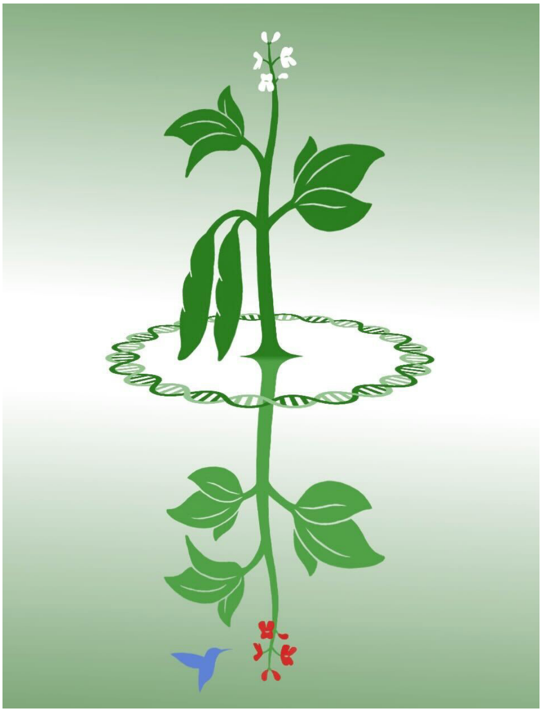
ACTA UNIVERSITATIS AGRICULTURAE SUECIAE

DOCTORAL THESIS NO. 2024:25
FACULTY OF NATURAL RESOURCES AND AGRICULTURAL SCIENCES

Adaptation in two species of *Phaseolus* with contrasting mating systems

LE YU



Adaptation in two species of *Phaseolus* with contrasting mating systems

Le Yu
Faculty of Natural Resources and Agricultural Sciences
Department of Plant Biology
Uppsala

SLU
SWEDISH UNIVERSITY
OF AGRICULTURAL
SCIENCES

DOCTORAL THESIS
Uppsala 2024

Acta Universitatis Agriculturae Sueciae
2024:25

Cover: Common bean and its outcrossing relative
(Drawing: Le Yu)

ISSN 1652-6880
ISBN (print version) 978-91-8046-314-0
ISBN (electronic version) 978-91-8046-315-7
https://doi.org/10.54612/ a.2b30e01u1h

Swedish University of Agricultural Sciences, Department of Plant Biology, Uppsala, Sweden
Print: SLU Grafisk Service, Uppsala 2024

Adaptation in two species of *Phaseolus* with contrasting mating systems

Abstract

For common bean, the loss of photoperiod sensitivity has been a major pre-requisite for domestication and further adaptation to European climates. Its self-pollinated mating system also makes it a good model for identifying genes that display enhanced levels of differentiation in gene expression, together with its closely related species, open-pollinated runner bean. We first use GWAS and haplotype structure tests to identify single nucleotide and structural variants strongly associated to the capacity to flower under neutral and long days. Gene-pool specific selective sweeps were identified that correspond to the independent domestication events in the Americas. A strong GWAS signal associated TFL-1 to the loss of photoperiod sensitivity in both Mesoamerican and Andean gene-pools, indicating its important function in this biological process. Then we investigated the genetic architecture of gene expression variation in common bean and runner bean, to determine if and how mating system has influenced patterns of adaptation and gene expression variation. We observe large differences in genomic characteristics between the two species. For the genetic architecture of gene expression, mating system appears to have minor effects and gene expression is likely under stabilizing selection in both species, regardless of mating system. We additionally found that in selfing species, connectivity in gene co-expression network played greater roles in regulating gene expression. To further investigate gene expression divergence and the underlying selection forces, we calculated the selection index. It is highly representative of gene expression divergence between two species. Although selective sweeps at the genomic level act less effectively on highly diverged genes, we found that DNA/RNA modification process may play an important role in conducting gene expression variation.

Keywords: Domestication, adaptation, mating system, gene expression variation, common bean, runner bean

Abstrakt

För trädgårdsbönan (*Phaseolus vulgaris*) har förlusten av fotoperiodkänslighet varit en viktig förutsättning för domesticering och ytterligare anpassning till ett europeiskt klimat. Dess självbefruktande parningssystem gör det till en bra modell för att identifiera gener som visar ökad differentiering i genuttryck jämfört med dess nära släkting rosenbönan (*Phaseolus coccineus*). I den här avhandlingen har vi oss av GWAS och haplotypstrukturtester för att identifiera enskilda nukleotid- och strukturella varianter som starkt är associerade med förmågan att blomma under neutrala och långa dagar. Genpoolsspecifika signaturer av naturlig selektion identifierades och som visade sig motsvara två oberoende domesticeringstillfällen i syd- respektive mellanamerika. En stark GWAS-signal visade sig koppla TFL-1 till förlusten av fotoperiodkänslighet i både den mellanamerikanska och sydamerikanska genpoolen, vilket indikerar dess viktiga funktion i denna biologiska process. Sedan undersökte vi den genetiska arkitekturen av genutrycksvariationen i trädgårds- och rosenböna för att avgöra om och hur parningssystemet har påverkat olika mönster av anpassning och genuttrycksvariation. Vi observerar stora skillnader i genomiska egenskaper mellan de två arterna. För den genetiska arkitekturen av genuttryck tycks parningssystemet ha mindre effekter och genuttryck är troligen under stabiliserande selektion i båda arterna, oavsett parningssystem. Vi fann dessutom att i självbefruktande arter spelar genuttrycksnätverkets konnektivitet större roller för regleringen av genuttryck. För att ytterligare undersöka genuttrycksskillnader och de underliggande selektionskrafterna beräknade vi ett sk. selektionsindex. Det är mycket representativt för genuttrycksdifferensen mellan två arter. Även om olika selektionsignaler på genomisk nivå verkar vara mindre effektiva inom högt divergerade gener, fann vi att DNA/RNA-modifieringsprocessen kan spela en viktig roll för att mediera genuttrycksvariation.

Nyckelord: Domesticering, anpassning, parningssystem, genuttrycksvariation, trädgårdsböna, rosenböna

Contents

List of publications

This thesis is based on the work contained in the following papers, referred to by Roman numerals in the text:

I. **Martha Rendón-Anaya, Greta Buinovskaja, Le Yu, Pär K. Ingvarsson.** (2023). Convergent evolution of photoperiod sensitivity during domestication and diversification in common bean (manuscript)

II. **Le Yu, Martha Rendón-Anaya, Pär K. Ingvarsson.** (2023). Adaptation and gene expression variation in selfing and outcrossing *Phaseolus* species (manuscript)

III. **Le Yu, Martha Rendón-Anaya, Pär K. Ingvarsson.** (2023). Patterns of gene expression divergence and selection background between two *Phaseolus* species (manuscript)

The contribution of Le Yu to the papers included in this thesis was as follows:

I. Phenotypic data and sample collection, DNA extraction. Contributed comments on the final draft of the manuscript.

II. Planned analysis, performed bioinformatics work and data analyses and took lead in writing the paper.

III. Planned some of the analysis, performed bioinformatics work and data analyses and took lead in writing the paper.

List of figures

Abbreviations

CV	Coefficient of expression variation
eQTL	Expression quantitative trait locus
GWAS	Genome-wide association study
LD	Linkage disequilibrium
PCA	Principal component analysis
QTL	Quantitative trait locus
RIL	Recombinant inbred line
SI	Selection index
SV	Structural variation
TWAS	Transcriptome-wide association studies
WGD	Whole-genome duplication

1. Introduction

1.1 Natural selection and adaptation

Charles Darwin came up with the idea of natural selection after a five-year voyage to study plants, animals and fossils in South America and the Pacific Islands and brought the idea to the attention of the world in his best-selling book ‘On the Origin of Species’ (1859). Natural selection is one of the basic mechanisms of evolution and is important to understand the origins of living things exhibiting their diversity and complexity. Natural selection can act on any heritable phenotypic trait, and any aspect of the environment can create selection pressure, including sexual selection and competition with members of the same or other species (Arnqvist & Rowe, 2005; Emlen & Zimmer, 2019; Miller, 2011).

Adaptation is the dynamic evolutionary process of natural selection that allows organisms to thrive in their local environments, thereby enhancing their evolutionary fitness. It can also be considered a phenotypic trait that has a functional role in every organism, maintained and evolved through natural selection. The rediscovery of Mendel's work in 1900 showed that heritable variation was controlled by discrete genes. Loci, genes and even individual mutations responsible for phenotypic variations are considered to be the potential targets of natural selection (Bomblies & Peichel, 2022). Therefore, understanding the genetic basis of adaptive traits has become a vital requirement for understanding various hypotheses regarding the genetics of adaptation (Phillips, 2005). The genes behind many ecologically relevant traits have been identified, such as skeletal traits associated with predation and foraging in sticklebacks (Cleves et al., 2014; Colosimo et al., 2005; Erickson, Baek, Hart, Cleves, & Miller, 2018; Shapiro et al., 2004), beak shape and size of Darwin's finches (Lamichhaney et al., 2015; Lamichhaney et al., 2016), and flowering time in *Arabidopsis* and its relatives (Gaudinier & Blackman, 2020).

Many aspects of plant growth and development show fine-tuned local adaptation, and this is particularly true for the transition from vegetative growth to flowering. Daylight is a highly predictable external cue that allows organisms to predict seasonal changes and regulate their biological functions accordingly. The length of the light cycle within a 24-hour time frame, also known as photoperiod, regulates many aspects of plant growth (Osnato, Cota, Nebhnani, Cereijo, & Pelaz, 2022; Roux, Touzet, Cuguen, & Le Corre, 2006). One of the important plant features affected by the duration of the day is flowering time. In many plant species, flowering requires exposure to specific photoperiods and/or temperatures and flowering may be delayed or prevented when these requirements are not met. As plants encounter environments that subject them to novel conditions, genetic adaptations that modify, relax or eliminate existing constraints on flowering can enable expansion across environmental gradients and/or climatic regimes. Although flowering time can be seen from one perspective as a relatively simple trait, the transition from vegetative to reproductive stages is accompanied by significant changes to a wide range of other developmental traits, including stem elongation, apical dominance, lateral branching, and resource allocation. Thus, a large number of genes that contribute to the control of flowering time have been identified in *Arabidopsis* and other model plant species (Huang et al., 2017; Huang & Nusinow, 2016), and it has been observed that genetic networks regulating flowering and the transitions between vegetative and reproductive stages are largely conserved across Angiosperms (Pin & Nilsson, 2012).

1.1.1 Directionality of selection

Selection can be classified in several different aspects, such as by its impact on traits, genetic diversity, the stage of the life cycle where it acts, by the unit of selection, or by the resource of competition. When a component of a trait is heritable, selection alters allele frequencies, or variants of the gene responsible for trait variations. Based on its impact on allele frequencies, selection can be categorized into three types: directional, stabilizing and diversifying selection (Rice, 2004). Directional selection is a type of natural selection, where a single phenotype is favored, leading to a continuous shift in the allele frequency to one specific direction. It happens when the environment changes and species phenotypes move to one end of the existing variation spectrum (*General Biology*, 2021). At the genomic level, directional selection makes the allele frequency in population tends to change towards a certain direction. This process may persist until the allele becomes fixed, resulting in the entire population sharing the fitter phenotype (Rieseberg, Widmer, Arntz, & Burke, 2002). Stabilizing selection is more prevalent; it favors intermediate variants by discriminating against individuals with extreme traits, resulting in the majority of the population being well adapted to their local environment. Stabilizing selection also preserves functional genetic elements, such as protein-coding genes or regulatory sequences, over time by selectively discouraging deleterious variants (Charlesworth, Lande, & Slatkin, 1982). Diversifying selection occurs when extreme trait values are favored over intermediate trait values, resulting in the increased phenotypic variation and further divergence, which may ultimately result in speciation.

1.1.2 Domestication and domestication syndrome

Domestication is a multi-generational reciprocal relationship that developed between humans and other organisms, where humans can take over control and care for a species in return for a steady supply of resources, including food. The domestication of plants began about 13,000 to 11,000 years ago with the emergence of grains such as wheat and barley, as well as crops such as lentils, peas, chickpeas and flax in the Middle East (Abbo et al., 2003; Q. Chen, Li, Tan, & Tian, 2021; Fuller, 2008; Weiss & Zohary, 2011). Domesticated food crops are derived from a phylogenetically diverse assemblage of wild ancestors through artificial selection for different traits (Meyer, DuVal, & Jensen, 2012). In plants, domestication has affected morphological genes, such as increasing seed size and preventing seed head shattering. Understanding the adaptive mechanisms under domestication is of critical importance to modern societies, and domestication of plants provides an ideal framework to study evolutionary responses to selection.

During plant domestication, some traits are common to domesticated plants, such as larger grain or fruit size, shorter height, changes in shattering and increasing yield. These shared traits are known as domestication syndrome. It can be subjected to many types of selection, ranging from strong directional selection for traits such as seed or fruit size to diversifying selection for traits like color or flavor (Wedger, Schumann,

& Gross, 2021). Some authors argue there is little evidence for parallelism at the genetic level and that similar traits underlying the domestication syndrome in different species are usually controlled by loci that are not homologous (Glémin & Bataillon, 2009; Martínez-Ainsworth & Tenaillon, 2016). However, there are also inter-species convergence reported. In soybean, a stay-green locus where the *G* gene (*GmG*), encoding a CAAX amino-terminal protease was identified, involving in controlling the pigmentation of the seed coat and is strongly associated with loss of seed dormancy (M. Wang et al., 2018). It is still inconclusive that associations among traits may suggest a common mechanism, or these correlations could result from sequential, rather than simultaneous, changes in traits during the process of domestication (Hansen Wheat, Fitzpatrick, Rogell, & Temrin, 2019).

