiscent pods, a GWAS by Di Vittori et al. (2020) identified an ortholog of *AtMYB26* from *Arabidopsis thaliana* as the best candidate for loss of

pod shattering, in a genomic region ∼11 kb downstream of the most highly associated peak.

Development and Maintenance of Common Bean Increase Intelligent Collections

Several drawbacks characterize conventional conservation management of PGRs. Among these, there is the problem that seed collections are assembled and maintained on an accession basis, with each accession usually constituting a mixture of genotypes that represents a population. As a consequence, the information collected at the phenotype level cannot be directly linked to a specific genotype. Furthermore, hundreds of accessions are conserved and maintained in gene banks with very little information available (i.e., there is a lack of comprehensive information regarding passport data and descriptors useful for users, combined with accession heterogeneity and unharmonized data), which makes their selection and use for specific purposes by researchers and breeders often difficult. Moreover, the available information is not easily accessible, being in databases that are centralized and were not designed to integrate data obtained by external users.

Recently, European Union Horizon 2020 funded Project INCREASE—Intelligent Collections of Food Legumes Genetic Resources for European Agrofood Systems—with the aim of overcoming such limitations and implementing a new approach for conservation, management, and characterization of PGRs (Bellucci et al., submitted).

Along with chickpea, lentil, and lupin (Kroc et al., in preparation; Guerra-García, Gioia, von Wettberg, Logozzo, & Bett, in preparation; Kumar et al., in preparation), common bean is one of the legume species that form the basis of the INCREASE project. With the purpose of conserving, managing, and making the best use of food-legume genetic resources, the INCREASE project is developing innovative conservation procedures through the integration of two complementary strategies (i.e., ex situ and in situ approaches) and the development of SSD-purified accessions based on single homozygous genotypes. SSD collections provide the possibility of associating phenotypes with reliable genotype information, which can then be used to advance and improve legume research, breeding, and cultivation.

To achieve this goal, the project plans to develop INCREASE Intelligent Collections ("intelligent" in terms of being able to memorize, learn, improve, and evolve) as a set of three nested core collections that will be characterized genetically and/or phenotypically at different levels, according to their size. These comprise:

- The Reference Core (R-CORE), as the largest collection. For common bean, this will rely on already available SSD lines from previous projects (e.g., BEAN_ADAPT, BRESOV) complemented with heterogeneous accessions conserved in situ and ex situ, from which SSD lines will be developed to provide a worldwide and highly diverse panel of wild accessions, landraces, and cultivars. This collection will include more than 10,000 SSD lines that will be genotyped using at least a low-coverage approach (e.g., genotyping by sequencing and exome capture), and will be maintained in gene banks for long-term conservation.
- The Training Core (T-CORE), representing a subsample of the R-CORE and comprising ∼450 lines. A deeper sequencing approach is planned for T-CORE (e.g., Illumina whole-genome sequencing), along with broad phenotypic characterization (e.g., classical, molecular, high-throughput phenotyping) under both controlled and field conditions.
- The Hyper-CORE (H-CORE), which will consist of 40 to 80 accessions that are carefully chosen on an evolutionary transect, with the primary aim being to sample the largest possible genotypic diversity of *P. vulgaris*, including closely related

species. These lines will be deeply phenotyped and genotyped, and a subsample of the H-CORE will be also sequenced to develop a pangenome.

The large amount of data produced will be analyzed to (i) investigate the level and structure of the genetic diversity of legume genetic resources; (ii) identify functional variants that might have major roles in determining the phenotypic variation for a large number of traits; and (iii) predict the phenotypes of PGRs in gene banks only on the basis of their genotypic characterization. This will provide an incredible tool in the hands of geneticists and breeders for the improvement of food legumes.

Here, we describe the procedures that will be applied in INCREASE to develop the Intelligent Collections, with a specific focus on the protocols that INCREASE has implemented and proposed for adoption by gene banks and research institutions to obtain SSD lines, to conserve and maintain their seeds, to make the seeds available for users, and to characterize them and integrate the data obtained into a centralized system, which will ultimately be accessible to anyone interested in these PGRs.

Figure 3 shows the workflow established by INCREASE in terms of the SSD development, conservation, maintenance, and characterization procedures. The protocols developed within INCREASE refer to activities related to Work Package 3 of the project, entitled "Sampling core collections, SSD development, DNA extraction and seed distribution." In particular, the three different phenotyping protocols described below were established to characterize the genetic resources of the project during SSD development and the subsequent seed increase cycles that will be performed under controlled conditions. These protocols were developed starting from IPGRI *P. vulgaris* descriptors (International Board for Plant Genetic Resources [IBPGR], 1982) and Crop Ontology (Bioversity International, 2011), and they have been modified specifically for purposes related to these SSD development activities.

CHARACTERIZING COMMON BEAN SEEDS FOR SEED TRAIT DESCRIPTORS

BASIC PROTOCOL 1

This protocol was developed to characterize seeds before the start of any cycle of seed increase. It allows for the characterization of the original phenotypes of seeds from heterogeneous materials, if the cycle to be carried out is the first (development of SSD lines from heterogeneous accessions), and for obtaining data on the seed morphology of each SSD line for each subsequent cycle of seed multiplication. Such data are important not only to obtain a characterization of seeds from different genetic resources but also to detect human and/or technical errors that can eventually occur during the chain of the selfing cycles.

NOTE: Record the seed traits as detailed below at the beginning of each primary seed increase cycle. The phenotypic characterization must be performed before using the seeds in any (and all) selfing cycles.

Materials

Seeds
Ruler
Analytical balance
Spreadsheet software

1. Take at least five seeds from each accession (heterogeneous materials) or line (previously developed SSD lines), and through visual observation, evaluate and classify the Seed Coat Pattern according to the following categories (Fig. 4):

   • 0 = absent
   • 1 = constant mottled, spotted

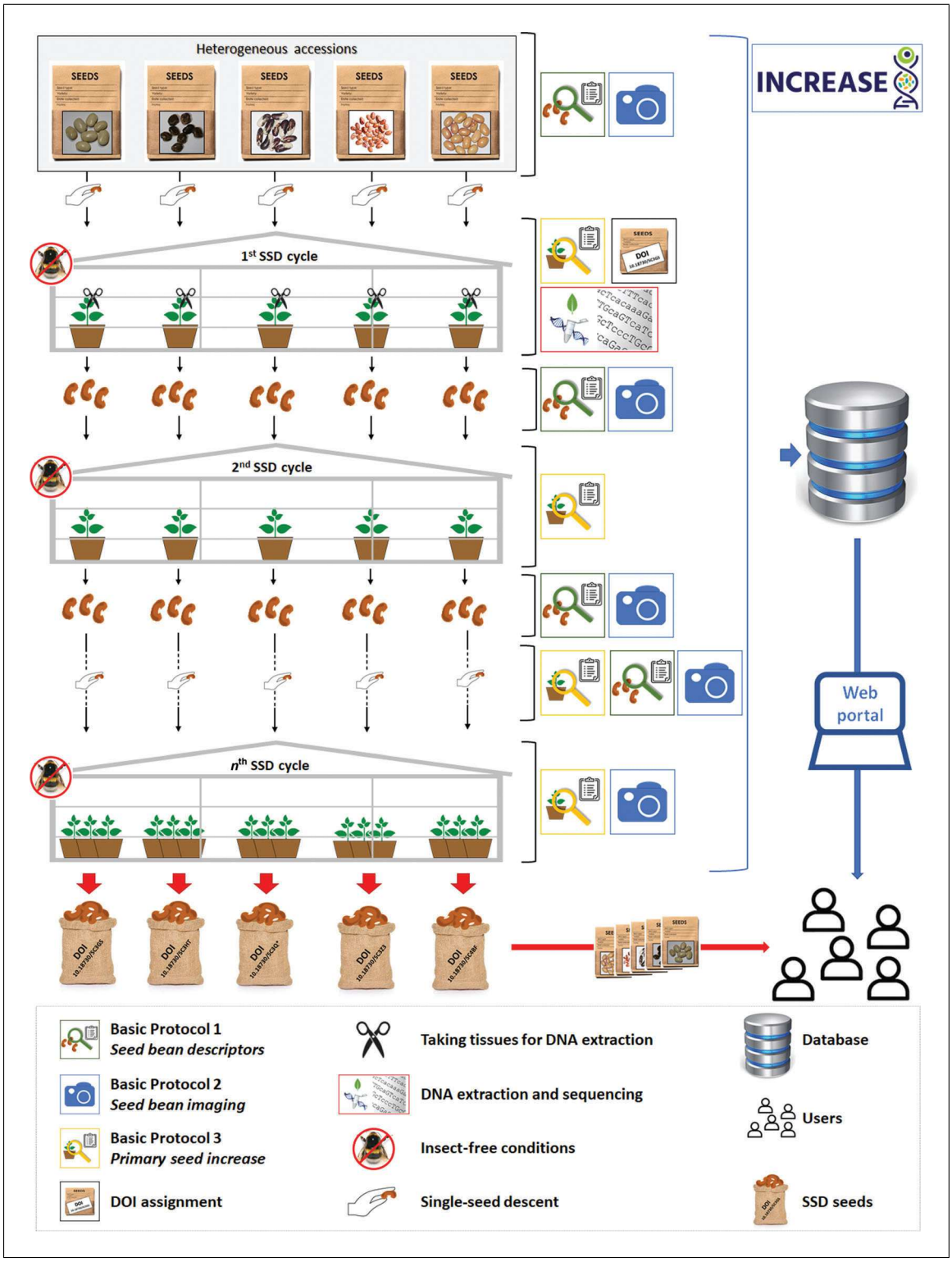


Figure 3 Summary of activities related to the development of single-cell descent (SSD) lines and subsequent selfing cycles.

- 2 = striped
- 3 = mottled
- 4 = constant mottled, marmorated
- 5 = spot near the hilum (swallow, geometric type)
- 6 = spot near the hilum (soldier, irregular spot)
- 7 = spot near the hilum (large, diffuse)

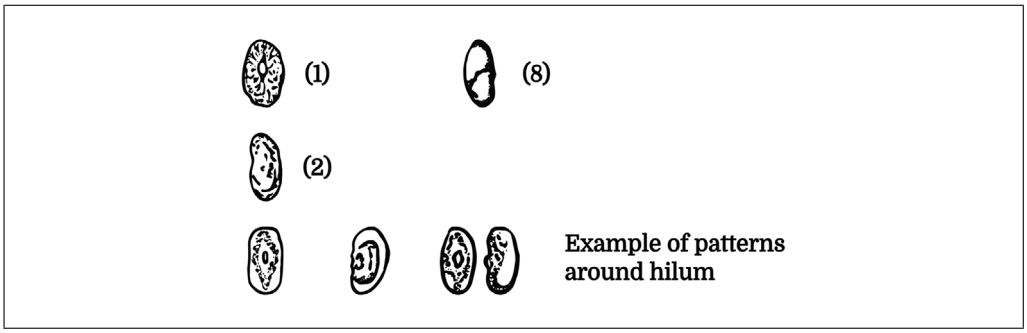


Figure 4 Examples of seed coat patterns.

- 8 = bipartite/tripartite
- 9 = covered
- 10 = coated
- 11 = other (specify in the Notes).

2. Take at least five seeds from each accession (heterogeneous materials) or line (previously developed SSD lines), and through visual observation, evaluate and classify the Seed Coat Coloring according to the following three categories:

   - 1 = single colored
   - 2 = two colored
   - 3 = three (or more than three) colored.

3. Take at least five seeds from each accession (heterogeneous materials) or line (previously developed SSD lines), and through visual observation, evaluate and classify the Seed Coat Ground Color (i.e., primary color) according to the following 24 categories:

   - 1 = white
   - 2 = greenish white
   - 3 = yellow
   - 4 = light cream
   - 5 = ochre
   - 6 = green
   - 7 = olive green
   - 8 = gray
   - 9 = light brown
   - 10 = dark brown
   - 11 = light purple
   - 12 = red purple
   - 13 = purple
   - 14 = blue-purple
   - 15 = black
   - 16 = red
   - 17 = mustard yellow
   - 18 = gray-yellow
   - 19 = red-brown
   - 20 = pink
   - 21 = black-purple
   - 22 = gray-black
   - 23 = blue
   - 24 = other (specify in the Notes).

4. Take at least five seeds from each accession (heterogeneous materials) or line (previously developed SSD lines), and through visual observation, evaluate and classify the Seed Coat Secondary Color (if any) according to the following 24 categories:

   • 1 = white
   • 2 = greenish white
   • 3 = yellow
   • 4 = light cream
   • 5 = ochre
   • 6 = green
   • 7 = olive green
   • 8 = gray
   • 9 = light brown
   • 10 = dark brown
   • 11 = light purple
   • 12 = red-purple
   • 13 = purple
   • 14 = blue-purple
   • 15 = black
   • 16 = red
   • 17 = mustard yellow
   • 18 = gray-yellow
   • 19 = red-brown
   • 20 = pink
   • 21 = black-purple
   • 22 = gray-black
   • 23 = blue
   • 24 = other (specify in the Notes).

5. Take at least five seeds from each accession (heterogeneous materials) or line (previously developed SSD lines), and through visual observation, evaluate the presence/absence of Seed Color Veining.

   • 0 = absent
   • 1 = present.

6. Take at least five seeds from each accession (heterogeneous materials) or line (previously developed SSD lines), and through visual observation, evaluate and classify the Seed Brilliance (i.e., seed shininess or opaqueness at harvest) according to one of the following three categories:

   • 1 = matte
   • 2 = medium
   • 3 = shiny.

7. Take five seeds from each accession (heterogeneous materials) or line (previously developed SSD lines), and using a ruler, measure the Seed Coat Length (mm) by performing a lateral measurement, parallel to the hilum (Fig. 5). Use the mean of the values obtained from five seeds as the final measure.

8. Take five seeds from each accession (heterogeneous materials) or line (already developed SSD lines), and using a ruler, measure the Seed Coat Height (mm) by performing a lateral measurement, measured from the hilum to the opposite side (Fig. 5). Use the mean of the values obtained from five seeds as the final measure.

9. Take five seeds from each accession (heterogeneous materials) or line (already developed SSD lines), and using a ruler, measure the Seed Coat Width (mm) by

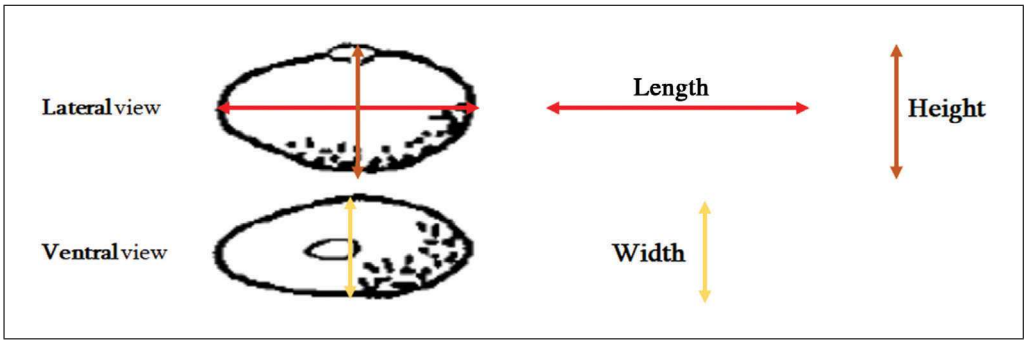


Figure 5 Seed length, height, and width measurements.

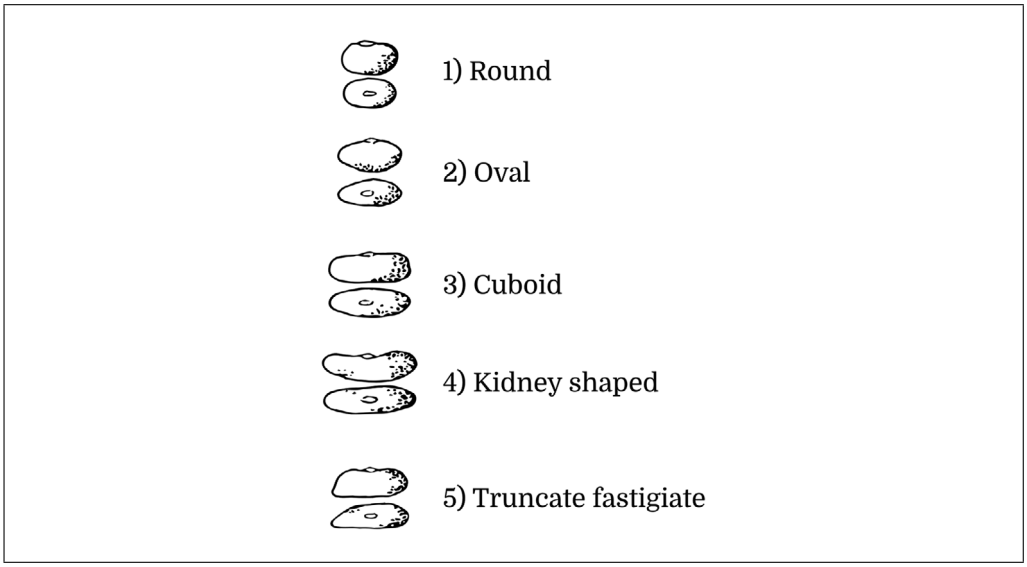


Figure 6 Seed shape descriptors.

performing a ventral measurement (Fig. 5). Use the mean of the values obtained from five seeds as the final measure.

10. Take at least five seeds from each accession (heterogeneous materials) or line (already developed SSD lines), and through visual observation, evaluate and classify the Seed Shape according to one of the following five categories (Fig. 6):

    • 1 = round
    • 2 = oval
    • 3 = cuboid
    • 4 = kidney shaped
    • 5 = truncate fastigiate.

11. Measure the 100-Seed Weight, preferably by measuring the seed weight of at least two samples of 100 seeds (if not possible, then the seed weight should be measured on 10 seeds, as three different samples). Immature and/or infected seeds should be excluded.

BEAN SEED IMAGING

BASIC PROTOCOL 2

The aim of SEED bean imaging is to document the beans during their respective propagation steps in regard to the following traits:

- Visual quality control of the propagation result;
- Reference images for shape, color, and size;

- Picture material for presenting the SSD line in various media;
- Basis for automated image analysis.

A comprehensive discussion on the digitization of collections based on Nelson, Paul, Riccardi, and Mast (2012) is given in de la Hidalga, van Walsun, Rosin, Sun, and Wijers (2019). Accordingly, the following requirements should be fulfilled as the best-practice standard.

NOTE: Image capture should be done using a defined process under comparable conditions, and any institution and/or research group involved in seed increase of shared SSD lines should implement a standardized system to take images of seeds.

Important considerations for photography are listed below.

- To obtain good separation of the beans from the background, it is advisable to place them on a background of a color not present in skins of the beans. Light gray is ideal here for a neutral picture impression; if this is too close to the color of the beans, a black or white background can be used. The background should be flat and should not have any structural pattern.
- Include a color chart in the photographs for quality control and postprocessing. This allows the lighting, white balance, and color accuracy of the image to be verified.
- Include a scale bar to allow measurement of the dimensions of the objects.

Materials

Uniform background (see above)
Color chart: e.g., ColorGauge Micro Target (Image Science Associates) or ColorChecker Classic Mini or Nano (X-Rite)
Scale bar
Photo lighting studio made of a lighting box (40 × 40 cm) with brightness-adjustable LED lamps (allowing gradual adjustment of light level)
Microscope camera of at least 20 megapixels and an appropriate macro lens mounted on a stable tripod above the box
Image-editing software for post-processing images that allows image normalization and resizing, as well as the subsequent addition of text or QR codes to the image

Imaging process

1. Select at least five beans to be photographed. Selected bean seeds should be typical of the respective line in terms of color and size, although any variability should also be captured.

2. Place the bean seeds, color chart, and scale bar on the background. Arrange the beans so that they are separated from each other, and position one bean near the rule, at scale position zero of the scale. Be sure to leave enough space for the subsequent insertion of the label information and barcode.

3. Begin photography, processing only one bean line per pass to avoid mixing or confusion of the material. Before recording, verify the identity of the material.

4. Within the project space, the name of the image should be set to start with the unique IncreaseID followed by the sample ID or plant ID, and a unique image number separated by underscores, as: <IncreaseID>_<SampleID>_<ImageID>

5. Set the image resolution and size. The resolution and size of an image depends on the intended use. The higher the initial values, the more versatile the application is, as smaller versions can be easily created. The following values in Table 1 provide the guidelines for the resolution. The color bit depth must always be 24-bit color.

Table 1 Guidelines for Image Resolution

Expected use	Resolution (ppi)
Research and preservation	600
Printing	300
Web publishing	72

6. Process the images (e.g., image enhancement, normalization, addition of label information, barcode). For the barcode, different variants are possible.

   a. Two-dimensional QR Code: This is preferred because further information can be stored in addition to the identifier. The data record for the examples given here is structured as follows:

      • Full Image-ID: <IncreaseID>_<SampleID>_<ImageID>
      • Image of:
      • DOI: <DOI of the SSD line that is shown>
      • Project-ID: <IncreaseID>
      • Sample-ID: <SampleID>
      • Species: *P. vulgaris* L.
      • Plant part: <Type of plant parts shown>
      • ITPGRFA Annex1 crop: Beans
      • Creation date: <yyyy-mm-dd> (e.g., 2020-10-26)

   b. Line codes (Code 128): These should only be used for the internal identification of image documents. They are the simplest form of identifier and allow the coding of an identifier for the image. The recommended format is as follows: <IncreaseID>_<SampleID>_ImageID

7. Transfer the finalized images to a central repository for the project and reference them in a management system.

   Working versions of the images should be clearly labeled. It is recommended that interim versions not be deleted before safe storage of the final images.

8. Integrate the image identification and description information as a label applied subsequent to the image processing. The human-readable information is limited to the identifiers of the depicted object, and the reference to the image rights license (here, the Creative Commons license CC BY-SA, *https://creativecommons.org*). Optionally, a logo can be included to boost brand recognizability.

   An example of photo documentation of an SSD cycle 1 bean harvest is given in Figure 7. A comparison of SSD harvest with the appearance of bean seed characteristics of the corresponding gene bank accession is shown in Figure 8. The usefulness of the light gray background for bean seed color and pattern depiction is demonstrated in Figure 9.

CHARACTERIZING BEAN LINES FOR PLANT TRAIT DESCRIPTORS SPECIFIC FOR COMMON BEAN PRIMARY SEED INCREASE

BASIC PROTOCOL 3

The main aim of this protocol is the phenotypic evaluation of the different lines grown in controlled conditions during the cycles of seed multiplication. These phenotypic data, along with those obtained by applying Basic Protocols 1 and 2, will be uploaded into the project database and integrated with those obtained for the same lines in other experiments (i.e., field trials). All of these data will represent precious information for future users, such as researchers, breeders, and so on.

Figure 7 Example of a photograph of a seed bean from one SSD line generated by IPK. Original: 5116 × 4042 pixels, 600 ppi; color depth, 24-bit; created on a photo station illuminated from left and right; camera: Kaiser Scando icoss.

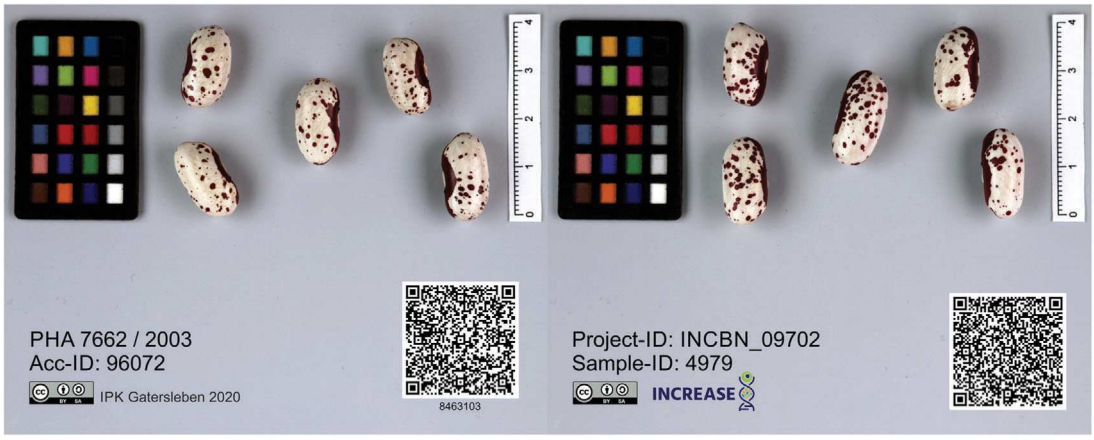


Figure 8 Example of visual validation of seed obtained: Two pictures from the original gene bank accession, with PHA 7662 at left and the corresponding SSD line INCBN_09702 at right.

Users should record the following Mandatory Traits (Priority 1, steps 1-27), which are considered as essential within INCREASE and must be recorded for each selfing cycle, along with Non-Mandatory Traits (Priority 2, steps 28-33). Non-Mandatory Traits are considered within INCREASE as non-essential, but preferable. The collection of Non-Mandatory Traits overlaps with the collection of Mandatory Traits, and thus the step numbers are not fully sequential in time.

NOTE: This protocol assumes that plants and seeds are evaluated by visual inspection and manual measurement.

NOTE: During Primary Seed Increase, one plant is grown and characterized for each line, for at least the first two selfing cycles (development of SSD seeds); in subsequent cycles it may be possible to grow more than one plant per line, depending on the available space in controlled conditions (see Fig. 3). If more than three plants are grown per line, traits can be recorded on a subsample (at least three randomly chosen plants).

NOTE: Before starting the protocol, collect the following information on the experimental site: data collector name, location of experimental trial, latitude of

Figure 9 Example of color compatibility with the chosen light gray background for beans. This allows color differences to be depicted, which can be compensated by normalization using the color chart.

experimental trial, longitude of experimental trial, altitude of experimental trial, and controlled-condition/insect-free measures applied (e.g., tunnel, greenhouse, grow chamber).

Materials

Seeds
Ruler
Analytical balance
Spreadsheet software

Recording Mandatory Traits (Priority 1)

1. Record the Sowing Date (i.e., date on which the seeds were sown).
2. Record the Days to Emergence (i.e., number of days after which the seedlings emerged, starting from the sowing day).

   Emergence is defined here as the time at which seedling cotyledons/leaves become visible.

3. Through visual observation, evaluate and classify the Hypocotyl Pigmentation (i.e., color of the hypocotyl) according to one of the following three categories:

   • 1 = purple
   • 2 = green
   • 3 = other (specify in the Notes).

4. Through visual observation, evaluate and classify the Leaf Color (chlorophyll) according to one of the following three categories of green:

   • 1 = pale green
   • 2 = medium green
   • 3 = dark green.

5. Through visual observation, evaluate the presence/absence of the Leaf Color (anthocyanin pigmentation, red-purplish or red color) according to one of the following categories:

   • 0 = absent
   • 1 = present.

6. Record the Days to Beginning of Flowering (i.e., number of days from sowing to the appearance of the first open flower). Record this based on the presence of one open flower at any node.

   Open flower refers to when any flower banner (standard petal) is visible.

7. Through visual observation, evaluate and classify the Flower Color—i.e., the color of the standard for freshly opened flowers—according to the following 14 categories:

   • 1 = white
   • 2 = greenish
   • 3 = pink
   • 4 = light purple
   • 5 = purple
   • 6 = dark purple
   • 7 = white with purple spots
   • 8 = white with red veins
   • 9 = white with green spots
   • 10 = pink with green spots
   • 11 = light purple with green spots
   • 12 = red
   • 13 = greenish with purple spots
   • 14 = other (specify in the Notes).

   It is important to evaluate a freshly opened flower, as flower colors are highly changeable after opening.

8. Through visual observation, evaluate and classify the Flower Color of Wings for freshly opened flowers according to the following nine categories:

   • 1 = white
   • 2 = greenish
   • 3 = pink
   • 4 = light purple
   • 5 = purple
   • 6 = dark purple
   • 7 = white or light purple with dark purple edges
   • 8 = white with red veins
   • 9 = other (specify in the Notes).

   Again, be sure to evaluate freshly opened flowers, as flower colors are highly changeable after opening.

9. Record the Days to Pod Formation (i.e., the number of days after planting until the plants have at least one visible pod).

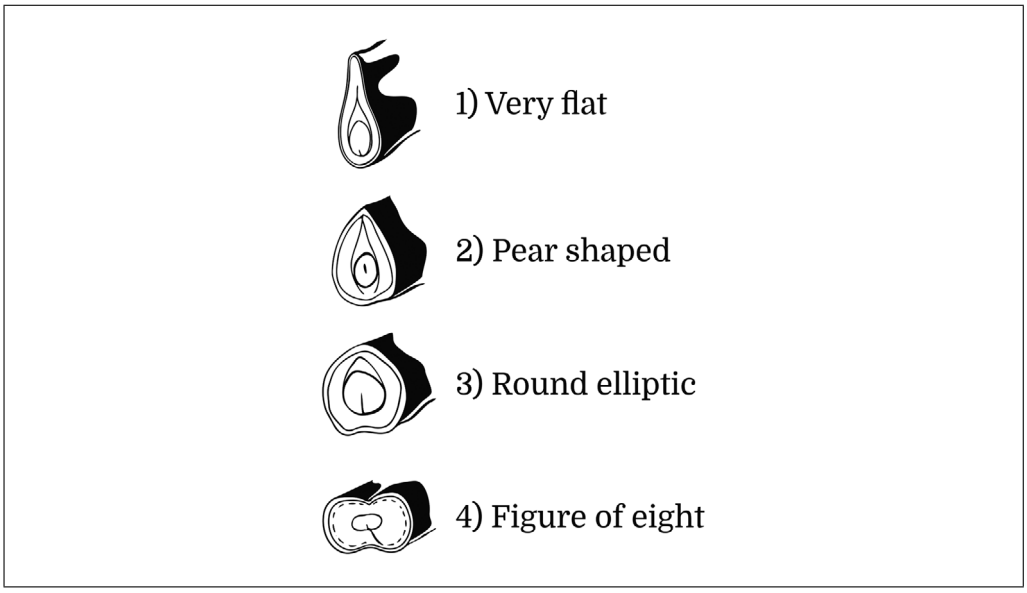


Figure 10 Pod cross-sections.

10. Through visual observation, evaluate and classify the Pod Cross-Section from fully expanded immature pods according to one of the following five categories (Fig. 10):

    • 1 = very flat
    • 2 = pear shaped
    • 3 = round elliptic
    • 4 = figure of eight
    • 5 = other (specify in the Notes).

11. Through visual observation, evaluate and classify the Pod Curvature for fully expanded immature pods according to one of the following four categories (Fig. 11):
    • 1 = straight
    • 2 = slightly curved
    • 3 = curved
    • 4 = recurving.

12. Through visual observation, evaluate and classify the Pod Color at Physiological Maturity (PM) according to one of the following seven categories (see Fig. 12 for an example of bean pods at physiological maturity):

    • 1 = light yellow
    • 2 = gold yellow/dark yellow
    • 3 = light green/gray-green
    • 4 = green/dark green
    • 5 = red
    • 6 = purple
    • 7 = other (specify in the Notes).

13. Through visual observation, evaluate and classify the Pattern of Pod Pigmentation at Physiological Maturity (PM) according to one of the following six categories:

    • 0 = none
    • 1 = speckled
    • 2 = mottled
    • 3 = striped
    • 4 = covered, coated

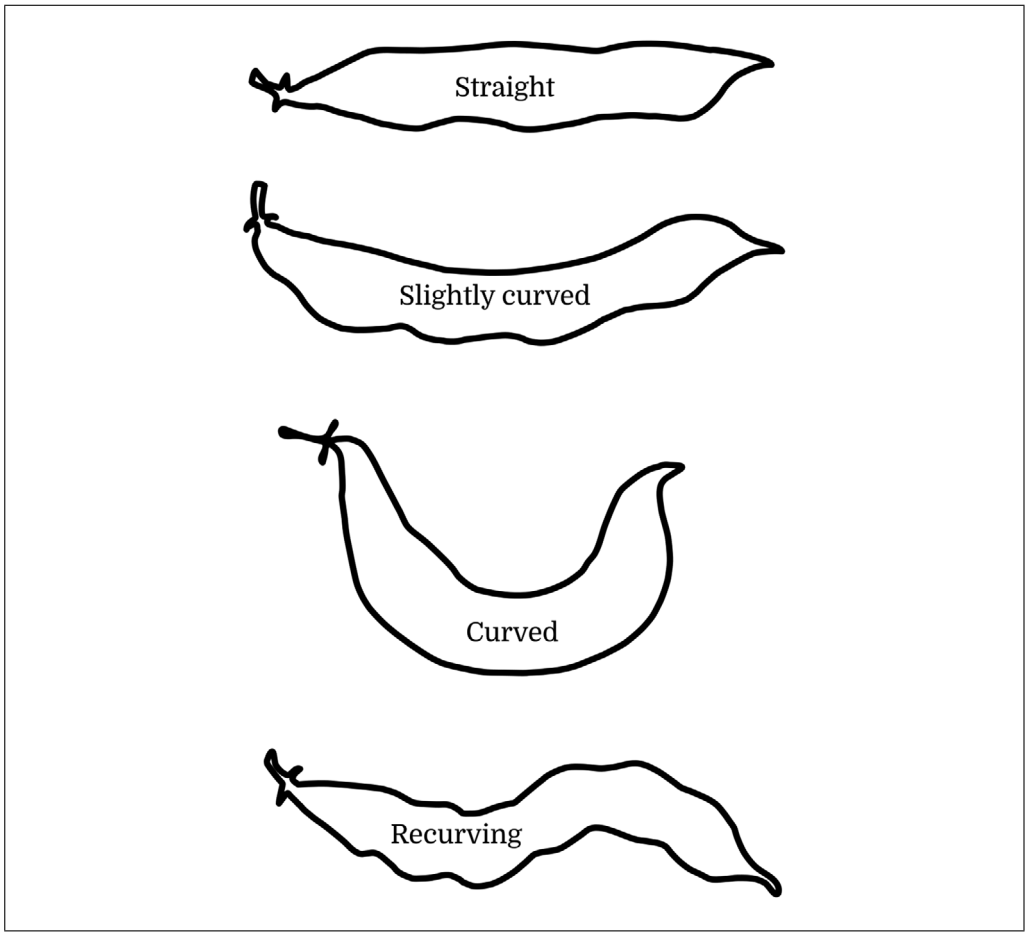


Figure 11 Pod curvature patterns.

Figure 12 Bean plant bearing pods showing mature color (i.e., at physiological maturity).

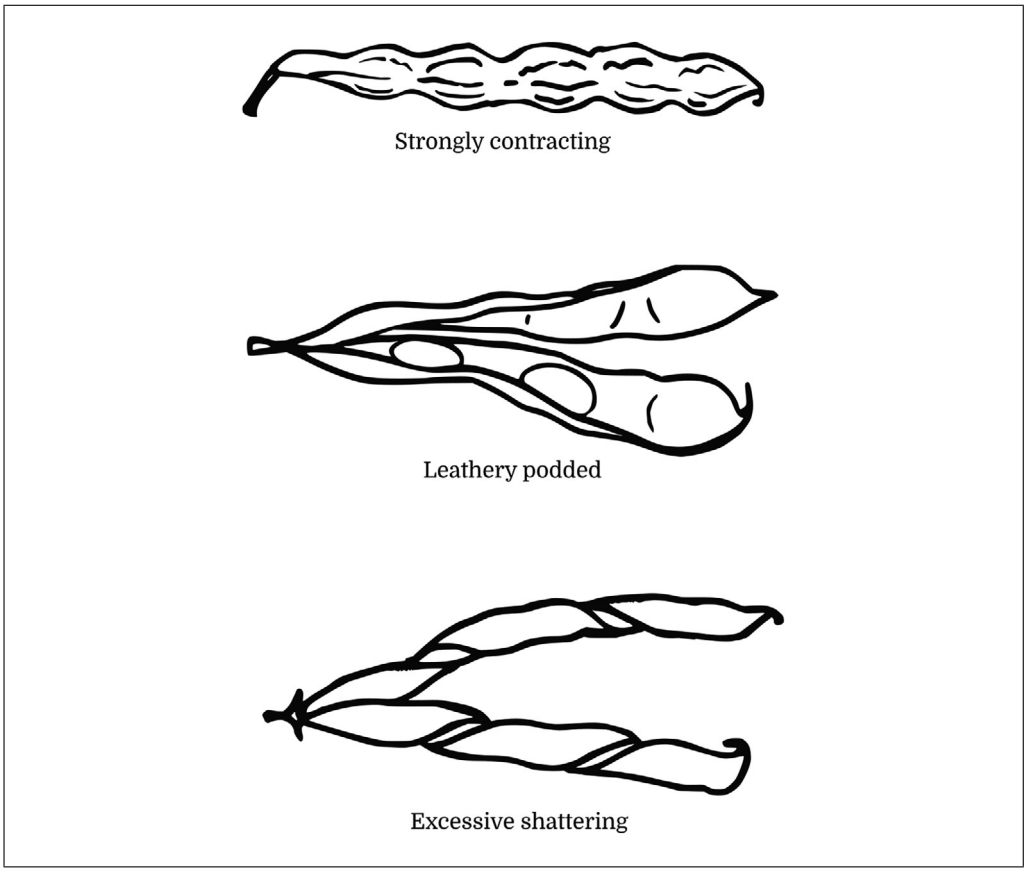


Figure 13 Pod wall fiber patterns.

- 5 = other (specify in the Notes).

14. Through visual observation, evaluate and classify the Pod Wall Fiber of the mature pods according to one of the following three categories (Fig. 13):

    - 1 = strongly contracting (at dry maturity, adhering around seed)
    - 2 = leathery podded (dry pods will not spontaneously open)
    - 3 = excessive shattering (with strong twisting of dry pods).

15. Record the Leaf Persistence at the time when 90% of the pods are dry according to one of the three categories:

    - 1 = all leaves dropped
    - 2 = intermediate
    - 3 = all leaves persistent.

16. Record the Plant Determinacy (i.e., determinate/indeterminate character) during flowering as follow:

    - 1 = determinate
    - 2 = indeterminate.

17. Record the Plant Length as the distance (cm) from the soil surface to the top of the plant, at a time when the plants have at least one open flower.

18. Measure the Stem Diameter (cm) just above the soil surface at plant maturity.

19. Record the time of Full Maturity (i.e., the number of days after planting after which 90% of the pods on the plant are golden-brown).

20. Record the Days to Harvest (i.e., the number of days from planting to harvest).

21. Measure the Pod Length (cm) at dry harvest maturity.

    This measurement should be made when the pod is completely dry.

22. Measure the Pod Width (cm).

    This observation should be made on well-developed pods, with the width assessed from suture to suture on unopened pods.

23. Record the Total Seed Mass (g) of all of the seeds harvested individually from each plant.

24. Record the Total Number of Seeds (count) harvested individually from each plant.

25. Count and record the 100-Seed Mass (g).

    Do not calculate this value from "Total seed mass" and "Total number of seeds."

26. Record Diseases, and if possible, describe or make notes regarding disease status.

    • 0 = no disease present
    • 1 = disease present
    • 2 = unsure.

    There is a "Disease-specific comments" column for making any notes related to diseases, including but not limited to the observation that many or specific diseases are present.

27. Record the Stress Susceptibility, and if possible, describe or make notes about this.

    • 0 = no stress present
    • 1 = stress present
    • 2 = unsure.

    There is a "Stress-specific comments" column for making any notes related to stress.

Recording Non-Mandatory Traits (Priority 2)

28. Through visual observation, evaluate and classify the Emerging Cotyledon Color according to the following six categories:

    • 1 = purple
    • 2 = red
    • 3 = green
    • 4 = very pale green
    • 5 = white
    • 6 = other (specify in the Notes).

29. Record the Leaf Shape of the terminal leaflet of the third trifoliate leaf according to the following three categories (Fig. 14):

    • 1 = triangular
    • 2 = quadrangular
    • 3 = round.

30. Record the Pod Color for fully expanded immature pods according to the following seven categories:

    • 1 = light yellow
    • 2 = golden yellow/dark yellow
    • 3 = green/dark green
    • 4 = light green/gray-green
    • 5 = red
    • 6 = purple
    • 7 = other (specify in the Notes).

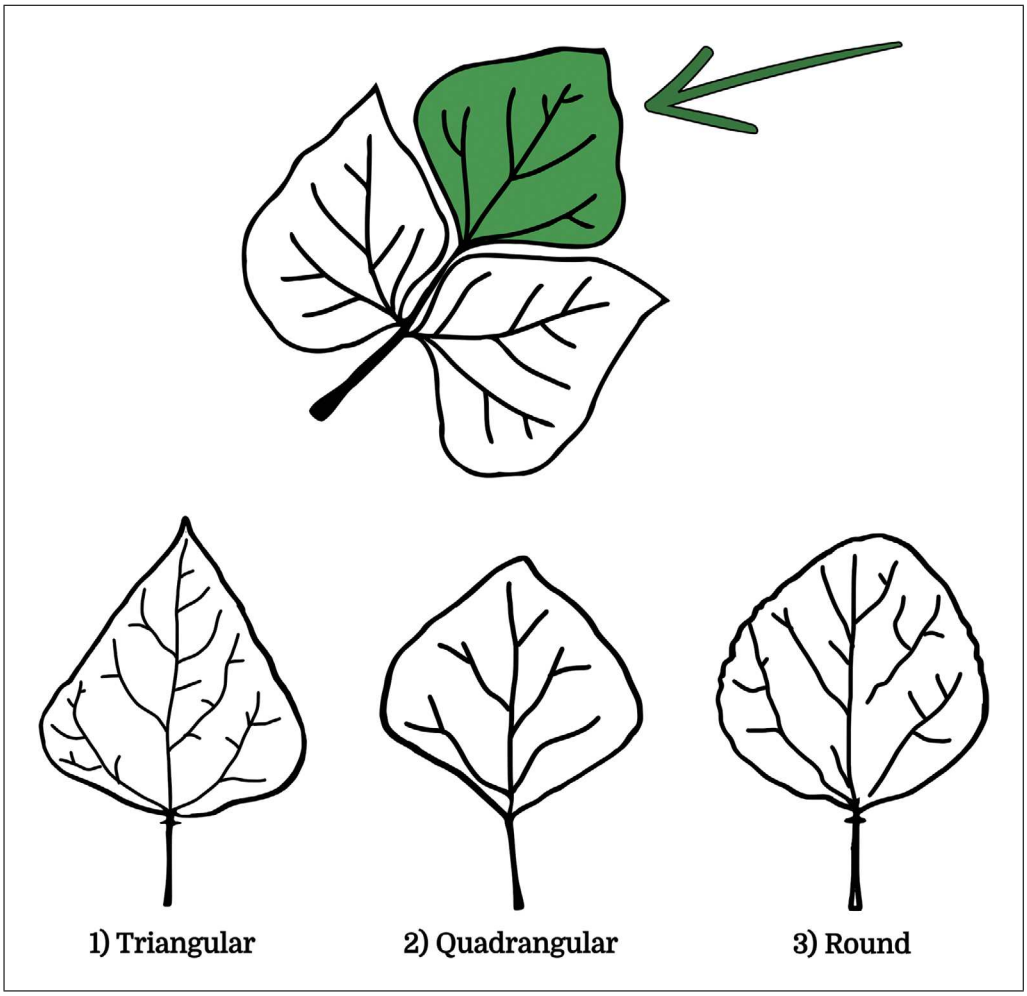


Figure 14 Leaf shapes.

31. Record the Pod Length (cm) of the largest fully expanded immature pod, from three randomly chosen pods.

32. Record the Pod Suture String status on fully expanded immature pods from three randomly chosen pods, but only if the seed multiplication will be not compromised. Classify into one of the following four categories:

    • 0 = stringless
    • 1 = few strings
    • 2 = moderately stringy
    • 3 = very stringy.

33. Through visual observation, evaluate and classify, if present, the Stress Susceptibility according to the following three categories:

    • 1 = low susceptibility
    • 2 = medium susceptibility
    • 3 = high susceptibility.

34. If a stress is present, record possible source of damage caused to aerial plant parts.

    • Low temperature
    • High temperature
    • Drought
    • High soil moisture

- Salinity
- Soil acidity
- Pests:
  - White fly
  - Thrips
  - Aphid
  - Red spider
  - Weevil
  - Other
- Fungi:
  - White mold
  - Anthracnose
  - Root rot
  - Powdery mildew
  - Ascochyta
  - Other
- Bacteria and viruses:
  - Bean common mosaic virus
  - Bean common necrotic mosaic virus
  - Halo blight.

COMMENTARY

Background Information

Here, we report an established system (i.e., phenotyping procedures and protocols) for the development of SSD lines starting from common bean genetic resources, for their maintenance and for the production of seeds for distribution. This system is based on a set of phenotyping protocols that can be applied by gene banks and research institutes that also want to use such resources. The strength of this system is that it is designed to produce both the genomic data associated with each pure line and an "open" platform for integration of data produced within INCREASE. This will also include data that are produced after the conclusion of INCREASE for the same materials. The idea is to set up a centralized system that can be used by gene banks and research institutes that can also integrate new data, and at the same time, be freely accessible by any user.

Acknowledgments

This study was conducted as part of the INCREASE project, funded from the European Union Horizon 2020 Research and Innovation program under grant agreement no. 862862.

Author Contributions

Gaia Cortinovis: Writing-original draft, writing-review and editing. **Markus Oppermann:** Validation, writing-original draft, writing-review and editing. **Kerstin Neumann:** Validation, writing-original draft, writing-review and editing. **Andreas Graner:** Validation, writing-original draft, writing-review and editing. **Tania Gioia:** Validation, writing-original draft, writing–review and editing. **Marco Marsella:** Validation, writing-original draft, writing-review and editing. **Saleh Alseekh:** Validation, writing-original draft, writing-review and editing. **Alisdair R. Fernie:** Validation, writing-original draft, writing-review and editing. **Roberto Papa:** Validation, writing-original draft, writing-review and editing. **Elisa Bellucci:** Supervision, writing-original draft, writing-review and editing. **Elena Bitocchi:** Supervision, writing-original draft, writing-review and editing.

Conflict of Interest

The authors declare no conflict of interest.

Data Availability Statement

Data sharing not applicable—no new data generated.

Literature Cited

Almeida, C. P., Paulino, J. F. C., Morais Carbonell, S. A., Chiorato, A. F., Song, Q., Di Vittori, V., … Benchimol-Reis, L. L. (2020). Genetic diversity, population structure, and Andean introgression in Brazilian common bean cultivars after half a century of genetic breeding. *Genes*, *11*, 1298. doi: 10.3390/genes11111298.

Angioi, S. A., Rau, D., Attene, G., Nanni, L., Bellucci, E., Logozzo, G., … Papa, R. (2010). Beans in Europe: Origin and structure of the European landraces of *Phaseolus vulgaris L. Theo-*

retical and Applied Genetics, *121*, 829–843. doi: 10.1007/s00122-010-1353-2.

Ariani, A., & Gepts, P. (2019). Signatures of environmental adaptation during range expansion of wild common bean (*Phaseolus vulgaris*) [Preprint]. *BioRxiv*, 571042. Retrieved from https://www.biorxiv.org/content/10.1101/571042v1. doi: 10.1101/571042.

Ariani, A., Mier y Teran, J. C., & Gepts, P. (2017). Spatial and temporal scales of range expansion in wild *Phaseolus vulgaris*. *Molecular Biology and Evolution*, *35*, 119–131. doi: 10.1093/molbev/msx273.

Azani, N., Babineau, M., Bailey, C. D., Banks, H., Barbosa, A. R., Pinto, R. B., … Zimmerman, E. (2017). A new subfamily classification of the Leguminosae based on a taxonomically comprehensive phylogeny. The Legume Phylogeny Working Group (LPWG). *Taxon*, *66*(1), 44–77.

Bellucci, E., Aguilar, M., Alseekh, S., Bett, K., Brezeanu, C., Cook, D., … Papa, R. (submitted). The INCREASE project: Intelligent collections of food-legume genetic resources for European agrofood systems. *Plant Journal*, submitted.

Bellucci, E., Bitocchi, E., Ferrarini, A., Benazzo, A., Biagetti, E., Klie, S., … Papa, R. (2014). Decreased nucleotide and expression diversity and modified co-expression patterns characterize domestication in the common bean. *The Plant Cell*, *26*, 1901–1912. doi: 10.1105/tpc.114.124040.

Bellucci, E., Bitocchi, E., Rau, D., Nanni, L., Ferradini, N., Giardini, A., … Papa, R. (2013). Population structure of barley landrace populations and gene-flow with modern varieties. *PLoS One*, *8*(12), e83891. doi: 10.1371/journal.pone.0083891.

Bellucci, E., Bitocchi, E., Rau, D., Rodriguez, M., Biagetti, E., Giardini, A., … Papa, R. (2014). Genomics of origin, domestication and evolution of *Phaseolus vulgaris*. In R. Tuberosa, A. Graner, & E. Frison (Eds.), *Genomics of plant genetic resources* (pp. 483–507). Berlin, Germany: Springer.

Bioversity International. (2011). Crop ontology curation and annotation tool. Generation Challenge Programme, Bioversity International (Rome) as project-implementing agency. Retrieved from https://www.cropontology.org/.

Bitocchi, E., Nanni, L., Bellucci, E., Rossi, M., Giardini, A., Spagnoletti Zeuli, P., … Papa, R. (2012). Mesoamerican origin of the common bean (*Phaseolus vulgaris* L.) is revealed by sequence data. *Proceedings of the National Academy of Sciences of the United States of America*, *109*, E788–E796. doi: 10.1073/pnas.1108973109.

Bitocchi, E., Nanni, L., Rossi, M., Rau, D., Bellucci, E., Giardini, A., … Papa, R. (2009). Introgression from modern hybrid varieties into landrace populations of maize (*Zea mays* ssp. *mays* L.) in central Italy. *Molecular Ecology*, *18*, 603–621. doi: 10.1111/j.1365-294X.2008.04064.x.

Bitocchi, E., Bellucci, E., Rau, D., Albertini, E., Rodriguez, M., Veronesi, F., … Nanni, L. (2015). European flint landraces grown in situ reveal adaptive introgression from modern maize. *PLoS ONE*, *10*(4), e0121381. doi: 10.1371/journal.pone.0121381.

Bitocchi, E., Rau, D., Bellucci, E., Rodriguez, M., Murgia, M. L., Gioia, T., … Papa, R. (2017). Beans (*Phaseolus* ssp.) as a model for understanding crop evolution. *Frontiers in Plant Science*, *8*, 722. doi: 10.3389/fpls.2017.00722.

Bitocchi, E., Rau, D., Benazzo, A., Bellucci, E., Goretti, D., Biagetti, E., … Papa, R. (2017). High level of nonsynonymous changes in common bean suggests that selection under domestication increased functional diversity at target traits. *Frontiers in Plant Science*, *7*, 2005. doi: 10.3389/fpls.2016.02005.

Broughton, W. J., Hernández, G., Blair, M., Beebe, S., Gepts, P., & Vanderleyden, J. (2003). Beans (*Phaseolus* spp.) — Model food legumes. *Plant and Soil*, *252*, 55–128. doi: 10.1023/A:1024146710611.

Cichy, K. A., Porch, T., Beaver, J. S., Cregan, P., Fourie, D., Glahn, R. P., … Miklas, P. N. (2015). A *Phaseolus vulgaris* diversity panel for Andean bean improvement. *Crop Science*, *55*, 2149–2160. doi: 10.2135/cropsci2014.09.0653.

Cichy, K. A., Wiesinger, J. A., & Mendoza, F. A. (2015). Genetic diversity and genome-wide association analysis of cooking time in dry bean (*Phaseolus vulgaris* L.). *Theoretical and Applied Genetics*, *128*, 1555–1567. doi: 10.1007/s00122-015-2531-z.

Cortinovis, G., Di Vittori, V., Bellucci, E., Bitocchi, E., & Papa, R. (2020). Adaptation to novel environments during crop diversification. *Current Opinion in Plant Biology*, *13*, 1–15. doi: 10.1016/j.pbi.2019.12.011.

Cortinovis, G., Frascarelli, G., Di Vittori, V., & Papa, R. (2020). Current state and perspectives in population genomics of the common bean. *Plants*, *9*, 330. doi: 10.3390/plants9030330.

de la Hidalga, N. A., van Walsun, M., Rosin, P., Sun, X., & Wijers, A. (2019). Quality management methodologies for digitisation operations. Retrieved from https://zenodo.org/record/3469521.

Desiderio, F., Bitocchi, E., Bellucci, E., Rau, D., Rodriguez, M., Attene, G., … Nanni, L. (2013). Chloroplast microsatellite diversity in *Phaseolus vulgaris*. *Frontiers in Plant Science*, *3*, 312. doi: 10.3389/fpls.2012.00312.

Di Vittori, V., Bitocchi, E., Rodriguez, M., Alseekh, S., Bellucci, E., Nanni, L., … Papa, R. (2020). Pod indehiscence in common bean is associated to the fine regulation of PvMYB26 and a non-functional abscission layer. *Journal of Experimental Botany*, *72*(5), 1617–1633. doi: 10.1093/jxb/eraa553.

Diamond, J. (2002). Evolution, consequences and future of plant and animal domestication. *Nature*, *418*, 700–707. doi: 10.1038/nature01019.

Dwivedi, S. L., Salvatore, C., Blair, M. W., Upadhyaya, H. D., Are, A. K., & Ortiz, R. (2016). Landrace germplasm for improving yield and

abiotic stress adaptation. *Trends Plant Science*, *21*, 31–41. doi: 10.1016/j.tplants.2015.10.012.

Elshire, R. J., Glaubitz, J. C., Sun, Q., Poland, J. A., Kawamoto, K., Buckler, E. S., & Mitchell, S. E. (2011). A robust, simple genotyping-by-sequencing (GBS) approach for high diversity species. *PLoS One*, *6*(5), e19379. doi: 10.1371/journal.pone.0019379.

Gepts, P. (2010). Crop domestication as a long-term selection experiment. *Plant Breeding Reviews*, *24*, 1–44.

Gioia, T., Logozzo, G., Attene, G., Bellucci, E., Benedettelli, S., & Negri, V. (2013). Evidence for introduction bottleneck and extensive inter-gene pool (Mesoamerica × Andes) hybridization in the European common bean (*Phaseolus vulgaris* L.) germplasm. *PLoS One*, *8*, e75974. doi: 10.1371/journal.pone.0075974.

Glémin, S., & Bataillon, T. (2009). A comparative view of the evolution of grasses under domestication. *The New Phytologist*, *183*, 273–290. doi: 10.1111/j.1469-8137.2009.02884.x.

González, M. Y., Philipp, N., Schulthess, A. W., Weise, S., Zhao, Y., Börner, A., … Reif, J. C. (2018). Unlocking historical phenotypic data from an ex situ collection to enhance the informed utilization of genetic resources of barley (*Hordeum* sp.). *Theoretical and Applied Genetics*, *131*, 2009–2019. doi: 10.1007/s00122-018-3129-z.

Graham, P. H., & Vance, C. P. (2003). Legumes: Importance and constraints to greater use. *Plant Physiology*, *131*, 872–877. doi: 10.1104/pp.017004.

Guerra-García, A., Gioia, T., von Wettberg, E., Logozzo, G., & Bett, K. E. Increasing lentil genetic resources: Evolutionary history, recent genomic characterization of germplasm, and the need for well characterized collections. *Current Protocols in Plant Biology*, in preparation.

Harlan, J., & de Wet, J. (1971). Towards a rational classification of cultivated plants. *Taxon*, *20*, 509–517. doi: 10.2307/1218252.

International Board for Plant Genetic Resources (IPGRI). (1982). *Phaseolus vulgaris descriptors*. Rome, Italy: International Board for Plant Genetic Resources.

Kamfwa, K., Cichy, K. A., & Kelly, J. D. (2015a). Genome-wide association analysis of symbiotic nitrogen fixation in common bean. *Theoretical and Applied Genetics*, *128*, 1999–2017. doi: 10.1007/s00122-015-2562-5.

Kamfwa, K., Cichy, K. A., & Kelly, J. D. (2015b). Genome-wide association study of agronomic traits in common bean. *Plant Genome*, *8*, 1–12. doi: 10.3835/plantgenome2014.09.0059.

Kleintop, A. E., Myers, J. R., Echeverria, D., Thompson, H. J., & Brick, M. A. (2016). Total phenolic content and associated phenotypic traits in a diverse collection of snap bean cultivars. *Journal of the American Society for Horticultural Science*, *141*(1), 3–11. doi: 10.21273/JASHS.141.1.3.

Kroc, M. et al. (submitted). Lupin INCREASE Intelligent Collections: Characterization and development of single seed descent genetic resources. *Current Protocols*, submitted.

Kumar, S. et al. (submitted). Chickpea INCREASE Intelligent Collections: Characterization and development of single seed descent genetic resources. *Current Protocols*, in preparation.

Lewis, G. P., Schrire, B. D., Mackinder, B. A., Rico, L., & Clark, R. (2013). A linear sequence of legume genera set in a phylogenetic context: A tool for collections management and taxon sampling. *South African Journal of Botany*, *89*, 76–84. doi: 10.1016/j.sajb.2013.06.005.

Luikart, G., Kardos, M., Hand, B. K., Rajora, O. P., Aitken, S. N., & Hohenlohe, P. A. (2018). Population genomics: Advancing understanding of nature. In O. P. Rajora (Ed.), *Population genomics: Concepts, approaches and applications* (pp. 3–79). Cham, Switzerland: Springer International Publishing.

McClean, P. E., Bett, K. E., Stonehouse, R., Lee, R., Pflieger, S., Moghaddam, S. M., … Mamidi, S. (2018). White seed colour in common bean (*Phaseolus vulgaris*) results from convergent evolution in the P (pigment) gene. *New Phytologist*, *219*, 1112–1123. doi: 10.1111/nph.15259.

McClean, P. E., Moghaddam, S. M., Lopéz-Millán, A. F., Brick, M. A., Kelly, J. D., Miklas, P. N., … Grusak, M. A. (2017). Phenotypic diversity for seed mineral concentration in North American dry bean germplasm of MA ancestry. *Crop Science*, *57*, 3129–3144. doi: 10.2135/cropsci2017.04.0244.

McCouch, S., Navabi, K., Abberton, M., Anglin, N. L., Barbieri, R. L., Baum, M., … Rieseberg, L. H. (2020). Mobilizing crop biodiversity. *Molecular Plant*, *13*, 1341–1344. doi: 10.1016/j.molp.2020.08.011.

Milner, S. G., Jost, M., Taketa, S., Mazon, E. R., Himmelbach, A., Oppermann, M., … Stein, N. (2019). Genebank genomics highlights the diversity of a global barley collection. *Nature Genetics*, *51*, 319–326. doi: 10.1038/s41588-018-0266-x.

Mir, R. A., Sharma, A., & Mahajan, R. (2020). Crop landraces: Present threats and opportunities for conservation. In R. Salgotra & S. Zargar (Eds.), *Rediscovery of genetic and genomic resources for future food security* (pp. 335–350). Singapore: Springer.

Moghaddam, S. M., Brick, M. A., Echeverria, D., Thompson, H. J., Brich, L. A., Lee, R., … McClean, P. E. (2017). The genetic architecture of dietary fiber and oligosaccharide content in a MA panel of edible dry bean (*Phaseolus vulgaris L.*). *Plant Genome*, *11*, 1. doi: 10.3835/plantgenome2017.08.0074.

Moghaddam, S. M., Mamidi, S., Osorno, J. M., Lee, R., Brick, M., Kelly, J., … McClean, P. E. (2016). Genome-wide association study identifies candidate loci underlying agronomic traits in a Middle American diversity panel of common bean. *Plant Genome*, *9*, 1–21. doi: 10.3835/plantgenome2016.02.0012.

Mousavi-Derazmahalleh, M., Bayer, P. E., Hane, J. K., Valliyodan, B., Nguyen, H. T., Nelson, M. N., … Edwards, D. (2019). Adapting legume crops to climate change using genomic approaches. *Plant, Cell & Environment*, *42*, 6–19. doi: 10.1111/pce.13203.

Murgia, M. L., Attene, G., Rodriguez, M., Bitocchi, E., Bellucci, E., Fois, D., … Rau, D. (2017). A comprehensive phenotypic investigation of the 'pod-shattering syndrome' in common bean. *Frontiers in Plant Science*, *8*, 251. doi: 10.3389/fpls.2017.00251.

Myers, J. R., Wallace, L. T., Moghaddam, S. M., Kleintop, A. E., Echeverria, D., Thompson, H. J., … McClean, P. E. (2019). Improving the health benefits of snap bean: Genome-wide association studies of total phenolic content. *Nutrients*, *11*, 2509. doi: 10.3390/nu11102509.

Nelson, G., Paul, D., Riccardi, G., & Mast, A. (2012). Five task clusters that enable efficient and effective digitization of biological collections. *ZooKeys*, *209*, 19–45. doi: 10.3897/zookeys.209.3135.

Oladzad, A., Porch, T., Rosas, J. C., Moghaddam, S. M., Beaver, J., Beebe, S. E., … McClean, P. E. (2019). Single and multi-trait GWAS identify genetic factors associated with production traits in common bean under abiotic stress environments. *G3: Genes, Genomes, Genetics*, *9*(6), 1881–1892. doi: 10.1534/g3.119.400072.

Oladzad, A., Zitnick-Anderson, K., Jain, S., Simons, K., Osorno, J. M., McClean, P. E., & Pasche, J. S. (2019). Genotypes and genomic regions associated with *Rhizoctonia solani* resistance in common bean. *Frontiers in Plant Science*, *10*, 956. doi: 10.3389/fpls.2019.00956.

Parker, T. A., Berny Mier, Y., Teran, J. C., Palkovic, A., Jernstedt, J., & Gepts, P. (2020). Pod indehiscence is a domestication and aridity resilience trait in common bean. *New Phytologist*, *225*, 558–570. doi: 10.1111/nph.16164.

Rau, D., Murgia, M. L., Rodriguez, M., Bitocchi, E., Bellucci, E., Fois, D., … Papa, R. (2019). Genomic dissection of pod shattering in common bean: Mutations at non-orthologous loci at the basis of convergent phenotypic evolution under domestication of leguminous species. *The Plant Journal*, *97*, 693–714. doi: 10.1111/tpj.14155.

Rendón-Anaya, M., Montero-Vargas, J. M., Saburido-Álvarez, S., Vlasova, A., Capella-Gutierrez, S., Ordaz-Ortiz, J. J., … Herrera-Estrella, A. (2017). Genomic history of the origin and domestication of common bean unveils its closest sister species. *Genome Biology*, *18*, 60. doi: 10.1186/s13059-017-1190-6.

Rodriguez, M., Rau, D., Bitocchi, E., Bellucci, E., Biagetti, E., Carboni, A., … Attene, G. (2016). Landscape genetics, adaptive diversity, and population structure in *Phaseolus vulgaris*. *New Phytologist*, *209*, 1781–1794. doi: 10.1111/nph.13713.

Schmutz, J., McClean, P. E., Mamidi, S., Wu, G. A., Cannon, S. B., Grimwood, J., … Jackson, S. A. (2014). A reference genome for common bean and genome-wide analysis of dual domestications. *Nature Genetics*, *46*, 707–713. doi: 10.1038/ng.3008.

Schröder, S., Mamidi, S., Lee, R., McKain, M. R., McClean, P. E., & Osorno, J. M. (2016). Optimization of genotyping by sequencing (GBS) data in common bean (*Phaseolus vulgaris* L.). *Molecular Breeding*, *36*, 6. doi: 10.1007/s11032-015-0431-1.

Singh, N., Wu, S., Raupp, W. J., Sehgal, S., Arora, A., Tiwari, V., … Poland, J. (2019). Efficient curation of genebanks using next generation sequencing reveals substantial duplication of germplasm accessions. *Scientific Reports*, *9*, 650. doi: 10.1038/s41598-018-37269-0.

Soltani, A., Moghaddam, S. M., Walter, K., Restrepo-Montoya, D., Mamidi, S., Schroder, S., … Osorno, J. M. (2017). Genetic architecture of flooding tolerance in the dry bean Middle-American diversity panel. *Frontiers in Plant Science*, *8*, 1183. doi: 10.3389/fpls.2017.01.

Song, Q., Jia, G., Hyten, D. L., Jenkins, J., Hwang, E. Y., Schroeder, S. G., … Cregan, P. B. (2015). SNP assay development for linkage map construction, anchoring whole-genome sequence, and other genetic and genomic applications in common bean. *G3: Genes, Genomes, Genetics*, *5*, 2285–2290. doi: 10.1534/g3.115.020594.

Tock, A. J., Fourie, D., Walley, P. G., Holub, E. B., Soler, A., Cichy, K. A., … Miklas, P. N. (2017). Genome-wide linkage and association mapping of halo blight resistance in common bean to race 6 of the globally important bacterial pathogen. *Frontiers in Plant Science*, *8*, 1170. doi: 10.3389/fpls.2017.01170.

Toro, O., Tohme, J., & Debouck, D. G. (1990). Wild bean (*Phaseolus vulgaris* L.): Description and distribution. In *Centro Internacional de Agricultura Tropical (CIAT)* Cali, Colombia: International Board for Plant Genetic Resources (IBPGR).

Vavilov, N. (1920). *The law of homologous series in variation*. Lecture at the 3rd All-Russian Breeding Conference, Saratov.

Vavilov, N. (1922). The law of homologous series in variation. *Journal of Genetics*, *12*, 47–89. doi: 10.1007/BF02983073.

Vlasova, A., Capella-Gutiérrez, S., Rendón-Anaya, M., Hernández-Oñate, M., Minoche, A. E., Erb, I., … Guigó, R. (2016). Genome and transcriptome analysis of the Mesoamerican common bean and the role of gene duplications in establishing tissue and temporal specialization of genes. *Genome Biology*, *17*, 32. doi: 10.1186/s13059-016-0883-6.

Wallace, L., Arkwazee, H., Vining, K., & Myers, J. R. (2018). Genetic diversity within snap beans and their relation to dry beans. *Genes*, *9*(12), 587. doi: 10.3390/genes9120587.

Wu, J., Wang, L., Fu, J., Chen, J., Wei, S., Zhang, S., … Wang, S. (2020). Resequencing of 683 common bean genotypes identifies yield component trait associations across a north–south cline. *Nature Genetics*, *52*, 118–125. doi: 10.1038/s41588-019-0546-0.

Zeven, A. C. (1998). Landraces: A review of definitions and classifications. *Euphytica*, *104*, 127–139.

Zhu, Y., Chen, H., Fan, J., Wang, Y., Li, Y., Chen, J., … Mundt, C. C. (2000). Genetic diversity and disease control in rice. *Nature*, *406*, 718–722.

Zuiderveen, G. H., Padder, B. A., Kamfwa, K., Song, Q., & Kelly, J. D. (2016). Genome-wide association study of anthracnose resistance in Andean beans (*Phaseolus vulgaris*). *PLoS One*, *11*, e0156391. doi: 10.1371/journal.pone.0156391.

Internet Resources

https://ssl.fao.org/glis/

Global Information System of the International Treaty on Plant Genetic Resources for Food and Agriculture. This service assigns Digital Object Identifiers (dois) to Plant Genetic Resources for Food and Agriculture (PGRFA) for reference in third-party systems and the scientific literature.

https://www.genesys-pgr.org/

Online platform hosting information about Plant Genetic Resources for Food and Agriculture (PGRFA) conserved in gene banks worldwide.

https://phytozome.jgi.doe.gov/pz/portal.html

Plant Comparative Genomics portal of the U.S. Department of Energy's Joint Genome Institute.

https://www.pulsesincrease.eu/

INCREASE Project website.

https://creativecommons.org

The Creative Commons license CC BY-SA.

http://www.cropontology.org

Crop ontology curation and annotation tool.

https://www.xrite.com/categories/calibration-profiling/colorchecker-classic-family/colorchecker-classic-mini

X-Rite, Inc. (2020) ColorChecker Classic Mini (retrieved 26 October 2020).

Chapter II

Current State and Perspectives in Population Genomics of the Common Bean

Gaia Cortinovis, Giulia Frascarelli, Valerio Di Vittori and Roberto Papa *

Dipartimento di Scienze Agrarie, Alimentari ed Ambientali (D3A), Università Politecnica delle Marche, Via Brecce Bianche, 60131 Ancona, Italy; g.cortinovis@pm.univpm.it (G.C.); g.frascarelli@pm.univpm.it (G.F.); v.divittori@staff.univpm.it (V.D.V.)

* Correspondence: r.papa@univpm.it; Tel.: +39-071-220-4984

DOI: https://doi.org/10.3390/plants9030330

plants

MDPI

Review

Current State and Perspectives in Population Genomics of the Common Bean

Gaia Cortinovis, Giulia Frascarelli, Valerio Di Vittori and Roberto Papa *

Dipartimento di Scienze Agrarie, Alimentari ed Ambientali (D3A), Università Politecnica delle Marche, Via Brecce Bianche, 60131 Ancona, Italy; g.cortinovis@pm.univpm.it (G.C.); g.frascarelli@pm.univpm.it (G.F.); v.divittori@staff.univpm.it (V.D.V.)

* Correspondence: r.papa@univpm.it; Tel.: +39-071-220-4984

Received: 24 January 2020; Accepted: 3 March 2020; Published: 5 March 2020

Abstract: Population genomics integrates advances in sequencing technologies, bioinformatics tools, statistical methods and software into research on evolutionary and population genetics. Its application has provided novel approaches that have significantly advanced our understanding of new and long-standing questions in evolutionary processes. This has allowed the disentangling of locus-specific effects from genome-wide effects and has shed light on the genomic basis of fitness, local adaptation and phenotypes. "-Omics" tools have provided a comprehensive genome-wide view of the action of evolution. The specific features of the *Phaseolus* genus have made it a unique example for the study of crop evolution. The well-documented history of multiple domestications in *Phaseolus vulgaris* L. (common bean) and its further adaptation to different environments have provided the opportunity to investigate evolutionary issues, such as convergent evolution in the same species across different domestication events. Moreover, the availability of the *P. vulgaris* reference genome now allows adaptive variations to be easily mapped across the entire genome. Here, we provide an overview of the most significant outcomes obtained in common bean through the use of different computational tools for analysis of population genomics data.

Keywords: population genomics; genetic diversity; evolutionary history of the common bean; adaptive selection

1. Introduction

According to neutral theory, the great majority of evolutionary changes at the molecular level involve random fixation of selectively neutral (or nearly neutral) alleles through cumulative effects of sampling drift and under the input of novel mutations [1,2]. Neutral theory also provides the theoretical framework to be able to disentangle the roles of different evolutionary forces in the shaping of the diversity within species and populations, in order to distinguish the effects of adaptation from those of demography and population history [3,4]. The beginning of the new century can be considered the start of the population genomics era. This refers to the use of high-density markers and genome-wide sampling to identify and separate locus-specific effects (e.g., selection) from genome-wide effects (e.g., drift, gene flow and inbreeding), with the aim being to improve our understanding of population microevolution [3].

Population genomics has not just been a conceptual advance but, rather, a larger-scale approach. Its applications can address questions that have long been studied using previous tools (e.g., effective population size, population structure, phylogeography and demography) [5]. The use of novel tools and statistical tests now allows previously inaccessible issues to be addressed, such as physical mapping of adaptive variations and of molecular variants that underlie genotype fitness and relevant phenotypic variations throughout the genome [6–9].

Over the last 50 years, several techniques have been developed to assess genetic diversity and to investigate molecular evolution and phylogeny of plants. In addition to the use of classical markers (e.g., Mendelian traits), the first estimates of the levels of genetic variation within and between natural populations at multiple loci were provided in the 1970 s, using allozyme analyses [10,11]. The arrival of DNA-based markers in the late 1970 s allowed deeper investigations into genetic diversity, with the possibility to observe patterns of variation directly in DNA sequences and to quantify the number of mutations between different alleles [12]. In particular, analyses of mitochondrial DNA laid the foundation for the phylogeography field through the use of restriction fragment length polymorphism (RFLP) markers; this provided a deeper look in time at the relationships and connectivity among populations [13]. Since 1983, the applications of the polymerase chain reaction (PCR) completely revolutionized the approaches for the development and screening of genetic markers through our enhanced ability to discover mutations, which has resulted in significant advantages, such as the opportunity to examine many polymorphic loci [14].

Several PCR-based genotyping methods fall under the general category of DNA fingerprinting [15]. Among these, the discovery of simple sequence repeat (SSR) markers (also known as microsatellites), which are loci with tandem repeats of two to six nucleotide motifs, has allowed the direct scoring of both homozygous and heterozygous loci. By the end of the twentieth century, SSRs generally became the markers of choice in different population genomics studies [16]. In parallel, the development of amplified fragment length polymorphism (AFLP), which can be considered as the first class of genome-wide markers, provided the possibility to co-amplify an unprecedented number of restriction fragments without a priori information about the nucleotide sequence [17].

Direct PCR-based DNA sequencing opened the path for new approaches to genomic characterization, most notably with the discovery of the so-called third-generation markers: single nucleotide polymorphisms (SNPs). SNPs are the most abundant bi-allelic and co-dominant markers, and they are characterized by simple mutational patterns [18], with their exploitation initially made possible using Sanger sequencing technology. The advent of genotyping microarrays and, in the last few decades, the development of high-throughput sequencing methods (e.g., next-generation sequencing), further enhanced the detection and characterization of molecular markers. Next-generation sequencing platforms include such systems as: the Illumina genome analyzer, including the HiSeq, MiSeq, NextSeq and NovaSeq systems; the 454 Life Sciences FLX genome sequencer; the Thermo Fischer Scientific SOLiD, Ion Torrent and Ion Proton systems; the PacBio real-time sequencer and, more recently, the Oxford nanopore technologies. These have provided the possibility to produce millions of DNA sequence reads in a massive, parallel and high-throughput way, which has made entire genomes and transcriptomes available for population genomics studies at an exponential pace [19].

The availability of high-throughput sequencing platforms and high-density DNA markers has allowed parallel analyses of many loci on relatively large sample sizes and exploiting several statistical methods [20]. These advances have thus expanded the detection and conservation of important genetic variations to also provide a comprehensive genome-scale view of the actions of evolution [21], even in non-model organisms [22]. Moreover, the possibility to explore molecular phenotypes (i.e., metabolomics and transcriptomics data) has allowed the development of molecular evolutionary phenomics approaches [23,24].

The basic population genomics approach is characterized by four steps: sampling of individuals with different phenotypes and/or from different environments, genome-wide genotyping with high-density molecular markers, testing for outlier loci in population datasets and validation of loci that are both neutral and under selection. Neutral loci can be used to infer population demography and history, while loci putatively under selection provide adaptive information, which can be used for biodiversity conservation and evolutionary inferences [4].

With the increasing number of genetic markers available and the greater computational capacity of computers, given a sample of genes, it has also become possible to simulate the evolutionary history of a population/species under different and realistic evolutionary scenarios. Reconstruction of the

genealogy that describes the descent relationships underlies the "backwards-in-time" models, for which the mathematical description is provided in the coalescent model [25]. In plant research, one of the first population genomics approaches dates back to 2005, with comprehensive studies on the whole genome of maize. In this regard, Wright et al. [26] and Yamasaki et al. [27] performed large-scale genomic screening for SNPs on 774 and 1095 randomly selected maize genes, respectively. With the aim to better understand the effects of artificial selection, these studies used a novel coalescent simulation approach and likelihood analysis, and they estimated that 2% to 4% of maize genes have been under selection during maize domestication and improvement. Moreover, the integration of quantitative trait locus (QTL) mapping and the analysis of the "selective sweep" effects allowed the target genomic regions that were under selection to be narrowed down [28].

All of these recent genomics advances described above were also widely applied in population genetics studies of the common bean, and a new epoch began for this crop. With the release of the high-quality reference genomes of the Andean G19833 [29] and the Mesoamerican BAT93 [30] genotypes, several evolutionary issues were clarified. This improved our understanding of the genome organization and its structural variations, as well as of the environmental adaptation, geographic origins, domestication and diversification of the common bean. Recent progress in population genomics of the common bean have also provided the opportunity to extend this knowledge to closely related legume species and more widely to other crops through comparative genomics studies.

Phaseolus spp. can be considered a unique model for the study of crop evolution. It comprises five domesticated species (*P. vulgaris*, *P. coccineus*, *P. dumosus*, *P. acutifolius* and *P. lunatus*), two of which were domesticated independently both in Mesoamerica and in the Andes (*P. vulgaris* and *P. lunatus*), which offered the opportunity to disentangle the genetic basis of the domestication process not only among species of the same genus but also between gene pools within the same species. Moreover, their recent divergence and their different mating systems make *Phaseolus* spp. an ideal system for comparative genomics studies [31]. Here we will review *P. vulgaris* studies that have addressed evolutionary questions.

2. Origin of the Common Bean in the Light of Different Molecular Markers

The extant genetic diversity of a population is the result of its complex evolutionary history and factors such as genetic drift, gene flow (including introgression from wild forms or closely related species), selection and mutation, along with the mating systems, are crucial in the shaping of the genetic diversity and structure of a pool of individuals.

Wild forms of *P. vulgaris* extend across the highlands of what is now Latin America, between Northern Mexico and Northwestern Argentina [32]. These are characterized by three main eco-geographic gene pools: Mesoamerican and Andean, which are the major gene pools, with both including wild and domesticated forms, and the population from Northern Peru-Ecuador (PhI), which has a relatively narrow distribution of wild individuals [33]. Phaseolin data [34,35], allozymes [36] and multi-locus markers [37–40] have together confirmed the structure of the diversity of the gene pools of the common bean, and have often highlighted the higher genetic variability of the Mesoamerican gene pool compared to the Andean one. Indeed, Rossi et al. [40] used a large set of AFLP markers to dissect out the internal structure within both the Mesoamerican and the Andean gene pools, and they always detected a higher proportion of polymorphic loci in the wild forms compared to the domesticated ones, with the Mesoamerican gene pool being much more diverse and structured compared to the Andean population. Rossi et al. [40] compared their AFLP data with SSR data from Kwak and Gepts [41], and they noted that the differences in genetic diversity between the Mesoamerican and Andean wild gene pools were highly associated with the mutation rates of the molecular markers; the higher the marker mutation rate, the lower the differences between the Mesoamerican and Andean gene pools (Figure 1). Based on this, and assuming the bottleneck model proposed by Nei et al. [42], Rossi et al. [40] suggested that only the Mesoamerican origin of *P. vulgaris* could explain the contrasting patterns of diversity for different molecular markers when comparing

the Andean and Mesoamerican gene pools. This hypothesis was supported and strongly consolidated by Bitocchi et al. [43], who showed a loss of nucleotide diversity (Lπ) of 90% in the Andean wild population compared to the Mesoamerican population (Figure 1).

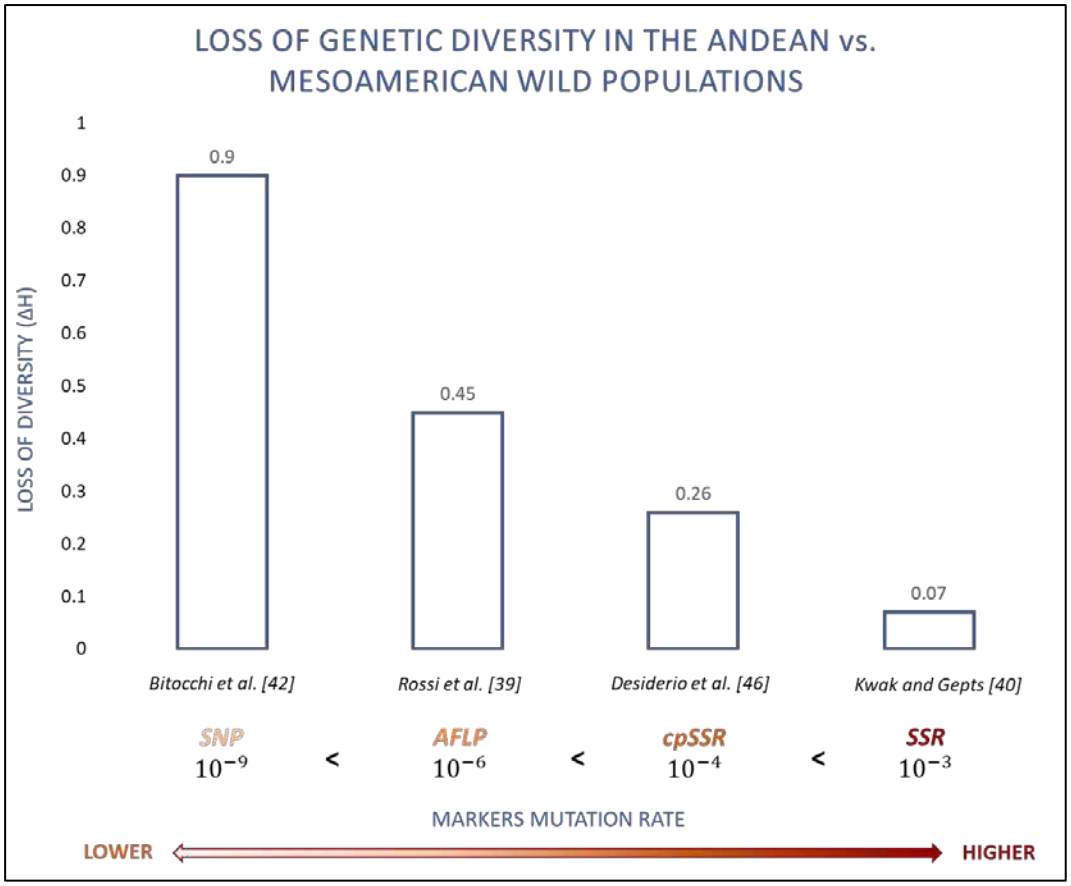


Figure 1. Loss of genetic diversity in the Andean versus Mesoamerican wild populations in the light of different molecular markers. SNP: single nucleotide polymorphisms, AFLP: amplified fragment length polymorphisms, cpSSR: chloroplast simple sequence repeats and SSR: simple sequence repeats.

Markers that differ in their mutation rates can highlight very different patterns of molecular diversity in the same species or population; this is because the number of generations needed for mutations to allow the recovery of the genetic diversity after a bottleneck is expected to be close to the reciprocal of the mutation rate of the markers [42,44,45]. Thus, the lower mutation rates of SNPs (for the *Fabaceae* family, this was estimated to be ~6.1×10^{-9} mutations per base pair per generation [46]) compared to other types of markers allowed Bitocchi et al. [43] to detect the occurrence of the Andean bottleneck with much greater resolution. Indeed, they detected a loss of genetic diversity of about two-fold, three-fold and 13-fold those observed in a comparable sample of *P. vulgaris* genotypes using AFLP (45%) [40], chloroplast (cp) SSRs (26%) [47] and SSRs (7%) [41], respectively (Figure 1).

The Mesoamerican origin hypothesis was also supported by additional data from Bitocchi et al. [43], who were the first to report a clear-cut population structure into four different groups for the wild Mesoamerican accessions. Moreover, their phylogenetic analysis revealed that both the Andean and the Northern Peru-Ecuador wild accessions were strongly related to two distinct Mesoamerican groups that were located in a wide area of Central Mexico. Thus, both the Andean and Northern Peru-Ecuador gene pools appeared to have originated through different migration events from the Mesoamerican populations of Central Mexico (Figure 2), as also confirmed by the work of Schmutz et al. [29] and supported by the approximate Bayesian computation analysis performed by Ariani et al. [48].

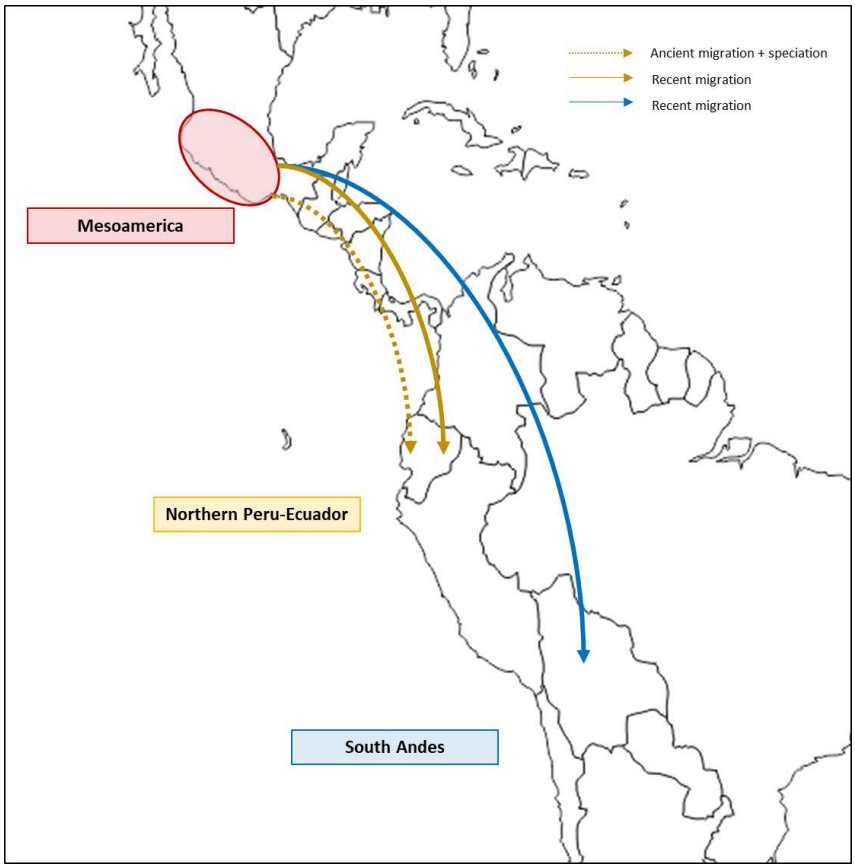


Figure 2. Graphical representation of the two different evolutionary hypotheses for wild *P. vulgaris* migration in America.

Recently, Rendón-Anaya et al. [49] proposed the hypothesis of a slightly different evolutionary history, with the suggestion that the introduction in Northern Peru-Ecuador originated from an ancestral form and occurred much earlier than the diversification of *P. vulgaris* within the *Vulgaris* group (Figure 2). They analyzed both the chloroplast and nuclear genomes by sequencing 18 *P. vulgaris* accessions that spanned the three gene pools (eight wild, two domesticated Mesoamerican accessions; one wild, two domesticated Andean accessions and five Northern Peru-Ecuador accessions) to reconstruct the phylogenetic relationships of this species. While they supported the Mesoamerican origin of common bean, Rendón-Anaya et al. [49] also proposed an early dissemination and speciation event in Mesoamerica before the split into the two current major gene pools (i.e., Mesoamerican and Andean). Using a maximum-likelihood approach, they performed a phylogenetic analysis based on genome-wide SNPs and on a 55-kb chloroplast genome fragment. They observed two distinct clades, one that included all of the wild Peru-Ecuador accessions and another that included all of the other *P. vulgaris* genotypes. These results led them to conclude that the Northern Peru-Ecuador population evolved from a speciation event that occurred before the separation of the Andean gene pool from the Mesoamerican gene pool. However, considering the results obtained by Desiderio et al. [46], more genotypes (preferably as wild accessions) should be analyzed to validate this hypothesis.

3. Domestication of the Common Bean

Phenotypic and genotypic information based on a variety of methods coherently support the occurrence of two independent domestication events, one in Mesoamerica and the other in the Andes, where the two major domesticated gene pools originated [35,40,41,50,51]. The genetic routes towards domestication were also confirmed recently by Schmutz et al. [29], who performed pooled resequencing

of 160 wild and domesticated accessions from the centers of origin, as well as releasing the first high-quality reference genome of *P. vulgaris*.

3.1. Mapping Domestication Traits

Domestication resulted in several phenotypic and genetic changes between the domesticated forms and their wild ancestors (i.e., the domestication syndrome), such as differences in growth habit and photoperiod sensitivity; variations in shape, color and size of the edible parts and reduction or loss of seed dormancy and seed shattering. Starting from the pioneering work of Koinange et al. [52] to date, molecular variations and QTLs associated with the modulation of domestication traits have been mapped and partially characterized [53], with recent major advances towards the identification of the genes responsible for seed shattering [54–56], determinate/indeterminate growth habits [57] and seed color [58].