1.2 Plant mating system divergence and selfing adaptation

Flowering plants exhibit unparalleled reproductive diversity, reflected in different pollination mechanisms and mating systems. One theme that has been consistently explored is the transition in mating systems from outcrossing to selfing, since this shift can significantly influence the genetic structure and adaptive capabilities of populations (Gutiérrez-Valencia, 2022; Karron et al., 2012). Most flowering plants are hermaphroditic, but the proportion of seeds fertilized by self-insemination and outcrossing pollen varies widely between species, ranging from primary self-fertilization to exclusive outcrossing (Whitehead, Lanfear, Mitchell, & Karron, 2018). Plants exhibit plenty of adaptations that promote cross-pollination and reduce the occurrence of self-fertilization, and understanding the mechanisms behind these phenomena is a key goal of evolutionary biology (Barrett & Harder, 1996). Early work on outcrossing rates was mainly based on imprecise estimates in morphological polymorphisms of single gene (Harding & Barnes, 1977; Schemske & Lande, 1985). Today the availability of molecular markers coupled with maximum likelihood multi-locus measures outcrossing rates more accurately, and facilitated a detailed, population-level view of distribution of outcrossing rate variation (Ritland, 2002).

At the genetic level, the reduction in effective population size induced by selfing shapes patterns of genetic diversity, reduce haploid gene flow and effective rates of recombination. These processes interact to reduce genetic variation in selfing species, often to levels substantially below outcrossing relatives (Ingvarsson, 2002; Laenen et al., 2018). However, mating system also alters the rate and dynamics of adaptation since selfing favors fixation of recessive mutations (and disfavours dominant mutations) compared to outcrossing species and also alters the expected fixation times (Hartfield & Glémin, 2014, 2016; Morran, Parmenter, & Phillips, 2009). ‘Haldane’s sieve’ states that dominant beneficial alleles are more likely to fix than recessive ones, because they are more exposed to selection. Selfing organisms are more likely to fix recessive types than outcrossers and not subject to Haldane’s sieve, because selfing rapidly creates homozygotes, increasing all selection acting on mutations. However, longer homozygous tracts in selfers also decrease the ability of recombination of creating new genotypes (Hartfield & Glémin, 2016). Furthermore, interference between different mutations under selection can further reduce the efficacy of selection in low-recombining genetic regions and high selfing rates amplifies the effects of background selection (Hartfield & Glémin, 2016).

Under polygenic adaptation, selfing can also qualitatively change the dynamics of adaptation compared to outcrossing. Below a given selfing threshold, a population attains an ‘outcrossing-like’ equilibrium. Here, highly inbred individuals are quickly removed, allowing only lineages with minimal or no recent history of selfing to endure. Above the threshold, genetic associations cannot be ignored, and the population achieves a 'purged equilibrium'. In this state, the genome predominantly consolidates with intense linkage disequilibrium, and the selection of a few superior haplotypes leads to a substantial reduction in genetic variation (Hartfield, Bataillon, & Glémin, 2017). There is another hypothesis that introducing a self-breeding mechanism to a population at outcross equilibrium is expected to initially transform the dominant and epistatic variance into additive genetic variances, consequently increasing the response to selection (Cockerham, 1984). Once the population reaches a purged equilibrium, the long-term responses to selection may be impaired due to the lack of available genetic variation (Hartfield et al., 2017; Lande & Porcher, 2015). These dynamics have recently been characterized in freshwater snails by experimentally manipulating outcrossing rates, with the surprising result that genetic variation can be lost within a few generations under the selfing regime (Burgarella et al., 2015). Considering the major impact of the transitioning from outcrossing to selfing on genetic diversity, population ecology and dynamics, it is reasonable to anticipate that the change in mating system would also change basic genomic processes, such as gene expression (Cutter, 2019; Thomas et al., 2012). In the *Capsella* genus, mating system change affected gene expression in three tissues. The impact of a transition in mating systems extends beyond reproductive tissues and encompasses both adaptation to selfing and a reduction in selection pressure on traits that were previously constrained (Z. Zhang et al., 2022). A study in Eichhornia species proved that regulatory changes in gene expression may contribute to the transition from outcrossing to selfing. The shift from outcrossing to high levels of self-fertilization may potentially influence the *cis*-regulatory genomic landscape of angiosperm species (Arunkumar, Maddison, Barrett, & Wright, 2016).

1.3 Methods to identify loci underlying adaptation

Phenotypic traits of agronomical, ecological and economical interest are often polygenic, with many elements outlining natural variation between individuals (Cubillos, Coustham, & Loudet, 2012). One central goal in classical genetics is to find how the distinct allelic variants can shape phenotypic diversity through multiple environmental and genetic interactions (Mackay, Stone, & Ayroles, 2009). Much of our understanding of the processes driving domestication comes from a subset of well-studied crops, particularly crops of major economic importance and model crops, i.e. crops that have had their genomes analysed and are transformable. These crops have been critical for developing our fundamental understanding of domestication as a continuum of ongoing processes. Understanding the ecological distribution, population structure, genetic diversity, origins and evolution of a species is of vital importance to germplasm collections conservation and management strategies, especially a crop wild relative (Ford-Lloyd et al., 2011). Most research begin with a phenotype of interest, and then identify underlying genomic regions via genetic analyses, such as quantitative trait locus (QTL) and association mapping (Ross-Ibarra, Morrell, & Gaut, 2007). By using these top-down genetic methods, we can identify potential genes and mutations behind the putatively adaptive phenotypic differences between populations and closely related species. Another approach is to start with the concept of adaptation, using population genetic methods to search for signals of adaptation in whole genome region, and along with traditional molecular methods to move from gene to phenotype (Figure 1).

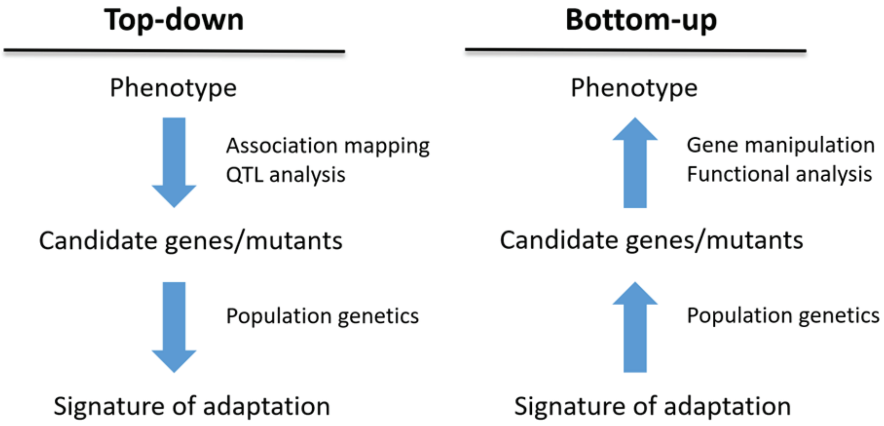


Figure 1. Phenotype–genotype hierarchy represented by top-down and bottom-up approaches (Modified from (Ross-Ibarra et al., 2007)).

1.3.1 Association mapping

Instead of focusing on two parental lines that are strongly differentiated in a phenotype, population-based association mapping searches for linkage between phenotype and genotype in unrelated individuals from natural populations (McCarthy et al., 2008). Theory predicts that evolution occurs primarily through polygenic quantitative traits (Barton & Keightley, 2002). Adaptation is therefore expected to involve rather subtle allele frequency changes at many small effect loci (Le Corre & Kremer, 2003; Pritchard, Pickrell, & Coop, 2010; Rockman, 2012). Classic association studies as well as recent whole genome-wide association studies (GWAS) have confirmed that variation at quantitative traits is due to a large number of loci and few high-effect loci (Mackay et al., 2009; Turner, Stewart, Fields, Rice, & Tarone, 2011; Yang et al., 2010). The development of new genotyping technologies has improved the mapping of quantitative trait loci (QTLs), which are controlled by multiple gene locus (Atwell et al., 2010). The underlying QTLs can be mapped by linkage to polymorphic marker loci with clear Mendelian segregation such as molecular polymorphisms (Cubillos et al., 2012). QTLs can be mapped in unrelated individuals from the same population (association mapping) or in families or segregating progeny of crosses between genetically divergent strains (linkage mapping). In both cases, a large number of individuals are required to detect and localize QTLs. The number of individuals and genotypes per individual needed increases as the precision of localization increases and the QTL effect size decreases (Mackay et al., 2009). GWAS and linkage mapping strategies have identified innumerable numbers of QTLs in plant species. However, only a small amount of essentially large-effect loci originally detected as QTLs have been characterized molecularly (Alonso-Blanco et al., 2009; Trontin, Tisné, Bach, & Loudet, 2011). Furthermore, many resolved QTLs correspond to changes that dramatically change the encoded protein sequence, possibly because transcriptional regulation differences are harder to identify, target and complement. The short of direct methods to relate gene expression variation to phenotypic variation has therefore left its actual contribution to natural variation unappreciated (Cubillos et al., 2012).

To date, we have seen the development of analysis of genome function, especially in genome variation. One prominent direction has been the studies on expression quantitative trait loci (eQTLs), which is the discovery of genetic variants that can explain variation in gene expression levels. eQTLs located close to their eGene (associated gene) are called local or *cis*-eQTLs, those eQTLs farther away from eGenes (usually located on different chromosomes) are called distant or *trans*-eQTLs (Figure 2). Such studies have offered promise not only for the characterization of functional sequence variation but for the understanding of gene regulation and interpretation on GWAS (Nica & Dermitzakis, 2013). Most eQTL mapping studies are aimed to find co-expression network of genes to identify candidate genetic elements contribute to complex phenotypes (Hammond et al., 2011; J. Wang et al., 2010). However, only a handful of studies were intended to find signatures of selection (Bullard, Mostovoy, Dudoit, & Brem, 2010; Fraser et al., 2011) and to understand the evolutionary forces acting on heritable expression traits (Rockman, Skrovanek, & Kruglyak, 2010), makes it an interesting question to explore.

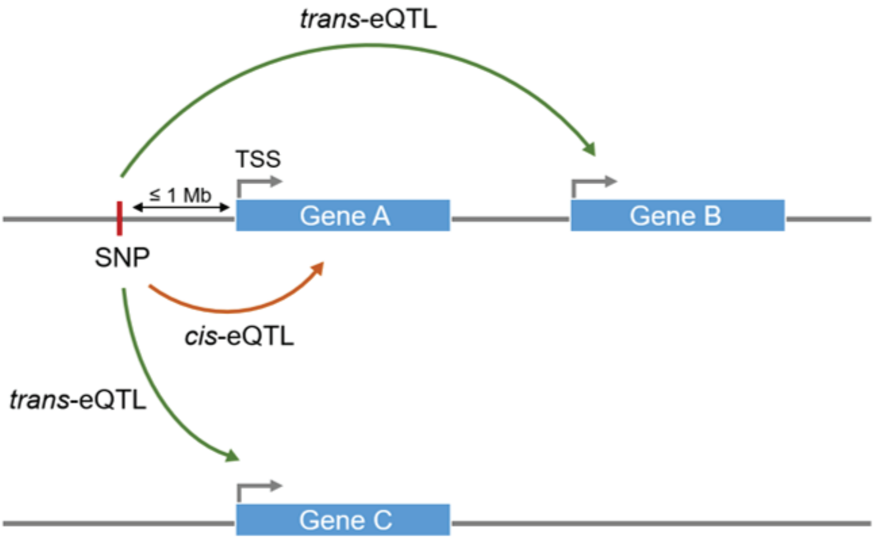


Figure 2. Schematic of *cis-*/*trans-* eQTL regulation.

1.3.2 Population genetics

Depending on the phenotypic-genotype association, many phenotypic changes are considered adaptive, especially in domesticated organisms where selection is strong, and the direction of selection can be inferred. However, many divergences between domesticated species and their progenitors may not be adaptive. For example, QTLs decreasing protein content in wheat are unlikely to have been directly selected during domestication (Ross-Ibarra et al., 2007; Uauy, Distelfeld, Fahima, Blechl, & Dubcovsky, 2006). Thus, it is important to test adaptive hypotheses by using population genetic methods.