One of the most important targets of domestication in the common bean was the seed shattering trait. Wild common bean is characterized by high seed shattering, which is crucial for propagation of the progeny and to ensure high fitness of the genotypes. Conversely, the domesticated forms have partially or completely lost this seed dispersal, as lower dehiscence ensures reduction of yield losses in the field [59,60]. Rau et al. [54] used an introgression line-mapping population and proposed a model with a major locus associated to the complete loss of pod shattering, which was localized to the distal part of chromosome Pv05, plus additional QTLs hypostatic to the major QTL, at which the cumulation of wild alleles increases the level and mode of seed shattering (i.e., number of shattered and twisted pods per plant). The major QTL was recently confirmed by Parker et al. [55], who performed association mapping on a panel of 208 Andean accessions. In the model proposed by Rau et al. [54] for modulation of seed shattering, at least other two loci on chromosomes Pv05 and Pv04 were proposed, and overall, the model explained 72.4% of the phenotypic variance for this trait. Additional loci with minor effects on chromosomes Pv09, Pv04 and Pv06 were proposed to explain the variability associated with the level of seed shattering [54]. The multi-locus control of seed shattering was confirmed by Parker et al. [55], who mapped loci for seed shattering also on chromosomes Pv03, Pv08 and Pv09. Several genes have been proposed to be responsible for the genetic control of the pod-shattering trait, but additional analyses are needed to further narrow down the list of candidates. With regards to the growth habit, Repinsky et al. [57] identified the functional orthologue to *AtTFL1* (*Terminal Flower 1*) in the common bean (*PvTFL1y*). The co-segregation of *PvTFL1y* with the *fin* locus [52] for the determinate growth habit, the strong decrease in mRNA abundance associated with two haplotypes at *PvTFL1y* locus and the rescue of the indeterminate phenotype in the *tfl1* mutant in *Arabidopsis* with the wild allele of *PvTFL1y* allowed Repinsky et al. [57] to establish that *PvTFL1y* gene controls the determinate/indeterminate habitus. McClean et al. [58] recently characterized the molecular structure of the gene responsible at the *P* locus for the presence/absence of seed color in the common bean, and they validated its function through virus-induced gene silencing. They identified four alleles at the basis of the pigmented seeds phenotype, while several independently derived *p* alleles for white seeds were detected, which suggested that a convergent evolution mechanism is at the basis of the white-seed phenotype. Recent and future advances in the development of population genomics tools and statistical approaches will have a crucial role in shedding light on the genetic basis of pod shattering and on other domestication/diversification traits.

3.2. Signature of Selection

As previously described, population genomics aims to disentangle the effects of selection from those of other evolutionary forces through analysis of the aberrant patterns of DNA polymorphisms and assuming a neutral scenario. The domestication process is usually associated with a reduction in genetic diversity [40,50,61,62] and with an increase in divergence between wild and domesticated populations, due to demographic factors that affect the entire genome and to natural and artificial selection at target loci [63]. Moreover, considering the parallel domestication in the Andes and

Mesoamerica, Bitocchi et al. [50] estimated a three-fold greater reduction in genetic diversity between the wild and domesticated Mesoamerican pools, with respect to the equivalent comparison in the Andean pool. These data can be explained as a consequence of the bottleneck that occurred in the wild Andean germplasm, which impoverished the genetic diversity before domestication and resulted in a lesser effect of the subsequent domestication in the Andes.

Papa et al. [64] used AFLP markers to identify several associations between the map locations of various domestication genes and QTLs and the regions of high divergence between the wild and domesticated genotypes. The potential of the use of population genomics in the common bean was also demonstrated by Papa et al. [65], where they used pooled DNA samples and analyzed 2506 AFLP loci to identify a large portion of the genome (16%) that had been affected by the domestication process, with many markers under selection associated to known loci for the domestication syndrome traits.

Bellucci et al. [23] exploited the potential of RNA sequencing (RNA-seq), which combines information from nucleotide diversity and gene expression, and demonstrated for the first time that common bean domestication in Mesoamerica was characterized not only by a significant reduction in the nucleotide diversity but also by deep impact on the architecture of gene expression and co-expression at the whole transcriptome level. In more detail, Bellucci et al. [23] adopted an approach based on de novo assembly of a reference transcriptome, and they mapped on it the RNA-seq data from 21 inbred genotypes: 10 wild and eight domesticated Mesoamerican genotypes, with one wild and two domesticated Andean genotypes as controls. The final dataset of 188,107 SNPs distributed across 27,243 contigs was used to study the domestication process of the common bean in Mesoamerica. Bellucci et al. [23] identified signatures of selection on contigs from RNA-seq data by testing the significance of two ad-hoc statistical indices and using a coalescent simulation approach that considered the absence of selection during domestication. Taking into consideration the demographic parameters available from previous studies [61,66], Bellucci et al. [23] revealed that 9% of the contigs were actively selected during common bean domestication and that the selection in these contigs induced further reductions (26%) in the diversity of gene expression. Generally, genes that are putatively under selection show greater genetic diversity in the wild alleles compared to the domesticated alleles. Indeed, most of the contigs affected by selection in Bellucci et al. [23] were monomorphic in the domesticated gene pool and polymorphic in the wild germplasm. However, diversifying selection was also detected, which was reflected in a small fraction (2.8%) of the contigs in which the wild forms were fixed monomorphic, while the domesticated accessions were highly polymorphic. In addition, looking at differentially expressed contigs, down-regulation was observed mainly in the domesticated accessions, compared to the wild accessions, which indicated the occurrence of loss-of-function mutations. These results suggested that domestication increased the functional diversity at a few target loci in parallel with an overall reduction in genetic diversity at the transcriptome-wide level. The results of Bellucci et al. [23] can be imputed to novel mutations that were selected for expansion and adaptation to new environments and agro-ecological growth conditions.

In parallel with the release of the common bean reference genome, Schmutz et al. [29] performed the first genome-wide analysis that considered both of the gene pools. They dissected out the effects of domestication at the genome-wide level by comparing wild and landrace accessions across 10-kb/2-kb sliding windows in the top 90% of the empirical distribution of the population for both $\pi_{wild}/\pi_{landrace}$ ratios and F_{ST} values. They analyzed the F_{ST} distribution and the loss of nucleotide diversity, and they defined genes and genomic regions under selection during domestication in each of the gene pools. Interestingly, only 7.2 Mb of the genome putatively under selection were shared between the Mesoamerican and Andean groups. Moreover, out of 1835 Mesoamerica and 748 Andean candidate genes, only 59 were common between the two domestication events.

Out of the total of 2364 PS contigs identified by Bellucci et al. [23], Di Vittori [56] identified 1642 PS genes that are the reference for 1935 PS contigs. According to the new mapping for the PS contigs of Bellucci et al. [23] and to the available information from Schmutz et al. [29], Figure 3 shows the genome-wide PS gene density in the Mesoamerican [23,29] and Andean [29] gene pools.

These maps were constructed using the RIdeogram package [67] in R. Interestingly, for all of the dataset, we identified only a few common regions with higher densities of PS genes as, for example, at the end of chromosomes Pv02 (~42–44 Mb) and Pv06 (from ~26 Mb to the end of the chromosome) and at the beginning of chromosomes Pv07 and Pv08. Moreover, in both of these studies, genomic regions with high-density of PS genes specific for the Mesoamerica gene pool were identified as, for example, at the end of chromosome Pv01 and around 10-12 Mb on Pv09. With regard to these regions, Parker et al. [55] recently identified a QTL for seed shattering at the beginning of chromosome Pv08, while the orthologue to *ATIND* [68] (*PvIND* in [69]) was localized to 44 Mb on Pv02 in the same region where the *St* locus for the pod string presence was mapped [52]. Interestingly, the major growth habit gene in the common bean, Phvul.001G189200 (*PvTFL1y*) [57], is located at ~45 Mb on Pv01, close to regions with relative high densities of selection signatures in the Mesoamerican gene pool. In addition to these regions, Figure 3 also highlights differences between Mesoamerican and Andean pools for the PS gene location, according to the data of Schmutz et al. [29], which agrees with the occurrence of at least two parallel domestication events.

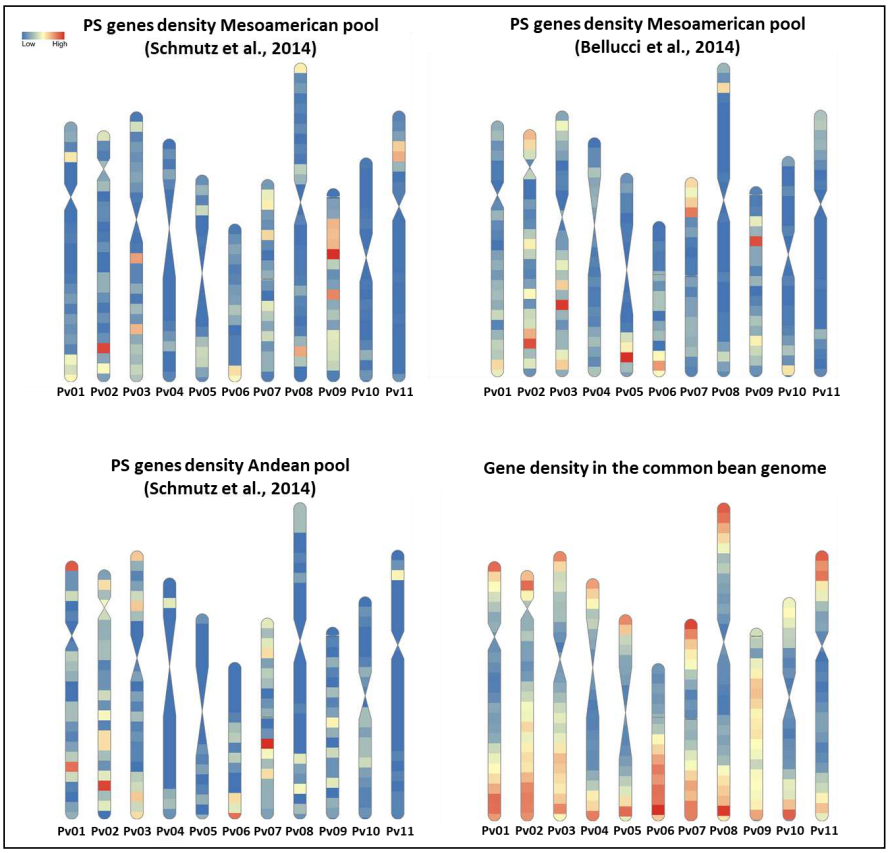


Figure 3. Genome-wide distribution of putatively under selection (PS) genes. **Top left**: distribution of PS genes in the Mesoamerica gene pool, according to Schmutz et al. [29]. **Top right**: distribution of PS genes in the Mesoamerica gene pool, according to Bellucci et al. [23]. **Bottom left**: distribution of PS genes in the Andean gene pool, according to Schmutz et al. [29]. **Bottom right**: gene density across the entire genome. Gene density is highlighted according to the color intensity in the legend at a 2-Mb window scale.

However, a similar density distribution of PS genes can be observed between the Mesoamerican [23] and Andean [29] gene pools with respect to the analysis of Schmutz et al. [29] of the Mesoamerican gene pool, as for example for the chromosome Pv02 (Figure 3). This last observation might suggest that integration of different datasets and approaches (e.g., genomic, transcriptomic and metabolomic analyses) can allow comparative studies and the identification of uncovered selection signatures,

although, at the same time, discordance between different analyses might arise from the different statistical approaches and sampling issues adopted for the detection of the selection signatures. Thus, deeper investigations are needed to understand the convergent evolution of domestication traits in common bean.

4. Diversification and Adaptation of P. Vulgaris to Different Agro-Ecological Conditions

As mentioned above, the wild common bean originated in Mesoamerica and, through migration events, it subsequently expanded from Northern Mexico to Northwestern Argentina, to encompass some 70 latitudinal degrees and elevations between 500 and 2000 m a.s.l [70–72]. This broad geographic distribution indicated that, through adaptive evolution, *P. vulgaris* colonized different agro-ecological locations compared to its area of origin. Several population genomics approaches can uncover the basis of genetic adaptation to different environments by looking at specific signatures at the genomic sequence level (e.g., outlier-loci detection strategy) or by looking for associations between genomic loci, phenotypic traits and environmental data [73,74].

Landscape genomics is an emerging discipline that combines population genetics, landscape ecology and spatial analytical techniques to identify environmental factors that have shaped adaptive variations that underlie local adaptation [4,75]. The number of landscape genomics studies has risen exponentially since 2003 [76], and among these, a study conducted by Rodriguez et al. [77] was the first particularly thought-provoking example of the potential of this method in *P. vulgaris*. In more detail, Rodriguez et al. [77] analyzed correlations between molecular markers and ecological variables at a continental scale. They used 131 SNPs in a population of 577 accessions (417 wild and 160 domesticated) that encompassed the wide geographic distribution of the common bean in America, with the aim being to examine the genetic–spatial patterns of the wild common bean. They reported the existence of well-defined wild genetic groups and variable degrees of diversity in both the Mesoamerican and Andean gene pools. Therefore, they investigated the spatial distribution of diversity using Mantel tests and multivariate analysis (using genetic and geographic information), which allowed them to determine the correlation between genetic and geographic distances. These analyses highlighted the presence of global structures (i.e., geographically closer individuals were also more genetically similar), which suggested that the effects of migration and genetic drift overlapped with selection effects in the same direction, with the consequent divergent selection as a result of local adaptation. Geographic and environmental data were combined with genetic diversity data to separate the effects of geography from those of ecology, and they reported a total of 26 loci (19.8%) that were putatively under selection for adaptation. Among these, different loci were shown to have compatible functions with adaptation features, such as chilling susceptibility, cold acclimation and mechanisms related to drought stress.

Recently, Ariani and Gepts [78] performed a similar analysis on 246 wild common bean accessions using a larger number of markers (~20,000 SNPs) that were widely distributed across the genome. Ariani and Gepts [78] coupled 19 bio-climatic variables with genome scan analysis for selection and genome-wide association analysis to identify which gene pools/genes were putatively under adaptive selection by temperature. Among the candidate genes identified by Ariani and Gepts [78], *Phvul.002G143100* appears to be particularly interesting; indeed, in the *Arabidopsis thaliana* model system, the homologous gene (*AtGRDP2*) is involved in flowering-time regulation, and its overexpression results in significant reduction in days to flowering [79]. The timing of important phenological stages is one of the most crucial diversification traits, in as much as it reflects the adaptation of a species through the tailoring of vegetative and reproductive growth phases to local climatic effects.

P. vulgaris became widespread not only in the Americas, as its cultivation extended worldwide, and it became the most important grain legume for direct human consumption [80]. The dissemination and introduction of the common bean into the Old World, as well as for other New World crops such as tomato, maize, squash, potato and tobacco, occurred after the 1492 voyage of Christopher Columbus. To investigate the evolutionary patterns of the common bean far from the New World, Angioi et al. [81] analyzed 94 and 307 *P. vulgaris* accessions from the Americas and Europe, respectively.

Several studies based on molecular and biochemical markers [69,81,82] reported that the European common bean populations include both Mesoamerican and Andean forms, and that the Andean germplasm was the most represented in Europe, even though the proportions of these gene pools can significantly differ across countries. Angioi et al. [81] also estimated that 44.2% of the European landraces derived from at least one hybridization event between Mesoamerican and Andean forms, which demonstrated the fundamental role of hybridization and recombination in the origin of the European common bean gene pool. This hybridization was promoted by the breakdown of the spatial isolation between the Mesoamerican and Andean accessions after their introduction into Europe and had a crucial impact on the maintenance of genetic diversity and common bean adaptation to highly variable environments. Novel combinations of genes/genomic regions probably arose in Europe after the introduction of the common bean and during its dissemination, on which adaptive selection acted (i.e., adaptive introgression). The new "-omics" technologies can help to fine-tune the molecular basis of these adaptation strategies, an aspect that is ongoing in the BEAN_ADAPT Project (funded through the second ERA-CAPS call; ERA-NET for Coordinating Action in Plant Sciences). This project is based on a multidisciplinary approach (i.e., genomics, transcriptomics, metabolomics, plant physiology, population/quantitative genetics and biochemistry) with the aim to extend the genetic basis of the phenotypic adaptation of *P. vulgaris* and its sister species *P. coccineus* in Europe and outside of their centers of origin.

5. Conclusions

Population genomics research is providing a more complete picture of genetic parameters across the entire genome in both model and non-model species [83]. The plummeting costs of DNA sequencing make genotyping feasible for hundreds to millions of individuals and loci and also allow the study of variations in gene expression, epigenetics and proteins. Furthermore, the combination of genome-wide data from sequencing tools, with improved coverage and resolution of metabolomic platforms, also allows mapping of several metabolites (mQTL mapping) [84]. For instance, Beleggia et al. [24] investigated for the first time the effects of selection on the accumulation of 51 primary metabolites and their relationships in the kernels of three *Triticum turgidum* L. subspecies, and they revealed domestication-associated changes in metabolite contents and in the metabolic correlation networks. Few metabolomics studies have been conducted so far in population genomics of the common bean [85]; however, even though there are sometimes still limits in the immediate translation of genetic variation into metabolic diversity [86], the integration of metabolic profiling with other "-omics" data might be highly effective for functional gene identification and elucidation of the common bean demographic history. One interesting study was carried out recently by Perez de Souza et al. [85], who combined a new approach for the annotation of specialized metabolites with transcriptomic sequencing data and phylogenetic analyses for several genotypes of *P. vulgaris* that belong to the Mesoamerican and Andean gene pools. Their data show that three classes of metabolites (i.e., hydroxycinnamates, flavonoids and triterpene saponins) accumulated differently across their accessions; moreover, the creation of a multi-omics dataset allowed them to identify with precision and accuracy a set of candidate genes that were responsible for important agronomical and ecological traits.

The main potentiality of population genomics approaches in *P. vulgaris* is emerging with the unraveling of the genetic bases of common bean domestication and adaptation to different environmental conditions. The discovery of advantageous genetic variants is fundamental not only to clarify the evolutionary history of a certain population but also to determine the heritability of simple and complex traits in order to design successful breeding strategies. Indeed, identification of the genetic architecture of plant adaptation to different environmental conditions appears to be a major element to address crucial societal challenges, such as mitigation and adaptation to climate changes [87]. Moreover, an excellent example of the potential population genomics approach was offered by Exposito-Alonso et al. [88] for the *A. thaliana* model system. As well as improving our knowledge of the genomic basis of past selection and adaptation to specific agro-ecosystems,

Exposito-Alonso et al. [88] were able to build genome-wide environmental selection models to predict how evolutionary pressures on species will work in inaccessible environments or even under future hypothetical climates [9].

Figure 1 shows the critical role of marker mutations in describing the diversity of plant populations; the higher the mutation rate, the lower the loss of diversity detectable. The occurrence of a bottleneck in the Andes before domestication was recovered from more quickly by markers with a high mutation rate compared with markers showing lower rates of mutation. For this reason, the lower mutation rate characteristic of SNP markers allowed Bitocchi et al. [43] to detect the effects of the bottleneck on the genetic diversity of the Andean wild germplasm with much higher resolution ($L\pi = 90\%$ compared to the Mesoamerican wild gene pool), which confirmed the Mesoamerican origin of the common bean.

Figure 2 shows the Mesoamerican origin of wild *P. vulgaris* proposed by Bitocchi et al. [43] consists of the hypothesis that the Andean and Northern Peru-Ecuador wild common bean populations originated from two independent migrations of the Mesoamerican wild population (Figure 2, blue and yellow solid arrows), which occurred about 110,000–165,000 years ago, prior to the domestication of the species. This hypothesis has also been supported by subsequent studies [29,47,66], and recently, it was confirmed by approximate Bayesian computation analysis [48]. Subsequently, two parallel and independent domestication events in Mesoamerica and in the Andes gave rise to the formation of the current two major domesticated gene pools. Rendón-Anaya et al. [49] proposed the hypothesis of a slightly different evolutionary history, supporting the occurrence of two migration events at different times. In particular, compared to the previous hypothesis, they suggested that the introduction of the wild ancestor "*Phaseolus protovulgaris*" into Northern Peru-Ecuador from Mesoamerica occurred much earlier (ancient migration, 0.9 Mya for plastid markers; Figure 2, dashed yellow arrow) than the diversification of *P. vulgaris* within the *Vulgaris* group.

Author Contributions: R.P. and G.C. designed and wrote the manuscript. R.P., G.C., G.F. and V.D.V. contributed to the drafting of the manuscript. All authors have read and agreed to the published version of the manuscript.

Funding: This work was conducted within the BRESOV Project funded by the European Union Horizon 2020 Research and Innovation Programme under Grant Agreement No. 774244.

Conflicts of Interest: The authors declare that they have no conflicts of interest.

References

1. Kimura, M. Evolutionary rate at the molecular level. *Nature* **1968**, *217*, 624–626. [CrossRef] [PubMed]
2. Kimura, M. Recent development of the neutral theory viewed from the wrightian tradition of theoretical population genetics. *Proc. Natl. Acad. Sci. USA* **1991**, *88*, 5969–5973. [CrossRef]
3. Black, W.C.; Baer, C.F.; Antolin, M.F.; DuTeau, N.M. Population genomics: Genome-wide sampling of insect populations. *Annu. Rev. Entomol.* **2001**, *46*, 441–469. [CrossRef] [PubMed]
4. Luikart, G.; England, P.R.; Tallmon, D.; Jordan, S.; Taberlet, P. The power and promise of population genomics: From genotyping to genome typing. *Nat. Rev. Genet.* **2003**, *4*, 981–994. [CrossRef] [PubMed]
5. Hohenlohe, P.A.; Hand, B.K.; Andrews, K.R.; Luikart, G. Population genomics provides key insights in ecology and evolution. In *Population Genomics: Concepts, Approaches and Applications*; Rajora, O.M.P., Ed.; Publisher: Gewerbestrasse, Switzerland, 2018; pp. 483–510. [CrossRef]
6. Hoban, S.; Kelley, J.L.; Lotterhos, K.E.; Antolin, M.F.; Bradburd, G.; Lowry, D.B.; Poss, M.L.; Reed, L.K.; Storfer, A.; Whitlock, M.C. Finding the genomic basis of local adaptation: Pitfalls, practical solutions, and future directions. *Am. Nat.* **2016**, *188*, 379–397. [CrossRef]
7. Hendricks, S.; Anderson, E.C.; Antao, T.; Bernatchez, L.; Forester, B.R.; Garner, B.; Hand, B.K.; Hohenlohe, P.A.; Kardos, M.; Koop, B.; et al. Recent advances in conservation and population genomics data analysis. *Evol. Appl.* **2018**, *11*, 1197–1211. [CrossRef]
8. Hunter, M.E.; Hoban, S.M.; Bruford, M.W.; Segelbacher, G.; Bernatchez, L. Next-generation, conservation genetics and biodiversity monitoring. *Evol. Appl.* **2018**, *11*, 1029–1034. [CrossRef]
9. Cortinovis, G.; Di Vittori, V.; Bellucci, E.; Bitocchi, E.; Papa, R. Adaptation to novel environments during crop diversification. *Curr. Opin. Plant Biol.* **2020**. [CrossRef]

10. Harris, H. Enzyme polymorphism in man. *Proc. Roy Soc. B* **1966**, *164*, 298–310.
11. Lewontin, R.C.; Hubby, J.L. A molecular approach to the study of genetic heterozygosity in natural populations. II. Amount of variation and degree of heterozygosity in natural populations of *Drosophila pseudoobscura*. *Genetics* **1966**, *54*, 595–609.
12. Schlötterer, C. The evolution of molecular markers-just a matter of fashion? *Nat. Rev. Genet.* **2004**, *5*, 63–69. [CrossRef] [PubMed]
13. Avise, J.C.; Lansman, R.A.; Shade, R.O. The use of restriction endonucleases to measure mitochondrial DNA sequence relatedness in natural populations. I. Population structure and evolution in the genus *Peromyscus*. *Genetics* **1979**, *92*, 279–295. [PubMed]
14. Allendorf, F.W. Genetics and the conservation of natural populations: Allozymes to genomes. *Mol. Ecol.* **2017**, *26*, 420–430. [CrossRef] [PubMed]
15. Khlestkina, E.K.; Salina, E.A. SNP markers: Methods of analysis, ways of development, and comparison on an example of common wheat. *Russ. J. Genet.* **2006**, *42*, 585–594. [CrossRef]
16. Morgante, M.; Olivieri, A.M. PCR-amplified microsatellites as markers in plant genetics. *Plant J.* **1993**, *3*, 175–182. [CrossRef]
17. Vos, P.; Hogers, R.; Bleeker, M.; Reijans, M.; van de Lee, T.; Hornes, M.; Friters, A.; Pot, J.; Paleman, J.; Kuiper, M.; et al. AFLP: A new technique for DNA fingerprinting. *Nucleic Acids Res.* **1995**, *23*, 4407–4414. [CrossRef]
18. Morin, P.A.; Luikart, G.; Wayne, R.K. The SNP workshop group. SNPs in ecology, evolution and conservation. *Trend. Ecol. Evol.* **2004**, *19*, 208–216. [CrossRef]
19. Sharma, T.R.; Devanna, B.N.; Kiran, K.; Singh, P.K.; Arora, K.; Jain, P.; Tiwari, I.M.; Dubey, H.; Saklani, B.; Kumari, M.; et al. Status and prospects of next-generation sequencing technologies in crop plants. In *Next-Generation Sequencing and Bioinformatics for Plant Science*; Bhadauria, V., Ed.; Caister Academic Press: Norfolk, UK, 2017; pp. 1–36. [CrossRef]
20. Akey, J.M.; Zhang, G.; Zhang, K.; Jin, L.; Shriver, M.D. Interrogating a high-density SNP map for signatures of natural selection. *Genom. Res.* **2002**, *12*, 1805–1814. [CrossRef]
21. Davey, J.W.; Hohenlohe, P.A.; Etter, P.D.; Boone, J.Q.; Catchen, J.M.; Blaxter, M.L. Genome-wide genetic marker discovery and genotyping using next-generation sequencing. *Nat. Rev. Genet.* **2011**, *12*, 499–510. [CrossRef]
22. Ellegren, H. Genome sequencing and population genomics in non-model organisms. *Trends Ecol. Evol.* **2014**, *29*, 51–63. [CrossRef]
23. Bellucci, E.; Bitocchi, E.; Ferrarini, A.; Benazzo, A.; Biagetti, E.; Klie, S.; Minio, A.; Rau, D.; Rodriguez, M.; Panziera, A.; et al. Decreased nucleotide and expression diversity and modified co-expression patterns characterize domestication in the common bean. *Plant Cell* **2014**, *26*, 1901–1912. [CrossRef] [PubMed]
24. Beleggia, R.; Rau, D.; Laidò, G.; Platani, C.; Nigro, F.; Fragasso, M.; De Vita, P.; Scossa, F.; Fernie, A.R.; Nikoloski, Z.; et al. Evolutionary metabolomics reveals domestication-associated changes in tetraploid wheat kernels. *Mol. Biol. Evol.* **2016**, *33*, 1740–1753. [CrossRef] [PubMed]
25. Kingman, J.F.C. The coalescent. *Stoch. Process Appl.* **1982**, *13*, 235–248. [CrossRef]
26. Wright, S.I.; Bi, I.V.; Schroeder, S.G.; Yamasaki, M.; Doebley, J.F.; McMullen, M.D.; Gaut, B.S. The effects of artificial selection on the maize genome. *Science* **2005**, *308*, 1310–1314. [CrossRef] [PubMed]
27. Yamasaki, M.; Tenaillon, M.I.; Bi, I.V.; Schroeder, S.G.; Sanchez-Villeda, H.; Doebley, J.F.; Gaut, B.S.; McMullen, M.D. A large-scale screen for artificial selection in maize identifies candidate agronomic loci for domestication and crop improvement. *Plant Cell* **2005**, *17*, 2859–2872. [CrossRef] [PubMed]
28. Yamasaki, M.; Wright, S.I.; McMullen, M.D. Genomic screening for artificial selection during domestication and improvement in maize. *Annu. Bot.* **2007**, *100*, 967–973. [CrossRef]
29. Schmutz, J.; McClean, P.E.; Mamidi, S.; Wu, G.A.; Cannon, S.B.; Grimwood, J.; Jenkins, J.; Shu, S.; Song, Q.; Chavarro, C.; et al. A reference genome for common bean and genome-wide analysis of dual domestications. *Nat. Genet.* **2014**, *46*, 707–713. [CrossRef]
30. Vlasova, A.; Capella-Gutiérrez, S.; Rendón-Anaya, M.; Hernández-Oñate, M.; Minoche, A.E.; Erb, I.; Camara, F.; Prieto Barja, P.; Corvelo, A.; Sanseverino, W.; et al. Genome and transcriptome analysis of the Mesoamerican common bean and the role of gene duplications in establishing tissue and temporal specialization of genes. *Genom. Biol.* **2016**, *17*, 32. [CrossRef]

31. Bitocchi, E.; Rau, D.; Bellucci, E.; Rodriguez, M.; Murgia, M.L.; Gioia, T.; Santo, D.; Nanni, L.; Attene, G.; Papa, R. Beans (*Phaseolus* ssp.) as a model for understanding crop evolution. *Front. Plant. Sci.* **2017**, *8*, 722. [CrossRef]
32. Toro, O.; Tohme, J.; Debouck, D.G. Wild bean (*Phaseolus vulgaris* L.): Description and distribution. In *Centro Internacional de Agricultura Tropical (CIAT)*; International Board for Plant Genetic Resources (IBPGR): Cali, CO, USA, 1990; p. 109, (CIAT publication no. 181).
33. Debouck, D.G.; Toro, O.; Paredes, O.M.; Johnson, W.C.; Gepts, P. Genetic diversity and ecological distribution of *Phaseolus vulgaris* in north-western South America. *Econ. Bot.* **1993**, *47*, 408–423. [CrossRef]
34. Gepts, P.; Bliss, F.A. F1 hybrid weakness in the common bean: Differential geographic origin suggests two gene pools in cultivated bean germplasm. *J. Hered.* **1985**, *76*, 447–450. [CrossRef]
35. Gepts, P.; Osborn, T.C.; Rashka, K.; Bliss, F.A. 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. [CrossRef]
36. Koenig, R.; Gepts, P. Allozyme diversity in wild *Phaseolus vulgaris*: Further evidence for two major centers of diversity. *Theor. Appl. Genet.* **1989**, *78*, 809–817. [CrossRef] [PubMed]
37. Becerra-Velásquez, V.L.; Gepts, P. RFLP diversity in common bean (*Phaseolus vulgaris* L.). *Genome* **1994**, *37*, 256–263. [CrossRef]
38. Freyre, R.; Ríos, R.; Guzmán, L.; Debouck, D.G.; Gepts, P. Ecogeographic distribution of *Phaseolus* spp. (Fabaceae) in Bolivia. *Econ. Bot.* **1996**, *50*, 195–215. [CrossRef]
39. Papa, R.; Gepts, P. Asymmetry of gene flow and differential geographical structure of molecular diversity in wild and domesticated common bean (*Phaseolus vulgaris* L.) from Mesoamerica. *Theor. Appl. Genet.* **2003**, *106*, 239–250. [CrossRef]
40. Rossi, M.; Bitocchi, E.; Bellucci, E.; Nanni, L.; Rau, D.; Attene, G.; Papa, R. Linkage disequilibrium and population structure in wild and domesticated populations of *Phaseolus vulgaris* L. *Evol. Appl.* **2009**, *2*, 504–522. [CrossRef]
41. Kwak, M.; Gepts, P. Structure of genetic diversity in the two major gene pools of common bean (*Phaseolus vulgaris* L., Fabaceae). *Theor. Appl. Genet.* **2009**, *118*, 979–992. [CrossRef]
42. Nei, M.; Maruyama, T.; Chakraborty, R. The bottleneck effect and genetic variability of populations. *Evolution* **1975**, *29*, 1–10. [CrossRef]
43. Bitocchi, E.; Nanni, L.; Bellucci, E.; Rossi, M.; Giardini, A.; Spagnoletti Zeuli, P.; Logozzo, G.; Stougaard, J.; McClean, P.; Attene, G.; et al. Mesoamerican origin of the common bean (*Phaseolus vulgaris* L.) is revealed by sequence data. *Proc. Natl. Acad. Sci. USA* **2012**, *109*, E788–E796. [CrossRef]
44. Nei, M.; Chakraborty, R.; Fuerst, P.A. Infinite allele model with varying mutation rate. *Proc. Natl. Acad. Sci. USA* **1976**, *73*, 4164–4168. [CrossRef] [PubMed]
45. Nei, M. Bottlenecks, genetic polymorphism and speciation. *Genetics* **2005**, *170*, 1–4. [PubMed]
46. Lynch, M.; Conery, J.S. The evolutionary fate and consequences of duplicate genes. *Science* **2000**, *290*, 1151–1155. [CrossRef] [PubMed]
47. Desiderio, F.; Bitocchi, E.; Bellucci, E.; Rau, D.; Rodriguez, M.; Attene, G.; Papa, R.; Nanni, L. Chloroplast microsatellite diversity in *Phaseolus vulgaris*. *Front. Plant. Sci.* **2013**, *3*, 312. [CrossRef] [PubMed]
48. Ariani, A.; Mier y Teran, J.C.; Gepts, P. Spatial and temporal scales of range expansion in wild *Phaseolus vulgaris*. *Mol. Biol. Evol.* **2017**, *35*, 119–131. [CrossRef] [PubMed]
49. Rendón-Anaya, M.; Montero-Vargas, J.M.; Saburido-Álvarez, S.; Vlasova, A.; Capella-Gutierrez, S.; Ordaz-Ortiz, J.J.; Aguilar, O.M.; Vianello-Brondani, R.P.; Santalla, M.; Delaye, L.; et al. Genomic history of the origin and domestication of common bean unveils its closest sister species. *Genome Biol.* **2017**, *18*, 60. [CrossRef]
50. Bitocchi, E.; Bellucci, E.; Giardini, A.; Rau, D.; Rodriguez, M.; Biagetti, E.; Santilocchi, R.; Spagnoletti Zeuli, P.; Gioia, T.; Logozzo, G.; et al. Molecular analysis of the parallel domestication of the common bean in Mesoamerica and the andes. *New Phytol.* **2013**, *197*, 300–313. [CrossRef]
51. Bellucci, E.; Bitocchi, E.; Rau, D.; Rodriguez, M.; Biagetti, E.; Giardini, A.; Attene, G.; Nanni, L.; Papa, R. Genomics of origin, domestication and evolution of *Phaseolus vulgaris*. In *Genomics of Plant Genetic Resources*; Tuberosa, R., Graner, A., Frison, E., Eds.; Springer: Berlin, Germany, 2014; pp. 483–507.
52. Koinange, E.M.K.; Singh, S.P.; Gepts, P. Genetic control of the domestication syndrome in common bean. *Crop Sci.* **1996**, *36*, 1037–1045. [CrossRef]

53. Di Vittori, V.; Bellucci, E.; Bitocchi, E.; Rau, D.; Rodriguez, M.; Murgia, M.L.; Nanni, L.; Attene, G.; Papa, R. Domestication and crop history. In *The Common Bean Genome*; Pérez de la Vega, M., Santalla, M., Marsolais, M., Eds.; Compendium of Plant Genomes; Springer International Publishing AG: New York, NY, USA, 2017; pp. 21–55. [CrossRef]
54. Rau, D.; Murgia, M.L.; Rodriguez, M.; Bitocchi, E.; Bellucci, E.; Fois, D.; Albani, D.; Nanni, L.; Gioia, T.; Santo, D.; et al. Genomic dissection of pod shattering in common bean: Mutations at non-orthologous loci at the basis of convergent phenotypic evolution under domestication of leguminous species. *Plant J.* **2019**, *97*, 693–714. [CrossRef]
55. Parker, T.A.; Berny Mier, Y.; Teran, J.C.; Palkovic, A.; Jernstedt, J.; Gepts, P. Pod indehiscence is a domestication and aridity resilience trait in common bean. *New Phytol.* **2019**, *225*, 558–570. [CrossRef]
56. Di Vittori, V. Common bean domestication (manuscript in preparation).
57. Repinsky, S.L.; Kwak, M.; Gepts, P. The common bean growth habit gene *PvTFL1y* is a functional homolog of *Arabidopsis* TFL. *Theor. Appl. Genet.* **2012**, *124*, 1539–1547. [CrossRef] [PubMed]
58. McClean, P.E.; Bett, K.E.; Stonehouse, R.; Lee, R.; Pflieger, S.; Moghaddam, S.M.; Geffroy, V.; Miklas, P.; Mamidi, S. White seed colour in common bean (*Phaseolus vulgaris*) results from convergent evolution in the P (*pigment*) gene. *New Phytol.* **2018**, *219*, 1112–1123. [CrossRef] [PubMed]
59. Di Vittori, V.; Gioia, T.; Rodriguez, M.; Bellucci, E.; Bitocchi, E.; Nanni, L.; Attene, G.; Rau, D.; Papa, R. Convergent evolution of the seed shattering trait. *Genes* **2019**, *10*, 68. [CrossRef] [PubMed]
60. Murgia, M.L.; Attene, G.; Rodriguez, M.; Bitocchi, E.; Bellucci, E.; Fois, D.; Nanni, L.; Gioia, T.; Albani, D.M.; Papa, R.; et al. A comprehensive phenotypic investigation of the "pod-shattering syndrome" in common bean. *Front. Plant Sci.* **2017**, *8*, 251. [CrossRef] [PubMed]
61. Mamidi, S.; Rossi, M.; Annam, D.; Moghaddam, S.; Lee, R.; Papa, R.; McClean, P. Investigation of the domestication of common bean (*Phaseolus vulgaris*) using multilocus sequence data. *Funct. Plant. Biol.* **2011**, *38*, 953–967. [CrossRef]
62. Nanni, L.; Bitocchi, E.; Bellucci, E.; Rossi, M.; Rau, D.; Attene, G.; Gepts, P.; Papa, R. Nucleotide diversity of a genomic sequence similar to shatterproof (PvSHP1) in domesticated and wild common bean (*Phaseolus vulgaris* L.). *Theor. Appl. Genet.* **2011**, *123*, 1341–1357. [CrossRef]
63. Glémin, S.; Bataillon, T. A comparative view of the evolution of grasses under domestication. *New Phytol.* **2009**, *183*, 273–290. [CrossRef]
64. Papa, R.; Acosta, J.; Delgado-Salinas, A.; Gepts, P. A genome-wide analysis of differentiation between wild and domesticated *Phaseolus vulgaris* from mesoamerica. *Theor. Appl. Genet.* **2005**, *111*, 1147–1158. [CrossRef]
65. Papa, R.; Bellucci, E.; Rossi, M.; Leonardi, S.; Rau, D.; Gepts, P.; Nanni, L.; Gioia, T. Tagging the signatures of domestication in common bean (*Phaseolus vulgaris*) by means of pooled DNA samples. *Annu. Bot.* **2007**, *100*, 1039–1051. [CrossRef]
66. Mamidi, S.; Rossi, M.; Moghaddam, S.M.; Annam, D.; Lee, R.; Papa, R.; McClean, P. Demographic factors shaped diversity in the two gene pools of wild common bean *Phaseolus vulgaris* L. *Heredity* **2013**, *110*, 267–276. [CrossRef]
67. Zhaodong, H.; Dekang, L.; Ying, G.; Jisen, S.; Dolf, W.; Guangchuang, Y.; Jinhui, C. RIdeogram: Drawing SVG graphics to visualize and map genome-wide data on idiograms. *Peer J. Comput. Sci.* **2020**, *6*, e251. [CrossRef]
68. Liljegren, S.J.; Roeder, A.H.K.; Kempin, S.A.; Gremski, K.; Østergaard, L.; Guimil, S.; Reyes, D.K.; Yanofsky, M.F. Control of fruit patterning in *Arabidopsis* by indehiscent. *Cell* **2004**, *116*, 843–853. [CrossRef]
69. Gioia, T.; Logozzo, G.; Attene, G.; Bellucci, E.; Benedettelli, S.; Negri, V. Evidence for introduction bottleneck and extensive inter-gene pool (Mesoamerica x Andes) hybridization in the European common bean (*Phaseolus vulgaris* L.) germplasm. *PLoS ONE* **2013**, *8*, e75974. [CrossRef]
70. Gepts, P. Origin and evolution of common bean: Past events and recent trends. *Hortscience* **1998**, *33*, 1124–1130. [CrossRef]
71. Cortés, A.J.; Monserrate, F.A.; Ramírez-Villegas, J.; Madriñán, S.; Blair, M.W. Drought tolerance in wild plant populations: The case of common beans (*Phaseolus vulgaris* L.). *PLoS ONE* **2013**, *8*, e62898. [CrossRef]
72. Porch, T.; Beaver, J.S.; Debouck, D.G.; Jackson, S.A.; Kelly, J.D.; Dempewolf, H. Use of wild relatives and closely related species to adapt common bean to climate change. *Agronomy* **2013**, *3*, 433–461. [CrossRef]
73. Schoville, S.D.; Bonin, A.; François, O.; Lobreaux, S.; Melodelima, C.; Manel, S. Adaptive genetic variation on the landscape: Methods and cases. *Annu. Rev. Ecol. Evolut. Syst.* **2012**, *43*, 23–43. [CrossRef]

74. Manel, S.; Joost, S.; Epperson, B.K.; Holderegger, R.; Storfer, A.; Rosenberg, M.S.; Scribner, K.T.; Bonin, A.; Fortin, M.J. Perspectives on the use of landscape genetics to detect genetic adaptive variation in the field. *Mol. Ecol.* **2010**, *19*, 3760–3772. [CrossRef]
75. Anderson, K.; Gaston, K.J. Lightweight unmanned aerial vehicles will revolutionize spatial ecology. *Front. Ecol. Environ.* **2013**, *11*, 138–146. [CrossRef]
76. Storfer, A.; Murphy, M.A.; Spear, S.F.; Holderegger, R.; Waits, L.P. Landscape genetics: Where are we now? *Mol. Ecol.* **2010**, *19*, 3496–3514. [CrossRef]
77. Rodriguez, M.; Rau, D.; Bitocchi, E.; Bellucci, E.; Biagetti, E.; Carboni, A.; Gepts, P.; Nanni, L.; Papa, R.; Attene, G. Landscape genetics, adaptive diversity, and population structure in phaseolus vulgaris. *New Phytol.* **2016**, *209*, 1781–1794. [CrossRef] [PubMed]
78. Ariani, A.; Gepts, P. Signatures of environmental adaptation during range expansion of wild common bean (*Phaseolus vulgaris*). *Biorxiv* **2019**. [CrossRef]
79. Ortega-Amaro, M.A.; Rodríguez-Hernández, A.A.; Rodríguez-Kessler, M.; Hernández-Lucero, E.; Rosales-Mendoza, S.; Ibáñez-Salazar, A.; Delgado-Sánchez, P.; Jiménez-Bremont, J.F. Overexpression of atgrdp2, a novel glycine-rich domain protein, accelerates plant growth and improves stress tolerance. *Front. Plant. Sci.* **2015**, *5*, 782. [CrossRef] [PubMed]
80. Broughton, W.J.; Hernández, G.; Blair, M.; Beebe, S.; Gepts, P.; Vanderleyden, J. Beans (*Phaseolus* spp.)—model food legumes. *Plant Soil* **2003**, *252*, 55–128. [CrossRef]
81. Angioi, S.A.; Rau, D.; Attene, G.; Nanni, L.; Bellucci, E.; Logozzo, G.; Negri, V.; Spagnoletti Zeuli, P.L.; Papa, R. Beans in europe: Origin and structure of the european landraces of *Phaseolus vulgaris* L. *Theor. Appl. Genet.* **2010**, *121*, 829–843. [CrossRef]
82. Santalla, M.; Rodiño, A.P.; De Ron, A.M. Allozyme evidence supporting southwester Europe as a secondary center of genetic diversity for common bean. *Theor. Appl. Genet.* **2002**, *104*, 934–944. [CrossRef]
83. Hohenlohe, P.A.; Phillips, P.C.; Cresko, W.A. Using population genomics to detect selection in natural populations: Key concepts and methodological considerations. *Int. J. Plant Sci.* **2010**, *171*, 1059–1071. [CrossRef]
84. Fiehn, O. Combining genomics, metabolome analysis, and biochemical modelling to understand metabolic networks. *Compar. Funct. Genom.* **2001**, *2*, 155–168. [CrossRef]
85. De Perez Souza, L.; Scossa, F.; Proost, S.; Bitocchi, E.; Papa, R.; Tohge, T.; Fernie, A.R. Multi-tissue integration of transcriptomic and specialized metabolite profiling provides tools for assessing the common bean (*Phaseolus vulgaris*) metabolome. *Plant J.* **2019**, *97*, 1132–1153. [CrossRef]
86. Schwahn, K.; Perez de Souza, L.; Fernie, A.R.; Tohge, T. Metabolomics-assisted refinement of the pathways of steroidal glycoalkaloid biosynthesis in the tomato clade. *J. Integr. Plant Biol.* **2014**, *56*, 864–875. [CrossRef]
87. Gerten, D.; Heck, V.; Jägermeyr, J.; Bodirsky, B.L.; Fetzer, I.; Jalava, M.; Kummu, M.; Wolfgang, L.; Rockström, J.; Schaphoff, S.; et al. Feeding ten billion people is possible within four terrestrial planetary boundaries. *Nat. Sustain.* **2020**. [CrossRef]
88. Exposito-Alonso, M.; Burbano, H.A.; Bossdorf, O.; Nielsen, R.; Weigel, D. Natural selection on the *Arabidopsis thaliana* genome in present and future climates. *Nature* **2019**, *573*, 126–129. [CrossRef] [PubMed]

© 2020 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (http://creativecommons.org/licenses/by/4.0/).

Chapter III

Adaptation to Novel Environments during Crop Diversification

Gaia Cortinovis, Valerio Di Vittori, Elisa Bellucci, Elena Bitocchi and Roberto Papa

Address

Dipartimento di Scienze Agrarie, Alimentari ed Ambientali, Universita`

Politecnica delle Marche, via Brecce Bianche, 60131 Ancona, Italy

Corresponding authors: Bitocchi, Elena (e.bitocchi@univpm.it),

Papa, Roberto (r.papa@univpm.it)

DOI: **https://doi.org/10.1016/j.pbi.2019.12.011**

Available online at www.sciencedirect.com

ScienceDirect

Current Opinion in Plant Biology

Adaptation to novel environments during crop diversification

Gaia Cortinovis, Valerio Di Vittori, Elisa Bellucci, Elena Bitocchi and Roberto Papa

In the context of the global challenge of climate change, mitigation strategies are needed to adapt crops to novel environments. The main goal to address this is an understanding of the genetic basis of crop adaptation to different agro-ecological conditions. The movement of crops during the Colombian Exchange that started with the travels of Columbus in 1492 is an example of rapid adaptation to novel environments. Many diversification-related traits have been characterised in multiple crop species, and association-mapping analyses have identified loci involved in these. Here, we present an overview of current knowledge regarding the molecular basis related to the complex patterns of crop adaptation and dissemination, particularly outside their centres of origin. Investigation of the genomic basis of crop expansion offers a powerful contribution to the development of tools to identify and exploit valuable genetic diversity and to improve and design novel resilient crop varieties.

Address
Dipartimento di Scienze Agrarie, Alimentari ed Ambientali, Università Politecnica delle Marche, via Brecce Bianche, 60131 Ancona, Italy

Corresponding authors: Bitocchi, Elena (e.bitocchi@univpm.it), Papa, Roberto (r.papa@univpm.it)

Current Opinion in Plant Biology 2020, **56**:203–217

This review comes from a themed issue on **AGRI**

Edited by **David Edwards**

For a complete overview see the Issue and the Editorial

Available online 10th February 2020

https://doi.org/10.1016/j.pbi.2019.12.011

1369-5266/© 2019 Elsevier Ltd. All rights reserved.

Introduction

Environmental change will result in strong ecological and genetic effects on gene and allele frequencies in many plant populations, as well as altering several aspects of agricultural systems, such as plant physiology and phenology, water availability, soil fertility, pathogen spread and host susceptibility. Many crops have evolved in response to climate change under increasingly stressful conditions in which their extinction is highly possible. However, there is more evidence for climate-driven range expansion than for range contraction [1]. This suggests that plants can cope with climate change through adaptive mechanisms, such as phenotypic plasticity and micro-evolution [2].

Genetic diversity represents the raw material on which adaptive selection acts, and as such, it has a fundamental role in both evolutionary history and future evolutionary pathways of a species [3]. Thus, persistent fluctuations in biotic and abiotic environmental factors provide a background of changing selection pressures to which species must respond, and in this way, genetic diversity is maintained within populations. Identification of the molecular basis of plant adaptation is needed to drive plant breeding into the development of novel varieties that can adapt to climate changes. Analysis of genetic diversity through population genomics and genotype–phenotype association approaches can be very useful tools to reach this aim [4], especially with the novel opportunities offered by the more recent advances in genomics and DNA sequencing technologies. The success of such studies critically depends on the type of plant material adopted. Moreover, if the search for the signature of selection is the objective, the populations must have an ancient and strong link with their growing environments.

Populations of wild plants and wild crop relatives can easily meet these prerequisites. There are several examples in the literature that have focused on wild germplasm to detect adaptive genetic control, along with studies on model species such as *Arabidopsis thaliana* [5•], with other examples available for crop species. Fustier *et al.* [6] investigated adaptation in 11 populations of teosinte, the wild progenitor of maize, along two elevation gradients in Mexico that showed continuous environmental changes over a short geographic scale. They evaluated 1664 individuals for 18 phenotypic traits and genotyped them for 38 microsatellite markers and 171 outlier single nucleotide polymorphisms (SNPs). These significantly differentiated between lowland and highland populations and/or correlated with environmental variables. They showed that >50% of the traits were differentiated due to local selection. A recent landscape genomics study of Rodriguez *et al.* [7•] reported on an analysis of correlations between molecular markers and ecological variables at a continental scale. They analysed a sample of 310 wild common bean georeferenced accessions that they genetically characterised at 131 SNPs. Geographic and environmental data were combined with

genetic diversity data to separate the effects of geography from those of ecology, and they reported a total of 26 loci (19.9%) that were putatively under selection for adaptation. Among these, different loci were shown to have compatible functions with adaptation features, such as chilling susceptibility, cold acclimation, and mechanisms related to drought stress [7•]. Recently, Mier y Teran *et al.* [8] characterised 112 wild common bean accessions that were representative of the geographic distribution of the Mesoamerican gene pool. This was applied at the molecular level (11 447 SNP markers) and the phenotypic level (root trait evaluation, comparison of control and drought stress), and considered environmental variables from the geographic coordinates of the origin of each accession. They defined genomic regions that were associated with productivity and drought adaptation in the wild germplasm.

Within the cultivated gene pools, the above-mentioned prerequisites for such studies are satisfied only by populations of landraces or, if available, by experimental populations, as composite crosses specifically developed over multiple generations of experimental evolution [88]. Landraces offer unique opportunities for integration of association mapping and signatures of selection analyses. Indeed, landraces are the product of an evolutionary interaction with the agro-ecosystem, and consequently, their genetic composition is determined by both stochastic and human-mediated or natural selection over decades of evolution, which means that they have maintained a considerable amount of genetic variability. Moreover, when multiple landrace populations grown in contrasting agroecological environments are compared, it is possible to tag the signatures of divergent selection [9–12]. This makes it possible to investigate the genes that are responsible for the 'genomic architecture' of the local adaptation of plants. After domestication, food crops spread widely between different geographic and cultural areas at different levels and to different extents, and this process ultimately contributes to the diversification of local agricultural subsistence.

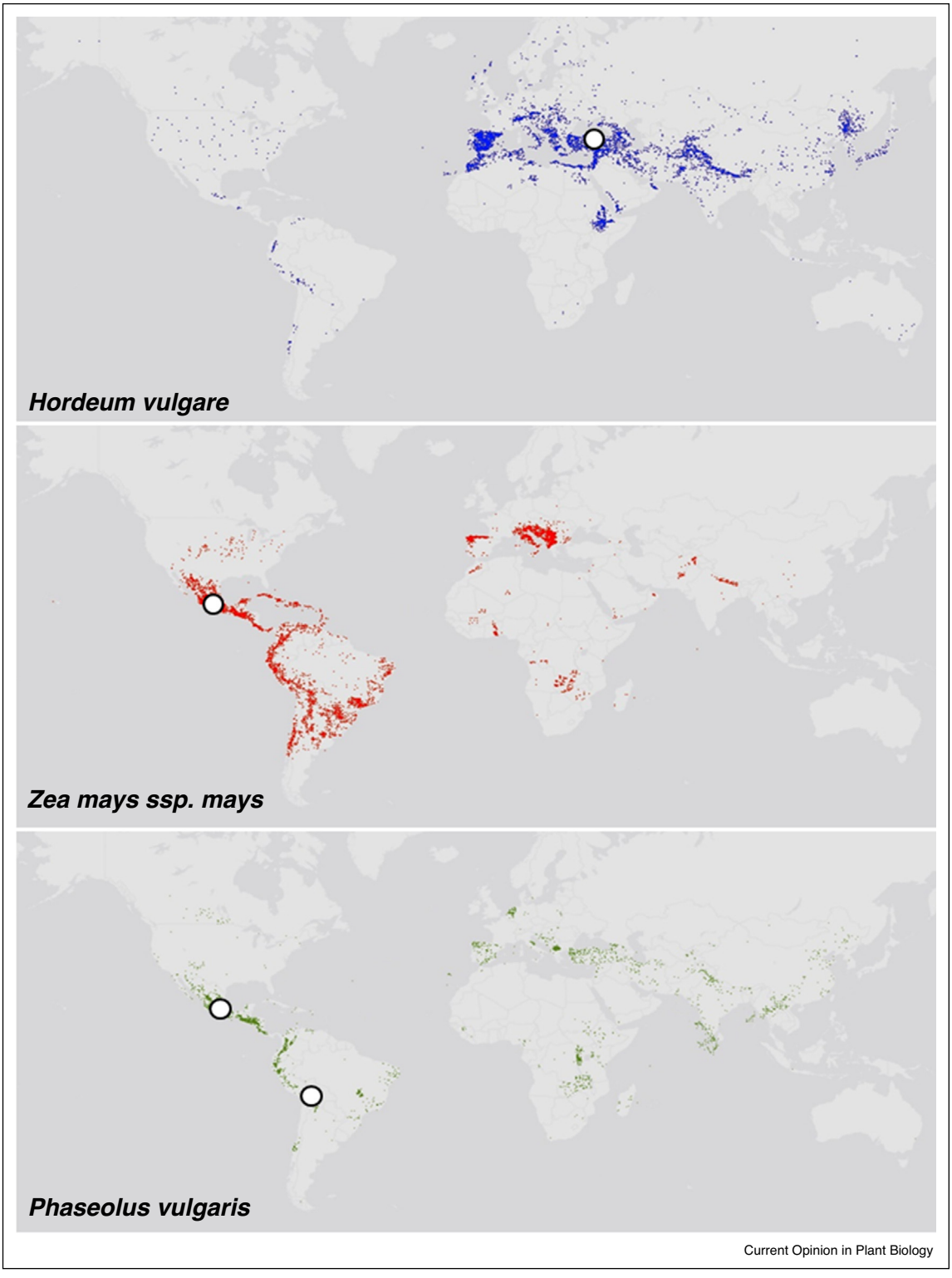
Among cereals, barley and maize are examples of crops that have achieved adaptive success worldwide (Figure 1). Barley is one of the primary plants that originated and was domesticated in the 'Fertile Crescent' about 11 000 years ago, and was later disseminated worldwide over a wide range of agro-climatic conditions [13]. Some of these conditions were particularly extreme, such as in Tibet, Nepal, Ethiopia and the Andes, where farmers cultivated barley on mountain slopes at altitudes higher than those for any other cereals [14,89]. Maize also has one of the broadest worldwide dissemination ranges. It was domesticated once in the Balsas region in the valley of Mexico about 9000 years ago, and it subsequently spread to geographically and ecologically diverse environments, from Canada to Chile [15]. Similarly, among legumes, the common bean can be considered as a crop that is now successfully widespread [16]. The post-domestication phase of crops outside their centres of origin (i.e. at regional, continental, worldwide levels) towards a wide range of agro-ecosystems has led to phenotypic and genetic divergence between domesticated forms. This process can be considered a fascinating model for the study of the adaptive evolution of crops, and it offers the possibility to discover new interesting genetic variants that have potential use in a climate alarm context, like that which we are currently in.

Diversification traits

Meyer and Purugganan [17•] reported on several observed traits in crops that accompanied their domestication and diversification, and their improvement phase. It is not particularly easy to clearly distinguish between genes that underlie domestication and those that control diversification traits. This is the case even if the genetic basis of adaptation might be more related to diversification traits that are related to post-domestication stages, such as for pigments, variations in size and chemical composition of edible parts, changes to the mating system (promoting allogamy or autogamy), resistance or tolerance to abiotic and biotic stresses, reduced vernalisation and photoperiod sensitivity, and changes to flowering time, the life cycle and dwarfism [17•]. It is important to consider that these traits can vary among crop species, considering also that they relate to crops that have adapted to specific agro-ecological conditions and cultures. In this regard, several examples can be found in the literature where the function of genes defined as putatively under selection during domestication of crops can be ascribed to diversification traits, thus traits upon which both natural and human selection have acted during crop expansion.

In common bean (*Phaseolus vulgaris*), Bitocchi *et al.* [18] compared selection analysis data obtained for the same genes in different studies of varying sizes, data types and methodologies. To study the effects of domestication at the genome level, they analysed nucleotide diversity at 49 gene fragments on a sample of 39 wild and domesticated Mesoamerican accessions of *P. vulgaris*. By applying population genomics approaches, they identified several genes that showed footprints of selection. At the same time, they used the SNP data of Rodriguez *et al.* [7•] to perform selection tests on a wider sample, which included 417 and 160 wild and domesticated accessions, respectively, of common bean. Finally, data were included from two further studies that focused on investigation of the domestication process in common bean [19,20•]. The final comparison of the data from these four studies provided independent evidence of selection for four genes: *AN-Pv33*, *AN-DNAJ*, *Leg223*, *AN-Pv69*. Gene-function investigations revealed that all of these genes are involved in plant resistance/tolerance to abiotic stresses, such as heat, drought and salinity. In this regard, adaptation of

Figure 1

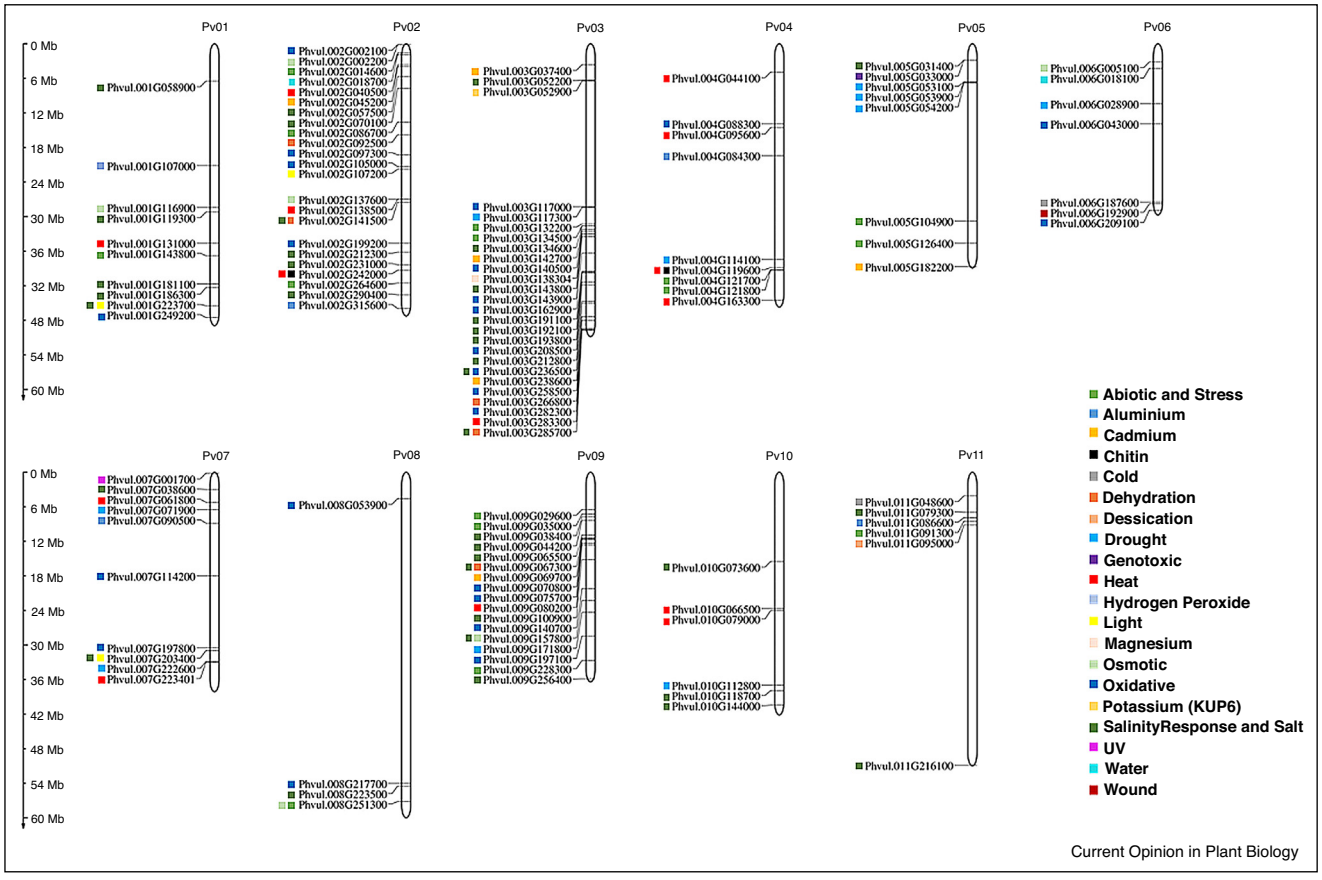


Geographic distribution of barley (top), maize (middle) and common bean (bottom) landraces from their centres of domestication.
The centres of domestication are represented by white dots with black borders. The distributions of the landraces/traditional cultivar accessions were obtained by plotting the geographic coordinates for where the seeds were collected. Data were extracted from the database of the Genesys platform (https://www.genesys-pgr.org/), which includes information from several genebanks.

plants to abiotic stresses is of crucial importance, because they are among the major environmental factors that affect plant productivity. By accessing the *A. thaliana* stress-responsive gene database (http://srgdb.bicpu.edu.in/) [21], we have identified a list of genes that were detected as functionally involved in abiotic stress responses in *Arabidopsis*. The OrthoFinder algorithm [22] and the 2.1 version of the *P. vulgaris* reference genome (https://phytozome.jgi.doe.gov/pz/portal.html) were then used to identify orthologous genes in common bean. A total of 770 common bean genes were found to be orthologous to *Arabidopsis* genes involved in abiotic stress responses (Table S1), 126 of which showed signature of selection during domestication in Bellucci *et al.* [20•] and/or Schmutz *et al.* [19] (Figure 2; Table S1). Among these genes, a very interesting candidate is a homologue of K^+ *uptake transporter6* (KUP6). The KUP6 gene has been shown to be a key factor in osmotic adjustment, through the balancing of potassium homeostasis in cell growth and drought stress responses in *A. thaliana* [23]. Its function is directly mediated by abscisic acid signalling, and under water-deficit stress this involves inhibition of cell expansion in both roots and guard cells, which is driven by decreased turgor. In Bellucci *et al.* [20•], KUP6 was also among the small fraction of outlier genes for which selection has increased the nucleotide diversity in the domesticated pool compared to the wild pool, which suggests selection due to crop expansion into the new environments with unexpected biotic and abiotic stresses (i.e. diversifying selection).

Meyer *et al.* [24] reported an example of geographic and environmental divergent adaptation between four populations of African rice (*Oryza glaberrima* Steud.). They sequenced the genomes of 93 landraces that spanned from west to central sub-Saharan Africa, to investigate the African rice post-domestication spread, its subsequent

Figure 2



Physical map of the 11 common bean chromosomes and genomic locations of genes putatively involved in abiotic stress responses and with selection signatures in common bean.

Common bean genes were identified based on orthology with those involved in abiotic stress responses in *Arabidopsis thaliana*, according to *The Arabidopsis Stress Responsive Gene Database* [21] and using the OrthoFinder algorithm [22]. The orthologous protein to the *A. thaliana* KUP6 is also shown in chromosome Pv03. For the map representation, we selected a subset of 126 common bean orthologues (see **Table S1** for the full list) that show selection signatures according to Schmutz *et al.* [19] and/or Bellucci *et al.* [20•]. Genes potentially associated to different stress responses based on the orthology with *A. thaliana* genes are highlighted according to the legend. The physical distances in the scale are reported in megabases (Mb).

adaptation to local environments, and the genes that were involved in these processes. They focused on salinity tolerance, as one of the major traits associated with geographic adaptation of African rice. The accessions were phenotyped for various salinity-associated fitness traits, and they found a significative loss of salinity tolerance in the southwest inland population. This adaptive phenotype was thus suggested to arise from the costs of maintaining tolerance in a geographic area associated with greater precipitation and decreased soil salinity. In contrast, no significant differences were seen for the northwest, northeast and southeast subpopulations. Genome-wide association studies allowed them to identify 11 loci that contained several genes that were putatively involved in salt-stress tolerance. Among the most significant of these was an orthologue of the *O. sativa* HAK5 gene. HAK5 has been shown to be a key component in the acquisition and transport of potassium, to improve salt resistance in potassium-deficient rice plants [25], and it might have had a crucial role in *O. glaberrima* adaptation along the western Atlantic coast.

The timing of important developmental stages (e.g. flowering time) is another main diversification trait that defines adaptation of plant populations to different environments. In maize, numerous studies have focused on identification of the genetic control of flowering time and on genetic variations at identified genes in different materials from diverse environments. Buckler *et al.* [26•] used a nested association mapping population of 200 recombinant inbred lines from 25 crosses, which resulted in a total of 5000 lines for identification of genes or genomic regions associated with flowering time. These lines were phenotyped in four locations over two years. Quantitative trait locus (QTL) mapping showed that the differences in flowering time were not caused by a few genes that had large effects, but rather by the cumulative effects of numerous QTLs (i.e. <100), each of which had only a small impact on this trait [26•]. However, to date, although a large body of mapping information on the QTLs that control maize flowering time is available [27,28], the molecular basis of these QTLs remains almost totally unknown, with the exception of four genes that have been demonstrated to be involved in flowering time: *Dwarf8* [29]; *ZmCCT* [30–32]; *Vgt1* [33]; and *ZCN8* [34].

The *Dwarf8* gene has been shown to be an orthologue of the gibberellic acid insensitive (*GAI*) gene, which is a transcription factor that negatively regulates gibberellin responses in *A. thaliana*. Association analysis has identified several interesting polymorphisms in maize *Dwarf8*. One of these is a 6-bp deletion in the C-terminal region of the open reading frame, and this showed strong association with flowering time [29]. With the aim to evaluate the contribution of *Dwarf8* to maize adaptation to temperate climates, Camus-Kulandaivelu *et al.* [35] analysed a wide collection of traditional landraces (144 from America, 131 from Europe) for indel polymorphisms in the *Dwarf8* gene. They reported a variation in the frequency of the *Dwarf8* deletion associated with altitude and latitude, which demonstrated that these features have an important role in driving local maize adaptation [15,35]. In particular, for American landraces, they showed that the frequency of the *Dwarf8* deletion was higher in northern Flint maize (83%) compared to maize groups from the tropical Caribbean (2%) and Mexican (4%). Instead, the Andean group that was represented by populations that originated from high altitudes (on average, 2200 m a.s.l.) showed a frequency of *Dwarf8* deletion of 58%. Similarly, in Europe, *Dwarf8* deletion has prevailed in landraces from northern Europe.

Vgt1 is also one of the major maize flowering-time QTLs, and a miniature transposon that is located ~70 kb upstream of *ZMRap2.7* was shown to be the causative variant of *Vgt1* that contributes to maize adaptation to temperate regions [33,36,37]. Ducrocq *et al.* [36] carried out an association mapping study on 375 maize inbred lines, which included inbred lines representative of the American and European diversity, with a wide range of flowering times. They reported that the *Vgt1* early allele showed higher frequency in the tropical materials. Moreover, the frequency of *Vgt1* alleles among the tropical populations varied with the altitude of the collection site, while the early allele was rare at low altitudes. These data support the hypothesis that adaptive selection followed domestication of maize, with early and late materials adapted to high altitude and low altitude cultivation systems, respectively.

Yang *et al.* [31] showed that a CACTA-like transposon insertion within the *ZmCCT10* promoter repressed *ZmCCT10* expression, which makes maize insensitive to long days. Likewise, Huang *et al.* [32] identified a Harbinger-like transposable element at ~57 kb upstream of *ZmCCT9* that functions as a *cis*-acting repressor of *ZmCCT9*, to enhance maize adaptation to higher latitudes. Comparisons of the gene sequence from teosinte and tropical and temperate maize revealed that both the adaptive insertions were completely absent in teosinte, and so they are likely to be *de-novo* mutations that occurred after the initial maize domestication [30–32].

Recently, Guo *et al.* [34] reported that two natural *cis*-variants in the promoter of *ZCN8* were gradually targeted by selection during the spread of maize from its tropical origin to northern North America, which led to earlier flowering plants that were adapted to the temperate growing regions. In more detail, *ZCN8* was proposed to be homologous to *A. thaliana* FLOWERING LOCUS (FT), and they considered it to be the maize florigen gene [38,39].

Another interesting example was the study of Vigouroux *et al.* [40••] on pearl millet. They analysed a total of 192 landraces that had been collected during two different periods (i.e. 1976, 2003) throughout Niger, in the

Sahel, which is one of the driest agro-ecosystems in Africa. This geographic area had undergone recurrent drought during this interval of 25 years. Along with the analysis of the phenological and morphological changes in the two samples evaluated in field experiments, they also investigated the genetic diversity across these two samples. In particular, they analysed the change in allele frequency at the *PHYC* flowering time locus [41], and showed that the allele that conferred earliness increased from 9.9% to 18.3% over this time frame. This study is an example of the strong adaptation of plants to changing environmental conditions even over relatively short evolutionary timescales. It also suggested that exploitation of genetic variability within landrace populations represents a strategy in response to future climate changes. However, they recommended the consideration of the mating system of the crop species, as they indicated that this strategy might be successful for allogamous species, such as pearl millet, but that further studies would be needed for autogamous species [40••].

SNPs are the markers of choice in different population genomics studies because they are the most abundant bi-allelic and co-dominant markers that are characterised by simple mutational patterns and by high-throughput and low-cost detection. Despite this, many other examples exist in literature that are based on structural variations, which refers to genomic changes in DNA segments of >1 kbp, such as insertions, deletions, inversions, or copy-number variations. It is highly possible that genes responsible for acclimatising and adaptation to different agro-ecological conditions and stress resistances will be identified in such genomic changes [42••]. As an example, Zhou *et al.* [43] reported the duplication and evolutionary history of the *COR15* gene that is involved in cold-stress defence, which was previously detected in two copies in several species of *Brassicaceae*. They cloned the homologous *COR15* sequences of 10 species of *Brassicaceae*, and when they performed evolutionary analyses they found significant inter-lineage differences in the evolutionary rates between the original and the duplicated genes. The most interesting data were perhaps observed for the analysis of the *COR15* genes of the *Draba* species, which contrary to the other lineages, is mainly present in cold-temperature, highly arid regions. Three important lines of evidence were observed: (i) the estimated non-synonymous and synonymous substitution ratio appeared to be higher among the duplicated genes; (ii) positive selection was detected for the duplicate *COR15* gene; and (iii) functional divergence was shown between the two groups of the proteins. Overall, these observations indicated that the functional differences in the *Draba* lineage between *COR15a*, as the original gene, and *COR15b*, as the duplicated gene, have been driven by adaptive evolution. This allowed its spread to cold locations during the Quaternary climatic oscillations, and subsequently its expansion to arid alpine and arctic regions. Similarly, De Bolt [44] examined whether *Arabidopsis* plants grown under different temperatures for several generations showed any differences in copy number variations relative to the control situation of growth under normal conditions. They showed that high temperatures promoted chromosomal segmental duplications.

Recent studies have also suggested that polyploids might have greater phenotypic flexibility for gene expression in response to environmental differences [45]. Ceccarelli *et al.* [46] showed that chromosome endoreduplication in *Sorghum bicolor* is a fundamental part of the adaptive response of plant genomes to salt stress. Their results showed that when exposed to salt-induced treatments, only competent genotypes underwent endopolyploidy of the root cortex cells, which allowed them to grow under sublethal salinity concentrations. The wide variability obtained as a result of polyploidy events was thus directly correlated with the tolerance increase of *S. bicolor* to salinity, which highlighted the important role of this mechanism in adaptive responses to different abiotic conditions. Similarly, Saleh *et al.* [47] reported that citrus tetraploid rootstock is more tolerant to salt stress than their corresponding diploid.

Selection for adaptation

Local adaptation occurs when populations that grow under heterogeneous environmental conditions evolve different phenotypic traits that provide a fitness advantage in their specific environment [48]. Selection acts on sequence variation, which can derive from the standing variation that has a long history of segregation within a crop before the advent of selection, or *de-novo* mutations that originate in populations (i.e. wild forms or landraces), or from hybridisation. Knowing the sources of variation on which selection for adaptation can act is important for several reasons, such as, for example, to understand how rapidly populations can adapt [3]. Exhaustive evidence that shows the relative role of standing variation or *de-novo* mutations after changes in the environment is still lacking. Adaptation is likely to be slower if selection acts on *de-novo* mutations, compared to what would be expected when it acts on standing variation, where beneficial alleles might already be available at higher frequencies [49]. Moreover, on average, adaptation from standing variation appears to occur through the fixing of more alleles with small effects [3,50], and can have greater potential for adaptation if the rate of environmental change is fast, rather than slow, by traversing larger distances in the phenotype space.

Along with useful standing variation and *de-novo* mutations, selection for adaptation can also act on new genotypic variations due to recombination after hybridisation [51,52]. In common bean, Bellucci *et al.* [20•] analysed RNA sequencing data from a set of Mesoamerican wild and domesticated accessions, and they showed that most of the genes detected as under selection during domestication showed reduced diversity in their domesticated

compared to their wild forms, as expected under positive selection from standing variation. However, 2.8% of the outlier genes showed no diversity in the wild form and polymorphism in the domesticated form. This thus suggested that in some cases the selection increased the nucleotide diversity of domesticated materials at target loci, the function of which was associated with adaptation traits, such as abiotic stress responses and flowering time [20•]. Interestingly, in the same species, Bitocchi *et al.* [18] analysed nucleotide data of 49 gene fragments in a sample of Mesoamerican wild and domesticated accessions, and they detected an excess of nonsynonymous mutations in the domesticated forms, particularly in the coding regions, compared to the non-coding regions. These mutations appeared to be recently derived mutations, and the investigations into the functions of their relative genes (responses to biotic and abiotic stresses) support a scenario where new functional mutations were selected for adaptation during diversification.

In maize, Guo *et al.* [34] asked whether the ZCN8 gene can affect natural variations in flowering time. They performed association analysis by sequencing ZCN8 and its upstream and downstream regions in segregant populations derived from a cross between W22, a temperate *Zea mays* ssp. *mays* inbred line, and 8759, a *Z. mays* ssp. *parviglumis* accession. They found a SNP in the promoter region of ZCN8 (i.e. SNP-1245) that coincided precisely with the allelic differences in flowering time between all of the parents of the teosinte–maize populations used in their study. They also sequenced the ZCN8 gene in a panel of 513 maize inbred lines and 45 teosinte lines (including lines of *Z. mays* ssp. *parviglumis*, the maize progenitor, and lines of its close relative species *Z. mays* ssp. *mexicana*). These data revealed that the early flowering allele of SNP-1245 was present in ~24% of the teosinte accessions, which suggested that this polymorphism was a standing variant in the maize wild progenitor selected during the early domestication of maize. Guo *et al.* [34] also detected a three-base-pair deletion variant (i.e. Indel-2339) about 1000 bases from SNP-1245 that was associated with flowering time and showed higher expression of ZCN8. Moreover, they did not find this allelic variant in the maize progenitor, although it was present in *Z. mays* ssp. *mexicana*, from which gene flow resulted in its introgression into maize [53]. Furthermore, low frequency of Indel-2339 (5%) was shown for South America germplasm (i.e. tropical maize), while it was selected at a higher frequency in northern United States accessions (30%; temperate maize). Overall, these data suggested that two independent associated mutations (i.e. *cis*-regulatory variants) in the promoter region of ZCN8 arose in a stepwise manner: SNP-1245 during the early domestication of maize, and subsequently Indel-2339 during maize diversification into the Mexican highlands. The discovery that ZCN8 has more than one functional mutation that segregates indicated that genes associated with crop domestication and diversification are subject to recurrent mutations that might be selective targets at different times during evolution.

Identification of adaptive introgression can be relatively easy when materials collected at different times are available, such as with historical collections. A recent example was seen by the study of Bitocchi *et al.* [11], where the effects were evaluated for hybridisation of modern maize and landraces over a relatively short period of 50 years. Bitocchi *et al.* [11] analysed and compared the genetic diversity of two samples of maize landraces from central Italy that were collected at two different times: an old collection that was carried out before the introduction of hybrid varieties, and a recent collection that had evolved in co-existence with modern maize. Population structure analysis allowed the detection of introgression from modern maize. Coupled to the data of selection analyses (i.e. detection of outlier loci in comparisons between historical and recent maize collections), these data indicated that selection pressures for adaptation have favoured new alleles that were introduced by migration from hybrids over the last 50 years. These data showed the crucial role of migration in the evolution of landrace populations grown on farms.

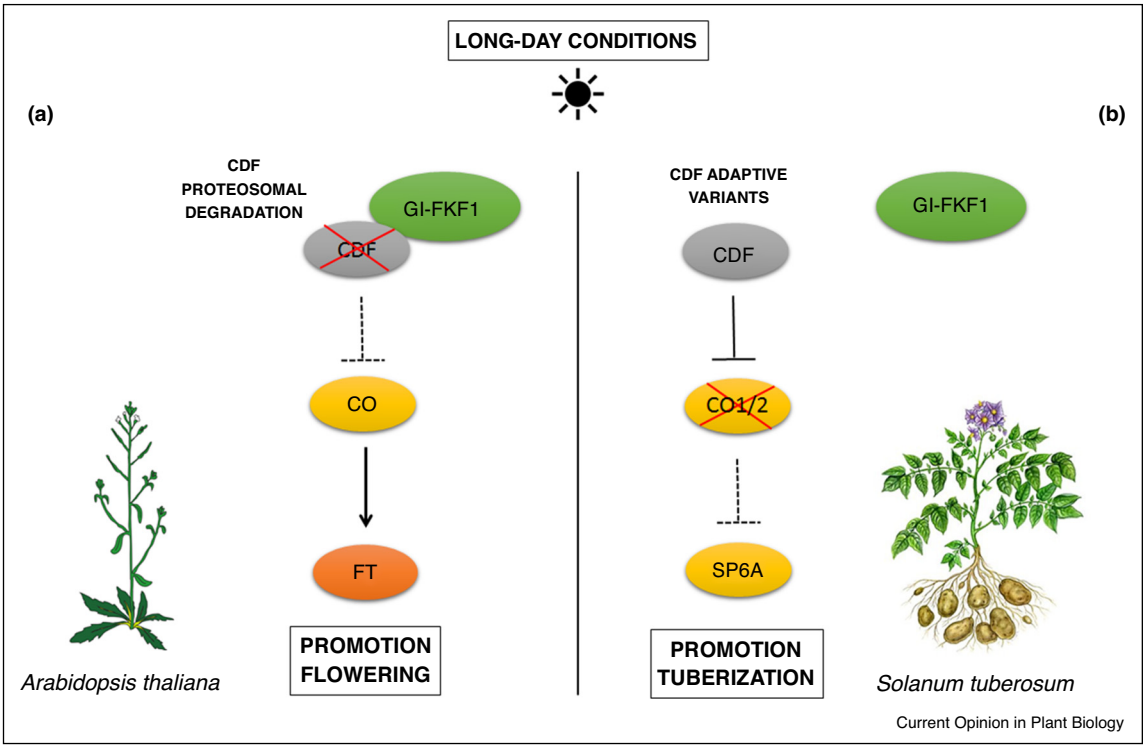
The Columbian Exchange: adaptation of crops from American homelands into Europe

The introduction of New World crops into Europe after the Columbus 1492 voyage was one of the most important evolutionary events related to agriculture, adaptation and biological changes, and more generally, to human society. In 1972, the historian Alfred Crosby coined the term 'Columbian Exchange', to designate the process of the biological diffusion triggered by the colonisation of the Americas by Europe. The benefits of the New World crops have resulted in their adoption in all parts of the world, which demonstrated that as the basis of this process, the plants underwent significant adaptation to the various agro-ecological conditions [54]. The growing knowledge about the adaptation of crops to new environments through the study of their introduction and expansion into Europe (i.e. a historically well-defined event of recent introduction and rapid adaptation) will be of great use for future major environmental and socio-economic changes, such as increases in temperature, variability of rainfall, and new consumer preferences. Several crops were introduced into Europe from the Americas (e.g. tomato, maize, beans, squash, potato, tobacco). This dissemination process occurred during the same historical period for several species, and it was characterised by diverse features (e.g. different mating systems and ploidy) that can be exploited to investigate their effects on genome diversity and to highlight the genetic control of adaptation. There are numerous studies in the literature for different crops that have highlighted the changes that occurred in their genomes due to colonisation of new

agro-ecosystems. Here, we present some examples for three crops that were involved in the Columbian Exchange, and which have been among the most important: potato, maize and common bean.

Following long debate during which most studies have suggested multiple domestications for *Solanum tuberosum L* (potato), Spooner *et al.* [55] demonstrated the monophyletic origin of cultivated potatoes through phylogenetic analysis and cladistic data. These showed that landraces of potato originated in the Andes of southern Peru, and subsequently became widespread throughout Chile, thus assuming the present-day distributions of the original cultivars. Potato was not brought to Europe by Columbus or others soon after the discovery of the New World in 1492; potato arrived later. The reason for this is that potato is a cool temperate crop of the high Andes of South America and was not discovered by the Spaniards until 1532 [56]. Potato cultivation in Europe spread rapidly, and also reached locations with significative growth and climate differences. For potato, the most important adaptation trait to European conditions – and a key event in its history – was to overcome the short-day dependency for tuberisation, due to the equatorial origin of potatoes [57,58,59••]. Indeed, when introduced into temperate zones, wild material forms tubers only during the shorter autumnal day lengths. The gradual arrival of winter, which is characterised by freezing temperatures, stops the correct maturation of the tubers, consequently killing the plant. In the *A. thaliana* model system, the pathway that controls flowering time is very complex, and the complexity of this regulation involves four intricate networks of signalling pathways (i.e. photoperiod, vernalisation, autonomous, gibberellins) [60] (Figure 3a). Among the proteins involved in this complex pathway, cycling Dof (DNA-binding with one finger) factors (CDFs) are a group of plant-specific transcription factors that repress flowering by downregulation of the expression of the *CONSTANS* (*CO*) gene, a central regulator of the photoperiod pathway [61]. In potato, the plant maturity phenotype has been reported as a major effect QTL that maps to chromosome 5, and this phenotype is a measure of several important secondary traits. These include development of the canopy, vegetative growth, onset of tuberisation, leaf senescence, life-cycle length and pathogen resistance [62]. Kloosterman *et al.* [57] used ultra-dense amplified fragment length polymorphism markers and two diploid segregant potato populations derived from crosses between wild and domesticated genotypes. In this way, they narrowed down the locus responsible for the plant maturity phenotype to a region of around 110 kb on chromosome 5. Screening for putative candidate genes, they identified the potato homologue of *CDF1* in this QTL region (*StCDF1*, *Solanum tuberosum CDF* gene 1). They sequenced *StCDF1* in the progenies of the mapping populations, which allowed identification of three *StCDF1* allelic variants: *StCDF1.1*, which was characteristic of short-day-dependent tuberisation descendants, and two insertion variants, *StCDF1.2* and *StCDF1.3*, that were typical of the early maturing/tuberising descendants. Kloosterman *et al.* [57] established that *StCDF* conserves its repressive function on the two potato *CONSTANS* genes (*StCO1/2*) that repress tuber formation during long days [63]. They also suggested that due to the loss of their C-terminal end, the *StCDF* Andean variants (i.e. *StCDF1.2*, *StCDF1.3*) led to accumulation of *StCO1/2* repressors. This interaction indirectly induced expression of *StSP6A*, the potato homologue of *FLOWERING TIME* (*FT*), which resulted in induction of tuber development under long days (Figure 3b). The absence of post-translational regulation of *StCDF1.2* and *StCDF1.3* allowed them to remain constant throughout the day, which formed the basis of potato diversification at different latitudes. A recent investigation explored haplotype diversity at the potato maturity locus *StCDF1* using a panel of 58 samples [58]. These included South American wild species, South American landraces, and North American cultivars derived from modern breeding programmes. Here, Hardigan *et al.* [58] reported 55 haplotypes for *StCDF1* that encoded 27 peptide variants. Four haplotype groups contained conserved deletions that affected the structure of the StCDF1 peptide. The DNA phylogeny of haplotypes at the *StCDF1* locus revealed that almost all long-day landraces/cultivars contained alleles that encoded shortened *StCDF1* proteins that were derived from introgression from wild species. This suggested a key role for the extant natural populations as essential sources of untapped adaptive potential. In the case of potato, *StCDF1* allele introgression from the wild species allowed potato cultivation in North America, and, probably, also subsequently in Europe. A very interesting study that focused on the origins and adaptation of European potatoes was carried out by Gutaker *et al.* [59••]. The strength of their work was the investigation of historical samples that spanned 350 years of potato evolution in Europe. Their materials included 29 historical herbarium specimens that they obtained from different European museums, which included three Chilean and 26 European historical samples. They also analysed 43 South American modern samples, and 16 European modern samples. An array-based targeted re-sequencing approach was used that allowed them to target the whole chloroplast genome and ~4.3 Mb of the nuclear genome, including *StCDF1* [57]. Analysis of these genetic data initially allowed Gutaker *et al.* [59••] to highlight the very complex scenario related to the introduction and wide spread of potatoes in Europe. These data indicated that the oldest European materials (i.e. collected between 1650 and 1750) derived from an ancestor of the Andean landraces, while in the subsequent 100 years there was introgression from newly introduced Chilean potatoes. The scenario is more complex considering that twentieth century European potatoes did not descend from their nineteenth century admixed predecessors, but are the result of introgression from wild potato species, as they were used in twentieth-century breeding programmes to introduce pathogen resistance [64]. It is also interesting that

Figure 3



Schematic representation of *CDF* gene function and interactions in the photoperiod pathway.
During long days, in the *A. thailana* model system **(a)**, the interaction between GIGANTEA (*GI*) and FLAVIN-BINDING KELCH REPEAT F-BOX 1 protein (*FKF1*) induces degradation of CYCLING DOF FACTOR (*CDF*), which is a repressor of CONSTANS (*CO*). *CO* promotes flowering by initiating transcription of the FLOWERING TIME (*FT*) gene. In *S. tuberosum L.* **(b)**, the *CDF* adaptive variant does not interact with the GI-FKF1 complex, which leads to repression of *CO1/2*. In contrast to *A. thaliana*, *CO1/2* act as repressors of *SP6A*, which is the potato homologue of *FT*. Repression of *CO1/2* allows expression of *SP6A* and promotion of potato tuberisation under long days, which forms the basis of potato diversification at different latitudes. Arrow, promotion of gene expression; truncated arrow, repression of gene expression; truncated dotted arrow, lack of repression due to pathway interruption.

Gutaker *et al.* [59••] highlighted the re-introduction of European potatoes into America, and that this impacted upon the Andean and Chilean potato diversity; indeed, European ancestry was detected in potatoes in the South American modern-day sample. Gutaker *et al.* [59••] also investigated the origins of the long-day adaptive alleles in the *StCDF1* gene. They reported the appearance of *StCDF1.2* and *StCDF1.3* adaptive alleles in Europe starting from 1810 only, with none of these insertion variants present in the oldest European samples of Andean descent (1650–1750), nor in the Andean landraces. For this reason, they excluded (with high confidence) the possibility that adaptation to long-day tuberisation had arisen from the Andean landraces standing variations. They showed the appearance of the adaptive alleles in Europe in correspondence with admixture with the newly introduced Chilean potatoes. However, there was no evidence of direct correlations between the adaptive variants and the historical samples from the lowlands of Chile. Gutaker *et al.* [59••] thus hypothesised that the adaptive insertions in the *StCDF1* gene originated *de novo* in Europe, and then became rapidly fixed due to their dominant inheritance and breeding advantage. However, they also stated that this hypothesis needs to be further confirmed, as their sampling of historical Chilean specimens is not particularly representative, and thus it did not allow clear rejection of the possibility of a Chilean origin of these adaptive insertions.