In a selective sweep, a new, beneficial mutation increases in frequency and eventually becomes fixed in the population, resulting in the reduction or elimination of genetic variation among nucleotide sequencies that are linked to the mutation. Selective sweeps are expected to generate a deviation in the allele frequency distribution from neutral expectations (Braverman, Hudson, Kaplan, Langley, & Stephan, 1995; Fay & Wu, 2000) and an increase of linkage disequilibrium (LD)(Kim & Nielsen, 2004; Przeworski, 2002; Thomson, 1977). Positive selection causes the new mutation to reach fixation so quickly that linked alleles can ‘hitchhike’ and also become fixed. Almost every plant or animal species has many partially isolated populations, and the genetic divergence between populations is always the result of divergent natural selection or genetic drift (Holsinger & Weir, 2009). The population genetics model of domestication predicts a decrease in diversity and an increased divergence between wild and domesticated populations, because of demographic factors that affect the whole genome, and selection at target loci (Bellucci et al., 2014). Selective sweeps can increase genetic divergence among populations and cause allele frequency spectrum to split from the expectation under neutrality, so it is important to identify regions of the genome that have been the target of selection. Many cross-population approaches also relies on allele frequency spectrum, e.g. the cross-population composite likelihood ratio method (XP-CLR)(H. Chen, Patterson, & Reich, 2010) or LD, e.g. the cross-population extended haplotype heterozygosity test (XP-EHH)(Sabeti et al., 2007; Tang, Thornton, & Stoneking, 2007). Since genomic patterns of natural selection include diverse types of selective sweeps, and existing methods often detect a specific type of sweep, it is reasonable to use multi-methods to detect diverse sweeps.

A critical aspect for the planning of association and population genomics studies is the population used for such an analysis and the level of LD that characterizes the species. The level of LD in natural and domesticated populations is mainly related to population size, effective recombination rate and mating system. A candidate gene approach is usually preferred when LD is low because too many markers will otherwise be needed to perform a whole-genome scan to cover the variation in the entire genomes. When LD is moderate or high, a whole-genome scan can be more appropriate (Rossi et al., 2009). As whole-genome sequencing becomes easier, identifying plenty of genetic markers is not a problem. However, it is still an issue to consider variable LD levels of different populations in these analyses.

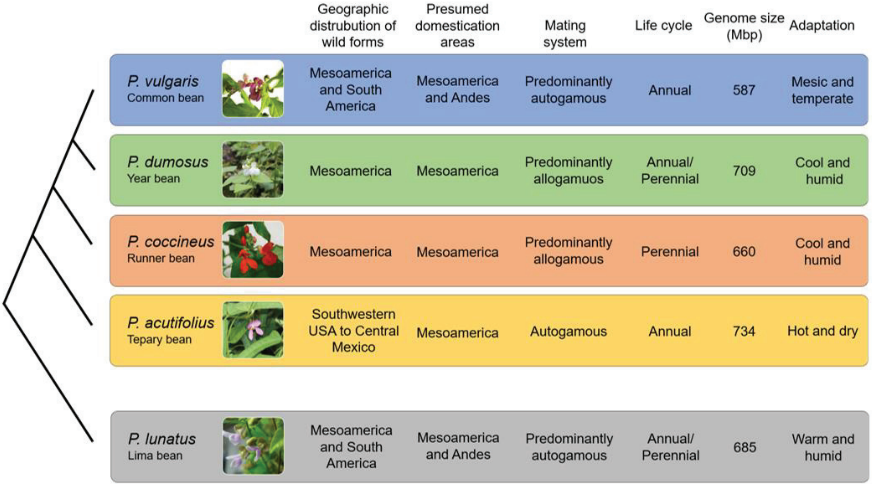
1.3.3 Gene co-expression network and connectivity

Genes and their protein products perform cellular processes, such as signal transmission, in the context of functional modules and are interrelated through a complex network of interactions (Hartwell, Hopfield, Leibler, & Murray, 1999). Understanding the network properties of individual genes or proteins in such networks are proved to be as important as understanding their isolated functions (Barabási & Oltvai, 2004). Gene connectivity in co-expression network is also proved to play an important role in buffering the effect of natural selection. Variation in gene expression has been linked to variation fitness and organismal function (Fraser, Hirsh, Giaever, Kumm, & Eisen, 2004; Z. Wang & Zhang, 2011). Transcriptional variance is consistent across tissues, and low variance genes have higher gene to gene connectivity (Scott, Diogo, Kristina, Luisa, & Julien, 2022). These results have been proved in many species (Cortijo, Bhattarai, Locke, & Ahnert, 2020; Scott et al., 2022). A previous study in *Populus tremula*, an outcrossing species, has also found a weak negative relationship between network connectivity and gene expression variance (Mähler et al., 2017).

1.4 Beans (*Phaseolus* spp.)

Legumes are important in agriculture as they form connections with bacteria that fix nitrogen from the air. This internal fertilization is the main reason that legumes are richer in proteins than other plants (Broughton et al., 2003). In many parts of the world, such as Latin America and parts of Africa, beans are the primary source of protein in human diets. *Phaseolus* is one of the most economically important legume genera containing around 70 plant species, from annual and perennial, herbaceous to woody vines. Species from the genus *Phaseolus* play key roles in mitigating the adverse environmental impacts of agricultural production and sustainable intensification (Vaz Patto et al., 2015).

At least seven domestication events have been documented within the genus *Phaseolus*, affecting five closely related *Phaseolus* spp. (Figure 3). Wild *P. vulgaris* and *P. lunatus* are distributed in Mesoamerica and South America, while *P. coccineus*, *P. dumosus* and *P. acutifolius* are restricted to Mesoamerica (Bitocchi et al., 2017). Two independently domestication have occurred for *P. vulgaris* and *P. lunatus*, different from other three species that have only been domesticated once each. Due to their complex and unique evolution history, *Phaseolus* species can be considered as an important model to study crop evolution. *Phaseolus vulgaris* (common bean) and *Phaseolus coccineus* (runner bean) are two closely related *Phaseolus* species, with different life history characteristics, such as mating system and life cycle. However, they have largely syntenic genomes and show low levels of genomic divergence, which makes the two species an ideal model system for investigating the impacts of mating system on patterns of selection during domestication.



	Geographic distrubution of wild forms	Presumed domestication areas	Mating system	Life cycle	Genome size (Mbp)	Adaptation
P. vulgaris Common bean	Mesoamerica and South America	Mesoamerica and Andes	Predominantly autogamous	Annual	587	Mesic and temperate
P. dumosus Year bean	Mesoamerica	Mesoamerica	Predominantly allogamuos	Annual/ Perennial	709	Cool and humid
P. coccineus Runner bean	Mesoamerica	Mesoamerica	Predominantly allogamous	Perennial	660	Cool and humid
P. acutifolius Tepary bean	Southwestern USA to Central Mexico	Mesoamerica	Autogamous	Annual	734	Hot and dry
P. lunatus Lima bean	Mesoamerica and South America	Mesoamerica and Andes	Predominantly autogamous	Annual/ Perennial	685	Warm and humid

Figure 3. Phylogenetic relationships of the five domesticated Phaseolus species, along with a comparison of their features (Modified from (Bitocchi et al., 2017)).

1.4.1 Common bean

Common bean (*Phaseolus vulgaris*) is the most important grain legume for direct human consumption worldwide. It is a food legume of social importance for humans, providing protein, valuable micronutrients and complex carbohydrates for more than 300 million people in the tropics (OECD, 2019). It is adapted to warm temperatures (temperate and mesic soil temperatures), and lower altitudes, with a rainfall of ~1,100 mm/year (Bitocchi et al., 2017). Common bean and soybean (*G. max*) diverged around 19.2 million years ago, and they share a whole-genome duplication (WGD) event around 56.5 million years ago (Lavin, Herendeen, & Wojciechowski, 2005). Soybean experienced an additional, independent WGD around 10 million years ago (Schmutz et al., 2010). Analysis of synonymous changes in coding sequences (Ks) within and between these genomes showed that common bean has evolved more rapidly than soybean since they split from their last common ancestor (Schmutz et al., 2014).

Early studies have shown that domestication in common bean induced a severe bottleneck. It has been reported that a around 20% reduction in genetic variation was found in the Mesoamerican common bean following domestication (Schmutz et al., 2014). This pattern has been described in other crops such as maize (Hufford et al., 2012), soybean (Li et al., 2013) and in cultivated Agave species (Eguiarte et al., 2013). The population of wild common bean in Andes is derived from the that in Mesoamerica, resulting in parallel domestication events in the two locations (Gaut, 2014). There was also an Andean-specific bottleneck that predated domestication. The wild Andean gene pool diverged from the wild Mesoamerican gene pool around 165 kya, coinciding with a strong bottleneck and a small founding population that lasted around 76 kya. Following the bottleneck, the species went through an exponential growth phase extending to the present day (Rendón-Anaya et al., 2017; Schmutz et al., 2014). The occurrence of a strong bottleneck in the Andean wild population before domestication suggest a Mesoamerican origin of common bean and support the occurrence of a single domestication in Mesoamerica and a similar scenario in the Andes (Rossi et al., 2009). In the early 1500s, through the Columbian exchange, common bean was introduced to Portugal and Spain and then quickly spread all over the Europe. These events made common bean to adapt to new agri-ecological conditions, where it had to adapt longer day lengths and cooler temperatures, and which also turned Europe into a secondary center of genetic diversity. The hybridization of Mesoamerican and Andean gene-pools in Europe, further enhanced genetic diversity and creating novel genetic variation (Bellucci et al., 2023). These events make common bean a perfect model to study mechanisms of domestication and adaptation to photoperiod.

1.4.2 Runner bean

Runner bean (*Phaseolus coccineus*) is a legume species closely related to common bean which is native to Mexico, Guatemala and Honduras (Salinas, 1988). Earlier studies used molecular markers to investigate the domestication history of scarlet runner bean, suggesting the occurrence of multiple domestication events. However, through the analysis of chloroplast and nuclear SSRs of runner bean accessions, including Mesoamerican landraces, wild samples from Mexico, Guatemala and Honduras, and European domesticated populations, the runner bean domestication event was proved to have happened in the Guatemala-Honduras area. Alternatively, another domestication event took place in Mexico, followed by extensive hybridization with the cultivated populations from Honduras and Guatemala (Angioi et al., 2009; Guerra-García, Suárez-Atilano, Mastretta-Yanes, Delgado-Salinas, & Piñero, 2017; Rodriguez et al., 2013; Spataro et al., 2011). One interesting feature of the runner bean domestication history is that the domestication process has not eroded levels of genetic variation to the same extent as in common bean, since both cultivated and wild populations in Central Plateau of Mexico and southern state of Chiapas maintain large and similar amounts of genetic variation (Escalante, Coello, Eguiarte, & Piñero, 1994). Similar results have also been reported in wild and cultivated populations from Mesoamerica (Rodriguez et al., 2013; Spataro et al., 2011). The predicted reduction in genetic diversity and an increased divergence between domesticated and wild forms was not found, suggesting that the domestication bottleneck either was not very severe or that there has been continued gene flow from wild population into the domesticated varieties (Meyer & Purugganan, 2013).

Common bean tends to be selfing and high levels of LD has been observed in both gene pools, mostly due to the gene pool structure, with a much higher LD in domesticated compared with wild populations. However, unlike common bean, runner bean is an open-pollinated species, with high levels of population structure and genetic diversity and which has not suffered a strong demographic bottleneck because of domestication. These features make them a good model to study how mating system affect adaptation and the probability of detecting selective sweep signal within the two species.

2. Aims of the study

The main aims of this thesis were to investigate the genetic architecture of adaptations in two closely related species, common bean and runner bean, to determine how mating system has influenced patterns of adaptation and gene expression variation. Specific objectives of the studies described in Papers I-III were to:

- Detect the signatures of domestication and adaptation and study the genomic background of photoperiod sensitivity in common bean (Paper I)
- Determine if and how mating system has influenced patterns of adaptation and gene expression variation in the two *Phaseolus* species (Paper II)
- Assess gene expression divergence between two species and determine to what extent this divergence has been driven by natural selection (Paper III)

3. Results and discussion

3.1 Parallel evolution of photoperiod sensitivity during domestication and diversification in common bean

The loss of photoperiod sensitivity has been a major pre-requisite for domestication and further adaptation of common bean to European climates. The complexity of such trait has been studied for decades, mostly under the optics of QTL mapping using recombinant inbred lines (RILs), and while independent origins of photoperiod sensitivity have been suggested in each gene pool, the genetic basis of the trait remains elusive. We collected 232 common bean accessions from different geographic origins, including Mesoamerica, the Andes, and Europe, and covering commercial accessions, land races and wild collected individuals from the centers of origin. All accessions were grown in phytotron chambers under the following regime: 10hr light/14hr darkness, 50% humidity, and temperature set at 18℃ in darkness and 20℃ in light. In parallel, the entire collection was grown in green-house conditions during the summer, which means that plants were exposed to very long days (>16hrs of light) and temperatures as high as >28℃. Phenotypic traits showed that wild accessions have small pods, dark flower colors and were all sensitive to daylength, i.e., they did not flower under long day regimes. In the domesticated accessions, we see an increase in pod size, lighter flower color and photoperiod insensitivity also starts to emerge (Paper I, Figure S1), displaying a very clear gradient of traits from wild to elite cultivars, in accordance with the expected behavior of wild vs. cultivated plants. However, the European accessions showed a dramatic increase in photoperiod insensitivity, as well as the prevalence of white flowers and very large pods. Overall, these results highlight the domesticated phenotypes of the accessions and formed the basis for the downstream association studies.