Another very important crop that became widespread in Europe during the Columbian exchange was *P. vulgaris* (common bean). This species originated in Mesoamerica, and wild forms became widespread by subsequent migration into South America; domestication took place independently in two geographically distant areas, Mesoamerica and the Andes, which represented the two main gene pools of the species [16]. The Mesoamerican common bean appears to have arrived in Europe through Spain and Portugal in 1506, following the first voyage of Columbus; then in 1528, the exploration of Peru by Pizarro opened the possibility of the introduction of the Andean common bean. *P. vulgaris* spread into the Old World over a very short time, and many common

bean landraces rapidly evolved in Europe as a result of its adaptation to new agro-ecological growth conditions. The dissemination of common bean into and across Europe followed very complex pathways, which involved different introductions from the Americas, and at the same time, direct exchanges among countries within Europe, and between European and other Mediterranean countries [16]. To investigate the evolutionary patterns of the common bean far from the Americas, Angioi *et al.* [54] analysed a wide sample of *P. vulgaris* accessions, as 94 from the Americas, and 307 from Europe. They included chloroplast simple sequence repeats (SSRs), and nuclear data (i.e. phaseolins, three indel-spanning markers of the PvSHATTERPROOF1, Pv*SHP1*, gene) and morphological data (i.e. coat pattern, seed size, colour and shape). In this way, Angioi *et al.* [65] showed that both the Mesoamerican and Andean gene pools were present in Europe and that the European germplasm was more prevalent as the Andean origin (67%). The trend was maintained at a smaller scale (i.e. a country level), whereby the Mesoamerican proportion was higher in the eastern parts of Europe, with a maximum of 46% in Greece, while the Andean type was most frequently found in three European macro areas: the Iberian Peninsula, Italy and central-northern Europe. Interestingly, and contrary to expectations, the European common bean did not show any strong reduction in genetic diversity due to the introduction bottlenecks and selection for adaptation to these new agro ecosystems and consumer preferences; indeed, Angioi *et al.* [54] and previous studies have shown very low reductions in diversity in common bean from Europe. These findings indicated a high level of gene flow among the different European geographic regions. Furthermore, they highlighted the role of the breakdown of the spatial isolation between the Mesoamerican and Andean accessions in Europe, with promotion of hybridisation, which had a significant impact on the maintenance of genetic diversity. By combining these chloroplast and nuclear data, they were able to identify hybridisation events, and they estimated that 44.2% of the European landraces derived from at least one hybridisation event between the Mesoamerican and Andean forms. Gioia *et al.* [66] complemented the dataset of Angioi *et al.* [65] with nuclear SSRs, and analysed a set of 89 American and 256 European landraces. Gioia *et al.* [66] combined the data from the recombination of the gene-pool-specific chloroplast SSRs, phaseolin and *PvSHP1* markers and the Bayesian assignments and admixture analysis based on nuclear SSRs, through which they were able to identify hybrids and distinguish them as 'pure' Mesoamerican and Andean genotypes. Novel combinations of genes/genomic regions thus arose in Europe after the common bean introduction and during its dissemination, on which adaptive selection acted (i.e. adaptive introgression). The new '-omics' technologies can help to fine-tune the molecular basis of these adaptations of the common bean in Europe, an aspect that is ongoing in the BEAN_ADAPT project (funded through the 2nd ERA-CAPS call, ERA-NET for Coordinating Action in Plant Sciences). This project is based on a multidisciplinary approach (i.e. genomics, population/quantitative genetics, biochemistry, plant physiology), with the aim being to dissect out the genetic basis and phenotypic consequences of the adaptation of *P. vulgaris* and its sister species *Phaseolus coccineus* from their centres of origin in the Americas into Europe and the new European agro-ecological environments.

Rebourg *et al.* [67] characterised a set of 131 European maize landraces according to morphological and genetic data (i.e. restriction fragment length polymorphism), and classified them into genetic groups that showed clear differentiation according to latitude. Six main European races were detected based on morphological and genetic differences: 'German flint', which included landraces mainly grown in Germany or the Alsace; 'north-eastern European flint', which included landraces mainly from France, and also Spain, Portugal and several eastern European countries; 'southern European flint', which was characterised by landraces from various countries which were mainly in southern Europe; 'Italian orange flint', as Italian landraces, with some others from southern Spain; 'Czechoslovakian type'; and 'Pyrenees-Galicia flint', which was characterised by two homogeneous subgroups, as the landraces from Galicia, and those from the Pyrenees and other regions of France. Then Rebourg *et al.* [68] included genetic data of 88 American landraces that were representative of the main American races in their previous dataset [56], to infer the genetic relationships among American and European maize populations. They showed signatures for the introduction of a bottleneck (European landraces retained overall 75% of the genetic diversity of those from America), and identified various types of American maize that were introduced into Europe at different times or in different places, which gave rise to distinctive European races [69]. Beyond confirming the importance of Caribbean germplasm, which was the first maize type to be introduced into Europe, they highlighted the close relationship between southern Spain and Caribbean populations, whereby the data revealed that introductions of North American flint populations had a key role in the adaptation of maize to the European climate. In particular, the data supported the hypothesis that present-day northern and eastern European flint germplasm was directly derived from North American flint populations. Northern flint populations were relatively insensitive to day length, and they had low temperature requirements for flowering. Earliness was a key factor for adaptation to the more temperate climates. Brandenburg *et al.* [70•] sequenced 67 genomes from both continents that covered 11 major groups, as representative of all of the American and European diversity. They used several population genomics and association mapping approaches to trace the origins of the European maize, and to investigate its demographic and selective history. One of the main outcomes of this study was the detection of admixture in the European maize materials. In particular, they

reported the admixed origins of the Italian flints from two contributions, the European flint and the southern European populations. This excluded the possibility of a third independent introduction, as had previously been suggested by Rebourg *et al.* [68], and instead emphasised the pivotal role of admixture in environmental maize adaptation. Moreover, the data of Brandenburg *et al.* [70•] highlighted the admixed origins of the European flints from the northern European flints and the tropical landraces. Interestingly, they also investigated the footprints of selection for adaptation to a wide range of climatic and ecological conditions, and they showed that numerous genes/gene networks were involved in flowering time, drought and cold tolerance, and in plant defence and starch properties. An example of the candidate genes for adaptation that were detected by associations between latitude and allele frequency was defined at GRMZM2G095955, a gene that is located in the vicinity of the maize floral activator, *ZCN8* [39]. They reported that in the *ZCN8* region there was a haplotype that was common to all temperate materials, and they showed segregation of this 'temperate' haplotype with a 'tropical' haplotype within the tropics, and to a lesser extent within the corn belt dents. Along with the previously characterised genes, they also revealed new candidates, including *ZCN5* (also known as *zen1* and *pebp5*), a gene from the same family as *ZCN8* that was recently reported to be associated with flowering time variations [71]. They also defined genes associated with plant responses to biotic and abiotic stresses, such as the *ZmASR2* gene (abscisic acid-induced, stress-induced, and ripening-induced protein 2), which was shown to have increased expression at the transcript and protein level under water-deficit conditions [72], and the *TPS23* gene that is involved in the control of the synthesis of a volatile sesquiterpene that attracts natural enemies of herbivores upon release [73].

Conclusions

Deeper understanding of the evolutionary processes and complex genetics mechanisms that form the basis of adaptation of plants to different environmental conditions is a very ambitious goal for evolutionary biologists, breeders and geneticists. It also has strong implications for overcoming the current challenges that agriculture has to face, such as to guarantee food security and quality, to adapt crops to marked variations in climate, and to protect and improve the environment. In this context, the identification of the genetic architecture both at genotype and population level that contribute to adaptive changes can strongly influence breeding targets and strategies. The potential applications are nearly infinite for the constitution of novel varieties in breeding programmes, but it will be crucial also for biodiversity conservation, to provide help in the implementation of the appropriate strategies. We have now in-hand novel tools and approaches that allow us to face this challenge through exploiting the unprecedented experimental power available. These include:

(i) Particularly advanced techniques that offer unique opportunities to scan a genome, not only to obtain genotypic information, but also to analyse the molecular phenotype of the whole genome, through analysis of the transcriptome, the metabolome, and the proteome [20,74,75,76••].

(ii) We canand approaches to analyse these data, which have also evolved to catch the complexity of these biological processes. Population genomics approaches allow the identification of candidate loci for adaptation using genotypic data without any prior information about phenotypes. Along with classical approaches aimed at detection of 'selective sweeps' [77], new methods and integrated approaches can be applied that take into account the concept that genes do not often actually operate as sole effectors, as they have roles in complex interactive systems, or gene networks that ultimately lead to a phenotype [78]. As an example of the impact that gene interactions can have on the determination of the phenotype, an *A. thaliana* genome-wide association analysis reported that for root length, epistatic effects can be so strong that they overcome the additive genetic variance [79•]. In soybean, Fang *et al.* [80] carried out a comprehensive genome-wide association studies that enabled identification of the underlying genetic loci, loci interaction, and genetic networks across important traits.

(iii) Multidisciplinary approaches can be applied and integrated to decipher the complexity of the genetic basis of adaptation. These can combine evidence from the signatures of selection analyses with association mapping to increase the power for the detection of regions that influence complex traits, while also reducing the number of false-positive signals [81,82]. Moreover, recently, different approaches have been developed based on the use of environmental variables that are treated as quantitative traits, and their association with molecular traits can be exploited as a tool to identify the loci that underlie local adaptation [12,83]. Similarly, network analyses can be used to investigate the roles of interactions between genes in local adaptation [84], using information on linkage disequilibrium shared between genome-wide multiple loci to perform linkage disequilibrium network analyses.

(iv) Landrace populations of crops are the 'perfect' model to apply all of these approaches to investigate adaptation features in the plant genome. They also allow the possibility to compare the effects of the same evolutionary process on the genome when this occurs as the following: independently on different populations of the same species (e.g. domestication in common bean occurred independently in Mesoamerica and the Andes) [18,19]; among different crop species within the same genus (e.g. different domesticated *Phaseolus* species) [16]; and/or among species of different genera (i.e. shattering trait in cereals)

[85] that are characterised by different features (e.g. diverse mating systems, diverse ploidy levels). These aspects offer great opportunities to go deeply into the molecular and developmental mechanisms at the basis of adaptation.

In this scenario, the Columbian Exchange represents a pivotal model. It offers a great opportunity to exploit all of these available tools and approaches, along with the plant genetic resources, to finally dissect out the genetic basis and phenotypic consequences of plant adaptation to new environments. This can now come through the study of their introduction from their respective centres of domestication in the Americas, and their expansion through Europe as a recent and historically well-defined event of rapid adaptation. Numerous crop species have been protagonists of these processes and have experienced adaptation in a relatively short period of time in the same geographic range (i.e. with the same environmental changes). What we need to do now is to investigate this process more deeply in different crops, and to compare and integrate the information obtained. A better understanding of variation in landscape structure across species and environments is also necessary to understand and predict how populations will adapt [86]. Moreover, advances in statistics and increased computing power already provide the possibility to develop predictive approaches, as demonstrated by Exposito-Alonso *et al.* [87••] who were able to build genome-wide environmental selection models to predict how evolutionary pressures on species will work in inaccessible environments, or even under future hypothetical climates.

Conflict of interest statement

Nothing declared.

Acknowledgements

This work was supported by grants from the ERA-NET for Coordinating Action in Plant Sciences-2nd ERA-CAPS call, BEAN_ADAPT project.

Appendix A. Supplementary data

Supplementary material related to this article can be found, in the online version, at doi:https://doi.org/10.1016/j.pbi.2019.12.011.

References and recommended reading

Papers of particular interest, published within the period of review, have been highlighted as:

- • of special interest
- •• of outstanding interest

1. Walther GR, Post E, Convey P, Menzel A, Parmesan C, Beebee TJC, Fromentin JM, Hoegh-Guldberg O, Bairlein F: **Ecological responses to recent climate change**. *Nature* 2002, **416**:389-395 http://dx.doi.org/10.1146/annurev-arplant-042817-040240.

2. Lorant A, Ross-Ibarra J, Tenaillon MI: **Genomics of long- and short-term adaptation in maize and teosinte**. *PeerJ* 2018, **6**: e27190v1 http://dx.doi.org/10.7287/peerj.preprints.27190v1 Preprints.

3. Barrett RDH, Schluter D: **Adaptation from standing genetic variation**. *Trends Ecol Evol* 2008, **23**:38-44 http://dx.doi.org/10.1016/j.tree.2007.09.008.

4. Mousavi-Derazmahalleh M, Bayer PE, Hane JK, Valliyodan B, Nguyen HT, Nelson MN, Erskine W, Varshney RK, Papa R, Edwards D: **Adapting legume crops to climate change using genomic approaches**. *Plant Cell Environ* 2019, **42**:6-9 http://dx.doi.org/10.1111/pce.13203.

5. • Fournier-Level A, Korte A, Cooper MD, Nordborg M, Schmitt J, Wilczek AM: **A map of local adaptation in *Arabidopsis thaliana***. *Science* 2011, **334**:86-89 http://dx.doi.org/10.1126/science.1209271.
This study provides complementary analyses of local adaptation to climate in *A. thaliana*, with the combination of genome-wide SNPs, fitness estimates in the field and continental-scale climate data.

6. Fustier MA, Martínez-Ainsworth NE, Aguirre-Liguori JA, Venon A, Corti H, Rousselet A, Dumas F, Dittberner H, Camarena MG, Grimanelli D *et al.*: **Common gardens in teosintes reveal the establishment of a syndrome of adaptation to altitude**. *bioRxiv* 2019:563585 http://dx.doi.org/10.1101/563585.

7. • Rodriguez M, Rau D, Bitocchi E, Bellucci E, Biagetti E, Carboni A, Gepts P, Nanni L, Papa R, Attene G: **Landscape genetics, adaptive diversity, and population structure in *Phaseolus vulgaris***. *New Phytol* 2016, **209**:1781-1794 http://dx.doi.org/10.1111/nph.13713.
This study represents one of the first examples of the application of landscape genomics to crops, to help in the understanding of adaptation on the scale of the natural landscape.

8. Mier Y, Teran BJC, Konzen ER, Medina V, Palkovic A, Ariani A, Tsai SM, Gilbert ME: **Root and shoot variation in relation to potential intermittent drought adaptation of Mesoamerican wild common bean (*Phaseolus vulgaris* L.)**. *Ann Bot* 2018, **124 (6)**:917-932 http://dx.doi.org/10.1093/aob/mcy221 UC Davis.

9. Pusadee T, Jamjod S, Chiang YC, Rerkasem B, Schaal BA: **Genetic structure and isolation by distance in a landrace of Thai rice**. *PNAS* 2009, **106**:13880-13885 http://dx.doi.org/10.1073/pnas.0906720106.

10. Bellucci E, Bitocchi E, Rau D, Nanni L, Ferradini N, Giardini A, Rodriguez M, Attene G, Papa R: **Population structure of barley landraces populations and geneflow with modern varieties**. *PLoS One* 2013, **8**:e83891 http://dx.doi.org/10.1371/journal.pone.0083891.

11. Bitocchi E, Bellucci E, Rau D, Albertini E, Rodriguez M, Veronesi F, Attene G, Nanni L: **European flint landraces grown *In-situ* reveal adaptive introgression from modern Maize**. *PLoS One* 2015, **10**: e0121381 http://dx.doi.org/10.1371/journal.pone.0121381.

12. Lasky JR, Upadhyaya HD, Ramu P, Deshpande S, Hash CT, Bonnette J, Juenger TE, Hyma K, Acharya C, Mitchell SE *et al.*: **Genome-environment associations in sorghum landraces predict adaptive traits**. *Sci Adv* 2015, **1**:e1400218 http://dx.doi.org/10.1126/sciadv.1400218.

13. Lister DL, Jones H, Hugo R, Oliveira HR, Petrie CA, Liu X, Cockram J, Kneale CJ, Kovaleva O, Jones MK: **Barley heads east: genetic analyses reveal routes of spread through diverse Eurasian landscapes**. *PLoS One* 2018, **13**:e0196652 http://dx.doi.org/10.1371/journal.pone.0196652.

14. Tanto Hadado T, Rau D, Elena B, Papa R: **Genetic diversity of barley (*Hordeum vulgare* L.) landraces from the central highlands of Ethiopia: comparison between the *Belg* and *Meher* growing seasons using morphological traits**. *Genet Resour Crop Evol* 2009, **56**:1131-1148 http://dx.doi.org/10.1007/s10722-009-9437-z.

15. Manchanda N, Snodgrass SJ, Ross-Ibarra J, Hufford MB: **Evolution and adaptation in the maize genome**. *The Maize Genome, Compendium of Plant Genomes*. Cham: Springer; 2018, 319-332 http://dx.doi.org/10.1007/978-3-319-97427-9_19.

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

17. • Meyer RS, Purugganan MD: **Evolution of crop species: genetics of domestication and diversification**. *Nat Rev Genet* 2013, **14**:840-852 http://dx.doi.org/10.1038/nrg3605.

This review provides a detailed and comprehensive overview of the process of domestication and the subsequent crop expansion.

18. Bitocchi E, Rau D, Benazzo A, Bellucci E, Goretti D, Biagetti E, Panziera A, Laidò G, Rodriguez M, Gioia T *et al.*: **High level of nonsynonymous changes in common bean suggests that selection under domestication increased functional diversity at target traits**. *Front Plant Sci* 2017, **7**:2005 http://dx.doi.org/10.3389/fpls.2016.02005.

19. Schmutz J, McClean PE, Mamidi S, Wu GA, Cannon SB, Grimwood J, Jenkins J, Shu S, Song Q, Chavarro C *et al.*: **A reference genome for common bean and genome-wide analysis of dual domestications**. *Nat Genet* 2014, **46**:707-713 http://dx.doi.org/10.1038/ng.3008.

20. • Bellucci E, Bitocchi E, Ferrarini A, Benazzo A, Biagetti E, Klie S, Minio A, Rau D, Rodriguez M, Panziera A *et al.*: **Decreased nucleotide and expression diversity and modified co-expression patterns characterize domestication in the common bean**. *Plant Cell* 2014, **26**:1901-1912 http://dx.doi.org/10.1105/tpc.114.124040.

This study is focused on investigation of the effects of the domestication process in common bean using RNA-seq data. It is one of the pioneering reports on the consequences of domestication, not only at the genome level, but also at the gene-expression level. It highlights a decrease in gene expression diversity in domesticated compared to wild forms, a pattern that appears to be general for all or most domesticated species.

21. Borkotoky S, Saravanan V, Jaiswal A, Das B, Selvaraj S, Murali A, Lakshmi PTV: **The *Arabidopsis* stress responsive gene database**. *Int J Plant Genomics* 2013, **2013**:949564 http://dx.doi.org/10.1155/2013/949564.

22. Emms DM, Kelly S: **OrthoFinder: solving fundamental biases in whole genome comparisons dramatically improves orthogroup inference accuracy**. *Genome Biol* 2015, **16**:157 http://dx.doi.org/10.1186/s13059-015-0721-2.

23. Osakabe Y, Arinaga N, Umezawa T, Katsura S, Nagamachi K, Tanaka H, Ohiraki H, Yamada K, Seo S, Abo M *et al.*: **Osmotic stress responses and plant growth controlled by potassium transporters in Arabidopsis**. *Plant Cell* 2013, **25**:609-624 http://dx.doi.org/10.1105/tpc.112.10570.

24. Meyer RS, Choi JY, Sanches M, Plessis A, Flowers JM, Amas J, Dorph K, Barretto A, Gross B, Fuller DQ *et al.*: **Domestication history and geographical adaptation inferred from a SNP map of African rice**. *Nat Genet* 2016, **48**:1083-1088 http://dx.doi.org/10.1038/ng.3633.

25. Yang T, Zhang S, Hu Y, Wu F, Hu Q, Chen G, Cai J, Wu T, Moran N, Yu L, Xu G: **The role of a potassium transporter $OsHAK_5$ in potassium acquisition and transport from roots to shoots in rice at low potassium supply levels**. *Plant Physiol* 2014, **166**:945-959 http://dx.doi.org/10.1104/pp.114.246520.

26. • Buckler ES, Holland JB, Bradbury PJ, Acharya CB, Brown PJ, Browne C, Ersoz E, Flint-Garcia S, Garcia A, Glaubitz JC *et al.*: **The genetic architecture of maize flowering time**. *Science* 2009, **325**:714-718 http://dx.doi.org/10.1126/science.1174276.

This study uses nested association mapping for QTL mapping in maize. This population is an extremely useful resource that was used to score the flowering time for nearly a million individual plants in four environments and over two years. Many QTLs with small additive effects on the flowering time were detected. This result is in contrast with other studies that investigated the genetic architecture of flowering time in other species, such as Arabidopsis and rice, where few loci of relatively large effects appear to control this trait.

27. Xu J, Liu Y, Liu J, Cao M, Wang J, Lan H, Xu Y, Lu Y, Pan G, Rong T: **The genetic architecture of flowering time and photoperiod sensitivity in maize as revealed by QTL review and meta-analysis**. *JIPB* 2012, **54**:358-373 http://dx.doi.org/10.1111/j.1744- 7909.2012.01128.x.

28. Navarro JAR, Willcox M, Burgueno J, Romay C, Swarts K, Trachsel S, Preciado E, Terron A, Vallejo Delgado H, Vidal V *et al.*: **A study of allelic diversity underlying flowering-time adaptation in maize landraces**. *Nat Genet* 2017, **49**:476-480 http://dx.doi.org/10.1038/ng.3784.

29. Thornsberry JM, Goodman MM, Doebley J, Kresovich S, Nielsen D, Buckler ESIV: ***Dwarf8* polymorphisms associate with variation in flowering time**. *Nat Genet* 2001, **28**:286-289.

30. Hung HY, Shannon LM, Tian F, Bradbury PJ, Chen C, Flint-Garcia SA, McMullen MD, Ware D, Buckler ES, Doebley JF, Holland JB: ***ZmCCT* and the genetic basis of day-length adaptation underlying the post-domestication spread of maize**. *PNAS* 2012, **109**:E1913-E1921 http://dx.doi.org/10.1073/pnas.1203189109.

31. Yang Q, Li Z, Li W, Ku L, Wang C, Ye J, Li K, Yang N, Li Y, Zhong T *et al.*: **CACTA-like transposable element in *ZmCCT* attenuated photoperiod sensitivity and accelerated the post-domestication spread of maize**. *PNAS* 2013, **111**:16969-16974 http://dx.doi.org/10.1073/pnas.1310949110.

32. Huang C, Sun H, Xu D, Chen Q, Liang Y, Wang X, Xu G, Tian CW, Li D, Wu L *et al.*: ***ZmCCT9* enhances maize adaptation to higher latitudes**. *PNAS* 2017:E334-E341 http://dx.doi.org/10.1073/PNAS.1718058115.

33. Salvi S, Sponza G, Morgante M, Tomes D, Niu X, Fengler KA, Meeley R, Ananiev EV, Svitashev S, Bruggemann E *et al.*: **Conserved noncoding genomic sequences associated with a flowering-time quantitative trait locus in maize**. *PNAS* 2007, **104**:11376-11381 http://dx.doi.org/10.1073/pnas.0704145104.

34. Guo L, Wang X, Zhao M, Huang C, Li C, Yang CJ, York AM, Xue W, Xu G, Liang Y *et al.*: **Stepwise *cis*-regulatory changes in *ZNC8* contribute to maize flowering-time adaptation**. *Curr Biol* 2018, **28**:3005-3015 http://dx.doi.org/10.1016/j.cub.2018.07.029.

35. Camus-Kulandaivelu L, Veyrieras JB, Madur D, Combes V, Fourmann M, Barraud S, Dubreuil P, Gouesnard B, Manicacci D, Charcosset A: **Maize adaptation to temperate climate: relationship between population structure and polymorphism of *Dwarf8* gene**. *Genetics* 2006, **172**:2449-2463 http://dx.doi.org/10.1534/genetics.105.048603.

36. Ducrocq S, Madur D, Veyrieras JB, Camus-Kulandaivelu L, Kloiber-Matiz M, Presterl T, Ouzunova M, Manicacci D, Charcosset A: **Key impact of *Vgt1* on flowering time adaptation in maize: evidence from association mapping and ecogeographical information**. *Genetics* 2008, **178**:2433-2437 DOI: 0.1534/genetics.107.084830.

37. Castelletti S, Tuberosa R, Pindo M, Salvi S: **A MITE transposon insertion is associated with differential methylation at the maize flowering time QTL *Vgt1***. *G3 Genes Genomes Genet* 2014, **4**:805-812 http://dx.doi.org/10.1534/g3.114.010686.

38. Lazakis CM, Coneva V, Colasanti J: ***ZCN8* encodes a potential orthologue of *Arabidopsis* FT florigen that integrates both endogenous and photoperiod flowering signal in maize**. *J Exp Bot* 2011, **62**:4833-4842 http://dx.doi.org/10.1093/jxb/err129.

39. Meng X, Muszynski MG, Danilevskaya ON: **The FT-like *ZCN8* gene functions as a floral activator and is involved in photoperiod sensitivity in maize**. *Plant Cell* 2011, **23**:942-960 http://dx.doi.org/10.1105/tpc.110.08140.

40. •• Vigouroux Y, Mariac C, De Mita S, Pham JL, Gérard B, Kapran I, Sagnard F, Deu M, Chantereau J, Ali A *et al.*: **Selection for earlier flowering crop associated with climatic variations in the Sahel**. *PLoS One* 2011, **6**:e19563 http://dx.doi.org/10.1371/journal.pone.0019563.

This study represents an excellent example of the usefulness of historical collections (i.e. materials collected at different times) of landraces to identify adaptive introgression. The results show strong adaptation of pearl millet landraces to changing environmental conditions, even over relatively short evolutionary timescales. It also suggests that exploitation of genetic variability within landrace populations represents a strategy in response to future climate changes.

41. Saïdou AA, Mariac C, Luong V, Pham JL, Bezançon G, Vigouroux Y: **Association studies identify natural variation at *PHYC* linked to flowering time and morphological variation in pearl millet**. *Genetics* 2009, **82**:899-910 http://dx.doi.org/10.1534/genetics.109.102756.

42. •• Khan AW, Garg V, Roorkiwal M, Golicz AA, Edwards D, Varshney RK: **Super-pangenome by integrating the wild side of a species for accelerated crop improvement**. *Trends Plant Sci* 2019 http://dx.doi.org/10.1016/j.tplants.2019.10.012.

This opinion paper is an updated overview of the state of understanding about the recent developments in crop pan-genomics. This study emphasises the importance of the dispensable genome as a repertoire of adaptive variability, which is useful for the elucidation of crop evolutionary

history and the development of new improved varieties that are resistant to stress and pathogens.

43. Zhou D, Zhou J, Meng L, Wang Q, Xie H, Guan Y, Ma Z, Zhong Y, Chen F, Liu J: **Duplication and adaptive evolution of the *COR15* genes within the highly cold-tolerant *Draba* lineage (Brassicaceae)**. *Gene* 2008, **441**:36-44 http://dx.doi.org/10.1016/j.gene.2008.06.024.

44. DeBolt S: **Copy number variation shapes genome diversity in *Arabidopsis* over immediate family generational scales**. *Genome Biol Evol* 2010, **2**:441-453 http://dx.doi.org/10.1093/gbe/evq033.

45. Mutti JS, Bhullar RK, Gill KS: **Evolution of gene expression balance among homeologs of natural polyploids**. *G3 Genes Genomes Genet* 2017, **7**:1225-1237 http://dx.doi.org/10.1534/g3.116.038711.

46. Ceccarelli M, Santantonio E, Marmottini F, Amzallag GN, Cionini PG: **Chromosome endoreduplication as a factor of salt adaptation in *Sorghum bicolor***. *Protoplasma* 2006, **227**:113-118 http://dx.doi.org/10.1007/s00709-005-0144-0.

47. Saleh B, Allario T, Dambier D, Ollitrault P, Morillon R: **Tetraploid citrus rootstocks are more tolerant to salt stress than diploid**. *C R Biol* 2008, **331**:703-710 http://dx.doi.org/10.1016/j.crvi.2008.06.007.

48. Le Corre V, Kremer A: **The genetic differentiation at quantitative trait loci under local adaptation**. *Mol Ecol* 2012, **21**:1548-1566 http://dx.doi.org/10.1111/j.1365-294X.2012.05479.x.

49. Innan H, Kim Y: **Pattern of polymorphism after strong artificial selection in a domestication event**. *PNAS* 2004, **101**:10667-10672 http://dx.doi.org/10.1073/pnas.0401720101.

50. Matuszewski S, Hermisson J, Kopp M: **Catch me if you can: adaptation from standing variation to a moving phenotypic optimum**. *Genetics* 2015, **200**:1255-1274 http://dx.doi.org/10.1534/genetics.115.178574.

51. Anderson E, Stebbins GL Jr: **Hybridization as an evolutionary stimulus**. *Evolution* 1954, **8**:378-388 http://dx.doi.org/10.2307/2405784.

52. Lewontin RC, Birch LC: **Hybridization as a source of variation for adaptation to new environments**. *Evolution* 1966, **20**:223-236 http://dx.doi.org/10.1111/j.1558-5646.1966.tb03369.x.

53. Hufford MB, Lubinksy P, Pyhäjärvi T, Devengenzo MT, Ellstrand NC, Ross-Ibarra J: **The genomic signature of crop-wild introgression in maize**. *PLoS Genet* 2013 http://dx.doi.org/10.1371/journal.pgen.1003477.

54. Crosby AW: *The Columbian Exchange: Biological and Cultural Consequences of 1492*. Westport, CT: Praeger Publishers; 2003.

55. Spooner DM, McLean K, Ramsay G, Waugh R, Bryan GJ: **A single domestication for potato based on multilocus amplified fragment length polymorphism genotyping**. *PNAS* 2005, **102**:14694-14699 http://dx.doi.org/10.1073/pnas.0507400102.

56. Hawkes JG, Francisco-Ortega J: **The early history of the potato in Europe**. *Euphytica* 1993, **70**:1-7.

57. Kloosterman B, Abelenda JA, Carretero Gomez MM, Oortwijn M, De Boer JM, Kowitwanich K, Horvath BM, Van Eck HJ, Smaczniak C, Prat S *et al.*: **Naturally occurring allele diversity allows potato cultivation in northern latitudes**. *Nature* 2013, **495**:246-250 http://dx.doi.org/10.1038/nature11912.

58. Hardigan MA, Parker F, Laimbeer E, Newton L, Crisovan E, Hamilton JP, Vaillancourt B, Wiegert-Rininger K, Wood JC, Douches DS *et al.*: **Genome diversity of tuber-bearing *Solanum* uncovers complex evolutionary history and targets of domestication in the cultivated potato**. *PNAS* 2017, **114(46)**: E9999-E10008 http://dx.doi.org/10.1073/pnas.1714380114.

59. •• Gutaker RM, Weib CL, Ellis D, Anglin NL, Knapp S, Fernandez-Alonso JL, Prat S, Burbano HA: **The origins and adaptation of European potatoes reconstructed from historical genomes**. *Nat Ecol Evol* 2019, **3**:1093-1101 http://dx.doi.org/10.1038/s41559-019-0921-3.

This recent study is an excellent example of the value of degraded and/or historic botanical specimens to study crop evolution. The work was based on an investigation of historical samples that spanned 350 years of potato evolution in Europe. Their materials included historical herbarium specimens that they obtained from different European museums and modern American and European potato samples. Next-generation sequencing data were used to highlight the power of combining contemporary and historical genomes to shed light on the complex evolutionary history of crops, and their adaptation to new environments.

60. Reeves PH, Coupland G: **Response of plant development to environment: control of flowering by daylength and temperature**. *Curr Opin Plant Biol* 2000, **3**:37-42 http://dx.doi.org/10.1016/S1369-5266(99)00041-2.

61. Putterill J, Robson F, Lee K, Simon R, Coupland G: **The CONSTANS gene of *Arabidopsis* promotes flowering and encodes a protein showing similarities to zinc finger transcription factors**. *Cell* 1995, **80**:847-857 http://dx.doi.org/10.1016/0092-8674(95)90288-0.

62. Visker M, Keizer L, Van Eck H, Jacobsen E, Colon L, Struik P: **Can the QTL for late blight resistance on potato chromosome 5 be attributed to foliage maturity type?** *Theor Appl Genet* 2003, **106**:317-325 http://dx.doi.org/10.1007/s00122-002-1021-2.

63. Gonzales-Schain ND, Diaz-Mendoza M, Zurczak M, Suarez-Lopez P: **Potato CONSTANS is involved in photoperiodic tuberization in a graft-transmissible manner**. *Plant J* 2012, **70**:678-690 http://dx.doi.org/10.1111/j.1365-313X.2012.04909.x.

64. Bradshaw JE, Bryan GJ, Ramsay G: **Genetic resources (including wild and cultivated *Solanum* species) and progress in their utilisation in potato breeding**. *Potato Res* 2006, **49**:49-65.

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

66. Gioia T, Logozzo G, Attene G, Bellucci E, Benedettelli S, Negri V: **Evidence for introduction bottleneck and extensive inter-gene pool (Mesoamerica x Andes) hybridization in the European common bean (*Phaseolus vulgaris* L.) germplasm**. *PLoS One* 2013, **8**:e75974 http://dx.doi.org/10.1371/journal.pone.0075974.

67. Rebourg C, Gouesnard B, Charcosset A: **Large scale molecular analysis of traditional European maize populations. Relationships with morphological variation**. *Heredity* 2001, **86**:574-587 http://dx.doi.org/10.1046/j.1365-2540.2001.00869.x.

68. Rebourg C, Chastanet M, Gouesnard B, Welcker C, Dubreuil P, Charcosset A: **Maize introduction into Europe: the history reviewed in the light of molecular data**. *Theor Appl Genet* 2003, **106**:895-903 http://dx.doi.org/10.1007/s00122-002-1140-9.

69. Tenaillon MI, Charcosset A: **A European perspective on maize history**. *C R Biol* 2011, **334**:221-228 http://dx.doi.org/10.1016/j.crvi.2010.12.015.

70. • Brandenburg JT, Mary-Huard T, Rigaill G, Hearne SJ, Corti H, Joets J, Vitte C, Charcosset A, Nicolas SD, Tenallion MI: **Independent introductions and admixtures have contributed to adaptation of European maize and its American counterparts**. *PLOS Genet* 2017, **13**:e1006666 http://dx.doi.org/10.1371/journal.pgen.1006666.

This study is the most recent report of the introduction of maize into Europe. Sixty-seven sequenced maize genomes that were representative of all of the American and European diversity were used to carry out several population genomics and association mapping approaches, to trace the origins of European maize and to investigate its demographic and selective history.

71. Li YX, Li CH, Bradbury PJ, Liu XL, Lu F, Romay CM, Glaubitz JC, Wu X, Peng B, Shi Y *et al.*: **Identification of genetic variants associated with maize flowering time using an extremely large multi-genetic background population**. *Plant J* 2016, **86**:391-402 http://dx.doi.org/10.1111/tpj.13174.

72. Virlouvet L, Jacquemot MP, Gerentes D, Corti H, Bouton S, Gilard F, Valot B, Trouverie J, Tcherkez G, Falque M *et al.*: **The ZmASR1 protein influences branched-chain amino acid biosynthesis and maintains kernel yield in maize under water-limited conditions**. *Plant Physiol* 2011, **157**:917-936 http://dx.doi.org/10.1104/pp.111.176818.

73. Köllner TG, Held M, Lenk C, Hiltpold I, Turlings TCJ, Gershenzon J, Degenhardt J: **A maize (E)-β-caryophyllene synthase implicated in indirect defense responses against herbivores is not expressed in most American maize varieties**. *Plant Cell* 2008, **20**:482-494 http://dx.doi.org/10.1105/tpc.107.051672.

74. Toubiana D, Fernie AR, Nikoloski Z, Fait A: **Network analysis: tackling complex data to study plant metabolism**. *Trends Biotechnol* 2013, **31**:29-36 http://dx.doi.org/10.1016/j.tibtech.2012.10.011.

75. Sade D, Shriki O, Cuadros-Inostroza A, Tohge T, Semel Y, Haviv Y, Willmitzer L, Fernie AR, Czosnek H, Brotman Y: **Comparative metabolomics and transcriptomics of plant response to *Tomato yellow leaf curl virus* infection in resistant and susceptible tomato cultivars**. *Metabolomics* 2015, **11**:81-97 http://dx.doi.org/10.1007/s11306-014-0670-x.

76. •• Beleggia R, Rau D, Laidò G, Platani C, Nigro F, Fragasso PDV, Scossa F, Fernie AR, Nikoloski Z, Papa R: **Evolutionary metabolomics reveals domestication-associated changes in tetraploid wheat kernels**. *Mol Biol Evol* 2016, **33**:1740-1753 http://dx.doi.org/10.1093/molbev/msw050.

In this paper, the authors developed and used for the first time a methodological pipeline to identify the signature of selection for molecular phenotypic traits (e.g., metabolites and transcripts) using a QST vs FST comparison combined with metabolic correlation networks. They investigated the effects of selection on the accumulation of 51 primary metabolites and revealed domestication-associated reduction in unsaturated fatty acids during domestication of emmer and changes in the amino acid content associated with the secondary domestication of durum wheat. This work represents the first example of a molecular evolutionary phenomics study.

77. Pavlidis P, Alachiotis NA: **Survey of methods and tools to detect recent and strong positive selection**. *J Biol Res Thessalon* 2017, **24**:7 http://dx.doi.org/10.1186/s40709-017-0064-0.

78. Lareau CA, McKinney BA: **Network theory for data-driven epistasis networks**. *Epistasis* 2015, **1253**:285-300 http://dx.doi.org/10.1007/978-1-4939-2155-3_15.

79. • Lachowiec J, Shen X, Queitsch C, Carlborg Ö: **A genome-wide association analysis reveals epistatic cancellation of additive genetic variance for root length in *Arabidopsis thaliana***. *PLoS Genet* 2015, **11**:e1005541 http://dx.doi.org/10.1371/journal.pgen.1005541.

This study is an excellent example of a genome-wide association analysis conducted in Arabidopsis that provides valuable insights into the genetic mechanisms that underlie complex quantitative traits and the influence of epistasis on evolutionary processes.

80. Fang C, Ma Y, Wu S, Liu Z, Wang Z, Yang R, Hu G, Zhou Z, Yu H, Zhang M *et al.*: **Genome-wide association studies dissect the genetic networks underlying agronomical traits in soybean**. *Genome Biol* 2017, **18**:161 http://dx.doi.org/10.1186/s13059-017-1289-9.

81. Stinchcombe JR, Hoekstra HE: **Combining population genomics and quantitative genetics: finding the genes underlying ecologically important traits**. *Heredity* 2008, **100**:158-170 http://dx.doi.org/10.1038/sj.hdy.6800937.

82. Schwarzenbacher H, Dolezal M, Flisikowski K, Seefried F, Wurmser C, Schlötterer C, Fries R: **Combining evidence of selection with association analysis increases power to detect regions influencing complex traits in dairy cattle**. *BMC Genomics* 2012, **13**:48 http://dx.doi.org/10.1186/1471-2164-13-48.

83. Eckert AJ, Heerwaarden JV, Wegrzyn JL, Nelson CD, Ross-Ibarra J, Gonzàles- Martìnez SC, Neale DB: **Patterns of population structure and environmental associations to aridity across the range of loblolly pine (*Pinus taeda* L., Pinaceae)**. *Genetics* 2010, **185**:969-982 http://dx.doi.org/10.1534/genetics.110.115543.

84. Kemppainen P, Knight CG, Sarma DK, Hlaing T, Prakash A, Maung Maung YN, Somboon P, Mahanta J, Walton C: **Linkage disequilibrium network analysis (LDna) gives a global view of chromosomal inversions, local adaptation and geographic structure**. *Mol Ecol Resour* 2015, **15**:1031-1045 http://dx.doi.org/10.1111/1755-0998.12369.

85. Lin Z, Li X, Shannon LM, Yeh CT, Wang ML, Bai G, Peng Z, Li J, Trick HN, Clemente TE *et al.*: **Parallel domestication of the Shattering1 genes in cereals**. *Nat Genet* 2012, **44**:720-724 http://dx.doi.org/10.1038/ng.2281.

86. Blanquart F, Bataillon T: **Epistasis and the structure of fitness landscapes: are experimental fitness landscapes compatible with fisher's geometric model?** *Genetics* 2009, **203**:847-862 http://dx.doi.org/10.1073/pnas.0906720106.

87. •• Exposito-Alonso M, Burbano HA, Bossdorf O, Nielsen R, Weigel D: **Natural selection on the *Arabidopsis thaliana* genome in present and future climates**. *Nature* 2019, **573**:126-129 http://dx.doi.org/10.1038/s41586-019-1520-9.

This study is based on an analysis of the extensive genome information available for *Arabidopsis thaliana* and the measures of the fitness on different rainfall conditions for 517 natural Arabidopsis lines grown in Germany and Spain. This experiment provides proof of concept for the use of genome-wide environment selection models for evolution-aware predictions of risks for biodiversity that are associated with climate changes.

88. Ceccarelli S, Grando S: **Participatory plant breeding: Who did it, who does it and where?** *Exp Agric* 2019:1-11 http://dx.doi.org/10.1017/S0014479719000127.

89. Tanto Hadado T, Rau D, Bitocchi E, Papa R: **Adaptation and diversity along an altitudinal gradient in Ethiopian barley (*Hordeum vulgare* L.) landraces revealed by molecular analysis**. *BMC Plant Biol* 2010, **10**:121 http://dx.doi.org/10.1186/1471-2229-10-121.

Chapter IV

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

*These first authors contributed equally to this work

**These last authors contributed equally to this work

#Corresponding author: R. Papa

[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 Sassare, Italy
[5] Max Planck Institute of Molecular Plant Physiology (MPI-MP), 14476 Potsdam-Golm, Germany
[6] School of Agricultural, Forestry, Food and Environmental Sciences, University of Basilicata, 85100 Potenza, Italy
[7] Department of Plant Sciences, University of California, 95616-8780 Davis, CA, USA
[8] Genebank Department, Leibniz Institute of Plant Genetics and Crop Plant Research, 06466 Gatersleben, Germany
[9] Department of Agronomy and Plant Genetics, University of Minnesota, St. Paul, MN, USA
[10] Regional Agrifood Research and Development Service (SERIDA), 33310, Villaviciosa, Asturias, Spain

Department of Agricultural, Food and Environmental Sciences
Marche Polytechnic University
Via Brecce Bianche, 60131 Ancona, Italy

In Submission

My contribution on this Chapter is:

- Data analysis
- Data processing
- Writing

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 Sassare, 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

28 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 (Crosby, 1972) 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 (Cortinovis et al., 2020, Brandenburg et al., 2017). Among the crops that originated in the Americas, common bean was rapidly adopted and successfully disseminated across Europe (Gepts 2002). It is now possible to identify local varieties of beans in Europe of both Andean and Mesoamerican origins Gepts and Bliss 1988, Santalla 2002, Sicard et al., 2005, Zeven et al., 1999, Angioi et al., 2010, 2011).

The introduction of common bean in Europe from two distinct centres of origin provided the opportunity for widespread inter-gene pool hybridisation and recombination (Angioi et al., 2010). 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 (Ariani et al. 2018; Bitocchi et al., 2017, Schmutz, et al, 2014) (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 (Tsitrone et al., 2003) 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. (2009) we investigated the common bean population structure within each gene pool from the Americas (see Supplementary Note 4.2). Using ADMIXTURE (Alexander et al 2009), 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. (1991): **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 (Trucchi et al., 2021). 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; Lawson et al., 2012) 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 (Alexander et al 2009) 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 (Gepts and Bliss 1988, Zeven, 1997). Globally, the inferred amount of per-accession introgressed material was different between EU_MES and EU_AND accessions (two sided K-S test, $P=3.3x10^{-3}$), 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 (Perale, 2001) 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. (2022). 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 (Perale, 2001). 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 (1988) and Bellucci et al., 2014, 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 (Bellucci et al., 2014, Wu et al. 2020) 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

208 from the Americas showed moderately higher nucleotide diversity than European ones (see
209 Supplementary Note 4.3), which appears to be due to the introduction bottleneck in Europe (Fig. 4a, b).
210 Differently, a contrasting pattern was seen when the American and European genetic diversities were
211 compared within their respective gene pools (AM_AND *vs* EU_AND; AM_MES *vs* EU_MES). Due to
212 admixture, the European diversity was always higher than that in the Americas with the unmasked
213 dataset, while the opposite was found using the masked dataset (Fig. 4b). In other words, particularly
214 the Andean common bean from Europe shows a higher diversity than those from America, because of
215 admixed ancestry with the Mesoamerican gene pool. This comparison of the estimated levels of genetic
216 diversity in Europe confirms the key role of inter-gene pool hybridisation and recombination in shaping
217 the diversity of the European common bean. Interestingly, as compared to the Mesoamerican gene pool,
218 the Andean gene pool showed higher contrast between unmasked and masked datasets, as the diversity
219 of the Andean germplasm in the centre of origin still reflects the bottleneck that occurred in the Andean
220 wild populations during the expansion into South America before domestication (Bitocchi et al., 2012)
221 that was reflected in the domesticated pool (Bitocchi et al., 2013). Indeed, here in non-coding regions
222 we have detected reduced diversity (θ_π/bp) of approximately 70% in the Andes compared to
223 Mesoamerica in the American accessions.

224 Considering the various genetic groups, there was significantly higher diversity in the Durango-
225 Jalisco race (i.e., M1) compared to race Mesoamerica (M2) and in race Peru (A2) compared to the
226 Nueva Granada race (A1). Moreover, a very low level of diversity was found in race Chile (A3). This,
227 along with the NJ tree shown in Fig. 1d, indicates that race Peru in the Andes and race Durango-Jalisco
228 in Mesoamerica were likely the first domesticated populations from which the other races were derived
229 by secondary domestication associated with the loss of photoperiod sensitivity (Fig. 1f). Indeed,
230 earliness and loss or reduction of photoperiod sensitivity were major traits under selection during
231 common bean expansion in Europe. Considering the Andean gene pool, this was connected to the
232 domestication pattern in the centre of origin. The earlier domesticated genotypes that are sensitive to
233 the photoperiod were less successfully disseminated in Europe. Indeed, the relationship between the
234 American and European genetic groups of Andean origin (as defined using the ChromoPainter
235 approach; Lawson et al., 2012; 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 (Meyer and Purugganan, 2013) 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 (Chakraborty and Weiss, 1988). 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 (Zeven, 1997), 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; Fariello et al., 2013), 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 (Yen et al., 2017; Suen et al., 2018). 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 (Adams et al., 2015) (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 (Heo et al., 2012). 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 (King et al., 2008; Seo et al., 2011). The paralogous *GNC* and *CGA1* genes act in a redundant way to promote greening downstream from the gibberellic acid signalling network (Richter et al., 2013).

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

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

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

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

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 Resources 9:197-201 doi: doi:10.1017/S1479262111000190

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

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.

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. Proceedings of the National Academy of Science, USA 109: E788–E796

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

433 Bitocchi E, Rau D, Bellucci E, Rodriguez M, Murgia ML, Gioia T, Santo D, Nanni L, Attene G, Papa
434 R (2017) Beans (*Phaseolus* ssp.) as a model for understanding crop evolution. Front Plant Sci.
435 8:722, doi: 10.3389/fpls.2017.00722
436 Brandenburg J-T, Mary-Huard T, Rigaill G, Hearne SJ, Corti H, Joets J, et al. (2017) Independent
437 introductions and admixtures have contributed to adaptation of European maize and its American
438 counterparts. PLoS Genet 13(3): e1006666. https://doi.org/10.1371/journal. pgen.1006666
439 Chakraborty R and Weiss KM (1988) Admixture as a tool for finding linked genes and detecting that
440 difference from allelic association between loci. Proceedings of the National Academy of Sciences,
441 85 (23) 9119-9123; doi: 10.1073/pnas.85.23.9119
442 Cortinovis G, Di Vittori V, Bellucci E, Bitocchi E, Papa R (2020) Adaptation to novel environments
443 during crop diversification. Current Opinion in Plant Biology, 56:218–222, doi:
444 10.1016/j.pbi.2019.12.011
445 Crosby AW (1972) The Columbian Exchange; Biological and Cultural Consequences of 1492.
446 Westport, Conn. :Greenwood Pub. Co.
447 Fariello MI, Boitard S, Naya H, SanCristobal M, Servin B (2013) Detecting signatures of selection
448 through haplotype differentiation among hierarchically structured populations. Genetics,
449 193(3):929-41
450 Gepts P and Bliss FA (1988) Dissemination pathways of common bean (*Phaseolus vulgaris*, Fabaceae),
451 deduced from phaseolin elettrophoretic variability. II Europe and Africa. Economic botany 42, 86-
452 104. https://doi.org/10.1007/BF02859038
453 Gepts P (2002) A Comparison between Crop Domestication, Classical Plant Breeding, and Genetic
454 Engineering. Crop Sci., 42: 1780-1790. https://doi.org/10.2135/cropsci2002.1780
455 Heo JB, Sung S, Assmann SM (2012) Ca^{2+}-dependent GTPase, extra-large G protein 2 (XLG2),
456 promotes activation of DNA-binding protein related to vernalization 1 (RTV1), leading to
457 activation of floral integrator genes and early flowering in *Arabidopsis*. J. Biol. Chem.
458 287(11):8242-8253, doi: 10.1074/jbc.M111.317412

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*). Journal of Experimental Botany 59(14):3811–3820, doi: 10.1093/jxb/ern231

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

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

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

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

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 Physiology 162(4):1992-2004; doi: 10.1104/pp.113.219238

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

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

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

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

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 Breeding, 124: 464-472. https://doi.org/10.1111/j.1439-0523.2005.01137.x

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

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

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

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

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

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

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

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

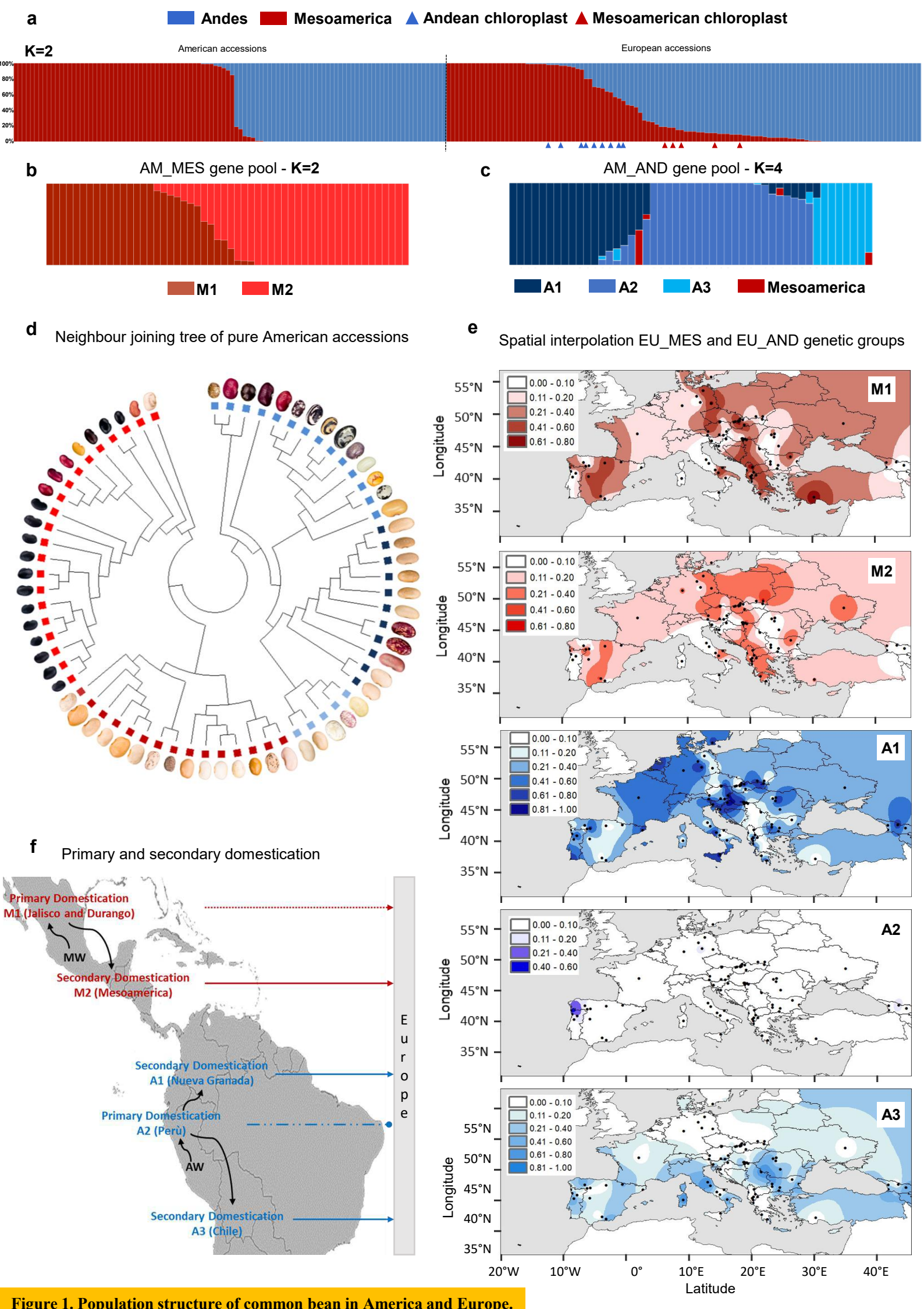
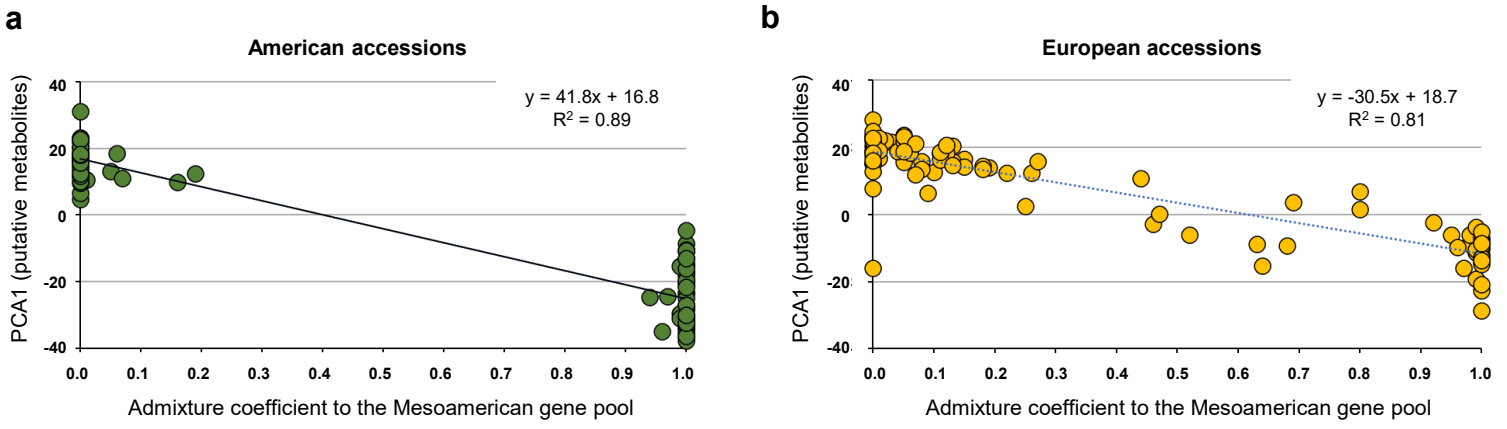
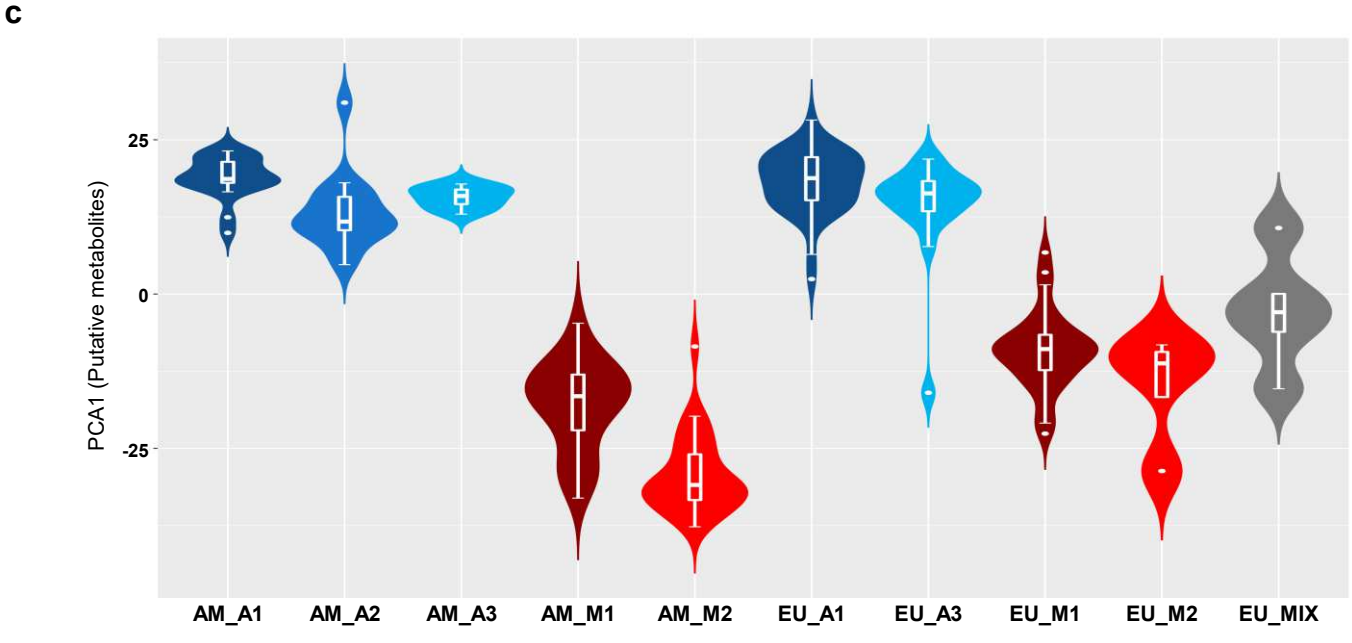


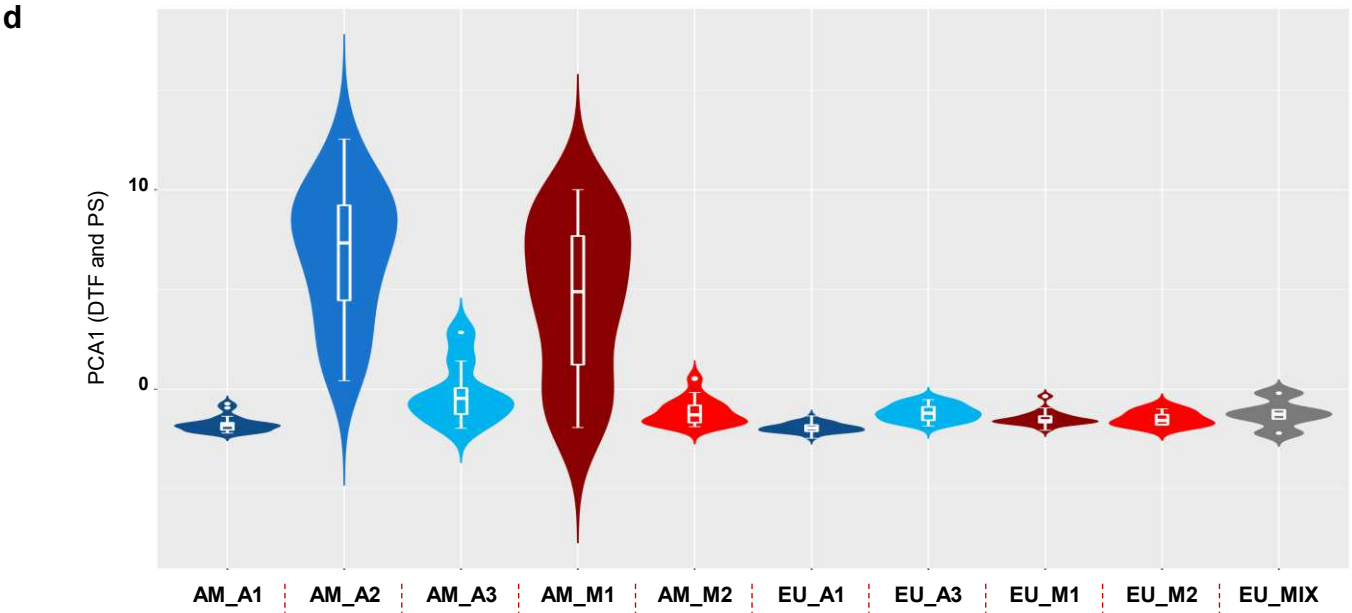
Figure 1. Population structure of common bean in America and Europe.

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

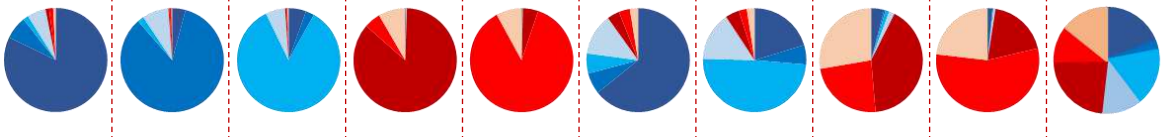
a, Admixture analysis (K=2) showing the inferred ancestry in the American (AM; left) and European (EU; right) accessions, with identification of two main gene pools, and several intermediates and admixed genotypes in Europe. **b**, Admixture plots for the AM Mesoamerican accessions (K=2) grouped by geographic 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).







e



	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.

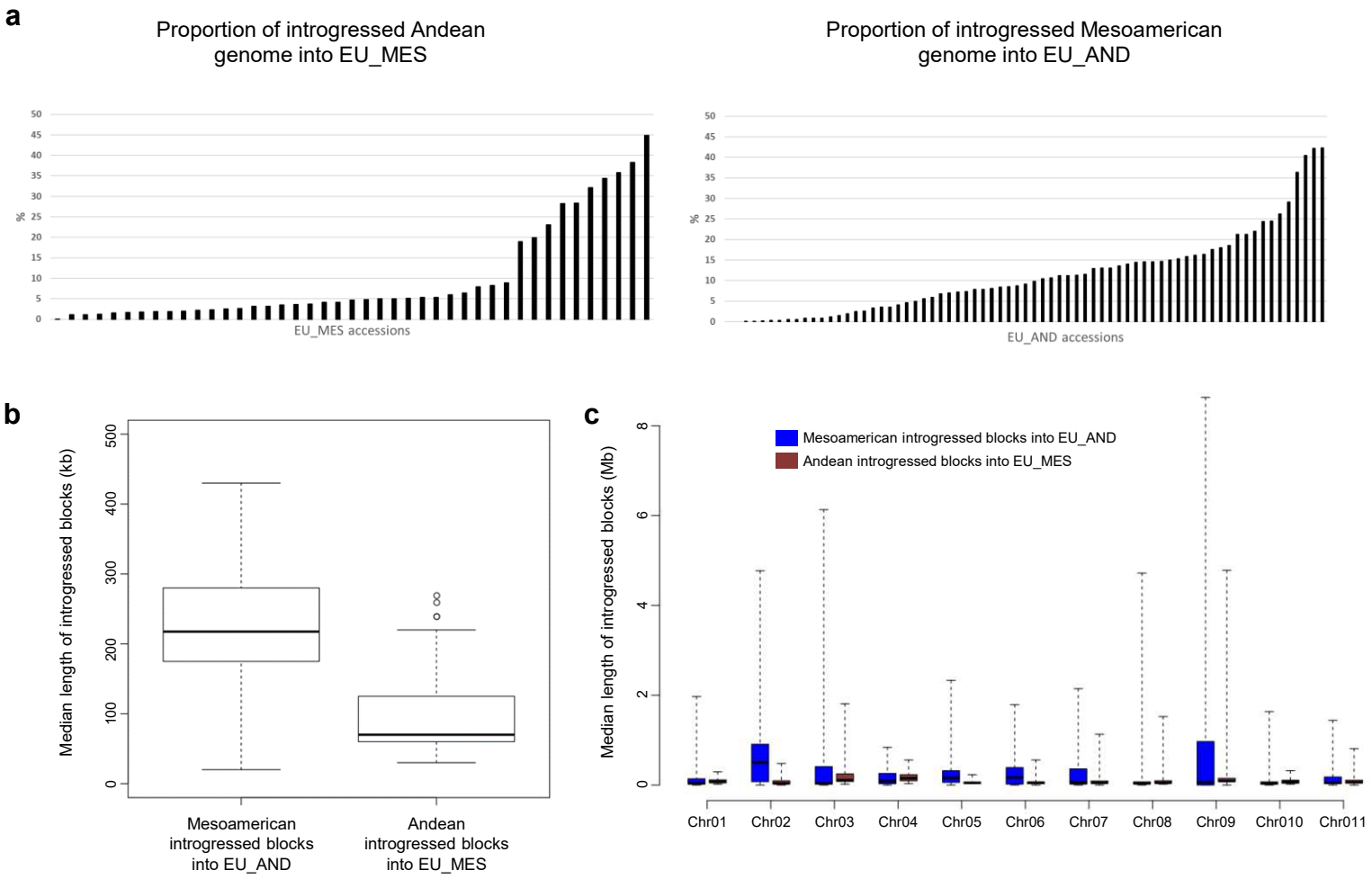


Figure 3. Mapping the introgression in the European common bean using Chromopainter.

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**).

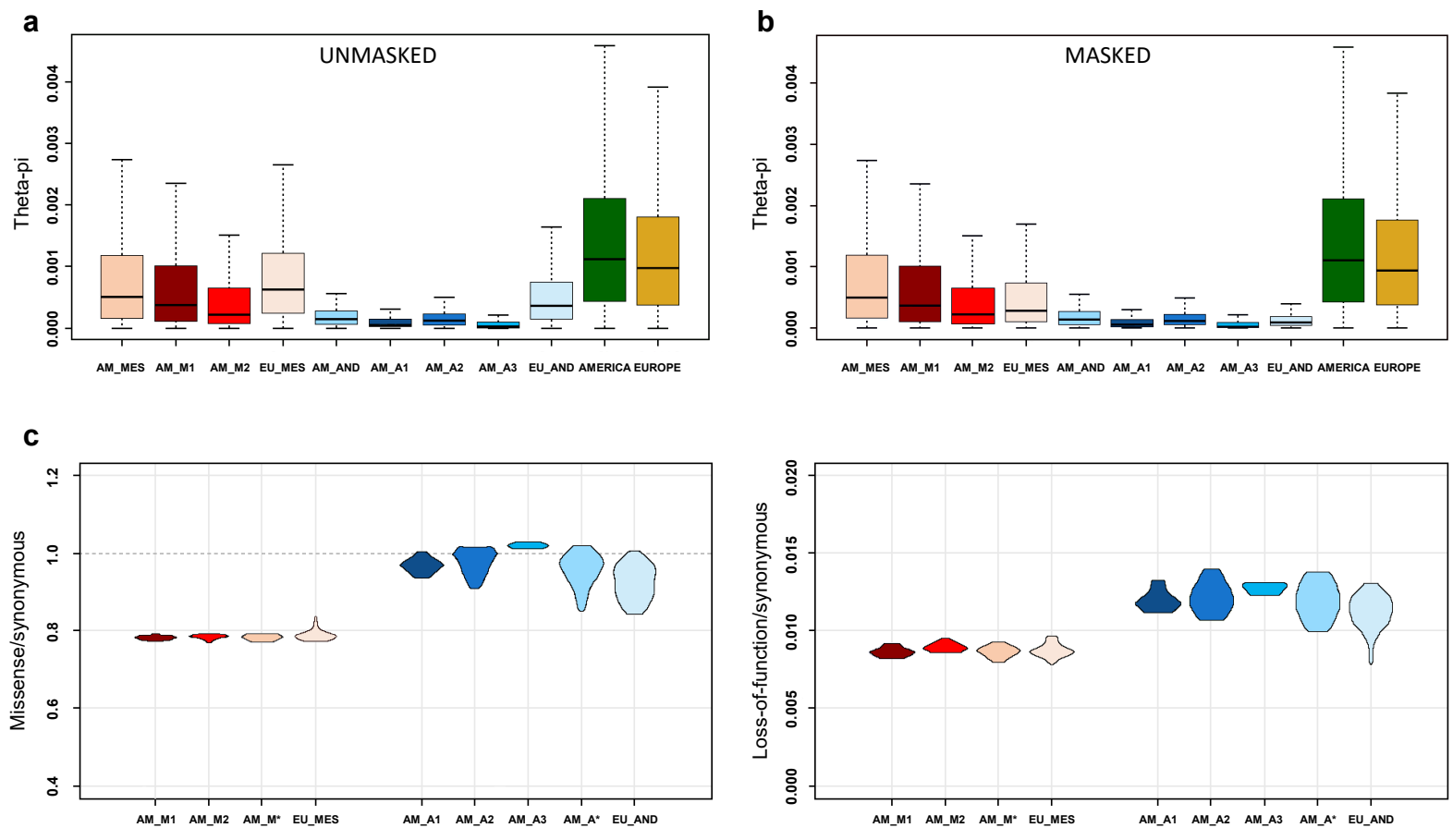


Figure 4. Boxplots of θπ averaged over 100-kb non-overlapping sliding windows, and genetic load.

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).

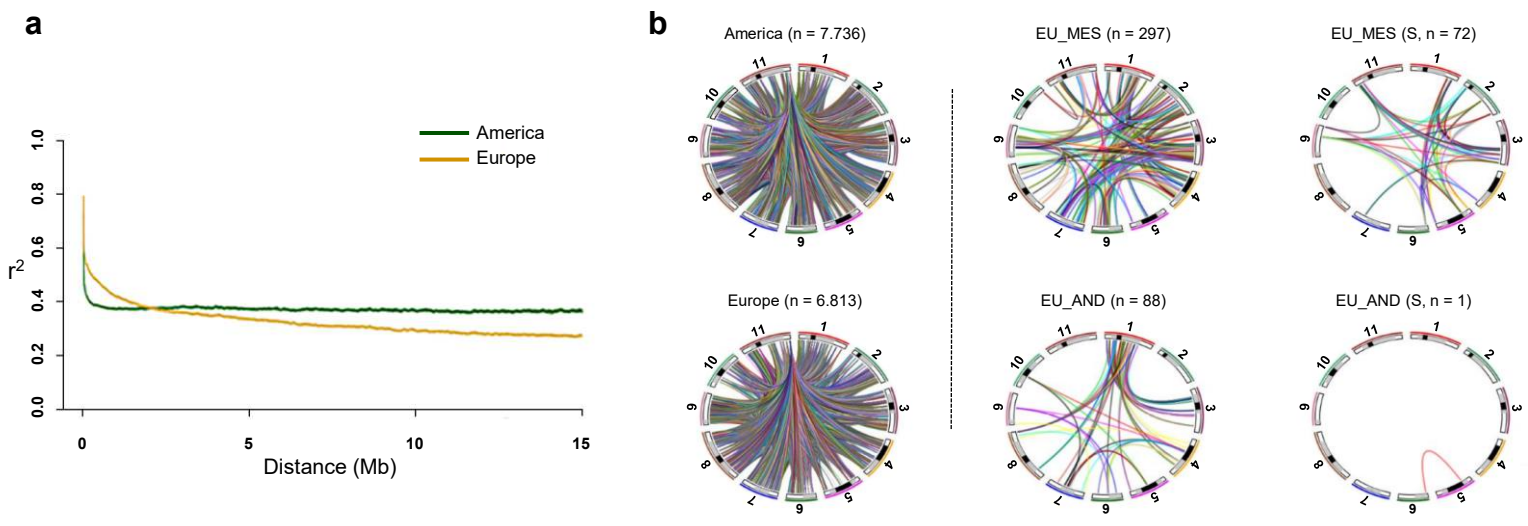


Figure 5. Linkage disequilibrium decay and inter-chromosomal linkage disequilibrium.

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).

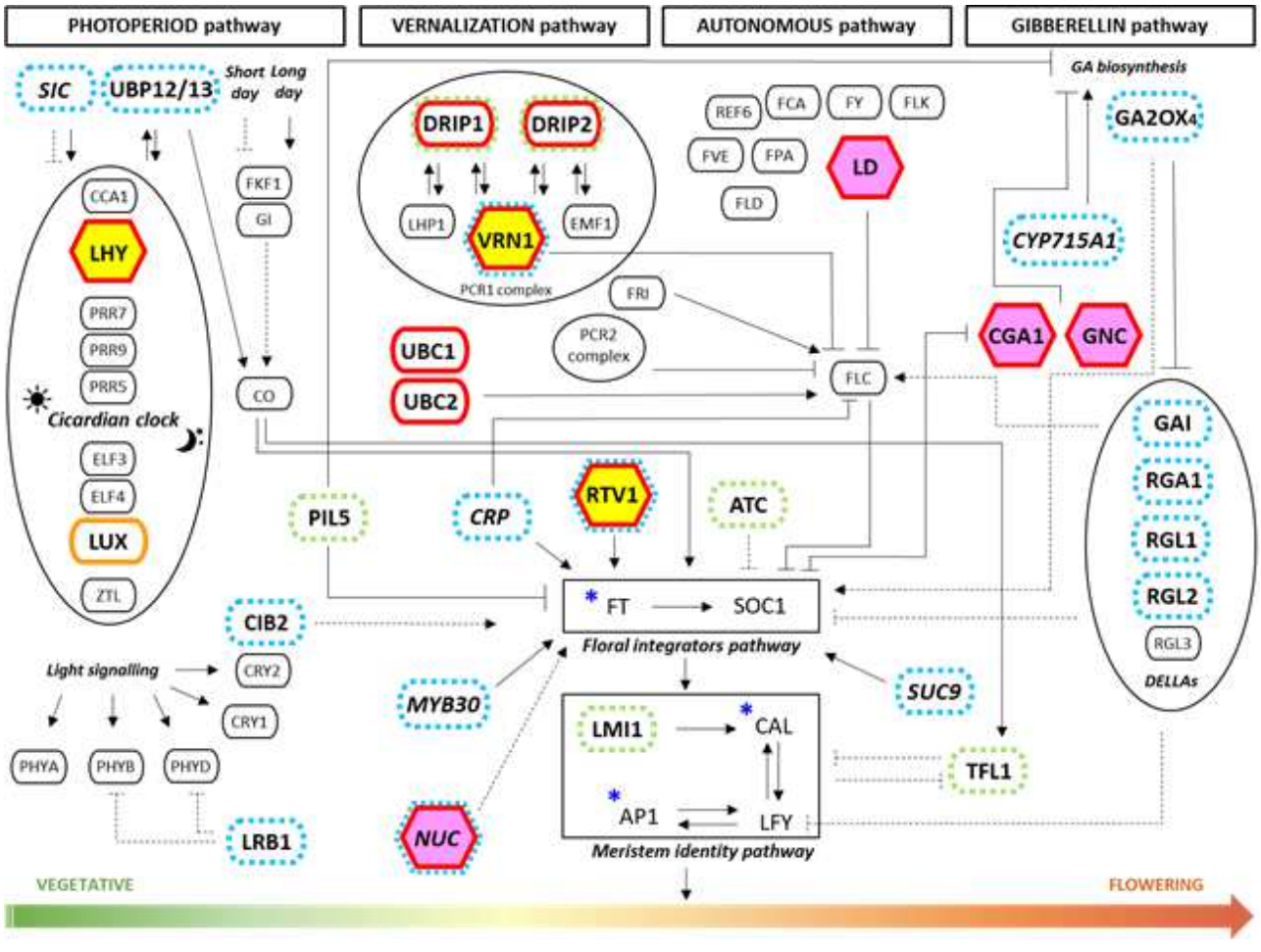


Figure 6. Candidate genes for adaptation.

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 (SAS Institute, Inc. 2012). 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 (Wei and Simko 2017; Peterson et al., 2020).

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. (2019). 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) (Schmutz et al., 2014) using BWA-mem (V0.7.15) (Li 2013). The unmapped reads were mapped to the *P. vulgaris* chloroplast genome (NCBI Entry: NC_009259). In both cases, SNP calling was performed using SAMtools (Li 2011) and GATK (V3.6) (McKenna et al., 2010; Van der Auwera et al., 2013). 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 (Danecek 2021). 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 (Danecek et al., 2011) (--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) (Cingolani et al., 2012).

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 (Browning & Browning 2007). ADMIXTURE v1.3 (Alexander et al., 2009) 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. (2009); Bitocchi et al. (2012, 2013) 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 (Corander et al., 2003, 2008). 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 (Kumar et al., 2018) using a bootstrap value of 1,000. Gaps and missing data were excluded.

Chromosome painting

The "chromosome painting" approach implemented in ChromoPainter v2.0 (Lawson et al., 2012) 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 (Quinlan et al., 2010) 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, Hijmans et al., 2005) for a total of 19 bioclimatic variables and 24 monthly variables (**SN4_ Supplemental Dataset 5**). The vegan R package (Oksanen et al., 2020) 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 ($\theta\pi$, Tajima 1983). 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 $\theta\pi$ 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 $\theta\pi$ 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 $\theta\pi$ 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 $\theta\pi$ estimate and the proportion of missing data for each position, respectively. The global within group $\theta\pi$ 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} (Weir and Cockerham 1984). 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 (Fariello et al., 2013) 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 (Scheet and Stephens 2006) 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 (Zhang et al., 2019) 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 (Zhang et al., 2013) 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 (Zhou and Stephens 2012; 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; Segura et al., 2012; 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 Di Vittori et al., 2021). 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 (Zhou et al., 2019; http://metascape.org).

References

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

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. Proceedings of the National Academy of Science, USA 109: E788–E796.

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.

Browning, S. R. & 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

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.

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

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

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.

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

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 M L, Ferreira J J, Campa A, Xu C, Fiorani F, Sampathkumar A, Fröhlich A, Attene G, Delledonne M, Usadel B, Fernie A R, Rau , Papa R, Pod indehiscence in common bean is associated with the fine regulation of PvMYB26, Journal of Experimental Botany, Volume 72, Issue 5, 27 February 2021, Pages 1617–1633, https://doi.org/10.1093/jxb/eraa553

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.

Hijmans, R. J., Cameron, S. E., Parra, J.L., Jones, P. G., Jarvis, A. (2005). Very high resolution interpolated climate surfaces for global land areas. International Journal of Climatology, 25, 1965–1978.

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

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

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.

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.

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

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 Research 20:1297-1303.

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

Peterson BG, Peter Carl P, 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

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

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

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.

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. The American Journal of Human Genetics, 78(4):629-44.

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

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. Nature genetics, 44(7), 825.

279 Tajima F (1983). Evolutionary relationship of DNA sequences in finite populations. Genetics, 105(2):437-60.
280 Van der Auwera GA, Carneiro MO, Hartl C, Poplin R, Del Angel G, Levy-Moonshine A, Jordan T, Shakir K, Roazen D,
281 Thibault J, Shakir K, Roazen D, Thibault J, Banks E, Garimella KV, Altshuler D, Gabriel S, DePristo MA (2013).
282 From FastQ data to high confidence variant calls: the Genome Analysis Toolkit best practices pipeline. Curr Protoc
283 Bioinformatics 43:11 10 11-33.
284 Wei T, Simko V (2017). R package "corrplot": Visualization of a Correlation Matrix (Version 0.84). Available from
285 https://github.com/taiyun/corrplot
286 Weir BS, Cockerham CC (1984). Estimating F-statistics for the analysis of population structure. evolution, 1358-1370.
287 Zhang H, Meltzer P, Davis S (2013). RCircos: an R package for Circos 2D track plots BMC bioinformatics, 14(1):244.
288 Zhang C, Dong SS, Xu JY, He WM, Yang TL (2019). PopLDdecay: a fast and effective tool for linkage disequilibrium
289 decay analysis based on variant call format files. Bioinformatics, Volume 35, Issue 10, 1786–1788,
290 https://doi.org/10.1093/bioinformatics/bty875
291 Zhou X, & Stephens M (2012). Genome-wide efficient mixed-model analysis for association studies. Nature genetics,
292 44(7), 821.
293 ZhouY, Zhou B, Pache L, Chang M, Khodabakhshi AH, Tanasichuk O, Benner C, Chanda SK (2019) Metascape provides
294 a biologist-oriented resource for the analysis of systems-level datasets. Nat Commun 10, 1523.
295 https://doi.org/10.1038/s41467-019-09234-6
296

Chapter V
The common bean pangenome

1. Pangenome concept

Genetic diversity represents the raw material on which adaptive selection acts, and as such, it has a fundamental role in both evolutionary history and future evolutionary pathways of a species (Barrett and Schluter, 2008). Analysis of the genetic diversity through population genomics and genotype-phenotype association approaches can be very useful tools to reach this aim (Mousavi-Derazmahalleh et al., 2019), especially now with the advent of next-generation sequencing (NGS) technologies. In the last years, reference genomes became available for several plant species (Badouin et al., 2017; Mascher et al., 2017; Varshney et al., 2017; Appels et al., 2018; Springer et al., 2018) and resequencing methodologies, wherein reads from each sample are aligned to a single reference genome, have been widely applied to capture the genetic diversity present in a species, generally in terms of single nucleotide polymorphisms (SNPs) (Cortinovis et al., 2020). With the assembly of increasing numbers of plant genomes, it is becoming accepted that a single reference does not reflect the complete genetic repertoire of a whole species (Springer et al., 2009; Saxena et al., 2014). Recent studies identified another source of diversity called structural variations (SVs), which consists of genomic variations in DNA segments of more than 1 kbp and include presence/absence variations (PAVs), copy number variations (CNVs), and other miscellaneous variations in the form of inversions, transversions, and inter/intrachromosomal translocations (Feuk et al., 2006; Cook et al., 2012; Qi et al., 2014; Wang et al., 2015). SVs are highly abundant in human genomes and their association with diseases has been established (Korbel et al., 2007; McCarroll and Altshuler 2007). The recent studies pertaining to SVs in plants have also demonstrated their importance in plant genetics as well (Muñoz-Amatriaín et al., 2013; Saxena et al., 2014). In detail, has been demonstrated the role of SVs in deciphering the phenotype and orchestrating the mechanism of defense response (McHale et al., 2012) and in driving plant adaptation to particular agro-ecological conditions (Gordon et al.,.2017). Being subject to selective pressure (Wang et al., 2021), SVs form an integral part of the evolutionary process of a given species. However, although SVs have been discovered to have a fundamental role in plants, their characterization is severely limited if a reference-centric approach is used (Tranchant-Dubreuil et al., 2019; Danilevicz et al., 2020; Golicz et al., 2016). This because mapping of the reads on the single reference genome tends to miss highly polymorphic regions and regions that are not present in the reference (Zhou et al., 2015; Varshney et al., 2017). Precise genomic relationships are only visible for those sequences that are similar enough to the reference genome to be alignable (Eizenga et al., 2020). This effect is known as reference bias: it is strongest for structural variations or sequences that are absent from the reference system, but it can be relevant even for SNPs (Hurgobin and Edwards, 2017). Pangenomic reference system can reduce this bias-effect by enabling the direct all-to-all comparisons (**Figure 1**) and the detection of all the genomic relationships that exist between

each genome and all the other accessions representative of the pangenomic system (Eizenga et al., 2020).

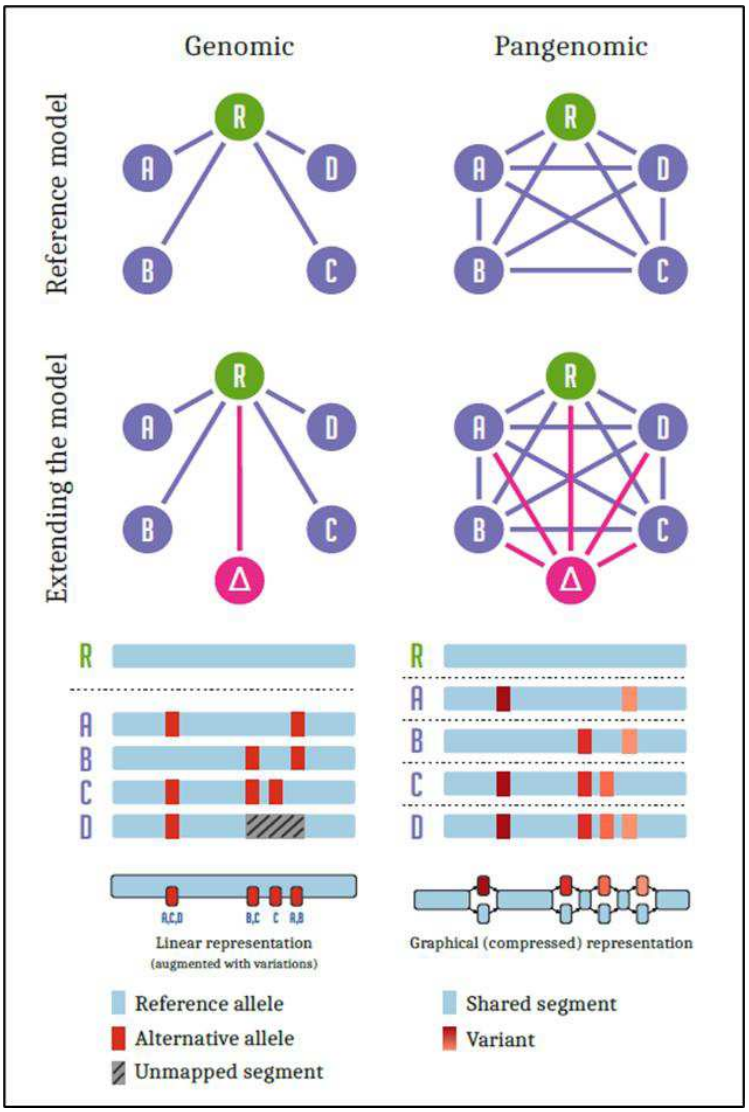


Figure 1: Pangenomic model. *Top left:* in reference based genomic analyses, all genomes (A: : :D) are compared to each other via their relationship to the reference genome R. *Top right:* in a pangenomic setting, we attempt to model direct relationships between all the genomes in our analysis, of which a particular reference R is chosen arbitrarily. *middle left:* When extending our analysis with a new genome, we add it to the genomic model by comparing it to reference R. *Middle right:* in contrast, adding a new genome to a pangenomic analysis compares it directly with all other genomes in the model. *Bottom left:* regions of some genomes are unalignable against the reference, and cannot be represented in a list of variants. *Bottom right:* a graphical model of the genomes allows direct all-to-all comparison, capturing all of their sequence relationships. *Figured adapted from Eizenga et al., 2020.*

For definition, the pangenome is the total genome architecture of a species developed by the sequencing and analysis of multiple accessions. From a bioinformatic point of view, pangenome reference combines non-redundant genome sequences into a single annotated file (Danilevicz et al., 2020). The production of pangenomes, which reflect the SVs and polymorphisms in genomes, enables more robust and in-depth comparisons of variation within species or higher taxonomic groups (Khan et al., 2019). The pangenome concept was first introduced by Tettelin

et al. (2005) for the genome analysis of multiple pathogenic isolates of bacteria and it was soon adopted in plant and animal researchers, resulting in over 20 eukaryotic pangenome studies performed to date (Golicz et al., 2016). A pangenome broadly comprises two parts: the core genome and the variable genome (**Figure 2**). The core genome is the common set of sequences shared by all individuals of a species and it is generally described as the minimal genome sequence required for a cell to live (Blaustein et al., 2019). On the other hand, the variable genome is composed of partially shared and/or nonshared DNA sequence elements and it is related to genes responsible for adaptation and survival in different agro environments, such as flowering time, diseases resistance and stress responses (Xu et al., 2006; Yu et al., 2014; Hardigan et al., 2016; Hoopes et al., 20189; Bayer et al., 2019). Plant genomes are highly dynamic, with gene-size insertions leading to the amplification of gene families and the generating of new genes through gene fusions, so the characterization of SVs in core and variable genes can provide a valuable approach to understand gene evolution (Krasileva, 2019). Pangenomic analyses open new ways to investigate and compare multiple genomes of closely related individuals at once, and more broadly new opportunities for optimizing breeding and studying evolution. This emerging concept combined with the power of the third-generation sequencing technologies gives unprecedented opportunities to uncover new genes associated to important agronomic traits, to fully explore genetic diversity, to advance knowledge about the evolutionary forces that shape genome organization and dynamics, and to increase our knowledge about evolutionary mechanisms that allow organisms to adapt quickly to new environments.

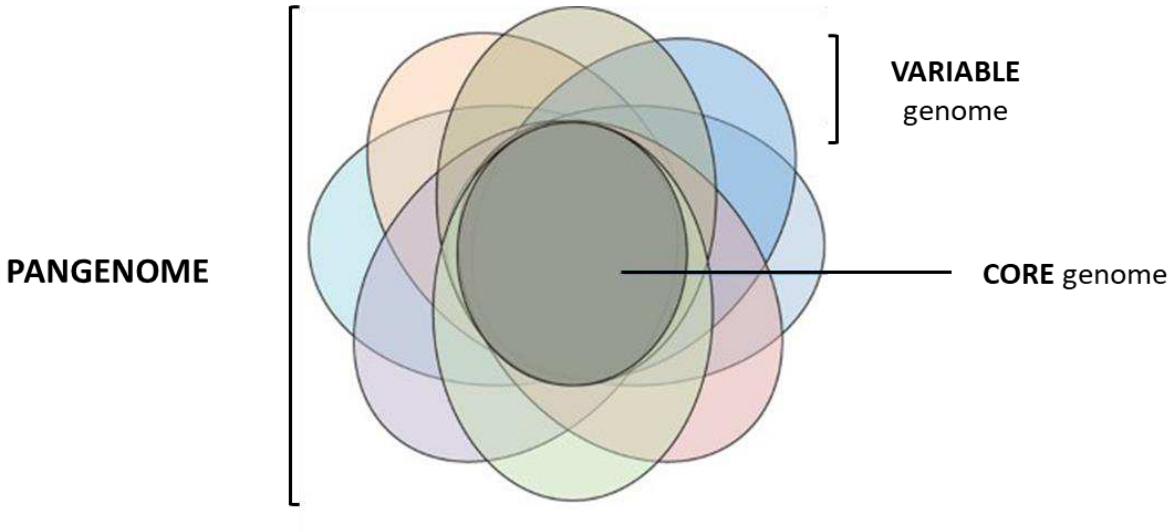


Figure 2: Schematic pangenome representation.