3.1.1 Population structure and selective sweep

We first dissected the signatures of domestication and adaptation to photoperiod through scans of positive selection. Nucleotide diversity, inbreeding coefficient (*F*) and LD decay patterns in each subpopulation (Paper I, Figure S3, S4A) indicate that domesticated accessions always have lower diversity, higher inbreeding rate, and extended linkage disequilibrium compared with their wild relatives. We then identified regions targeted by positive selection during domestication in the Americas and adaptation to European conditions by using both site frequency spectrum and linkage disequilibrium-based approaches and achieved potential windows under positive selection. We found no overlap between selective sweep regions between gene-pools. Functional enrichment analyses performed show that DNA repair and cellular response to DNA damage stimulus, as well as protein dephosphorylation were significantly enriched in both domesticated gene pools (Paper I, Figure 2C). In the outliers associated with adaptation to Europe, we observed similar categories such as carbohydrate, nitrogen and sulfur compound metabolic processes (Figure 2D). Only accessions with a Mesoamerican domesticated background were enriched for protein phosphorylation at XP-CLR outliers, probably because of the low power to detect selective sweeps in the genomes of plants with an Andean origin, due to the substantially reduced diversity seen in these accessions.

3.1.2 Photoperiod associated haplotypes

We studied the genomic architecture of photoperiod sensitivity by means of GWAS and haplotype structure analyses. The resolution of our comparisons allowed us to identify genomic signals specific to each gene-pool and to identify the genetic basis of convergent evolution behind the loss of photoperiod sensitivity. We identified an event of convergent evolution around a key regulator of photoperiod sensitivity, *TERMINAL FLOWER 1 (TFL1)*, on chromosome 1. *TFL1* is a floral repressor which is closely related to the florigen gene *FLOWERING LOCUS T* (*FT*)(Corbesier et al., 2007). Furthermore, since structural variation (SV) have been shown to play a role in plant resistance and immunity, we also looked for gene models that overlap with gene pool specific predictions of SVs and searched for functional enrichments among those genes to further understand the role of SVs in the domestication and adaptation processes. We were able, for the first time, to associate structural variants to traits that have been important during the domestication process by identifying gene pool-specific SVs in the common bean.

Overall, our gene pool-specific signals of selection indicate that the loss of photoperiod sensitivity is a progressive event that launched with domestication in the Americas, followed by recent genomic innovations after the introduction of common bean in Europe. Although there might be some inadequacy in our experiments where, for example, temperature could also be a factor influencing flowering time, we still have provided interesting results that highlights the importance of convergent evolution behind the emergence of such a complex trait as photoperiod insensitivity and provide further evidence on the use of similar genomic strategies to achieve a common phenotype.

3.2 Adaptation and gene expression variation in selfing and outcrossing species

To determine if and how mating system has influenced patterns of adaptation and gene expression variation in common bean and runner bean, we investigate the genetic architecture of gene expression variation in the two species. We selected 98 unrelated common bean accessions, together with 96 runner bean accessions to carry out these analyses. With the help of high-throughput sequencing technology, we performed both re-sequencing and RNA-sequencing analyses, to gain a global insight on how mating system affect adaptation in two species.

3.2.1 Divergence at the genome level

Self-fertilization is known to have many consequences for a species such as reducing the effective population size, effective recombination rates, altering emergence probabilities of new mutations and can give rise to different adaptative trajectories in selfing and outcrossing species (Hartfield et al., 2017). As two gametes are inherited from the same parent in selfing species, homozygosity increases and weakens the mixing effects of recombination (Nordborg, 2000). Thus, at the genomic level, we calculated linkage disequilibrium (LD), inbreeding coefficient, the proportion of long runs of homozygosity (F_{roh}) and recombination rate, which can represent the extent of inbreeding in each species, respectively. LD, inbreeding coefficient and F_{roh} were calculated for each accession, whereas recombination rate was calculated in 20 kb windows across the genome. We observed different patterns between the two species. In common bean, a substantially slower LD decay was detected. It also displayed higher F_{roh} and inbreeding coefficients and hence lower effective recombination rates compared to the outcrossing runner bean (Paper II, Figure 1). These results are consistent with our expectations and that these differences reflect the difference of their mating system. Furthermore, we found a positive correlation between local recombination rates between two species. Homologous recombination rates varied extensively across chromosome and was positively correlated with gene density, but negatively correlated with transposon density, GC content and distribution of structural variation (Ma et al., 2023).

The reduction in effective population size induced by selfing shapes patterns of genetic diversity, reduce haploid gene flow and effective rates of recombination. These processes interact to reduce genetic variation in selfing species to levels substantially below outcrossing relatives (Ingvarsson, 2002; Laenen et al., 2018). A study on two groups of freshwater snails with different mating system proved that in the selfer compared with the outcrosser, a strongly reduced genetic diversity, decreased efficacy of purifying selection, slower rate of adaptive evolution, and weakened codon usage bias/GC-biased gene conversion in full agreement with theoretical expectations (Burgarella et al., 2015). In plants, a very limited genetic variation was found in selfing species *C. rubella*, compared with closely related self-compatible relative (Y.-L. Guo et al., 2009). We calculated nucleotide diversity within each gene region in the two species and as expected we detect a higher nucleotide diversity level in the outcrossing species (Paper II, Figure 4G). Furthermore, a positive correlation between recombination rate and nucleotide diversity was found in both species. This result is in accordance with findings in many other species and this pattern is often attributed to the action of natural selection, suggesting that natural selection plays a vital role in the maintenance of genetic variation in both species.

3.2.2 Expression divergence at the transcriptomic level and gene connectivity

Further we wanted to investigate the impact of mating system on gene expression levels as plant mating systems may also impact the adaptation of polygenic traits. Under directional selection, genetic variance is mostly conserved, and the trait mean changes linearly through time in outcrossing species, while genetic variance of selfing species initially increases much more rapidly than in the outcrossing population (Hartfield & Glémin, 2016). We planted 196 bean accession (98 in common bean and 96 in runner bean) in the greenhouse. To minimize the effects of any morphological divergence between the two species, we selected primary leaves at the seedling stage for RNA-sequencing. After processing raw sequencing reads, we calculated gene expression levels by mapping reads to the common bean genome (v2.1). To account for the divergence between runner bean and the common bean reference genome, gap opening and gap extension penalty parameters were increased to allow more mismatches and gaps when mapping sequencing reads for runner bean. We obtained 98.3% average mapping rate on common bean and 98.1% on runner bean. Transcripts Per Kilobase Million (TPM) were counted for mapped reads, considering the comparisons within/between species. After filtering, 23,625 expressed genes were detected in common bean and 24,007 in runner bean out of a total of 27,433 genes and these were used for further analyses. Principal component analysis (PCA) result (Paper III, Figure 1A) and correlation heatmap (Paper II, Figure S3) shows the divergence of global gene expression between the two species. However, gene expression data display less divergence between species and due to population structure compared with that observed at the genomic level (Paper II, Figure S1). These results indicate that gene expression is likely under stabilizing selection and that the effects of natural selection might be buffered by the organization of the gene co-expression network.

We calculated the coefficient of expression variation (CV) of each gene, which is defined as the standard deviation normalized by the expression mean, to assess divergence of genetic variance between the two species. The distribution of CVs in two species shows that the divergence between two species is not significant (Paper II, Figure S6). We separately built co-expression networks in two species. Based on the gene modules and correlations, we estimated the connectivity of each gene, which is calculated as the sum of correlation coefficients with all other genes within the co-expression network. We detected a negative correlation of CV and connectivity in both species (Paper II, Figure 3). To further investigate the difference in CV, core genes were identified as the top 10% of genes in a module with the highest normalized k_{diff}, the difference between intra- and inter-modular connectivity, while also having an intra-modular connectivity >1 (Mähler et al., 2017). Using this definition, 31 of 35 modules in common bean and 40 of 44 modules in runner bean contained at least one core gene. We compared connectivity and CV of core versus non-core genes and found a higher CV in non-core genes compared to core genes in runner bean (P-value < 0.05), but not in common bean (P-value >0.05)(Paper II, Figure S7). These results may suggest that although gene expression is under stabilizing selection in both species, the structure of co-the expression network may play different roles in regulating gene expression level in the two species.

3.2.3 The challenge of eQTL mapping in selfing and outcrossing species

Expression quantitative trait locus (eQTL) mapping has become as a powerful new approach to study the genetic architecture of gene expression variation in many organisms and provides molecular links between genetic variation and phenotypic diversity (Aguet et al., 2017; Hormozdiari et al., 2018; Lappalainen et al., 2013; Zhu et al., 2016). It has been shown that trait-associated SNPs are three times more likely to be associated with gene expression (Fehrmann et al., 2011; Hernandez et al., 2012; Nica et al., 2010; Nicolae et al., 2010). Such significant enrichment indicates that many SNP-trait associations may be mediated through gene expression. Over the past decade, eQTL mapping has been used to study the genetic architecture of gene expression regulatory variation in major crops and model plants (J. Wang et al., 2010; X. Wang et al., 2018; West et al., 2007; L. Zhang et al., 2017; Xu Zhang, Cal, & Borevitz, 2011). To figure out how eQTLs control gene expression in the two species, we performed eQTL mapping using bi-allelic SNPs with a minor allele frequency (MAF) >0.1 and the gene expression dataset from each species separately. We then grouped significant SNPs for each gene that fell within a maximum distance of 10 kb between two sequential SNPs and only retained clusters with more than three significant SNPs as putative eQTLs. The most significant SNP in the cluster was chosen to represent the eQTL for that region and gene. In previous studies in crops, especially selfing species, LD-clumping of SNPs was a prerequisite before eQTL mapping. Since we have found out that as an outcrossing species, runner bean displays a quite rapid LD decay, we decided to skip this step for runner bean. We further carried out LD clumping for the remaining eQTLs using r^2 >0.2 within 5 Mb regions to account for long-range LD. To further minimize the effect of population structure and hidden confounders during RNA sequencing and to identify the optimal number of covariates, we made a test by carrying out eQTL mapping with 1 to 10 PCs (principal components) derived from population structure and gene expression values to account for hidden confounding variation. Considering those hidden confounders during RNA-seq, such as technical factors, may decrease the number of eQTL identified (Mähler et al., 2017), we ultimately decided to use first 2 PCs from gene expression (explaining 36.94% of variation in common bean and 33.45% of variation in runner bean) as covariates, in which case the maximum amount of eQTLs could be detected. Meanwhile, in order to remove the possibly confounding effects of population structure (Paper II, Figure S1), we use also first included the first two PCs from population structure (explaining 19.24% of variation in common bean and 12.21% of variation in runner bean) as covariates for in our eQTL mapping analyses (Paper II, Figure S4).

It is a challenge to carry out eQTL mapping in both selfing and outcrossing species and there are many aspects to be aware of. By removing potentially hidden confounders, we identified a similar number of eQTLs in two species. In total, we identified 72,230 eQTLs in common bean and 69,108 in runner bean, respectively, at a 5% FDR. Figure 2A and 2B in Paper II display the global map of eQTLs distributed over all 11 chromosomes in two species, respectively. eQTLs can be classified as acting in *cis* or *trans* where *cis*-eQTLs act in an allele-specific manner and are generally considered to be local while *trans* acting eQTLs affect both alleles of a gene and are usually located far away from that gene. Interest in eQTLs has been strong due to their ability to identify mechanistic links between phenotype and genotype (Josephs, Lee, Stinchcombe, & Wright, 2015; Lappalainen, 2015). We find no evident *trans*-eQTL signal in either species, or this is similar to an earlier study in *Populus tremula* (Mähler et al., 2017). Considering the density variation between eQTLs and the corresponding eGene for eQTLs located on the same chromosome as the target eGene, we categorized eQTLs located on the same chromosome as the target gene either as local-eQTL (≤1 Mb) or distant-eQTL (>1 Mb), assuming that most detectable eQTLs located within a chromosome were local (Mähler et al., 2017)(Paper II, Figure 2D). Previous studies have also shown that *cis*-regulation of gene expression plays important roles in the evolution of plant mating systems. For example, in *Capsella rubella*, *cis*-regulatory mutations contributed to the evolution of the selfing syndrome (Steige, Reimegård, Koenig, Scofield, & Slotte, 2015). Strictly speaking, local-eQTL are not the same set as *cis*-eQTL and the same is true for distant-eQTLs and *trans*-eQTLs, but in our case, this could be representative, assuming that most detectable eQTLs located within a chromosome were local (Paper II, Figure 2D). Thus, we carried out Gene Ontology (GO) enrichment analysis for all eGenes to further understand their function, and we found that eGenes regulated by local/distant eQTL were enriched for response to stimulus and signaling (Paper II, Figure S5).

The gene co-expression network structure, *cis*-regulatory variation and tissue specificity have all been shown to play important roles in explaining gene expression variation between individuals, populations and species (Burban, Tenaillon, & Le Rouzic, 2022; Guo et al., 2016; Mähler et al., 2017). First, we found that in both species, CV is negatively correlated with gene connectivity, indicating that genes located in the hub of the network show more stable gene expression levels. We then compared connectivity and CV of local-, distant- and non-eGene and we found that non-eGenes have higher connectivity and lower CV compared with eGenes in both species, indicating that eGenes are under-represented in network module cores, but display a 'stability' in expression variation. Local-eGenes present lower connectivity than distant-eGenes, and this is not detected in CV (Paper II, Figure 3C, 3D). Combining with the result that only in runner bean core genes displayed lower connectivity, we assume that gene connectivity and co-expression network divergence in two species is the main reason of expression variation between two species.

3.2.4 Co-expression networks may influence gene expression variation

Since differential gene expression signatures may be partially explained by genetically encoded gene-specific features, we wonder if nucleotide diversity has any detectable influence on gene expression variance. In both common bean and runner bean, nucleotide diversity shows a weak negative correlation with gene connectivity but a positive correlation with CV (Paper II, Figure

4A-D). Although we detected higher nucleotide diversity in the outcrossing species, its effect on gene expression is limited. Considering CV as an informative metric of the phenotype, we tested measurements that may have an influence on CV, including gene connectivity, core vs. non-core gene status, eGene vs. non-eGene status, expression level, nucleotide diversity and recombination rate. Owing to the collinearity of various characteristics (Paper II, Figure S9), we performed PCA to examine the extent to which these measurements were interdependent. We found that in both species, PC1, which explained the most variance in both species, consisted of a mix of the different measurements. Gene expression level showed a strong influence on PC3 while connectivity of genes in the co-expression network and whether they are located within network module cores or not contributed largely to PC5 (Paper II, Figure 4A, 4C). To determine the correlation between CV and potential genetic and gene expression measurements, we calculated Spearman's rank correlation between CV and PCs. CV was significantly negatively correlated with PC1 in both species, which is not specifically represented by any measurement. PC3, which largely reflected gene expression level, showed significant negative correlations with CV, in accordance with results in other studies (Burban, Tenaillon, & Le Rouzic, 2022), indicating that lowly expressed genes spanned a wide spectrum of CV while highly expressed genes generally had low CV, suggesting that highly expressed genes are under stronger stabilization selection. In addition, in common bean, connectivity and core status of genes in the co-expression network (PC4) is negatively related with CV, which is not observed in runner bean. Although we detected significantly higher recombination rate and nucleotide diversity in outcrossing runner bean, the function of these measures has limited effects in shaping gene expression variation. These results suggest that mating system per se have limited effects, and it's rather the divergence of the structure the gene co-expression network structure that has a greater impact on shaping expression variations in the two species, further demonstrating that gene expression is under stabilizing selection.

Taken together, we built up a global map of gene expression variation at the genomic and transcriptomic level in two closely related *Phaseolus* species with different mating systems, to gain insights into genetic architecture of adaptation and the possible consequences of mating system variation. We observed that mating system has an impact at the genomic level and that selfing tends to increase linkage disequilibrium and homozygosity, reduce nucleotide diversity and effective recombination rate. However, at the transcriptome level, the impacts of mating system variation are limited**.** We calculated CV as a measurement of gene expression variance, and we also calculated gene connectivity and identified eQTLs and pointed out the importance of taking mating system in to account in the eQTL mapping analysis. We measured correlations between these measures and found a positive correlation between nucleotide diversity and recombination rate, representing the effect of natural selection and a negative correlation between connectivity and CV, suggesting expression stabilization of hub genes. We also observed lower connectivity and higher CV in eGenes indicating that they might be either be key regulators or their direct targets such regulators. All these results show the same pattern in both two species. For the gene co-expression networks, we identified core genes in each module but a divergence in CV between core genes and non-core genes was only observed in runner bean, possibly representing the outcome of structural and functional divergence of the co-expression networks in the two species. When we put everything together to assess out which factors that contribute most to gene expression variation, we, as expected, observed that the outcrossing species has higher nucleotide diversity and recombination rates compared with selfing species but thattheir contribution to gene expression variation is nevertheless limited, indicating that gene expression is mostly evolving under stabilizing selection. Although there is no significant CV divergence between core/non-core genes in common bean, the co-expression network structure seems to contribute more to gene expression variation. It will be very interesting to further dig into the co-expression network structure of the two species in order to identify potential genes that may be responsible for any differences in network structure.

3.3 Patterns of between-species divergence in gene expression and the underlying selection signals

Previous studies have shown that gene expression levels and patterns are primarily under stabilizing selection (Blekhman, Marioni, Zumbo, Stephens, & Gilad, 2010; Khaitovich, Enard, Lachmann, & Pääbo, 2006; Romero, Ruvinsky, & Gilad, 2012; Somel et al., 2009) and our results in Paper II also demonstrate this. Next, we decided to compare gene expression-derived estimates of Q_{ST} with genome-wide estimates of nucleotide diversity to assess to what extent genes displaying divergent gene expression between the two species also show signs of the action of natural selection, through e.g. selective sweeps. Q_{ST} is a statistic intended to measure the degree of genetic differentiation among populations or species with regard to a quantitative trait (Spitze, 1993). In our study, we use gene expression levels to calculate estimates of Q_{ST} for gene expression. Q_{ST} was calculated by taking the ratio of genetic variance between species divided by the variance between species plus two times variance within species. Since gene expression estimates at the individual level were non-replicated, we are not able to separate the true estimate of variance within species from residual error variation due to e.g. measurement errors. Including unaccounted error variation will thus lead to slightly inflated estimates of variance within species, but the overall effect of this is to reduce estimates of Q_{ST}, making our Q_{ST} estimates conservative. In total, we derived Q_{ST} values for 22,769 genes with detectable expression. We quantified gene expression divergence between the two species using Q_{ST} and identified 228 outlier genes, falling in the top 1% of Q_{ST} values (Paper III, Figure 1B, 1C). These genes were considered to show divergent expression patterns between the two species. The outlier genes are spread across all 11 chromosomes with no obvious clustering. Furthermore, GO enrichment analysis showed that genes with highly diverged gene expression were mainly related to lipid biosynthetic process, silencing by RNA and negative regulation of gene expression (Paper III, Supplementary Table 1).

Then we wanted to assess selective sweep signatures at the genomic level of these highly diverged genes, and we therefore calculated F_{ST}, the cross-population composite likelihood ratio test (XP-CLR) (Chen, Patterson, & Reich, 2010) and the cross-population extended haplotype homozygosity (XP-EHH, (Sabeti et al., 2007; Tang, Thornton, & Stoneking, 2007) to identify and quantify the extent of selective sweeps acting on these genes. We compared highly diverged genes with background variation across the genome using similar sized sets of randomly sample genes. We observed that the Q_{ST} outlier genes did not deviate significantly from the genome-wide background variation for either F_{ST} or XP-EHH value (Paper III, Figure 2A, 2C) but have significantly lower XP-CLR scores compared to the genome-wide background (Paper III, Figure 2B). These results suggest that genes displaying enhanced differentiation in gene expression between the two species are in fact less likely than randomly selected genes to be influenced by selective sweeps at the genomic level, indicating that other mechanisms than selective sweeps may be responsible for their expression divergence. To assess these results, we further investigated D_{XY} of the Q_{ST} outlier genes and observed that these genes have significantly higher D_{XY} values compared with the genomic background, indicating that absolute sequence divergence may impact expression differentiation. Our results indicate that selective sweeps act less frequently on genes with high gene expression divergence. Meanwhile our result demonstrated that patterns of D_{XY} are more conserved than variation in within-species diversity, in accordance with a study in two flycatcher species (Chase, Ellegren, & Mugal, 2021). We therefore suggest that D_{XY} could be a more representative statistic for assessing divergence in gene expression in some cases. We also calculated an index for selection on gene expression (SI), which is based on the widely used McDonald-Kreitman and adapted for assessing the mode and strength of selection operating on gene expression levels (Warnefors & Eyre-Walker, 2012). This index is also calculated based on gene expression variance between species and variance within species while also taking D_{XY} and nucleotide diversity into account. We found that the SI statistic was significantly correlated with Q_{ST} (Paper III, Figure 3B) and this correlation likely results from the two statistics being based on the same estimates for gene expression variation within and between species. However, the two index have slightly different interpretations. A negative SI value indicates stabilizing selection, whereas a positive SI value suggests directional selection on gene expression (Warnefors & Eyre-Walker, 2012). We found that most genes exhibited negative SI values which are again indicative of stabilizing selection acting on gene expression in these genes, providing further evidence that gene expression is generally evolving under strong stabilizing selection. The genes displaying positive SI values were considered as candidates for evolving under directional selection and they are largely overlapped with genes displaying high divergence in Q_{ST}, including the genes shown to be related to lipid biosynthetic process, silencing by RNA and negative regulation of gene expression. Together these results indicate that directional selection may play an important role in the divergence of gene expression between the two species. This result is in accordance with previous observations in rice, that a large fraction of differential expressed genes between two species evolve under directional selection (J. Guo et al., 2016).

With the assumption that directional selection plays important roles for gene expression divergence in two species, we further investigated if directional selection acted specifically on regulatory regions. Since *cis*-regulatory mutation have been shown to play important roles in mediating gene expression variation between individuals, populations and species, we choose to estimated ratios of polymorphism to divergence (π/D_{XY}) for 5′ flanking and 3′ flanking regions and compared these to values from the coding regions of corresponding genes. Assuming that regions falling in the lowest 5% tail of the π/D_{XY} ratio have signatures indicating recent directional selection (Guo et al., 2016; Holloway, Lawniczak, Mezey, Begun, & Jones, 2007), we found no significant enrichment of regions under directional selection in either species for genes that show strong expression divergence (Paper III, Figure 4). To further confirm this result, we performed the same comparison for F_{ST}, XP-CLR and XP-EHH, separately. Again, we did not find significance between highly diverged genes and background genes, no matter whether we looked at gene regions or *cis*-regulatory regions (Paper III, Figure S4). Taken together, although we find evidence for gene expression divergence between the two *Phaseolus* species being induced by directional selection, we cannot identify any clear selective signatures at the genomic level, neither in coding regions nor on the corresponding *cis*-regulatory regions. Based on the function enrichment result of highly diverged genes, we have an idea that modifications at the epigenetic level might give some further insights, and this clearly deserves further studies.

4. Future perspectives

Common bean, an important food legume, is of great significance to agricultural production. Meanwhile, its migration and domestication history make it a great model system for studying the underlying mechanisms of plant adaptation. Flowering time is an important life history feature that coordinate life cycle and local environment conditions, and it is a prerequisite for production of seeds or fruits for harvest. Research on the genetic basis of flowering time in *Arabidopsis* has revealed a complex genetic network on acting environmental and internal signals. It will be meaningful to focus on pleiotropic effects on agronomically important characters, and any downstream regulators. In addition, further verifications of the candidate genes we have identified, using from controlled experiment would be an obvious next step.

We have demonstrated that gene expression is under stabilizing selection and that mating system has minor effect on gene expression variance and natural selection on gene expression. However, at the genomic level it would be important to take further factors of population genetics into consideration, such as the demographic history of the two species, including bottlenecks, genetic drift and gene flow between populations, in order to build a more detailed overview on the population genetic processes affecting variation in the two species. Our eQTL mapping results could also be extended by, for example, performing transcriptome-wide association studies (TWAS). TWAS is an instrumental post-analysis designed to detect significant gene-trait associations by focusing on the link between transcription-level regulation and specific traits. By using information on the regulation of eQTLs, TWAS can detect how functional genes are regulated as have been shown for disease-associated variants in humans which has provided important insights into the mechanisms of several diseases (Mai, Lu, Gao, Zeng, & Xiao, 2023). Our results provide an opportunity for this integrative approach of TWAS. Furthermore, our RNA-sequencing analysis was carried out by only focusing on the global divergence between two species. However, RNA-sequencing from different tissues, environmental conditions or even across time can provide further information, and lead to a more thorough understanding of the underlying genes and mechanisms. With the help of new RNA-seq technologies, e.g. Iso-seq, gene expression divergence between species can also be detected more precisely.

Since our result shows that RNA silencing might be an important function inducing expression divergence between the two species, further studies at the epigenetic level would also be fruitful. DNA methylation is an important epigenetic modification involved in the regulation of gene expression. Most genes have higher degree of methylation within gene regions but it has been demonstrated that some transcriptionally silenced genes also have higher methylation in the 5’ flank regions (Lister et al., 2009; Xiaoyu Zhang et al., 2006; Zilberman, Gehring, Tran, Ballinger, & Henikoff, 2007). With the help of methylation sequencing, we expect a more comprehensive understanding of genomic regulation and genetically regulated genes underlying complex traits.

References

Abbo, S., Shtienberg, D., Lichtenzveig, J., Lev, x, Yadun, S., & Gopher, A. (2003). The Chickpea, summer cropping, and a new model for pulse domestication in the ancient Near East. *The Quarterly Review of Biology, 78*(4), 435-448. doi:10.1086/378927

Aguet, F., Brown, A. A., Castel, S. E., Davis, J. R., He, Y., Jo, B., . . . Biospecimen Collection Source Site—, N. (2017). Genetic effects on gene expression across human tissues. *Nature, 550*(7675), 204-213. doi:10.1038/nature24277

Alonso-Blanco, C., Aarts, M. G. M., Bentsink, L., Keurentjes, J. J. B., Reymond, M., Vreugdenhil, D., & Koornneef, M. (2009). What has natural variation taught us about plant development, physiology, and adaptation? *The Plant Cell, 21*(7), 1877. doi:10.1105/tpc.109.068114

Angioi, S. A., Desiderio, F., Rau, D., Bitocchi, E., Attene, G., & Papa, R. (2009). Development and use of chloroplast microsatellites in *Phaseolus* spp. and other legumes. *Plant Biology, 11*(4), 598-612. doi:10.1111/j.1438-8677.2008.00143.x

Arnqvist, G., & Rowe, L. (2005). *Sexual Conflict*: Princeton University Press.