With the aim to better understand the genetic basis and phenotypic consequences of the parallel common bean domestications and its adaptation to novel and different agro ecosystems, we developed and analyzed the first common bean pangenome.

2. Materials and methods

The pangenome construction and PAVs calling were performed in collaboration with the University of Verona - Department of Biotechnology - Prof. Massimo Delledonne.

Starting data

The reference genome of *Phaseolus vulgaris* (Pvulgaris_442_v2.0; PV442) and those of other four accessions of the same species (i.e., MIDAS, G12873, BAT93, and JaloEPP558), belonging to both the Mesoamerican and Andean gene pools, were used. The reference genome PV442 was downloaded from Phytozome (https://phytozome-next.jgi.doe.gov/info/Pvulgaris_v2_1), the genomes of MIDAS and G12873 were already sequenced and assembled by the University of Verona, while BAT93 and JaloEPP558 have been provided by the INRA (Institut National De La Recherche Agronomique -France). In addition to the five high-quality genomes (**Table 1**), a panel of 339 low coverage WGS common accessions (wild and domesticated forms representative of the entire *P. vulgaris* species) were included in the pangenome construction.

	PV442 Andean	MIDAS Andean	G12873 Mesoamerican	BAT93 Mesoamerican	JALOEPP558 Andean
Total Assembly Size (bp)	537,218,636	509,180,482	584,993,346	637,803,808	606,487,223
Number of Contigs	478	1,913	6,293	1,441	1,061
Contigs Average Lenght (bp)	1,123,888	266,168	92,959	442,611	571,618
Contigs N50 (bp)	49,670,989	3,412,857	2,176,347	11,017,447	13,928,440
Contigs N90 (bp)	31,236,378	211,500	57,415	1,083,210	2,001,031
Longest Contigs (bp)	63,048,260	24,636,533	20,321,960	36,599,242	45,469,680

Table 1: Genome assembly statistics. The table presents the total assembly size, the total number of contigs for each genome, the contigs average length, the N50 and the N90 and the size of the longest contig for the five high-quality genomes.

Pangenome construction

The construction of the common bean pangenome was performed with a non-iterative approach. In detail, PV442 was independently aligned on each of the other four high-quality genomes with minimap2 (v.2.17) and the deletions, representing the Non-Reference Regions (NRR), have been extracted using Assemblytics (v.1.2.1). Uncovered contigs were identified using samtools depth (v1.1). Deletions and uncovered contigs were then filtered for a minimum length of 1Kb and clustered with CD-HIT-EST (v.4.8.1) with a sequence identity of 90%. Then, reads from 339 low coverage WGS of *P. vulgaris* were trimmed with fastp (v. 0.21.0) and aligned

to the pangenome using Bowtie2 (v. 2.3.5.1) with default parameters. After the alignment, the unmapped reads from each accession have been extracted with samtools (v1.10), assembled with MASURCA (v 3.4.2) with default parameters and clustered with CD-HIT-EST (v.4.8.1) with a sequence identity of 90%. As well as for the five high-quality genomes, the threshold of clustering NRRs was defined on the bases of the similarity analysis performed on orthologous and paralogous genes, in order to maintain in the pangenome only one orthologous sequence among the different accessions and, at the same time, to maintain all the paralogous. In detail, sequence similarity of the orthologous genes was calculated by aligning with minimap2 (v 2.17) MIDAS and G12873 genome assemblies on 2330 complete single copy BUSCO genes (ORFs) recovered from the annotation of PV442. The percentage of identity was calculated for each ORF by dividing the number of matches of the alignments by the reference gene ORF length. On average, MIDAS and G12873 genes shared with their corresponding orthologous in PV442 (BUSCO genes), the 99.59% and 98.71% of sequence identity, respectively. Sequence similarity between paralogous copies was carried out on three gene families (OG0000273, OG0000328 and OG0000085) of PV442 including respectively 26, 37 and 62 genes. Each gene within the same family was used as reference and all the other paralogous copies were aligned against it with minimap2 (v 2.17) and the percentage of ORF identity was calculated for each gene family by dividing the number of matches of the alignments by the reference gene ORF length. On average, gene members belonging to the same family shared a percentage of identity of 44%, 62% and 60% in the three families tested. Considered the results obtained by the identity analysis of orthologous and paralogous genes, a threshold of 90% of gene identity was selected for NRR clustering. After the clustering, NRR sequences were compared with NCBI non-redundant nucleotide databases using BLASTn, considering a minimum identity and coverage of 80% and 25% respectively, to remove the contigs matching with organelles and contaminants.

Pangenome annotation

Repetitive sequences were identified and softmasked using RepeatMasker v 4.1.2-p1. In order to use Augustus v 3.3.3 for ab initio gene prediction, the extrinsic evidence have been aligned on the pangenome using Genome Threader v1.7.1. Proteins of *P. Vulgaris, M.truncatula* e *G.soja* were considered as external evidences to annotate introns and CDS regions. In addition, RNA-seq evidences between domesticated and wild accessions (unpublished data and Bellucci et al., 2014) were aligned on the pangenome using Hisat2 v.2.2.1 and converted into intronic hints with bam2hints v.3.2.1 to maintain only those with an intronic coverage higher than 20. The training of the gene predictor of Augustus was performed with BUSCO genes obtained by "Fabales_odb10" database. The next step has consisted in concatenating all the obtained predictions and extracting the predicted proteins. Subsequently, the functional annotation was performed scanning predicted genes with InterProscan v 5.46.81.0 to detect presence of protein domains and then filtering 'repeated element'-related domains. The filtered proteins were blasted against the pangenome with BLASTp v 2.12.0 and filtered by the best hits. The clustering of the predicted genes was performed with the proteins of all the species considered in the annotation using OrthoFinder v 2.5.4 and the functional annotation results were obtained through a custom script.

Subsequently, regions of PV442 corresponding to each gene of UNIVR annotation were compared against PV442 assembly using BLASTn v 2.9 and best hits were extracted. BLAST coordinates of PV442 were then intersected with the gff file of the official annotation provided by Phytozome (https://phytozome-next.jgi.doe.gov/) with bedtools intersect (v.2.29.2). The correspondence between gene IDs of the two annotations was then inferred.

PAV calling

PAV calling was carried out by align the reads from the 339 accessions on the new reference-pangenome. PAV threshold was defined on the bases of coverage values of 1000 BUSCO genes (ORFs). BUSCO genes are orthologous genes that should be present in all the accessions considered, but to avoid the bias that a few poorly represented genes in the reads of an accession may introduce, the 1% lower values (the 10 less covered genes) were discarded. PAV presence absence was performed on genes annotated on PV442 according to the official annotation available on Phytozome website (https://phytozome-next.jgi.doe.gov/). The average depth of coverage was calculated for each accession using Bedtools genomecov (v 2.29.2). The presence/absence of the annotated genes in the different accessions was called using samtools (v. 1.11) coverage. Based on the "MIN" threshold, the number of genes present in each accession was calculated. According to their gene presence frequencies, PAVs have been categorized as core or variable.

PAV analysis

Here, we used the 14,074 variable genes for performing population genomics analyses within a representative subset of 97 common bean accessions out of a total of 339 that have been used for both the pangenome construction and PAV calling. The panel considered in the present investigation is composed of 31 wild forms (i.e., 16 Mesoamerican and 15 Andean) and 66 pure domesticated accessions (i.e., 30 Andean and 36 Mesoamerican) (**Figure 3**)**.** In detail, the panel of 66 domesticated pures accessions was the material of the Bean Adapt Project (Chapter IV of this thesis) and so, well-characterized genetically and phenotypically (i.e., day to flowering and growth habit).

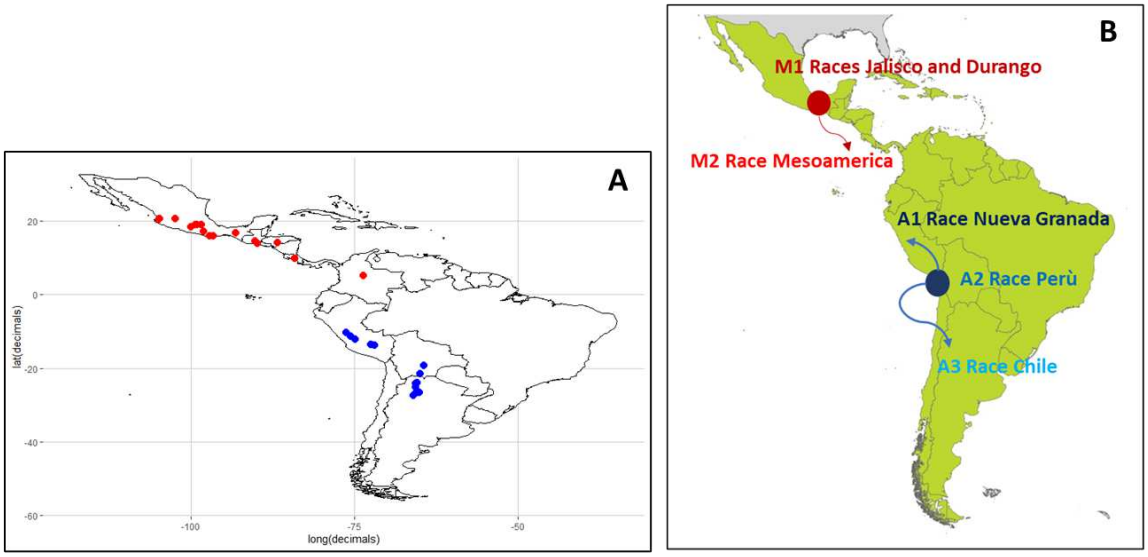


Figure 3: Geographic distribution. **A)** Geo-plot of the 31 wild common bean accessions. **B)** Schematic representation of the geographic distribution of the 66 domesticated pures accessions: 15 accessions belong to the M1 subgroup (race Jalisco-Durango), 21 accessions belong to the M2 subgroup (race Mesoamerica), 11 accessions belong to the A1 subgroup (race Nueva Granada), 14 accessions belong to the A2 subgroup (race Peru), and 5 accessions belong to the A3 subgroup (race Chile).

Principal Component Analysis (PCA)

PCA is a mathematical algorithm that reduces the dimensionality of the data while retaining most of the variation in the data set (Jolliffe et al., 2002). It accomplishes this reduction by identifying directions, called principal components, along which the variation in the data is maximal. By using a few components, each sample can be represented by relatively few numbers instead of by values for thousands of variables (e.g., pangenome data). Samples can then be plotted, making it possible to visually assess similarities and differences between accessions and determine whether accessions can be grouped (Ringnér et al., 2008). Here, PAV-based PCA was conducted analyzing the PAV matrix (14,074 gene PAVs) through the logisticPCA R Package (Landgraf and Lee, 2015).

Neighbor-Joining (NJ) Phylogenetic Tree

The PAV matrix (14,074 binary data) was used to calculate the p-distance matrix in order to determine the genetic distance among individuals by using the R *"ape"* Package. Based on the Jaccard distance and 10,000 bootstraps resampling, the NJ-tree was constructed by using the R *"NJ"* function and visualized with Figtree - http://tree.bio.ed.ac.uk/software/figtree/.

Fst Analysis (Wright's Fixation Index)

Fst is a measure of the genetic differentiation between populations varying between zero (i.e., no difference) and one (i.e., a fixed genetic difference). It involves comparing how similar two individuals from the same subpopulation are to the total population, thus providing a measure of the amount of genetic variance that can be explained by the population structure. The 14,074 variable genes were analyzed separately in the Mesoamerican and Andean gene pools, comparing the PAV frequency of each gene between the wild and domesticated forms. The monomorphic PAVs between wild and domesticated accessions were filtered and the final datasets consisted of 9,458 Mesoamerican PAVs and 8,152 Andean PAVs. Each PAVs were considered as a single locus (0/1) and Fst analysis was performed by applying the following formula for each gene PAVs: Fst = (H total - H within) / H total, where H is the heterozygosity. Negative Fst values have been converted into 0. The normality condition of the data was verified based on visual inspections (i.e., histogram and QQ-plot) and formal examination (i.e., the Anderson-Darling normality test). Only PAVs being in the top 5% of the Fst distribution between wild and domesticated forms (i.e., Fst ≥ 0.30) were considered as putatively under selection.

Gene Ontology (GO) Enrichment Analysis

All the 14,074 variable genes and those genes with high-Fst (i.e., Fst ≥ 0.30) between wild and domesticated common bean accessions were considered and subjected to GO term enrichment analysis by using the enrichment analysis available on the Metascape tool (Zhou et al., 2019).

The identification of the orthologous genes in *A. thaliana* was performed using the Orthofinder tool (Emms and Kelly, 2019) and comparing the entire protein sequences from *P. vulgaris* (v2.1) and *A. thaliana* (TAIR10).

Manual Gene Function Investigation

All the *P. vulgaris* gene PAVs with high-Fst between wild and domesticated forms and for which we had the orthologous in *A. thaliana* were subjected to manual gene function investigation.

3. Results

Pangenome development and PAVs discovery

We constructed the first common bean pangenome using a non-iterative approach. Overall, five high-quality genomes (i.e., PV442, MIDAS, G12873, BAT93, and JaloEPP558) plus a set of 339 low coverage WGS data were used in the pangenome development. Once the common bean pangenome was constructed, we performed the functional annotation and the PAV calling as described in the materials and methods. We also examined the coverage of each gene for each individual and categorized the gene PAVs according to their gene presence frequencies. The common bean pangenome, including reference (PV442) and non-reference sequences (NRRs), consists of ~779.99 Mb and contains 35,016 predicted protein-coding genes. In detail, the common bean pangenome identified approximately 242.78 Mb of novel sequences that were absent from the reference genome and that encode an additional 7,583 genes (**Table2** and **Table 3**). Of these, the 8% of non-reference genes (598 genes) are shared by all the individuals (i.e., core genes), while the 92% of non-reference genes (6,985 genes) are partially shared and/or nonshared among the individuals (i.e., variable genes). Overall, in the common bean pangenome, a total number of 20,942 genes (60%) were called as present in all the accessions, representing the core genes, while the variable genes were 14,074, representing 40% of the total annotated genes (**Table 3**).

	Reference - PV442 -	**DELETIONS**	**UNCOVERED CONTIGS**	**ASSEMBLED ACCESSIONS**	**PANGENOME**
SUM (bp)	537,218,636	36,444,092	68,815,641	137,515,934	**779,994,303**
N50 (bp)	49,670,989	11,367	36,315	2,874	**40,923,498**
N CONTIG (bp)	478	5,182	3,831	55,161	**64,652**
MEAN CONTIG LENGHT (bp)	1,123,888	7,033	17,963	2,493	**12,065**

Table 2: The table shows the statistics of clustered NRR and final dimension of the total pangenome (PV442 + NRR), subdivided between NRR extracted as deletions, NRR extracted from the uncovered contigs, and NRR extracted from the unmapped reads of the accessions assembled. The sum of the length of the regions, the N50, the number of contigs, and the mean length of the contigs are reported.

		PAVs calling MIN TRESHOLD	
	N° of annotated genes	**CORE genes**	**VARIABLE genes**
PV442 - REFERENCE -	27,433	20,344 **74%**	7,089 **26%**
NRRs	7,583	598 **8%**	6,985 **92%**
TOT.	35,016	20,942 **60%**	14,074 **40%**

Table 3: Number of annotated genes and proportion of the core and variable genes identified in the common bean pangenome

Investigation of the common bean evolutionary history

The 14,074 variable gene PAVs in a representative subset of 97 common bean accessions (**Figure 4**) were used for performing population genomics analyses. The 97 accessions were grouped by gene pool (i.e., Andean and Mesoamerican), biological status (i.e., wild and domesticated), and subgroup (i.e., A1, A2, A3, M1, or M2). Visually, we observed a broad gene PAV distribution within different genetic groups, with substantial variation between the Mesoamerican and Andean gene pools (**Figure 4)**.

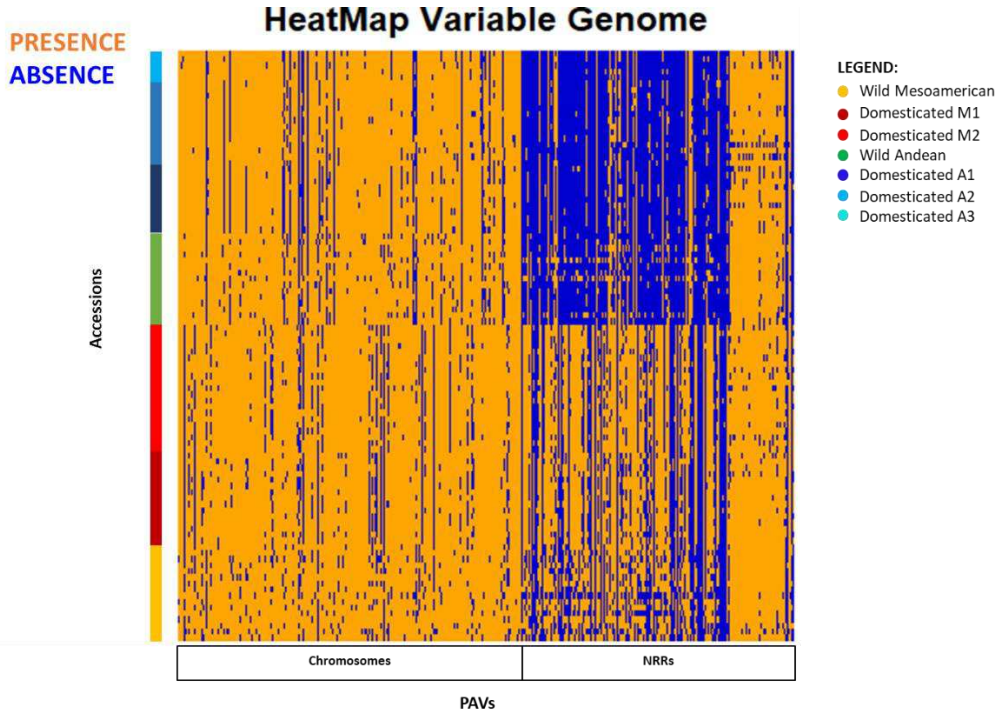


Figure 4: Heatmap. Gene presence and absence variations (gene PAVs) along the common bean pangenome (PV442 + NRRs) for the 97 common bean accessions.

When we examined the number of total gene PAVs per individual, we observed that accessions coming from the same gene pool clustered together in separate groups, with the Mesoamerican accessions showing a higher number of variants compared to the Andean accessions (**Figure 5**, **Figure 6**, and **Table 4**). The lower number of genes statistically significant in the Andean gene pool compared to the Mesoamerican one could reflect the different evolutionary trajectory of these two populations. Indeed, the Andean domestication arose from wild germplasm that, in contrast to the Mesoamerican, was highly impoverished in genetic variability as a result of the bottleneck that have occurred in the Andes before domestication (Bitocchi et al., 2013). Thus, this result suggests that the sequential bottlenecks led to a reduction in the genetic diversity in terms of structural variants in the Andean gene pool, resulting to have fewer gene PAVs than in the Mesoamerican gene pool. Statistically significant differences in genes number per individuals were registered also between wild and domesticated Mesoamerican accessions, with the wild forms showing a lower number of gene PAVs compared to both the domesticated M1 (p-value 2.90E-05) and M2 (p-value 7.10E-05) subgroups (**Figure 6** and **Table 4**). On the other

hand, no significant differences in genes number were registered between wild and domesticated Andean accessions. With regard to the domesticated accessions only, for the Andean gene pool, the subgroup A1 is statistically different (p-value 0.03) from the subgroup A2. No differences have been detected between the Mesoamerican domesticated subgroups (**Figure 6** and **Table 4**).

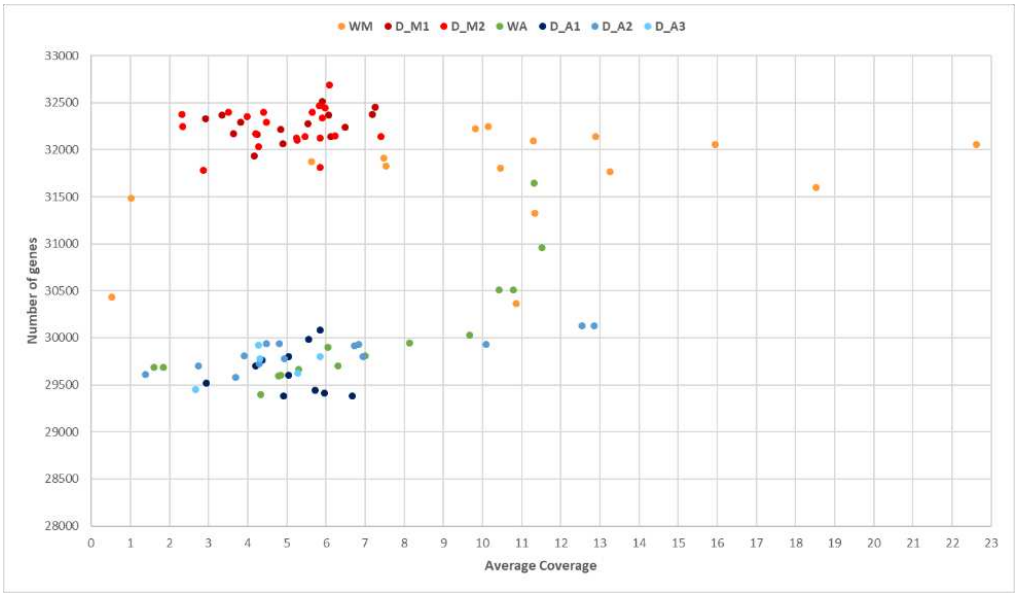


Figure 5: Average coverage vs total number of present genes per each individuals. The x axis shows the average coverage, while the y axis is the total number of present genes. The plot represents the 97 common bean accessions (i.e., 31 wild and 66 domesticated pures). The accessions are colored for gene pool (i.e., M=Mesoamerican; A=Andean), biological status (i.e., W=Wild; D=Domesticated), and subgroup (i.e., M1=race Durango-Jalisco; M2=race Mesoamerica; A1=race Nueva Granada; A2=race Peru; A3=race Chile).

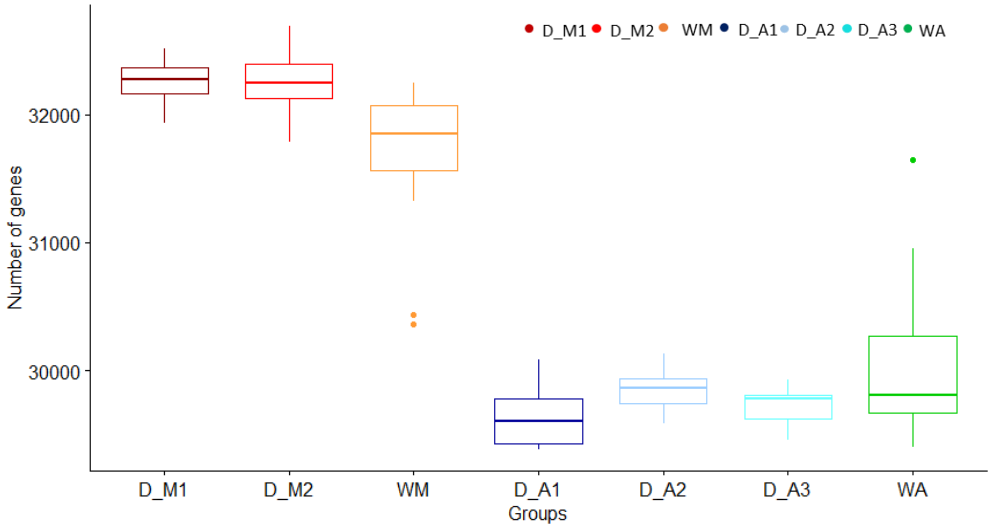


Figure 6: Box plot total number of present genes per individuals. The x axis shows the accessions grouped for gene pool (i.e., M=Mesoamerican; A=Andean), biological status (i.e., W=Wild; D=Domesticated), and subgroup (i.e., M1=race Durango-Jalisco; M2=race Mesoamerica; A1=race Nueva Granada; A2=race Peru; A3=race Chile). The y axis is the total number of present genes per groups. The box-plot represents 97 common bean accessions (i.e., 31 wild and 66 domesticated pures).

Pairwise Wilcoxon Test + p-value adjustment method: Benjamini-Hochberg						
	D_A1	**D_A2**	**D_A3**	**D_M1**	**D_M2**	**WA**
D_A2	0.03448	-	-	-	-	-
D_A3	0.43643	0.22968	-	-	-	-
D_M1	6.80E-07	9.00E-08	0.00019	-	-	-
D_M2	6.50E-08	9.10E-09	5.80E-05	0.86577	-	-
WA	0.08579	1	0.43643	6.50E-08	7.50E-09	-
WM	4.60E-07	6.50E-08	0.00016	2.90E-05	7.10E-05	3.00E-06

Table 4: Statistics related to the number of genes per individuals. The table shows the results of the Pairwise Wilcox test corrected by the Benjamini-Hochberg method.

In addition to the number of genes per accession, the number of genes per genetic group was calculated (**Figure 7 and Table 5**). As expected, since the domestication process is usually associated with a reduction in the genetic diversity, we detected that for both the Mesoamerican and Andean gene pools the wild forms have a higher number of gene PAVs compared to the domesticated accessions. Moreover, we observed that the Mesoamerican subgroup M1 and the Andean subgroup A2 have a higher number of genes than the other subgroups belonging to the same gene pool. This result is in line with our previous hypothesis (Chapter IV of this thesis) according to which, the subgroups M1 and A2 experienced primary domestications, while the other subgroups undergone secondary domestications from M1 and A2, thus resulting with a lower level of diversity in terms of PAVs.

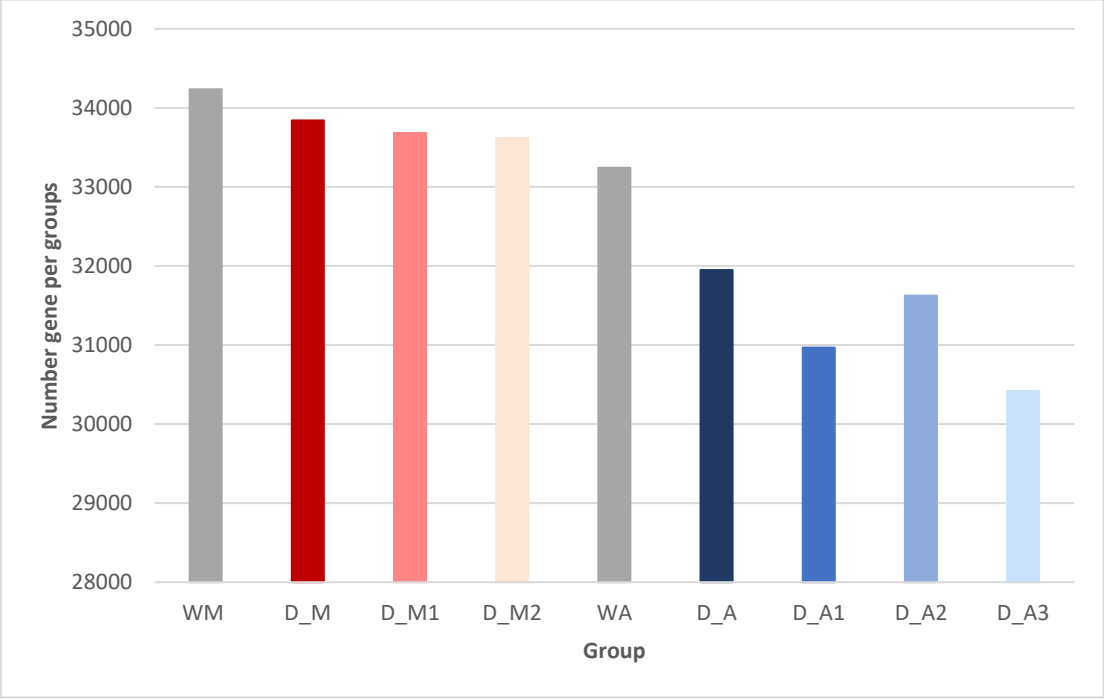


Figure 7: Histograms. Total number of present genes per groups.

Pairwise Chi-square test + p.value adj method -BH-							
	WM	D_M1	D_M2	WA	D_A1	D_A2	D_A3
WM	1.00E+00	5.17E-33	1.16E-39	1.47E-87	0.00E+00	0.00E+00	0.00E+00
D_M1	5.17E-33	1.00E+00	4.69E-01	8.41E-15	2.075076e-32	5.19E-209	0.00E+00
D_M2	1.16E-39	4.69E-01	1.00E+00	6.96E-11	1.99E-304	4.47E-194	0.00E+00
WA	1.47E-87	8.41E-15	6.96E-11	1.00E+00	6.21E-211	3.32E-119	1.54E-299
D_A1	0.00E+00	2.075076e-32	1.99E-304	6.21E-211	1.00E+00	8.41E-15	2.26E-09
D_A2	0.00E+00	5.19E-209	4.47E-194	3.32E-119	8.41E-15	1.00E+00	2.50E-45
D_A3	0.00E+00	0.00E+00	0.00E+00	1.54E-299	2.26E-09	2.50E-45	1.00E+00

Table 5: Statistics related to the number of genes per group. The table shows the results of the Pairwise Wilcox test corrected by the Benjamini-Hochberg method.

PAV-based PCA (**Figure 8**) and PAV-based phylogenetic analysis (**Figure 9**) confirmed the population structure of the *P. vulgaris* species (Chapter IV of this thesis). In detail, PAV-based PCA (**Figure 8**) showed that the first (i.e., PC1) and the second (i.e., PC2) components explained cumulatively the 47% of the total PAVs diversity across different groups and subgroups. In particular, we observed that the PC1 explains mainly the genetic diversity that exist among the Mesoamerican and Andean gene pools, whereas the PC2 splits the groups and subgroups within each gene pools.

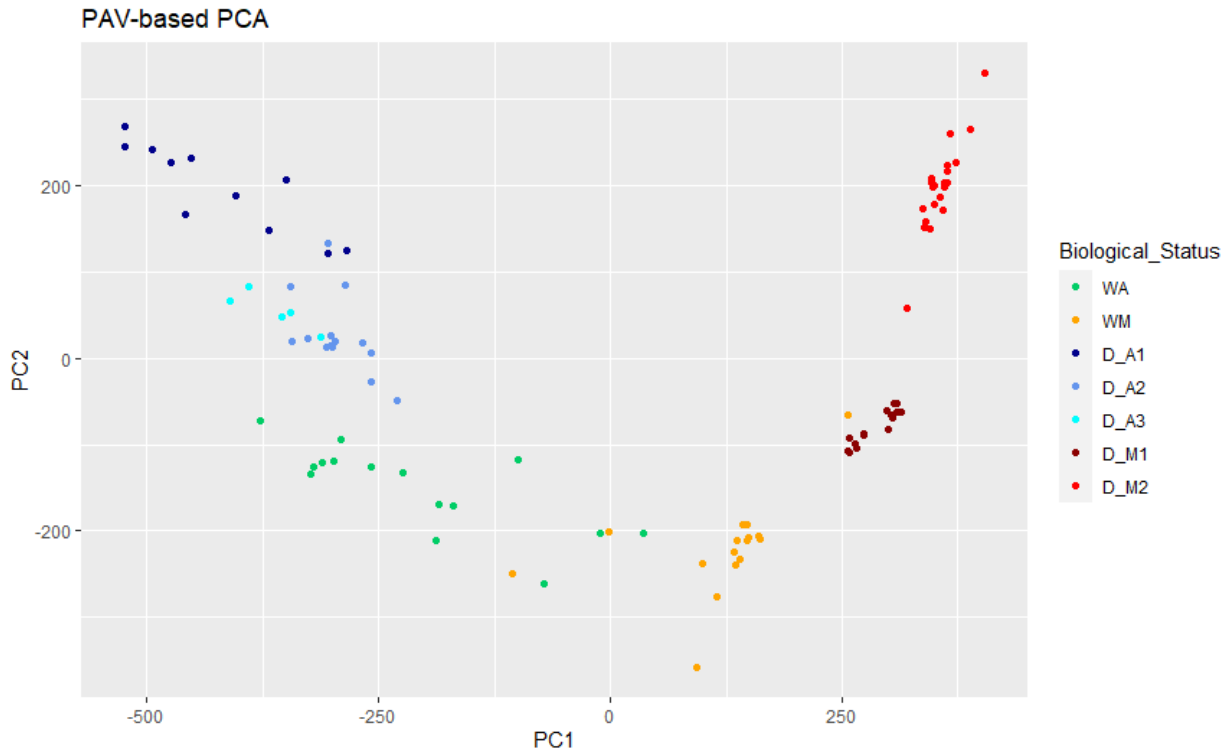


Figure 8: PAV-based PCA. The biplots summarize the distribution of the common bean accessions in the space of the PC1 and PC2. The amount of cumulative variance explained by the two first components is the 47%.

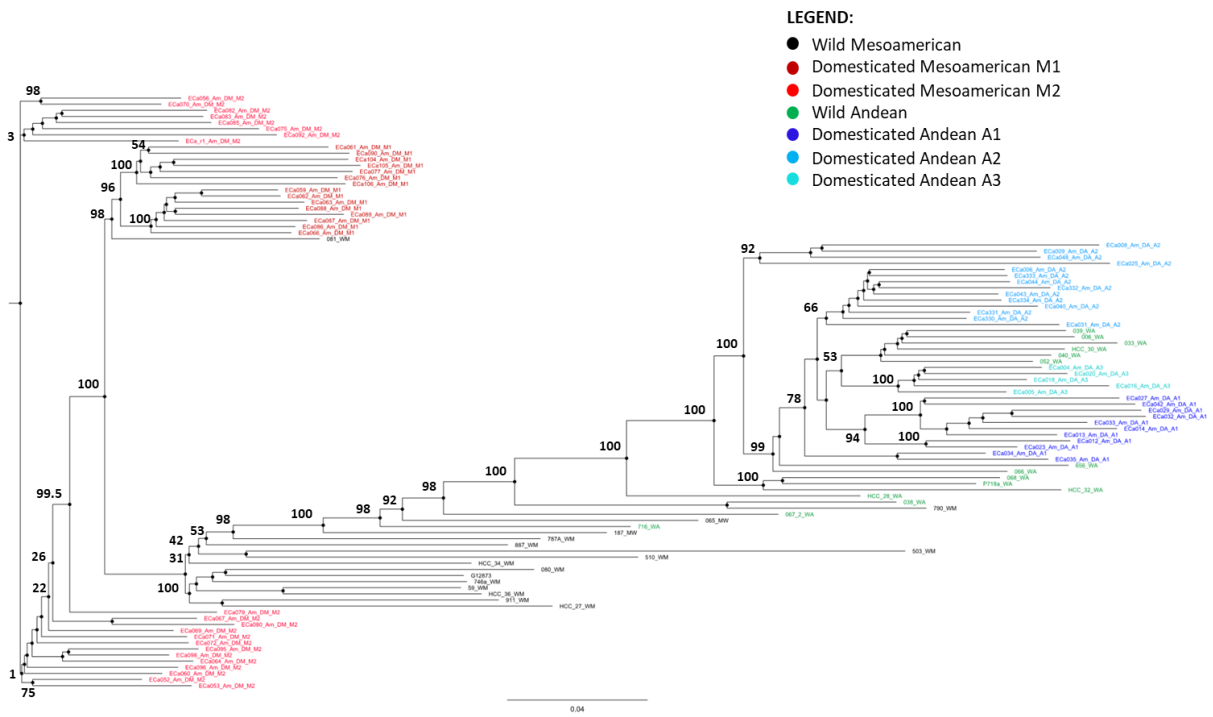


Figure 9: Gene PAV-based phylogenetic tree of the 97 common bean accessions (wild and domesticated). The phylogenetic tree shows the genetic relationship among the 97 common bean accessions. **Black:** wild Mesoamerican accessions; **dark red:** domesticated Mesoamerican accessions belonging to the subgroup M1; **red:** domesticated Mesoamerican belonging to the subgroup M2; **green:** wild Andean accessions; **dark blue:** domesticated Andean accessions belonging to the subgroup A1; **blue:** domesticated Andean accessions belonging to the subgroup A2; **light blue:** domesticated Andean accessions belonging to the subgroup A3.

In order to detect gene PAVs putatively under selection during common bean domestications, the 14,074 variable genes were subjected to Fst analysis between the wild and domesticated forms. Globally, 626 and 380 gene PAVs have been detected as putatively under selection in the Mesoamerican and Andean gene pools, respectively. Of these genes, 284 Mesoamerican and 186 Andean map on the reference genome (PV442), while the rest (i.e., 342 Mesoamerican and 194 Andean genes) is located on the NRR sequences of the common bean pangenome (**Figure 10**). Overall, the Mesoamerican and the Andean gene pools have in common 48 PAVs with high Fst values (Fst≥0.30). By comparing their frequencies between wild and domesticated forms for each gene pool, we found that the gene PAV-48 is fixed (frequency=1) and nearly fixed (frequency=0.92) in the wild and domesticated Mesoamerican accessions, respectively. PAV-48 has a significantly high frequency (frequency=0.90) also in the domesticated Andean accessions; however, it is nearly absent in the wild Andean accessions (frequency=0.27). Except for the gene PAV-48, all the other 47 genes showed frequency changes in the same direction in the two main gene pools. In detail, this means that in both the Mesoamerican and Andean gene pools PAV-3, PAV-8, PAV-22, and PAV-46 have a higher frequency in the domesticated than in the wild forms, while the remaining 43 gene PAVs show a higher frequency in the wild compared to the domesticated accessions (**Figure 11**). These preliminary results suggest the possibility that convergent evolution phenomena have occurred within the Mesoamerican and Andean gene pools during their parallel and independent domestications, Further analyses are needed to clarify this issue.

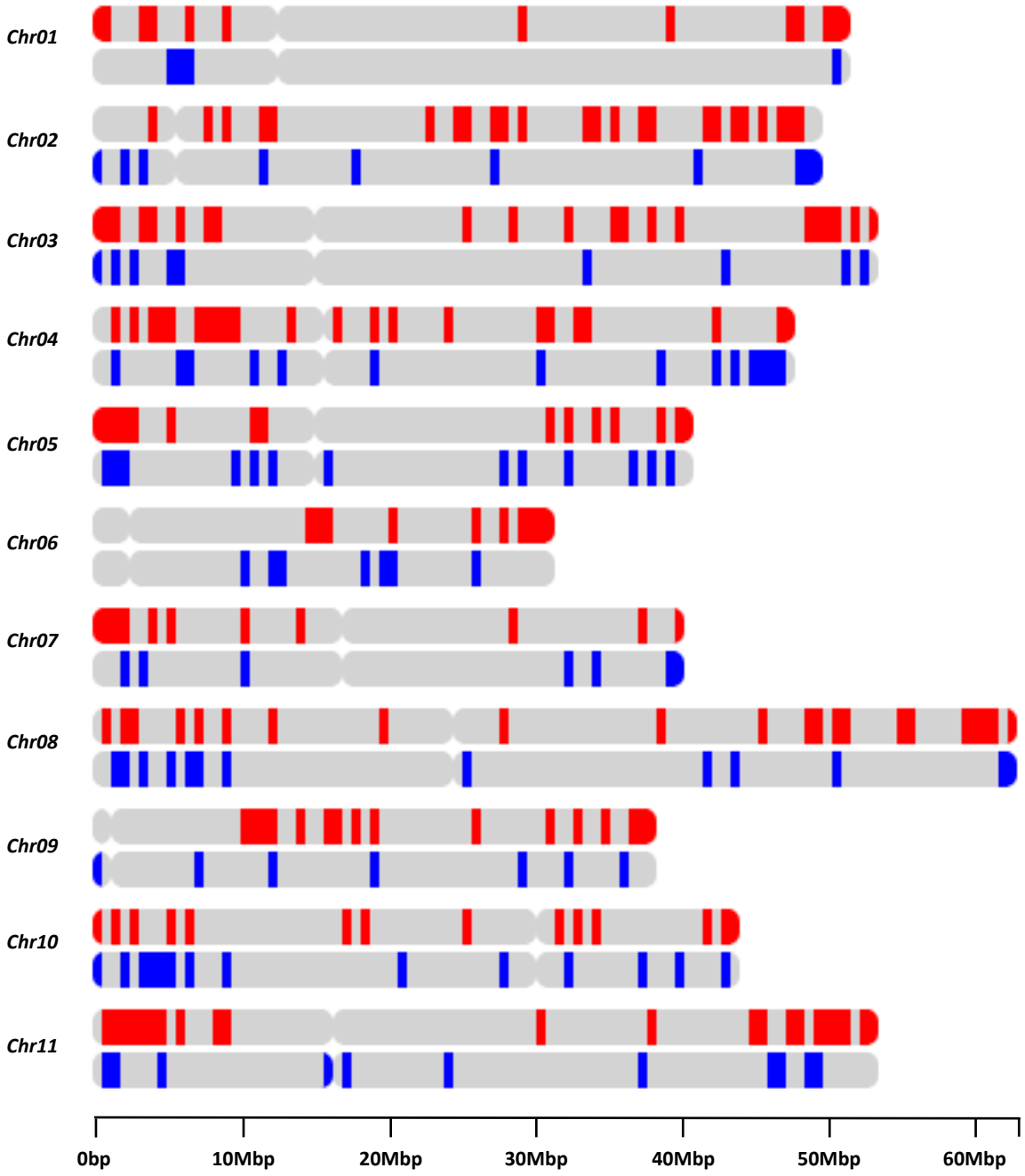


Figure 10: Genome-wide data on ideograms. The genome-wide plot shows the variable PAVs with high-Fst along the 11 chromosomes of the common bean reference genome (PV442). **Red**: 284 Mesoamerican gene PAVs with Fst≥0.30; **Blue**: 186 Andean gene PAVs with Fst≥0.30.

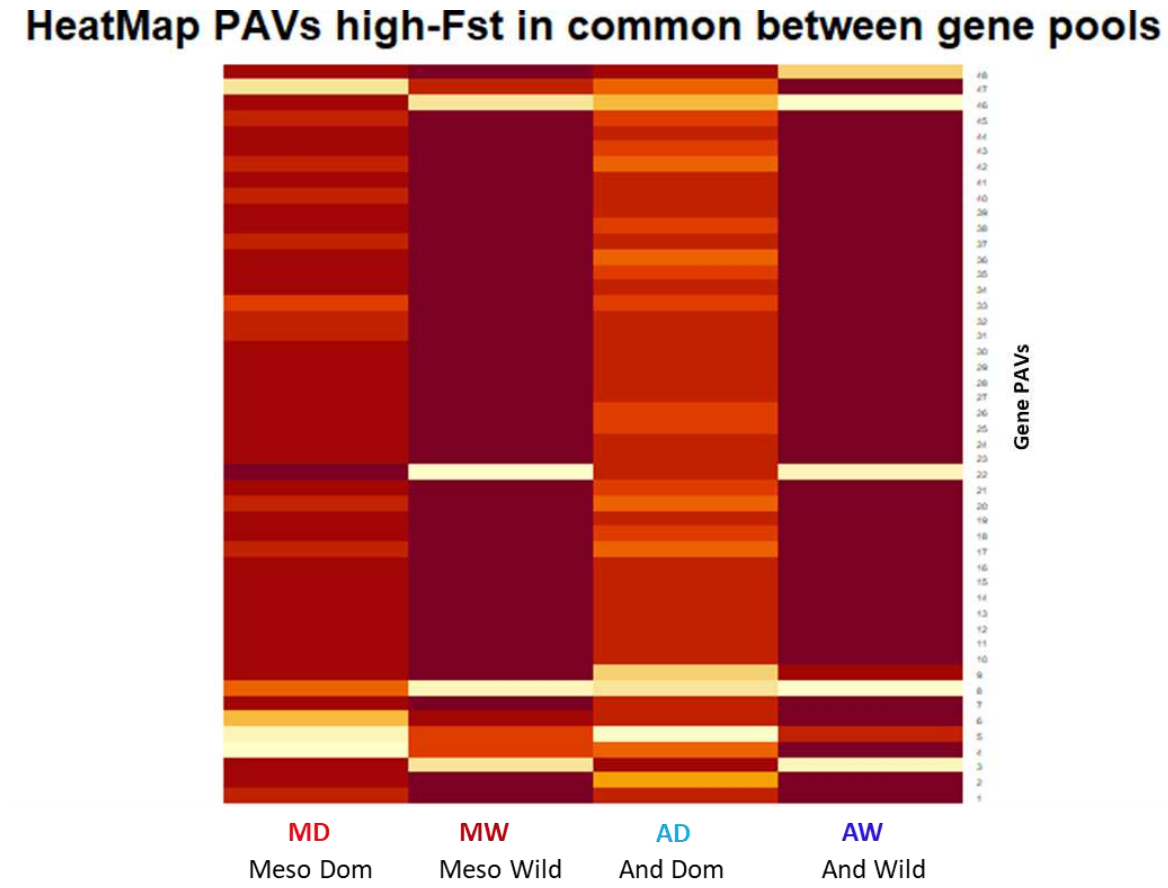


Figure 11: The heatmap shows the frequency for each of the 48 gene PAVs with Fst≥0.30 in common between the Mesoamerican and the Andean gene pools, by grouping for gene pool (i.e., Mesoamerican and Andean) and biological status (i.e., wild and domesticated).