Arunkumar, R., Maddison, T. I., Barrett, S. C. H., & Wright, S. I. (2016). Recent mating-system evolution in *Eichhornia* is accompanied by *cis*-regulatory divergence. *New Phytologist, 211*(2), 697-707. doi:https://doi.org/10.1111/nph.13918

Atwell, S., Huang, Y. S., Vilhjálmsson, B. J., Willems, G., Horton, M., Li, Y., . . . Nordborg, M. (2010). Genome-wide association study of 107 phenotypes in *Arabidopsis thaliana* inbred lines. *Nature, 465*(7298), 627-631. doi:10.1038/nature08800

Barabási, A.-L., & Oltvai, Z. N. (2004). Network biology: understanding the cell's functional organization. *Nature Reviews Genetics, 5*(2), 101-113. doi:10.1038/nrg1272

Barrett, S. C. H., & Harder, L. D. (1996). Ecology and evolution of plant mating. *Trends in Ecology & Evolution, 11*(2), 73-79. doi:https://doi.org/10.1016/0169-5347(96)81046-9

Barton, N. H., & Keightley, P. D. (2002). Understanding quantitative genetic variation. *Nature Reviews Genetics, 3*(1), 11-21. doi:10.1038/nrg700

Bellucci, E., Benazzo, A., Xu, C., Bitocchi, E., Rodriguez, M., Alseekh, S., . . . Papa, R. (2023). Selection and adaptive introgression guided the complex evolutionary history of the European common bean. *Nature communications, 14*(1), 1908. doi:10.1038/s41467-023-37332-z

Bellucci, E., Bitocchi, E., Ferrarini, A., Benazzo, A., Biagetti, E., Klie, S., . . . Papa, R. (2014). Decreased nucleotide and expression diversity and modified coexpression patterns characterize domestication in the common bean. *The*

Plant Cell, 26(5), 1901-1912. doi:10.1105/tpc.114.124040

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

Bomblies, K., & Peichel, C. L. (2022). Genetics of adaptation. *Proceedings of the National Academy of Sciences, 119*(30), e2122152119. doi:10.1073/pnas.2122152119

Braverman, J. M., Hudson, R. R., Kaplan, N. L., Langley, C. H., & Stephan, W. (1995). The hitchhiking effect on the site frequency spectrum of DNA polymorphisms. *Genetics, 140*(2), 783-796. doi:10.1093/genetics/140.2.783

Broughton, W. J., Hernández, G., Blair, M., Beebe, S., Gepts, P., & Vanderleyden, J. (2003). Beans (*Phaseolus* spp.) – model food legumes. *Plant and Soil, 252*(1), 55-128. doi:10.1023/A:1024146710611

Bullard, J. H., Mostovoy, Y., Dudoit, S., & Brem, R. B. (2010). Polygenic and directional regulatory evolution across pathways in *Saccharomyces*. *Proceedings of the National Academy of Sciences, 107*(11), 5058. doi:10.1073/pnas.0912959107

Burban, E., Tenaillon, M. I., & Le Rouzic, A. (2022). Gene network simulations provide testable predictions for the molecular domestication syndrome. *Genetics, 220*(2), iyab214. doi:10.1093/genetics/iyab214

Burgarella, C., Gayral, P., Ballenghien, M., Bernard, A., David, P., Jarne, P., . . . Glémin, S. (2015). Molecular evolution of freshwater snails with contrasting mating systems. *Molecular biology and evolution, 32*(9), 2403-2416. doi:10.1093/molbev/msv121

Charlesworth, B., Lande, R., & Slatkin, M. (1982). A neo - Darwinian commentary on macroevolution. *Evolution, 36*(3), 474-498. doi:10.1111/j.1558-5646.1982.tb05068.x

Chen, H., Patterson, N., & Reich, D. (2010). Population differentiation as a test for selective sweeps. *Genome research, 20*(3), 393-402.

Chen, Q., Li, W., Tan, L., & Tian, F. (2021). Harnessing knowledge from maize and rice domestication for new crop breeding. *Molecular Plant, 14*(1), 9-26. doi:10.1016/j.molp.2020.12.006

Cleves, P. A., Ellis, N. A., Jimenez, M. T., Nunez, S. M., Schluter, D., Kingsley, D. M., & Miller, C. T. (2014). Evolved tooth gain in sticklebacks is associated with a *cis*-regulatory allele of Bmp6. *Proceedings of the National Academy of Sciences, 111*(38), 13912-13917. doi:10.1073/pnas.1407567111

Cockerham, C. C. (1984). Additive by additive variance with inbreeding and linkage. *Genetics, 108*(2), 487-500. doi:10.1093/genetics/108.2.487

Colosimo, P. F., Hosemann, K. E., Balabhadra, S., Villarreal, G., Dickson, M., Grimwood, J., . . . Kingsley, D. M. (2005). Widespread parallel evolution in sticklebacks by repeated fixation of ectodysplasin alleles. *Science, 307*(5717), 1928-1933. doi:10.1126/science.1107239

Corbesier, L., Vincent, C., Jang, S., Fornara, F., Fan, Q., Searle, I., . . . Coupland, G. (2007). FT protein movement contributes to long-distance signaling in floral induction of *Arabidopsis*. *Science, 316*(5827), 1030-1033. doi:10.1126/science.1141752

Cortijo, S., Bhattarai, M., Locke, J. C. W., & Ahnert, S. E. (2020). Co-expression networks from gene expression variability between genetically identical seedlings can reveal novel regulatory relationships. *Frontiers in Plant Science, 11*. doi:10.3389/fpls.2020.599464

Cubillos, F. A., Coustham, V., & Loudet, O. (2012). Lessons from eQTL mapping studies: non-coding regions and their role behind natural phenotypic variation in plants. *Current Opinion in Plant Biology, 15*(2), 192-198. doi:https://doi.org/10.1016/j.pbi.2012.01.005

Cutter, A. D. (2019). Reproductive transitions in plants and animals: selfing syndrome, sexual selection and speciation. *New Phytologist, 224*(3), 1080-1094. doi:https://doi.org/10.1111/nph.16075

Eguiarte, L. E., Aguirre-Planter, E., Aguirre, X., Colín, R., González, A., Rocha, M., . . . Souza, V. (2013). From isozymes to genomics: population genetics and conservation of *Agave* in México. *The Botanical Review, 79*(4), 483-506. doi:10.1007/s12229-013-9123-x

Emlen, D. J., & Zimmer, C. (2019). *Evolution: Making Sense of Life*: Macmillan Learning.

Erickson, P. A., Baek, J., Hart, J. C., Cleves, P. A., & Miller, C. T. (2018). Genetic dissection of a supergene implicates Tfap2a in craniofacial evolution of threespine sticklebacks. *Genetics, 209*(2), 591-605. doi:10.1534/genetics.118.300760

Escalante, A. M., Coello, G., Eguiarte, L. E., & Piñero, D. (1994). Genetic structure and mating systems in wild and cultivated populations of *Phaseolus coccineus* and *P. vulgaris* (Fabaceae). *American Journal of Botany, 81*(9), 1096-1103. doi:10.1002/j.1537-2197.1994.tb15603.x

Fay, J. C., & Wu, C.-I. (2000). Hitchhiking under positive Darwinian selection. *Genetics, 155*(3), 1405-1413. doi:10.1093/genetics/155.3.1405

Fehrmann, R. S. N., Jansen, R. C., Veldink, J. H., Westra, H.-J., Arends, D., Bonder, M. J., . . . Franke, L. (2011). *Trans*-eQTls reveal that independent genetic variants associated with a complex phenotype converge on intermediate genes, with a major role for the HLA. *PLoS genetics, 7*(8), e1002197. doi:10.1371/journal.pgen.1002197

Ford-Lloyd, B. V., Schmidt, M., Armstrong, S. J., Barazani, O., Engels, J., Hadas, R., . . . Maxted, N. (2011). Crop wild relatives—undervalued, underutilized and under threat? *BioScience, 61*(7), 559-565. doi:10.1525/bio.2011.61.7.10

Fraser, H. B., Babak, T., Tsang, J., Zhou, Y., Zhang, B., Mehrabian, M., & Schadt, E. E. (2011). Systematic detection of polygenic *cis*-regulatory evolution. *PLoS genetics, 7*(3), e1002023-e1002023. doi:10.1371/journal.pgen.1002023

Fraser, H. B., Hirsh, A. E., Giaever, G., Kumm, J., & Eisen, M. B. (2004). Noise minimization in eukaryotic gene expression. *PLOS Biology, 2*(6), e137. doi:10.1371/journal.pbio.0020137

Fuller, D. (2008). Recent lessons from Near Eastern archaeobotany: wild cereal use, pre-domestication cultivation and tracing multiple origins and dispersals.

Pragdhara, 18, 105-134.

Gaudinier, A., & Blackman, B. K. (2020). Evolutionary processes from the perspective of flowering time diversity. *New Phytologist, 225*(5), 1883-1898. doi:https://doi.org/10.1111/nph.16205

Gaut, B. S. (2014). The complex domestication history of the common bean. *Nature Genetics, 46*(7), 663-664. doi:10.1038/ng.3017

General Biology. (2021). LibreTexts.

Glémin, S., & Bataillon, T. (2009). A comparative view of the evolution of grasses under domestication. *New Phytologist, 183*(2), 273-290. doi:https://doi.org/10.1111/j.1469-8137.2009.02884.x

Guerra-García, A., Suárez-Atilano, M., Mastretta-Yanes, A., Delgado-Salinas, A., & Piñero, D. (2017). Domestication genomics of the open-pollinated scarlet runner bean (*Phaseolus coccineus* L.). *Frontiers in Plant Science, 8*, 1891. doi:https://doi.org/10.3389/fpls.2017.01891 doi:10.3389/fpls.2017.01891

Guo, J., Liu, R., Huang, L., Zheng, X.-M., Liu, P.-L., Du, Y.-S., . . . Ge, S. (2016). Widespread and adaptive alterations in genome-wide gene expression associated with ecological divergence of two *Oryza* species. *Molecular biology and evolution, 33*(1), 62-78. doi:10.1093/molbev/msv196

Guo, Y.-L., Bechsgaard, J. S., Slotte, T., Neuffer, B., Lascoux, M., Weigel, D., & Schierup, M. H. (2009). Recent speciation of *Capsella rubella* from *Capsella grandiflora*, associated with loss of self-incompatibility and an extreme bottleneck. *Proceedings of the National Academy of Sciences, 106*(13), 5246-5251. doi:10.1073/pnas.0808012106

Gutiérrez-Valencia, J. (2022). *Genomic studies of mating system variation in flowering plants.* (PhD dissertation), Stockholm University, Retrieved from https://urn.kb.se/resolve?urn=urn:nbn:se:su:diva-210927

Hammond, J. P., Mayes, S., Bowen, H. C., Graham, N. S., Hayden, R. M., Love, C. G., . . . Broadley, M. R. (2011). Regulatory hotspots are associated with plant gene expression under varying soil phosphorus supply in *Brassica rapa*. *Plant Physiology, 156*(3), 1230. doi:10.1104/pp.111.175612

Hansen Wheat, C., Fitzpatrick, J. L., Rogell, B., & Temrin, H. (2019). Behavioural correlations of the domestication syndrome are decoupled in modern dog breeds. *Nature communications, 10*(1), 2422. doi:10.1038/s41467-019-10426-3

Harding, J., & Barnes, K. (1977). Genetics of Lupinus. X. Genetic variability, heterozygosity and outcrossing in colonial populations of *Lupinus succulentus*. *Evolution, 31*(2), 247-255. doi:10.2307/2407746

Hartfield, M., Bataillon, T., & Glémin, S. (2017). The evolutionary interplay between adaptation and self-fertilization. *Trends in Genetics, 33*(6), 420-431. doi:10.1016/j.tig.2017.04.002

Hartfield, M., & Glémin, S. (2014). Hitchhiking of deleterious alleles and the cost of adaptation in partially selfing species. *Genetics, 196*(1), 281-293. doi:10.1534/genetics.113.158196

Hartfield, M., & Glémin, S. (2016). Limits to adaptation in partially selfing species. *Genetics, 203*(2), 959. doi:10.1534/genetics.116.188821

Hartwell, L. H., Hopfield, J. J., Leibler, S., & Murray, A. W. (1999). From molecular to modular cell biology. *Nature, 402*(6761), C47-C52. doi:10.1038/35011540

Hernandez, D. G., Nalls, M. A., Moore, M., Chong, S., Dillman, A., Trabzuni, D., . . . Cookson, M. R. (2012). Integration of GWAS SNPs and tissue specific expression profiling reveal discrete eQTLs for human traits in blood and brain. *Neurobiology of Disease, 47*(1), 20-28. doi:https://doi.org/10.1016/j.nbd.2012.03.020

Holsinger, K. E., & Weir, B. S. (2009). Genetics in geographically structured populations: defining, estimating and interpreting FST. *Nature Reviews Genetics, 10*(9), 639-650. doi:10.1038/nrg2611

Hormozdiari, F., Gazal, S., van de Geijn, B., Finucane, H. K., Ju, C. J. T., Loh, P.-R., . . . Price, A. L. (2018). Leveraging molecular quantitative trait loci to understand the genetic architecture of diseases and complex traits. *Nature Genetics, 50*(7), 1041-1047. doi:10.1038/s41588-018-0148-2

Huang, H., Gehan, M. A., Huss, S. E., Alvarez, S., Lizarraga, C., Gruebbling, E. L., . . . Nusinow, D. A. (2017). Cross-species complementation reveals conserved functions for EARLY FLOWERING 3 between monocots and dicots. *Plant Direct, 1*(4), e00018. doi:https://doi.org/10.1002/pld3.18

Huang, H., & Nusinow, D. A. (2016). Into the evening: complex interactions in the *Arabidopsis* circadian clock. *Trends in Genetics, 32*(10), 674-686. doi:10.1016/j.tig.2016.08.002

Hufford, M. B., Xu, X., van Heerwaarden, J., Pyhäjärvi, T., Chia, J.-M., Cartwright, R. A., . . . Ross-Ibarra, J. (2012). Comparative population genomics of maize domestication and improvement. *Nature Genetics, 44*(7), 808-811. doi:10.1038/ng.2309

Ingvarsson, P. (2002). A metapopulation perspective on genetic diversity and differentiation in partially self-fertilizing plants. *Evolution, 56*(12), 2368-2373.

Josephs, E. B., Lee, Y. W., Stinchcombe, J. R., & Wright, S. I. (2015). Association mapping reveals the role of purifying selection in the maintenance of genomic variation in gene expression. *Proceedings of the National Academy of Sciences, 112*(50), 15390.

Karron, J. D., Ivey, C. T., Mitchell, R. J., Whitehead, M. R., Peakall, R., & Case, A. L. (2012). New perspectives on the evolution of plant mating systems. *Annals of botany, 109*(3), 493-503. doi:10.1093/aob/mcr319

Kim, Y., & Nielsen, R. (2004). Linkage disequilibrium as a signature of selective sweeps. *Genetics, 167*(3), 1513-1524. doi:10.1534/genetics.103.025387

Laenen, B., Tedder, A., Nowak, M. D., Toräng, P., Wunder, J., Wötzel, S., . . . Slotte, T. (2018). Demography and mating system shape the genome-wide impact of purifying selection in *Arabis alpina*. *Proceedings of the National Academy of Sciences of the United States of America, 115*(4), 816-821. doi:10.1073/pnas.1707492115

Lamichhaney, S., Berglund, J., Almén, M. S., Maqbool, K., Grabherr, M., Martinez-Barrio, A., . . . Andersson, L. (2015). Evolution of Darwin’s finches and their beaks revealed by genome sequencing. *Nature, 518*(7539), 371-375. doi:10.1038/nature14181

Lamichhaney, S., Han, F., Berglund, J., Wang, C., Almén, M. S., Webster, M. T., . . . Andersson, L. (2016). A beak size locus in Darwin's finches facilitated character displacement during a drought. *Science, 352*(6284), 470-474. doi:10.1126/science.aad8786

Lande, R., & Porcher, E. (2015). Maintenance of quantitative genetic variance under partial self-fertilization, with implications for evolution of selfing. *Genetics, 200*(3), 891-906. doi:10.1534/genetics.115.176693

Lappalainen, T. (2015). Functional genomics bridges the gap between quantitative genetics and molecular biology. *Genome research, 25*. doi:10.1101/gr.190983.115

Lappalainen, T., Sammeth, M., Friedländer, M. R., 't Hoen, P. A. C., Monlong, J., Rivas, M. A., . . . The Geuvadis, C. (2013). Transcriptome and genome sequencing uncovers functional variation in humans. *Nature, 501*(7468), 506-511. doi:10.1038/nature12531

Lavin, M., Herendeen, P. S., & Wojciechowski, M. F. (2005). Evolutionary rates analysis of Leguminosae implicates a rapid diversification of lineages during the tertiary. *Systematic Biology, 54*(4), 575-594. doi:10.1080/10635150590947131

Le Corre, V., & Kremer, A. (2003). Genetic variability at neutral markers, quantitative trait loci and trait in a subdivided population under selection. *Genetics, 164*(3), 1205-1219. doi:10.1093/genetics/164.3.1205

Li, Y.-h., Zhao, S.-c., Ma, J.-x., Li, D., Yan, L., Li, J., . . . Qiu, L.-j. (2013). Molecular footprints of domestication and improvement in soybean revealed by whole genome re-sequencing. *BMC Genomics, 14*(1), 579. doi:10.1186/1471-2164-14-579

Lister, R., Pelizzola, M., Dowen, R. H., Hawkins, R. D., Hon, G., Tonti-Filippini, J., . . . Ecker, J. R. (2009). Human DNA methylomes at base resolution show widespread epigenomic differences. *Nature, 462*(7271), 315-322. doi:10.1038/nature08514

Ma, X., Fan, L., Zhang, Z., Yang, X., Liu, Y., Ma, Y., . . . Tian, Z. (2023). Global dissection of the recombination landscape in soybean using a high-density 600K SoySNP array. *Plant Biotechnology Journal, 21*(3), 606-620. doi:https://doi.org/10.1111/pbi.13975

Mackay, T. F. C., Stone, E. A., & Ayroles, J. F. (2009). The genetics of quantitative traits: challenges and prospects. *Nature Reviews Genetics, 10*(8), 565-577. doi:10.1038/nrg2612

Mähler, N., Wang, J., Terebieniec, B. K., Ingvarsson, P. K., Street, N. R., & Hvidsten, T. R. (2017). Gene co-expression network connectivity is an important determinant of selective constraint. *PLoS genetics, 13*(4), e1006402. doi:10.1371/journal.pgen.1006402

Mai, J., Lu, M., Gao, Q., Zeng, J., & Xiao, J. (2023). Transcriptome-wide association studies: recent advances in methods, applications and available databases. *Communications Biology, 6*(1), 899. doi:10.1038/s42003-023-05279-y

Martínez-Ainsworth, N. E., & Tenaillon, M. I. (2016). Superheroes and masterminds of plant domestication. *Comptes Rendus Biologies, 339*(7), 268-273. doi:https://doi.org/10.1016/j.crvi.2016.05.005

McCarthy, M. I., Abecasis, G. R., Cardon, L. R., Goldstein, D. B., Little, J., Ioannidis, J. P. A., & Hirschhorn, J. N. (2008). Genome-wide association studies for complex traits: consensus, uncertainty and challenges. *Nature Reviews Genetics, 9*(5), 356-369. doi:10.1038/nrg2344

Meyer, R. S., DuVal, A. E., & Jensen, H. R. (2012). Patterns and processes in crop domestication: an historical review and quantitative analysis of 203 global food crops. *New Phytologist, 196*(1), 29-48. doi:https://doi.org/10.1111/j.1469-8137.2012.04253.x

Meyer, R. S., & Purugganan, M. D. (2013). Evolution of crop species: genetics of domestication and diversification. *Nature Reviews Genetics, 14*(12), 840-852. doi:10.1038/nrg3605

Miller, G. (2011). *The Mating Mind: How Sexual Choice Shaped the Evolution of Human Nature*: Knopf Doubleday Publishing Group.

Morran, L. T., Parmenter, M. D., & Phillips, P. C. (2009). Mutation load and rapid adaptation favour outcrossing over self-fertilization. *Nature, 462*, 350. doi:10.1038/nature08496

Nica, A. C., & Dermitzakis, E. T. (2013). Expression quantitative trait loci: present and future. *Philosophical Transactions of the Royal Society B: Biological Sciences, 368*(1620), 20120362. doi:10.1098/rstb.2012.0362

Nica, A. C., Montgomery, S. B., Dimas, A. S., Stranger, B. E., Beazley, C., Barroso, I., & Dermitzakis, E. T. (2010). Candidate causal regulatory effects by integration of expression QTLs with complex trait genetic associations. *PLoS genetics, 6*(4), e1000895. doi:10.1371/journal.pgen.1000895

Nicolae, D. L., Gamazon, E., Zhang, W., Duan, S., Dolan, M. E., & Cox, N. J. (2010). Trait-associated SNPs are more likely to be eQTLs: annotation to enhance discovery from GWAS. *PLoS genetics, 6*(4), e1000888. doi:10.1371/journal.pgen.1000888

Nordborg, M. (2000). Linkage disequilibrium, gene trees and selfing: an ancestral recombination graph with partial self-fertilization. *Genetics, 154*(2), 923-929. doi:10.1093/genetics/154.2.923

OECD. (2019). *Safety Assessment of Foods and Feeds Derived from Transgenic Crops, Volume 3*.

Osnato, M., Cota, I., Nebhnani, P., Cereijo, U., & Pelaz, S. (2022). Photoperiod control of plant growth: flowering time genes beyond flowering. *12*. doi:10.3389/fpls.2021.805635

Phillips, P. C. (2005). Testing hypotheses regarding the genetics of adaptation. *Genetica, 123*(1), 15-24. doi:10.1007/s10709-004-2704-1

Pin, P. A., & Nilsson, O. (2012). The multifaceted roles of FLOWERING LOCUS T in plant development. *Plant, Cell & Environment, 35*(10), 1742-1755. doi:https://doi.org/10.1111/j.1365-3040.2012.02558.x

Pritchard, J. K., Pickrell, J. K., & Coop, G. (2010). The genetics of human adaptation: hard sweeps, soft sweeps, and polygenic adaptation. *Current Biology, 20*(4), R208-R215. doi:https://doi.org/10.1016/j.cub.2009.11.055

Przeworski, M. (2002). The signature of positive selection at randomly chosen loci. *Genetics, 160*(3), 1179-1189. doi:10.1093/genetics/160.3.1179

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*(1), 60. doi:10.1186/s13059-017-1190-6

Rice, S. H. (2004). *Evolutionary Theory: Mathematical and Conceptual Foundations*: Sinauer.

Rieseberg, L. H., Widmer, A., Arntz, A. M., & Burke, J. M. (2002). Directional selection is the primary cause of phenotypic diversification. *Proceedings of the National Academy of Sciences, 99*(19), 12242-12245. doi:10.1073/pnas.192360899

Ritland, K. (2002). Extensions of models for the estimation of mating systems using n independent loci. *Heredity, 88*(4), 221-228. doi:10.1038/sj.hdy.6800029

Rockman, M. V. (2012). The QTN program and the alleles that matter for evolution: all that's gold does not glitter. *Evolution, 66*(1), 1-17. doi:10.1111/j.1558-5646.2011.01486.x

Rockman, M. V., Skrovanek, S. S., & Kruglyak, L. (2010). Selection at linked sites shapes heritable phenotypic variation in *C. elegans*. *Science, 330*(6002), 372. doi:10.1126/science.1194208

Rodriguez, M., Rau, D., Angioi, S. A., Bellucci, E., Bitocchi, E., Nanni, L., . . . Attene, G. (2013). European *Phaseolus coccineus* L. landraces: population structure and adaptation, as revealed by cpSSRs and phenotypic analyses. *PLOS ONE, 8*(2), e57337. doi:10.1371/journal.pone.0057337

Ross-Ibarra, J., Morrell, P. L., & Gaut, B. S. (2007). Plant domestication, a unique opportunity to identify the genetic basis of adaptation. *Proceedings of the National Academy of Sciences, 104*(suppl 1), 8641. doi:10.1073/pnas.0700643104

Rossi, M., Bitocchi, E., Bellucci, E., Nanni, L., Rau, D., Attene, G., & Papa, R. (2009). Linkage disequilibrium and population structure in wild and domesticated populations of *Phaseolus vulgaris* L. *Evolutionary Applications, 2*(4), 504-522. doi:10.1111/j.1752-4571.2009.00082.x

Roux, F., Touzet, P., Cuguen, J., & Le Corre, V. (2006). How to be early flowering: an evolutionary perspective. *Trends in Plant Science, 11*(8), 375-381. doi:10.1016/j.tplants.2006.06.006

Sabeti, P. C., Varilly, P., Fry, B., Lohmueller, J., Hostetter, E., Cotsapas, C., . . . The, S. N. P. C. (2007). Genome-wide detection and characterization of positive selection in human populations. *Nature, 449*(7164), 913-918. doi:10.1038/nature06250

Salinas, A. D. (1988). Variation, taxonomy, domestication, and germplasm potentialities in *Phaseolus coccineus*. *Genetic Resources of Phaseolus Beans: Their maintenance, domestication, evolution and utilization* (pp. 441-463). Dordrecht: Springer Netherlands.