All the 14,074 variable genes and those genes with high-Fst (i.e., Fst ≥ 0.30) were considered and subjected to GO term enrichment analysis. Overall, 3,881 out of 14,074 gene PAVs have the orthologous in *A. thaliana*, resulting in 121 Mesoamerican and 93 Andean PAVs with high Fst and the *A. thaliana* orthologous for the GO enrichment. Results on the entire variable genome (**Figure 12**) showed that the top highly enriched pathways were involved in response to stimulus, hormones, organic compounds, immune system process, cell wall modifications, biotic and abiotic stress, and regulation of biological process involved in symbiotic interaction. Moreover, the GO enrichment analysis on those PAVs potentially under selection (Fst≥0.30 between wild and domesticated accessions), in both the Mesoamerican (**Figure 13**) and Andean (**Figure 14**) gene pools, highlighted with a higher resolution the important role of genes related to adaptation and survival in different agro-ecological conditions. For instance, gene ontology categories on these genes include the establishment/maintenance of cell polarity (GO:0007163), response to salicylic acid (GO:0009751), plant-pathogen interactions (ath04626), defense response to bacterium (GO:0042742), methylation (GO:0032259), flavonoid biosynthesis (ath00941), vitamin B6 metabolism (ath00750), anthocyanin-containing compound metabolic process (GO:0046283), and response to wounding (GO:0009611).

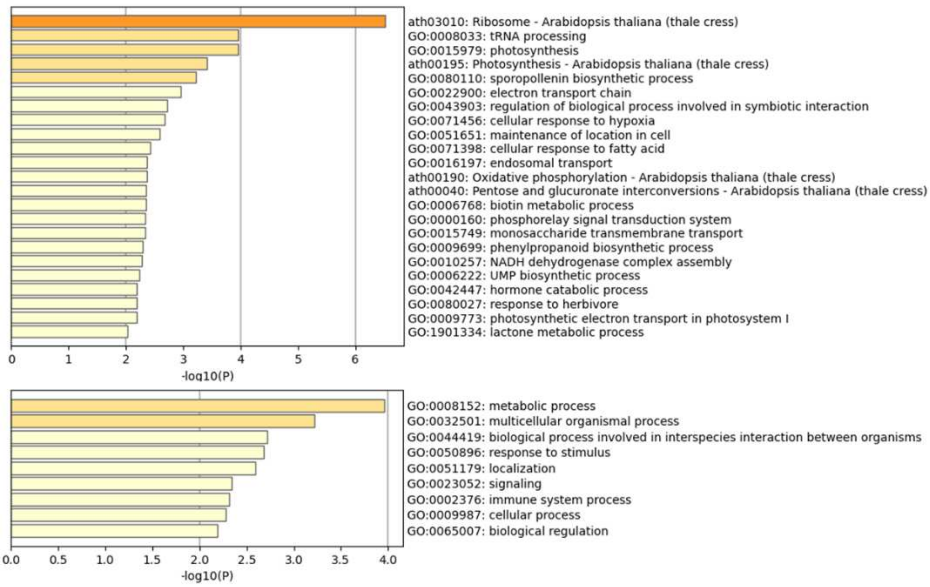


Figure 12: Bar graphs representing the result of the gene ontology enrichment analysis. GO classes representativity was investigated for all the 14,074 variable genes: 3,881 out of 14,074 gene PAVs have the orthologous in *A. thaliana* for the GO analysis. Enriched terms across input gene lists are colored by p-values. *Up bar graph*: the detailed GO terms; *Down bar graph*: the broad GO terms.

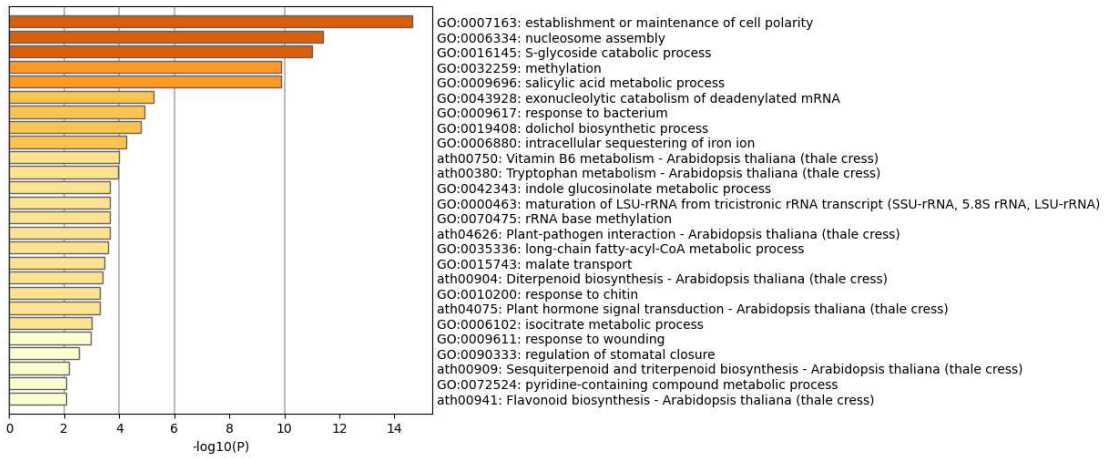
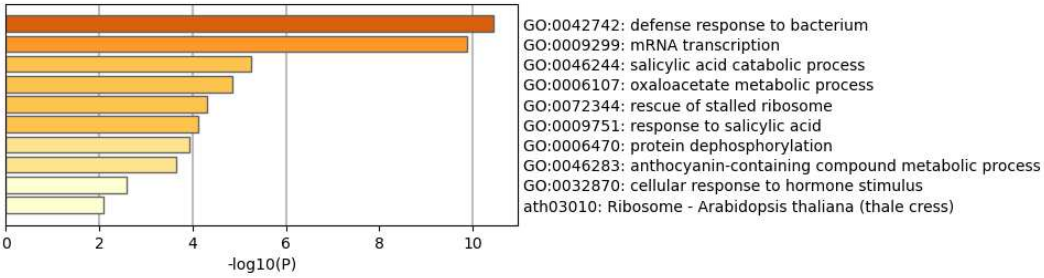


Figure 13: Bar graphs representing the result of the gene ontology enrichment analysis in the Mesoamerican gene pool. GO classes representativity was investigated for the genes with high-Fst values between wild and domesticated: 121 out of 626 PAVs have the orthologous in *A. thaliana* for the GO analysis. Enriched terms across input gene lists are colored by p-values.



Bar graphs representing the result of the gene ontology enrichment analysis in the Andean gene pool. GO classes representativity was investigated for the genes with high-Fst values between wild and domesticated: 93 out of 380 PAVs have the orthologous in *A. thaliana* for the GO analysis. Enriched terms across input gene lists are colored by p-values.

All the *P. vulgaris* gene PAVs with high Fst between wild and domesticated forms and for which we had the orthologous in *A. thaliana* were subjected to gene function investigation. Overall, 121 and 93 genes have been investigated out of a total of 626 Mesoamerican and 380 Andean candidates, respectively. The gene function investigation clearly validated the GO enrichment analysis and confirmed that variable genes potentially under selection are mainly associated to the domestication syndrome and adaptation traits. Namely, with regard to the Mesoamerican gene pool, we found two gene PAVs (Phvul.002G205200 and Phvul.003G265200) whose function in common bean is putatively associated to symbiotic interactions and 59 gene PAVs with an interesting function in *A. thaliana* model system. Among these 59 genes, Phvul.011G030200 is the orthologous to the PEROXIDASE 4 gene (PER4 or PRX4) in *A. thaliana*. PRX4 is a basic peroxidase regulated by day length with an important role in lignification (Fernández-Pérez et al., 2015). Lignins are covalently associated with polysaccharides in plant cell walls; they impart water impermeability, including resistance against tensile forces of the water columns and confer structural support and flexural stiffness to the aerial organs. These polyphenolic compounds are deposited mainly in tracheids, vessels, fibres of the xylem and phloem and schlerenchyma (Boerjan et al., 2003). In the common bean, the seed dispersal mechanism, one of the main target traits selected during crop domestication, is associated to the content and location of the fibres in the pods with a strict positive correlation between the shattering ability and increased lignin content (Prakken 1934; Murgia et al., 2017). Phvul.005G098800 is another promising candidate gene that is orthologous to the lignin biosynthetic gene AT5G48930 (HCT). HCT encodes an hydroxycinnamoyl-Coenzyme A shikimate/quinate hydroxycinnamoyltransferase that is involved in the phenylpropanoid pathway. Interestingly, Besseau et al. (2007) demonstrated that flavonoid accumulation in *A. thaliana* repressed in lignin synthesis affects auxin transport and plant growth. Another main target trait that was selected during the domestication process is the seed dormancy. The control of dormancy plays a fundamental adaptive role in nature by optimizing germination to the most suitable time. Here, we found Phvul.011G030200 that is orthologous to the HISTONE 2B gene (H2B or also known as HTB9). Interestingly, Liu et al. (2007) suggested the hypothesis that H2B monoubiquitination is required for chromatin remodeling in *A. thaliana* seed dormancy. We also detected various gene PAVs involved in the flowering network. Flowering time is strictly dependent on the photoperiod sensitivity, and it is one of the major diversification traits that defines adaptation of plant populations to different agro-ecological conditions. Phvul.008G247500 is orthologous to the AT2G30470 gene (VAL1). VAL1 is a member of a novel family of B3 domain proteins, and it is a transcriptional repressor in the silencing mechanism. Recently, Qüesta et al. (2016) identified a single point mutation at an intragenic nucleation site within FLOWERING LOCUS C (FLC) locus that blocks nucleation of plant homeodomain-Polycomb repressive complex 2 (PHD-PRC2) and allows VAL1 to promote histone deacetylation and FLC transcriptional silencing, thereby accelerating flowering. Another interesting candidate for flowering is Phvul.011G066200 that is orthologous to two different *A. thaliana* genes: H2B and HTB2. Both these genes encode a histone 2B protein. The histone H2B deubiquitination is required for transcriptional activation of FLC and for proper control of flowering (Schmitz et al., 2009). Also, the absence of the histone H2B monoubiquitination in the *Arabidopsis* hub1 (rdo4)

mutant reveals a role for chromatin remodeling in seed dormancy (Liu et al., 2017). Moreover, the histone H2B is also involved in regulating the dynamics of microtubules during the defense response to *Verticillium dahliae* toxins in *Arabidopsis* (Hu et al., 2014). Also, Phvul.003G260800 is orthologous to the AT2G18040 gene (PIN1AT). PIN1AT is a peptidyl-prolyl cis/trans isomerase that regulates root gravitropism and affects auxin transport (Xi et al., 2016). In addition, PIN1AT controls floral transition by accelerating cis/trans isomerization of the phosphorylated Ser/Thr-Pro motifs in two MADS-domain transcription factors, SUPPRESSOR OF OVEREXPRESSION OF CO 1 (SOC1) and AGAMOUS-LIKE 24 (AGL24) (Wang et al., 2010). We also identified several *P. vulgaris* candidate PAVs involved in responses to environmental stress. Among those, Phvul.011G066800 is orthologous to the AT3G46090 gene (ZAT7). The EAR-motif of ZAT7 is directly involved in enhancing the tolerance of transgenic *A. thaliana* plants to salinity stress (Ciftci-Yilmaz et al., 2007). Moreover, Phvul.011G066800 is also orthologous to the AT5G59820 gene (ZAT12). ZAT12 expression is activated at the transcriptional level during different abiotic stresses and in response to a wound-induced systemic signal. Using ZAT12 gain- and loss-of-function lines, Davletova et al. (2005) demonstrated the keyrole of ZAT12 in reactive oxygen and abiotic stress signaling (i.e., osmotic, salinity, high light, and heat stresses). Phvul.002G042400 is the orthologous to the AT3G22600 gene. AT3G22600 encodes a glycosylphosphatidylinositol (GPI)-anchored LTPg protein (LTPg5) that is part of the *A. thaliana* resistance mechanism against pathogens. LTPg5 has probably no direct antimicrobial activity but could perhaps act by associating with a receptor-like kinase, leading to the induction of defense genes (Ali et al., 2020).

Concerning the Andean gene pool, we detected one PAV (Phvul.005G008100) putatively involved in symbiotic interaction in common bean. Moreover, we identified 30 gene PAVs with an interesting function in *A. thaliana*. For instance, Phvul.003G003800 is orthologous to the TRANSPARENT TESTA16 gene (TT16). TT16 encodes a MADS-box protein responsible for the proper development and pigmentation of the seed coat (Nesi et al., 2002). Another promising candidate is Phvul.003G048000 that is orthologous to the AT1G49770 gene (RGE1 also known as ZOU). ZOU encodes a member of the basic-helix-loop-helix family of transcription factors and plays a role in determining the depth of primary dormancy in *A. thaliana* through the regulation of abscisic acid levels (McGregor et al., 2019). Phvul.007G273000 is also an interesting candidate that is orthologous to the ETHYLENE RESPONSE FACTOR 1 gene (ERF1). ERF1 is an upstream component in both jasmonate and ethylene signaling and is involved in pathogen resistance. It plays a positive role in salt, drought, and heat stress tolerance by stress-specific gene regulation, which integrates jasmonate, ethylene, and abscisic acid signals (Cheng et al., 2013). Phvul.005G010500 is also an interesting candidate, and its orthologous BIOTIN F (BIOF) in *A. thaliana* encodes a 7-keto-8-aminopelargonic acid (KAPA) synthase, the first committed enzyme of the biotin synthesis pathway (Pinon et al., 2005). In addition to its essential metabolic functions, BIOF is involved in survival pathways by modulating the defense genes expression and spontaneous cell death (Li et al., 2012). Another promising gene is Phvul.005G010700 that is orthologous to two *A. thaliana* MYST histone acetyltransferases HAM1 (AT5G64610) and HAM2 (AT5G09740). Xiao et al. (2013) by using an artificial microRNA strategy in *A. thaliana* uncovered a novel putative function of HAM1 and HAM2 in controlling the flowering time by epigenetic

modification of FLC and MADS-box Affecting Flowering genes 3/4 (MAF3/4) chromatins at histone H4 lysine 5 (H4K5) acetylation. Finally, Phvul.007G273400 is orthologous to the myb domain proteins 14 and 15. In detail, MYB14 encodes a nuclear protein that functions as an R2R3-MYB transcription activator. Knock-down of *AtMYB14* by artificial microRNA increased the tolerance to cold stress, demonstrating their involvement in freezing tolerance in *Arabidopsis* by affecting expression of *CBF* genes, a class of transcription factors that play important roles in cold response (Chen et al., 2013).

Our investigation also confirmed several genes previously proposed by other studies (i.e., Schmutz et al., 2014 and Bellucci et al., 2014) as candidates for common bean domestications.

4. Conclusions

We constructed the first common bean pangenome using five high-quality genomes and the whole-genome sequence reads of 339 genetically diverse common bean accessions. Overall, the common bean pangenome, including the reference (PV442) and non-reference sequences (NRRs), consists of ~779.99 Mb and contains 35,016 predicted protein-coding genes. The common bean pangenome identified approximately 242.78 Mb of novel sequences that were absent from the reference genome and that encode an additional 7,583 genes. The discovery of novel variants on the NRRs confirmed that the pangenomic approach is a robust and comprehensive method to capture with a greater resolution the variation present in a certain species. In particular, we found that the Mesoamerican gene pool has a higher number of variants (i.e., PAVs) per individual compared to the Andean gene pool, which can be explained by the sequential bottlenecks that occurred privately in the Andean gene pool (i.e., one before domestication and one with the domestication). These outcomes confirmed that the evolutionary history and trajectory of a species is a crucial factor that influences the level and structure of the current genetic diversity in crop species. Thus, our findings highlight the potential of the common bean as a model for the study of crop domestication and adaptation. This arises from its two geographically distinct and partially isolated gene pools with independent and parallel domestications, offering the almost unique opportunity to look at the domestication process as a replicate experiment. In addition to the differences between Mesoamerican and Andean gene pools in terms of the number of PAVs per individual, we detected that the Mesoamerican wild forms have fewer variants than the domesticated forms. However, as expected, since the domestication process is usually associated with a reduction in the genetic diversity, when we analyzed the number of gene PAVs per group, we detected that the wild forms have a higher number of variants compared to the domesticated accessions, for both the Mesoamerican and Andean gene pools. All the predicted genes identified in the common bean pangenome were categorized as core (60%) and variable (40%) genes, based on the frequency of the presence and absence variations. Interestingly, the variable genes were enriched with genes in response to stimulus, hormones, organic compounds, immune system process, cell wall modifications, biotic and abiotic stress, and regulation of biological processes involved in symbiotic interactions. Moreover, Fst analysis between wild and domesticated forms, on a representative subset composed of 97 accessions, detected 626 Mesoamerican and 380 Andean gene PAVs putatively under selection. Manual gene function investigation in *A. thaliana* orthologous confirmed that these candidate genes are mainly associated with the domestication syndrome and adaptation traits. These preliminary results provide a good starting point for common bean population genomics studies, as well as for identifying functional variants of agronomically and economically important traits useful for future legume breeding programs. The present study is still ongoing, further analyses are needed to better characterize the gene PAVs within common bean groups and subgroups in light of their broad agro-ecological adaptation. For instance, we are planning to perform PAV-based landscape pangenomics to identify relationships between environmental factors and genetic adaptation. We are also

finalizing the SNPs calling on the common bean pangenome in order to proceed with PAV-based GWAS and selection scan analyses.

5. References

Ali, M.A., Abbas, A., Azeem, F., Shahzadi, M., Bohlmann, H. The *Arabidopsis* GPI-anchored LTPg5 encoded by *At3g22600* has a role in resistance against a diverse range of pathogens. *Int. J. Mol. Sci.* **2020**, 21(5):1774, doi: 10.3390/ijms21051774.

Appels, R. et al. Shifting the limits in wheat research and breeding using a fully annotated reference genome. *Science* **2018**, 361:661, doi: 10.1126/science.aar7191.

Badouin, H., Gouzy, J., Grassa, C. et al. The sunflower genome provides insights into oil metabolism, flowering and asterid evolution. *Nature* **2017**, 546:148-152, doi: 10.1038/nature22380.

Barrett, R.D.H. and Schluter, D. Adaptation from standing genetic variation. *Trends Ecol. Evol.* **2008**, 23:38-44, doi: 10.1016/j.tree.2007.09.008.

Bayer, P.E. et al.: Variation in abundance of predicted resistance genes in the *Brassica* oleracea pangenome. *Plant Biotechnol J* **2019**, 17:789-800, doi: 10.1111/pbi.13015.

Bellucci, E., Bitocchi, E., Ferrarini, A., et al. Decreased nucleotide and expression diversity and modified co-expression patterns characterize domestication in the common bean. *Plant Cell* **2014**, 26:1901-1912, doi: 10.1105/tpc.114.124040.

Besseau, S., Hoffmann, L., Geoffroy, P., Lapierre, C., Pollet, B., Legrand, M. Flavonoid accumulation in *Arabidopsis* repressed in lignin synthesis affects auxin transport and plant growth. *Plant Cell.* **2007**, 19(1):148-62, doi: 10.1105/tpc.106.044495.

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

Blaustein, R.A., McFarland, A.G., Ben Maamar, S., Lopez, A., Castro-Wallace, S., Hartmann, E.M. Pangenomic approach to understanding microbial adaptations within a model-built environment, the international Space Station, Relative to Human Hosts and Soil. *mSystems* **2019**, 4(1):e00281-18, doi:10.1128/mSystems.00281-18.

Boerjan, W., Ralph, J., Baucher, M. Lignin biosynthesis. *Annu. Rev. Plant. Biol.* **2003**, Vol. 54:pp.519-546, doi: 10.1146/annurev.arplant.54.031902.134938.

Chen, Y., Chen, Z., Kang, J., Kang, D., Gu, H., Qin, G.. AtMYB14 regulates cold tolerance in *Arabidopsis. Plant Mol Biol Report.* **2013**, 31(1):87-97, doi: 10.1007/s11105-012-0481-z.

Cheng, M.C., Liao, P.M., Kuo, W.W., Lin, T.P. The *Arabidopsis* ETHYLENE RESPONSE FACTOR1 regulates abiotic stress-responsive gene expression by binding to different cis-acting elements in response to different stress signals. *Plant Physiol.* **2013**, 162(3):1566-82, doi: 10.1104/pp.113.221911.

Ciftci-Yilmaz, S., Morsy, M.R., Song, L., Coutu, A., Krizek, B.A., Lewis, M.W., Warren, D., Cushman, J., Connolly, E.L., Mittler, R. The EAR-motif of the Cys2/His2-type zinc finger protein Zat7 plays a

key role in the defense response of *Arabidopsis* to salinity stress. *J. Biol. Chem.* **2007**, 282(12):9260-8, doi: 10.1074/jbc.M611093200.

Cook, D.E. et al. Copy number variation of multiple genes at Rhg1 mediates nematode resistance in soybean. *Science* **2012**, 338:1206-1209, doi: 10.1126/science.1228746.

Cortinovis, G., Frascarelli, G., Di Vittori, V., Papa, R. Current state and perspectives in population genomics of the common bean. *Plants* **2020**, 9(3):330, doi: 10.3390/plants9030330.

Danilevicz, M.F., Tay Fernandez, C.G., Marsh, J.I., Bayer, P.E., Edwards D. Plant pangenomics: approaches, applications and advancements. *Current Opinion in Plant Biology* **2020**, 54:18-25, doi: 10.1016/j.pbi.2019.12.005.

Davletova, S., Schlauch, K., Coutu, J., Mittler, R. The zinc-finger protein Zat12 plays a central role in reactive oxygen and abiotic stress signaling in *Arabidopsis. Plant Physiol.* **2005**, 139(2):847-56, doi: 10.1104/pp.105.068254.

Eizenga et al. Pangenome graphs. *Annual Review of Genomics and Human Genetics* **2020**, 21:139-162, doi: 10.1146/annurev-genom-120219-080406

Emms, D.M., Kelly, S. OrthoFinder: phylogenetic orthology inference for comparative genomics. *Genome Biol*. **2019**, 20:238, doi: 10.1186/s13059-019-1832-y.

Fernández-Pérez, F., Vivar, T., Pomar, F., Pedreño, M.A., Novo-Uzal, E. Peroxidase 4 is involved in syringyl lignin formation in *Arabidopsis thaliana*. *J Plant Physiol.* **2015**, 175:86-94, doi: 10.1016/j.jplph.2014.11.006.

Feuk, L. Marshall, C.R., Wintle, R.F., Scherer, S.W. Structural variants: changing the landscape of chromosomes and design of disease studies. *Hum. Mol. Genet.* **2006**, 15:R57-R66, doi: 10.1093/hmg/ddl057.

Golicz, A.A., Batley, J., Edwards, D. Towards plant pangenomics. *Plant Biotechnology Journal* **2016** 14: 1099-1105, doi: 10.1111/pbi.12499.

Gordon, S.P. et al. Extensive gene content variation in the Brachypodium distachyon pangenome correlates with population structure. *Nature Communications*, **2017**, 8(1):2184, doi: 10.1038/s41467-017-02292-8.

Hardigan, M.A. et al.: Genome reduction uncovers a large dispensable genome and adaptive role for copy number variation in asexually propagated *Solanum tuberosum*. *Plant Cell* **2016**, 28:388, doi: 10.1105/tpc.15.00538.

Hoopes, G.M. et al.: An updated gene atlas for maize reveals organ-specific and stress-induced genes. *Plant J* **2019**, 97:1154-1167, doi: 10.1111/tpj.14184.

Hu, M., Pei, B.L., Zhang, L.F., Li, Y.Z. Histone H2B monoubiquitination is involved in regulating the dynamics of microtubules during the defense response to *Verticillium dahliae* toxins in *Arabidopsis*. *Plant Physiol.* **2014**, 164(4):1857-65, doi: 10.1104/pp.113.234567.

Hurgobin, B. and Edwards, D.SNP discovery using a pangenome: has the single reference approach become obsolete? *Biology* **2017**, 6: 21; doi: 10.3390/biology6010021**.**

Jolliffe, I.T. Principal component analysis for special types of data. In: *Principal Component Analysis. Springer Series in Statistics*. Springer, New York, NY, **2002**, doi: /10.1007/0-387-22440-8_13.

Jung, J.H., Lee, H.J., Ryu, J.Y., Park, C.M. SPL3/4/5 integrate developmental aging and photoperiodic signals into the FT-FD module in *Arabidopsis* flowering. *Mol Plant.* **2016**, 9(12):1647-1659, doi: 10.1016/j.molp.2016.10.014.

Korbel, J.O., Urban, A.E., Affourtit, J.P., Godwin, B-, Grubert, F., Simons, J.F., Kim, P,.M., Palejev, D., Carriero, N.J., Du, L., Taillon, B.E., Chen, Z., Tanzer,A., Saunders, A.C.E., Chi, J., Yang, F., Carter, N.P., Hurles, M.E., Weissman, S.M., Harkins, T.T., Gerstein, M.B., Egholm, M., Snyde, M. Paired-end mapping reveals extensive structural variation in the human genome. *Science* **2007**, 318:420-426, doi: 10.1126/science.1149504.

Krasileva, K.V. The role of transposable elements and DNA damage repair mechanisms in gene duplications and gene fusions in plant genomes. *Curr Opin Plant Biol* **2019**, 48:18-25, doi:

Landgraf, A.J., Lee, Y. Dimensionality Reduction for Binary Data through the Projection of Natural Parameters. *arXiv:1510.06112,* **2015***,* doi: 10.1016/j.jmva.2020.104668.

Li, J., Brader, G., Helenius, E., Kariola, T., Palva, E.T. Biotin deficiency causes spontaneous cell death and activation of defense signaling. *Plant J.* **2012**, 70(2):315-26, doi: 10.1111/j.1365-313X.2011.04871.x.

Liu, Y., Koornneef, M., Soppe, W.J. The absence of histone H2B monoubiquitination in the *Arabidopsis* hub1 (rdo4) mutant reveals a role for chromatin remodeling in seed dormancy. *Plant Cell.* **2007**;19(2):433-44, doi: 10.1105/tpc.106.049221.

Liu, Y., Koornneef, M., Soppe, W.J. The absence of histone H2B monoubiquitination in the Arabidopsis hub1 (rdo4) mutant reveals a role for chromatin remodeling in seed dormancy. *Plant Cell*. **2007**, 19(2):433-44, doi: 10.1105/tpc.106.049221.

MacGregor, D.R., Zhang, N., Iwasaki, M., Chen, M., Dave, A., Lopez-Molina, L., Penfield, S. ICE1 and ZOU determine the depth of primary seed dormancy in *Arabidopsis* independently of their role in endosperm development. *Plant J.* **2019**, 98(2):277-290, doi: 10.1111/tpj.14211.

Makkena, S. and Lamb, R.S. The bHLH transcription factor SPATULA regulates root growth by controlling the size of the root meristem. *BMC Plant Biol.* **2013**, 13:1, doi: 10.1186/1471-2229-13-1.

Mascher, M., Gundlach, H., Himmelbach, A. et al. A chromosome conformation capture ordered sequence of the barley genome. *Nature* **2017,** 544:427-433, doi: 10.1038/nature22043.

McCarroll, S.A. and Altshuler, D.M. Copy number variation and association studies of human disease. *Nat. Genet.* **2007**, 39:S37-S42, doi: 10.1038/ng2080.

McHale, L.K., Haun, W.J., Xu, W.W., Bhaskar, P.B., Anderson, J.E., Hyten, D.L., Gerhardt, D.J., Jeddeloh, J.A., Stupar R.M. Structural variants in the soybean genome localize to clusters of biotic stress-response genes. *Plant Physiology* **2012**, 159(4):1295-1308, doi: 10.1104/pp.112.194605.

Mousavi-Derazmahalleh, M.; Bayer, P.E.; Hane, J.K.; Valliyodan, B.; Nguyen, H.T.; Nelson, M.N.; Erskine, W.; Varshney, R.K.; Papa, R.; Edwards, D. Adapting legume crops to climate change using genomic approaches. *Plant Cell Environ*. **2019**, 42:6-9, doi: 10.1111/pce.13203.

Muñoz-Amatriaín, M., Eichten, S.R., Wicker, T. et al. Distribution, functional impact, and origin mechanisms of copy number variation in the barley genome. *Genome Biol.* **2013**, 14:R58, doi: 10.1186/gb-2013-14-6-r58.

Murgia, M.L., Attene, G., Rodriguez, M., Bitocchi, E., Bellucci, E., Fois, D., Nanni, L., Gioia, T., Albani, D.M., Papa, R., Rau, D. A comprehensive phenotypic investigation of the "pod-shattering syndrome" in common bean. *Front. Plant Sci.* **2017**, 8:251, doi: 10.3389/fpls.2017.00251.

Nesi, N., Debeaujon, I., Jond, C., Stewart, A.J., Jenkins, G.I., Caboche, M., Lepiniec, L. The TRANSPARENT TESTA16 locus encodes the ARABIDOPSIS BSISTER MADS domain protein and is required for proper development and pigmentation of the seed coat. *Plant Cell.* **2002,** 14(10):2463-79, doi: 10.1105/tpc.004127.

Pinon, V., Ravanel, S., Douce, R., Alban, C. Biotin synthesis in plants. The first committed step of the pathway is catalyzed by a cytosolic 7-keto-8-aminopelargonic acid synthase. *Plant Physiol.* **2005**, 139(4):1666-76, doi: 10.1104/pp.105.070144.

Prakken, R. Inheritance of colours and pod characters in *Phaseolus vulgaris* L. *Genetica* **1934**, 16:177-296, doi: 10.1007/BF02071498.

Qi, X., Li, MW., Xie, M. et al. Identification of a novel salt tolerance gene in wild soybean by whole-genome sequencing. *Nat. Commun.* **2014**, 5:4340, doi: 10.1038/ncomms5340.

Qüesta, J.I., Song, J., Geraldo, N., An, H., Dean, C. *Arabidopsis* transcriptional repressor VAL1 triggers Polycomb silencing at FLC during vernalization. *Science* **2016**; 353(6298):485-8, doi: 10.1126/science.aaf7354.

Ringnér, M. What is principal component analysis?. *Nat Biotechnol* **2008**, 26:303-304, doi: 10.1038/nbt0308-303.

Saxena, R.K., Edwards, D., Varshney, R.K. Structural variations in plant genomes. *Brief. Funct. Genomics* **2014**, 13(4):296-307, doi: 10.1093/bfgp/elu016.

Schmitz, R.J., Tamada, Y., Doyle, M.R., Zhang, X., Amasino, R.M. Histone H2B deubiquitination is required for transcriptional activation of FLOWERING LOCUS C and for proper control of flowering in *Arabidopsis*. *Plant Physiol.* **2009**, 149(2):1196-204, doi: 10.1104/pp.108.131508.

Schmutz, J.; McClean, P.E.; Mamidi, S. et al. A reference genome for common bean and genome-wide analysis of dual domestications. *Nat. Genet.* **2014**, 46:707-713, doi: 10.1038/ng.3008.

Sidaway-Lee, K., Josse, E.M., Brown, A., Gan, Y., Halliday, K.J., Graham, I.A., Penfield, S. SPATULA links daytime temperature and plant growth rate. *Curr Biol.* **2010**, 20(16):1493-7, doi: 10.1016/j.cub.2010.07.028.

Singh, S.P., Gepts, P., Debouck, D.G. Races of common bean (*Phaseolus vulgaris*, Fabaceae). Econ Bot. **1991**, 45:379396, doi: 10.1007/BF02887079.

Springer, N.M., Anderson, S.N., Andorf, C.M. et al. The maize W22 genome provides a foundation for functional genomics and transposon biology. *Nat. Genet.* **2018**, 50:1282-1288, doi: 10.1038/s41588-018-0158-0.

Springer, N.M., Ying, K., Fu, Y., Ji, T., Yeh, C.T. et al. Maize inbreds exhibit high levels of copy number variation (CNV) and presence/absence variation (PAV) in genome content. *PLoS Genet.* **2009**, 5(11):e1000734, doi: 10.1371/journal.pgen.1000734.

Tettelin, H. et al. Genome analysis of multiple pathogenic isolates of streptococcus agalactiae: Implications for the microbial "pan-genome". *Proceedings of the National Academy of Sciences* **2005**, 102(39):13950-13955, doi: 10.1073/pnas.0506758102.

Tranchant-Dubreuil, C., Rouard, M., Sabot, F. Plant pangenome: impacts on phenotypes and evolution. *Annual Plant Reviews Online* **2019**, 2(2), doi: 10.1002/9781119312994.apr0664.

Vaistij, F.E., Gan, Y., Penfield, S., Gilday, A.D., Dave, A., He, Z., Josse, E.M., Choi, G., Halliday, K.J., Graham, I.A. Differential control of seed primary dormancy in *Arabidopsis* ecotypes by the transcription factor SPATULA. *Proc. Natl. Acad. Sci. USA* **2013,** 110(26):10866-71, doi: 10.1073/pnas.1301647110.

Varshney, R., Saxena, R., Upadhyaya, H. et al. Whole-genome resequencing of 292 pigeonpea accessions identifies genomic regions associated with domestication and agronomic traits. *Nat. Genet.* **2017**, 49:1082-1088, doi: 10.1038/ng.3872.

Varshney, R., Shi, C., Thudi, M. et al. Pearl millet genome sequence provides a resource to improve agronomic traits in arid environments. *Nat. Biotechnol.* **2017**, 35:969-976, doi: 10.1038/nbt.3943.

Wang, K., Hu, H., Tian, Y. et al. The chicken pan-genome reveals gene content variation and a promoter region deletion in IGF2BP1 affecting body size. *Mol. Biol. Evol.* **2021**, 38(11):5066-5081 doi: 10.1093/molbev/msab231.

Wang, Y., Liu, C., Yang, D., Yu, H., Liou, Y.C. Pin1At encoding a peptidyl-prolyl cis/trans isomerase regulates flowering time in *Arabidopsis*. *Mol Cell.* **2010,** 37(1):112-22, doi: 10.1016/j.molcel.2009.12.020.

Wang, Y., Xiong, G., Hu, J. et al. Copy number variation at the GL7 locus contributes to grain size diversity in rice. *Nat. Genet.* **2015**, 47:944-948, doi: 10.1038/ng.3346.

Xi, W., Gong, X., Yang, Q., Yu, H., Liou, Y.C. Pin1At regulates PIN1 polar localization and root gravitropism. *Nat Commun.* **2016**, 7:10430, doi: 10.1038/ncomms10430.

Xiao, J., Zhang, H., Xing, L., Xu, S., Liu, H., Chong, K., Xu, Y. Requirement of histone acetyltransferases HAM1 and HAM2 for epigenetic modification of FLC in regulating flowering in *Arabidopsis*. *J. Plant Physiol.* **2013**, 170(4):444-51, doi: 10.1016/j.jplph.2012.11.007.

Xing, L., Liu, Y., Xu, S., Xiao, J., Wang, B., Deng, H., Lu, Z., Xu, Y., Chong, K. *Arabidopsis O*-GlcNAc transferase SEC activates histone methyltransferase ATX1 to regulate flowering. *EMBO J.* **2018**, 37(19):e98115, doi: 10.15252/embj.201798115.

Xu, K., Xu, X., Fukao, T. et al.: Sub1A is an ethylene-response-factor-like gene that confers submergence tolerance to rice. *Nature* **2006**, 442:705-708, doi: 10.1038/nature04920.

Yu, J., Tehrim, S., Zhang, F. et al.: Genome-wide comparative analysis of NBS-encoding genes between *Brassica* species and *Arabidopsis thaliana*. *BMC Genomics* **2014**, 15:3, doi: 10.1186/1471-2164-15-3.

Zhou, Y., Zhou, B., Pache, L., Chang, M., Khodabakhshi, A.H., Tanaseichuk, O., Benner, C., Chanda, S.K. Metascape provides a biologist-oriented resource for the analysis of systems-level datasets. *Nat Commun.* **2019**, doi: 10.1038/s41467-019-09234-6.

Zhou, Z., Jiang, Y., Wang, Z. et al. Resequencing 302 wild and cultivated accessions identifies genes related to domestication and improvement in soybean. *Nat. Biotechnol.* **2015**, 3: 408-414, doi: 10.1038/nbt.3096.

General Conclusion

The meteoric increase in sequencing with high throughput next-generation sequencing technologies (NGS) has dramatically changed our understanding of genomes. Their application has provided novel approaches that have significantly advanced our understanding of new and long-standing questions in evolutionary processes. With a particular focus on *P. vulgaris*, we characterized a panel of 218 WGS landrace accessions from both the American centers of diversity and the secondary centers of domestication and diversification in Europe through a large dataset of more than 2 million SNPs distributed genome-wide. We apply the most recent '-omics' technologies using a multidisciplinary approach (genomics, population/ and quantitative genetics, biochemistry, and plant physiology) to highlight the complex relationship between the genotypic and phenotypic diversity in common bean populations, as well as characterize the effect of selection for the development of the European common bean. The most striking result that we observed was the high amount of introgression between Andean and Mesoamerican gene pools also promoted by selection suggesting the hypothesis that the complex evolutionary history of the European common bean was probably guided by adaptive introgression. To fully explore the genetic diversity of *P. vulgaris*, we also developed and characterized the first common bean pangenome. Overall, we found a total of 7,583 new genes do not present in the reference genome, confirming that the pangenomic approach is a robust and comprehensive method to capture with a greater resolution the variation present in a certain species. Here, the most interesting outcome that we observed was the higher number of gene PAVs per individual in the Mesoamerican gene pool than the Andean one. This could be interpretated as the direct consequence of the sequential bottlenecks that occurred privately in the Andean gene pool and led to the impoverishment of its genetic diversity (Bitocchi et al., 2013), which result genetically more depauperated in terms of number of gene PAVs per individual compared to the Mesoamerican gene pool. Moreover, the variations identified through the common bean pangenome analysis highlighted the presence of genes mainly associated to the domestication syndrome and adaptation traits, such as dormancy, flowering, and defense responses to biotic and abiotic stress. The pangenome concept is offering a great opportunity to uncover new genes and increase our knowledge about evolutionary mechanisms that allow organisms to adapt quickly to new environments. Discovering genes and genetic mechanisms that contribute to phenotypic adaptation associated with environmental conditions and their mapping along the reference pangenome will provide a useful genetic tool for geneticists and breeders for

the constitution of novel varieties. This is a crucial aspect towards future major environmental and socio-economic changes, such as increases in temperature, differences in rainfall, and new consumer preferences. These outcomes will also be a step towards complete identification of all the functional elements encoded in the plant genome, which is one of the major scientific targets of plant research. Pangenome concept also provided the opportunity to extend this knowledge to closely related legume for comparative genomics studies (Khan et al., 2019). As different species in a genus are available for a given crop, useful genes can be transferred from one species to another either simply by a crossing mechanism, especially with species from the secondary gene pool, or by wide hybridization or modern chromosome/ genome engineering approaches for species belonging to other/distantly related gene pools. The almost unique situation that characterizes the *Phaseolus* genus is that five of its ~70 species have been domesticated (i.e., *Phaseolus vulgaris*, *P. coccineus*, *P. dumosus*, *P. acutifolius*, and *P. lunatus*), and in addition, for *P. vulgaris* and *P. lunatus*, the wild forms are distributed in both Mesoamerica and South America, where at least two independent and isolated episodes of domestication occurred. For this purpose, the implementation of the super-pangenome, that aims to represent the complete genetic architecture of a genus by combining the different pangenomes from all of the species of the given genus, would ultimately have the capacity to analyze gene evolution within and between species. The characterization, maintenance and the exploitation of food legume genetic resources in pre-breeding form the core development of both more sustainable agriculture and healthier food products. Indeed, in 2022 the IPCC report (https://www.ipcc.ch/report/ar6/wg2/downloads/report/IPCC_AR6_WGII_FinalDraft_FullReport.pdf) entitled “Impact, Adaptation and Vulnerability” indicated that the transition to novel plant-based diets could present major opportunities for adaptation and mitigation while generating significant co-benefits in terms of human health. Moreover, most of legume species can establish symbiotic association with nitrogen fixing bacteria, collectively known as rhizobia. Nitrogen fixation underlies the high protein content of legume seeds, and it is also of immense economic and ecologic importance, because it returns vital reduced nitrogen to the soil, thereby enhancing (agro)ecosystem productivity and sustainability. Historically, legumes were a primary source of agricultural nitrogen because they were grown in rotation with cereals. In most modern intensive agricultural systems, however, including those of Europe and North America, nitrogen fertilizer originates from industrial processes that require immense quantities of fossil fuel to reduce N2 to NH4. Thus, production of industrial fertilizers contributes ~3% of global CO2 and is a primary source of pollutant NO2. Moreover, runoff from fertilizer is among the world's most serious environmental

pollutants, causing eutrophication of marine systems. Therefore, the exploitation of the common bean genetic resources through the use of the pangenome will have a major impact on sustainable agriculture and the world's economic, social and environmental health.

References

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

Broughton, W.J.; Hernández, G.; Blair, M.; Beebe, S.; Gepts. P.; Vanderleyden, J. Beans (*Phaseolus* spp.) - model food legumes. *Plant and Soil* **2003**, 252:55-128, doi: 10.1023/A:1024146710611.

Cortinovis, G., Frascarelli, G., Di Vittori, V., Papa, R. Current state and perspectives in population genomics of the common bean. *Plants* **2020**, 9(3):330, doi: 10.3390/plants9030330.

Eizenga et al. Pangenome graphs. *Annual Review of Genomics and Human Genetics* **2020**, 21:139-162, doi: 10.1146/annurev-genom-120219-080406.

Graham, P.H. and Vance, C.P. Legumes: Importance and constraints to greater use. *Plant Physiology* **2003**, 131:872-877, doi: 10.1104/pp.017004.

Khan A.W., Garg, V., Roorkiwal, M., Golicz, A.A., Edwards, D., Varshney, R.K. Super-Pangenome by integrating the wild side of a species for accelerated crop improvement. *Trends in Plant Science* **2019**, doi: 10.1016/j.tplants.2019.10.012

Mousavi-Derazmahalleh, M.; Bayer, P.E.; Hane, J.K.; Valliyodan, B.; Nguyen, H.T.; Nelson, M.N.; Erskine, W.; Varshney, R.K.; Papa, R.; Edwards, D. Adapting legume crops to climate change using genomic approaches. *Plant Cell Environ*. **2019**, 42:6-9, doi: 10.1111/pce.13203.

Ray, D., Ramankutty, N., Mueller, N. et al. Recent patterns of crop yield growth and stagnation. *Nat Commun* **2012**, 3:1293, doi: 10.1038/ncomms2296.

Schmutz, J.; McClean, P.E.; Mamidi, S. et al. A reference genome for common bean and genome-wide analysis of dual domestications. *Nat. Genet.* **2014**, 46:707-713, doi: 10.1038/ng.3008.

Vlasova, A., Capella-Gutiérrez, S., Rendón-Anaya, M. et al. Genome and transcriptome analysis of the Mesoamerican common bean and the role of gene duplications in establishing tissue and temporal specialization of genes. *Genom. Biol.* **2016**, 17:32, doi: 10.1186/s13059-016-0883-6.