Schemske, D. W., & Lande, R. (1985). The evolution of self‐fertilization and inbreeding depression in plants. II. empirical observations. *Evolution, 39*(1), 41-52. doi:10.1111/j.1558-5646.1985.tb04078.x

Schmutz, J., Cannon, S. B., Schlueter, J., Ma, J., Mitros, T., Nelson, W., . . . Jackson,

S. A. (2010). Genome sequence of the palaeopolyploid soybean. *Nature, 463*(7278), 178-183. doi:10.1038/nature08670

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*(7), 707-713. doi:10.1038/ng.3008

Scott, W., Diogo, M., Kristina, M. G., Luisa, F. P., & Julien, F. A. (2022). Characterizing the landscape of gene expression variance in humans. *bioRxiv*, 2022.2011.2015.516646. doi:10.1101/2022.11.15.516646

Shapiro, M. D., Marks, M. E., Peichel, C. L., Blackman, B. K., Nereng, K. S., Jónsson, B., . . . Kingsley, D. M. (2004). Genetic and developmental basis of evolutionary pelvic reduction in threespine sticklebacks. *Nature, 428*(6984), 717-723. doi:10.1038/nature02415

Spataro, G., Tiranti, B., Arcaleni, P., Bellucci, E., Attene, G., Papa, R., . . . Negri, V. (2011). Genetic diversity and structure of a worldwide collection of *Phaseolus coccineus* L. *Theoretical and Applied Genetics, 122*(7), 1281-1291. doi:10.1007/s00122-011-1530-y

Spitze, K. (1993). Population structure in *Daphnia obtusa*: quantitative genetic and allozymic variation. *Genetics, 135*(2), 367-374. doi:10.1093/genetics/135.2.367

Steige, K. A., Reimegård, J., Koenig, D., Scofield, D. G., & Slotte, T. (2015). *Cis*-regulatory changes associated with a recent mating system shift and floral adaptation in *Capsella*. *Molecular biology and evolution, 32*(10), 2501-2514. doi:10.1093/molbev/msv169

Tang, K., Thornton, K. R., & Stoneking, M. (2007). A new approach for using genome scans to detect recent positive selection in the human genome. *PLOS Biology, 5*(7), e171. doi:10.1371/journal.pbio.0050171

Thomas, C. G., Li, R., Smith, H. E., Woodruff, G. C., Oliver, B., & Haag, E. S. J. C. B. (2012). Simplification and desexualization of gene expression in self-fertile nematodes. *Curr Biol, 22*(22), 2167-2172.

Thomson, G. (1977). The effect of a selected locus on linked neutral loci. *Genetics, 85*(4), 753-788. doi:10.1093/genetics/85.4.753

Trontin, C., Tisné, S., Bach, L., & Loudet, O. (2011). What does Arabidopsis natural variation teach us (and does not teach us) about adaptation in plants? *Current Opinion in Plant Biology, 14*(3), 225-231. doi:https://doi.org/10.1016/j.pbi.2011.03.024

Turner, T. L., Stewart, A. D., Fields, A. T., Rice, W. R., & Tarone, A. M. (2011). Population-based resequencing of experimentally evolved populations reveals the genetic basis of body size variation in *Drosophila melanogaster*. *PLoS genetics, 7*(3), e1001336. doi:10.1371/journal.pgen.1001336

Uauy, C., Distelfeld, A., Fahima, T., Blechl, A., & Dubcovsky, J. (2006). A NAC gene regulating senescence improves grain protein, zinc, and iron content in wheat. *Science, 314*(5803), 1298-1301. doi:10.1126/science.1133649

Vaz Patto, M. C., Amarowicz, R., Aryee, A. N. A., Boye, J. I., Chung, H.-J., Martín-Cabrejas, M. A., & Domoney, C. (2015). Achievements and challenges in improving the nutritional quality of food legumes. *Critical Reviews in Plant*

Sciences, 34(1-3), 105-143. doi:10.1080/07352689.2014.897907

Wang, J., Yu, H., Xie, W., Xing, Y., Yu, S., Xu, C., . . . Zhang, Q. (2010). A global analysis of QTLs for expression variations in rice shoots at the early seedling stage. *The Plant Journal, 63*(6), 1063-1074. doi:10.1111/j.1365-313X.2010.04303.x

Wang, M., Li, W., Fang, C., Xu, F., Liu, Y., Wang, Z., . . . Tian, Z. (2018). Parallel selection on a dormancy gene during domestication of crops from multiple families. *Nature Genetics, 50*(10), 1435-1441. doi:10.1038/s41588-018-0229-2

Wang, X., Chen, Q., Wu, Y., Lemmon, Z. H., Xu, G., Huang, C., . . . Tian, F. (2018). Genome-wide analysis of transcriptional variability in a large maize-teosinte population. *Molecular Plant, 11*(3), 443-459. doi:https://doi.org/10.1016/j.molp.2017.12.011

Wang, Z., & Zhang, J. (2011). Impact of gene expression noise on organismal fitness and the efficacy of natural selection. *Proceedings of the National Academy of Sciences, 108*(16), E67-E76. doi:10.1073/pnas.1100059108

Warnefors, M., & Eyre-Walker, A. (2012). A selection index for gene expression evolution and its application to the divergence between humans and chimpanzees. *PLOS ONE, 7*(4), e34935. doi:10.1371/journal.pone.0034935

Wedger, M. J., Schumann, A. C., & Gross, B. L. (2021). Candidate genes and signatures of directional selection on fruit quality traits during apple domestication. *American Journal of Botany, 108*(4), 616-627. doi:https://doi.org/10.1002/ajb2.1636

Weiss, E., & Zohary, D. (2011). The neolithic southwest asian founder crops: their biology and archaeobotany. *Current Anthropology, 52*(S4), S237-S254. doi:10.1086/658367

West, M. A. L., Kim, K., Kliebenstein, D. J., van Leeuwen, H., Michelmore, R. W., Doerge, R. W., & St. Clair, D. A. (2007). Global eQTL mapping reveals the complex genetic architecture of transcript-level variation in *Arabidopsis*. *Genetics, 175*(3), 1441-1450. doi:10.1534/genetics.106.064972

Whitehead, M. R., Lanfear, R., Mitchell, R. J., & Karron, J. D. (2018). Plant mating systems often vary widely among populations. *Frontiers in Ecology and Evolution, 6*.

Yang, J., Benyamin, B., McEvoy, B. P., Gordon, S., Henders, A. K., Nyholt, D. R., . . . Visscher, P. M. (2010). Common SNPs explain a large proportion of the heritability for human height. *Nature Genetics, 42*(7), 565-569. doi:10.1038/ng.608

Zhang, L., Su, W., Tao, R., Zhang, W., Chen, J., Wu, P., . . . Kuang, H. (2017). RNA sequencing provides insights into the evolution of lettuce and the regulation of flavonoid biosynthesis. *Nature communications, 8*(1), 2264. doi:10.1038/s41467-017-02445-9

Zhang, X., Cal, A. J., & Borevitz, J. O. (2011). Genetic architecture of regulatory variation in *Arabidopsis thaliana*. *Genome research, 21*(5), 725-733. doi:10.1101/gr.115337.110

Zhang, X., Yazaki, J., Sundaresan, A., Cokus, S., Chan, S. W. L., Chen, H., . . . Ecker, Joseph R. (2006). Genome-wide high-resolution mapping and functional analysis of DNA methylation in *Arabidopsis*. *Cell, 126*(6), 1189-1201. doi:https://doi.org/10.1016/j.cell.2006.08.003

Zhang, Z., Kryvokhyzha, D., Orsucci, M., Glémin, S., Milesi, P., & Lascoux, M. (2022). How broad is the selfing syndrome? Insights from convergent evolution of gene expression across species and tissues in the *Capsella* genus. *New Phytologist, 236*(6), 2344-2357. doi:https://doi.org/10.1111/nph.18477

Zhu, Z., Zhang, F., Hu, H., Bakshi, A., Robinson, M. R., Powell, J. E., . . . Yang, J. (2016). Integration of summary data from GWAS and eQTL studies predicts complex trait gene targets. *Nature Genetics, 48*(5), 481-487. doi:10.1038/ng.3538

Zilberman, D., Gehring, M., Tran, R. K., Ballinger, T., & Henikoff, S. (2007). Genome-wide analysis of Arabidopsis thaliana DNA methylation uncovers an interdependence between methylation and transcription. *Nature Genetics, 39*(1), 61-69. doi:10.1038/ng1929

Popular science summary

Adaptation is like a nature's way of fine-tuning species for success, describing how living organisms evolve to better survive and reproduce in their environments. Individuals with traits that enhance their ability to thrive and reproduce are more likely to pass those traits to the next generation. Through thousands of years, human cultivate and tame plants and animal, benefit by teaming up with certain wild plants and animals. The domestication events in plants resulted in huge amount of crop and vegetables cultivars, diverged from their wild relatives, both in phenotype and genetics. Common bean is one of the most important grain legumes in the world, domesticated around 8,000 years ago. Two independent domestication events and afterwards immigration events make it a perfect model to study mechanism of domestication and adaptation. Although there are arguments that domestication traits have been found to result from unrelated loci, our study on parallel evolution of photoperiod on common bean detected a strong signal associated with photoperiod sensitivity in both Mesoamerica and Andes gene-pool, provided evidence to the use of similar genomic strategies to achieve a common phenotype.

The evolution of plant mating system is also an interesting question. According to Stebbins hypothesis, selfing is considered to be an 'evolutionary dead end', leading to detrimental effect on plant traits and fitness. However, selfing is the principal mating system in many crops and vegetables. To understand the molecular genetic mechanism of selfing adaptation, we compared common bean with its closely related relative, the outcrosser runner bean, on both genomic and transcriptomic levels, we found that their gene expression under stabilizing selection, regardless with their mating system. Gene expression variation between species are most likely to their difference in topology of co-expression network. Furthermore, directional selection and DNA methylation may play important roles in the expression diversification of two species.

Populärvetenskaplig sammanfattning

Naturlig anpassning är naturens sätt att finjustera arter för ökad framgång och är en process beskriver hur levande organismer utvecklas för att bättre överleva och föröka sig i sina miljöer. Individer med egenskaper som förbättrar deras förmåga att trivas och föröka sig är mer benägna att föra vidare dessa egenskaper till nästa generation. Genom tusentals år har människor odlat och domesticerat växter och. Domesticeringen av olika växter har resulterat i en enorm mängd grödor och grönsaker som ofta skiljer sig avsevärt från sina vilda släktingar, både utseendemässigt och genetiskt. Trädgårdsbönan är en av de viktigaste baljväxterna i världen. Den domesticerades redan för cirka 8 000 år sedan och senare forskning har visat att. Den har genomgått två oberoende domesticeringstillfällen. Detta tillsammans med intrikuktionen av trädgårdsbönan till Europa på 1500-talet gör den till en utmärkt modell för att studera olika mekanismer som verkar under domesticering och anpassning. Det finns många studier som pekar på att även om olika grödor delar liknande domesticeringsegenskaper så ör den genetiska grunden för detta ofta helt oberoende. Trots detta visar vår studie om parallell anpassning av fotoperiod på att trädgårdsbönor från de olika domesticeringsområdena delar en stark signal associerad med fotoperiodkänslighet, vilket typer på att liknande genomiska strategier har varit i fokus för att uppnå en liknande fenotyp i de två olika populationerna.

Evolutionen av växters parningssystem är också en intressant fråga. Enligt Stebbins hypotes betraktas självbefruktning ofta som en 'evolutionär återvändsgränd' och som leder till skadliga effekter på växters egenskaper och överlevandsförmåga. Trots detta är självbefruktning det huvudsakliga parningssystemet hos många av våra vanligaste grödor och grönsaker. För att förstå de molekylärgenetiska mekanismerna för anpassning till självbefruktning jämförde vi trädgårdsbönor med dess nära släkting, rosenbönan, på både genom- och transkriptomnivå. Vi fann att genuttrycket oftast verkar vara under stabiliserande selektion, oavsett vilket parningssystem vi tittar på. Genuttrycksvariation mellan arter beror i stället mest sannolikt på deras skillnad konnektiviteten i genutrycksnätverking i de två arterna. Dessutom kan riktad selektion och DNA-metylering vara ytterligare viktiga faktorer som styr skillnader i uttrycksvariation mellan de två arterna.

Acknowledgements

Five years passed and finally it's my time to say something. We went through the dark days, waiting for the dawn in the silent night. But night was especially long for people like me, who is not good at 'adaptation'. I tried to grow plants at home in winter days, but they always turned yellow and withered quickly because of lacking sunlight. But finally I survived, and courage is my best cure.

I can't make it without people's help. So first, I'd like to thank my supervisor Pelle, who provided me the chance to continue my research career. Thank you for providing me enough freedom on my research. When I noticed that I have both re-sequencing data and RNA-sequencing data in my hand, a thousand ideas came to my mind. I tried whatever I have interest. Some of the results finally appeared in my manuscripts, but more never succeed and were kinda 'waste of time'. But I'm sure they will become my precious experience. And I've always been surprised by your hit-the-bullseye suggestions and ideas. I will never forget that you told me 'No difference is also a result' when I was struggling with my not significant statistics, dispelled the clouds over me, and stopped me when I was wrapped up in adjusting parameter details. Not a lot of supervision, but all meaningful to me. I always wish I could learn more from you.

I also want to thank my co-supervisor, Martha. I like your straight character, but since I'm not an expressive person, I didn't express my feelings to you much. Thank you for coming to me when I was in the darkest time and letting me know that you are always there. I admire your ability of scientific research and will always look up to you.

My co-supervisor Christina, thank you for providing me supports on the lab work. Members in Pelle's group: Girma, Mimmi, Xi, Rami, Yousef and Silvana, and everyone worked together, I won't forget the time we dine together, and help each other. We may only have a brief encounter, but I will always remember your kindness, make this group more like a family. Thanks administrative staff in the department, helping me with my not so smooth study so that I can move on.

My Chinese fellows, I'm very appreciate it to have you guys around me. Xi 师姐, I'm very glad to have you as my senior, both in my study and in my life. I will never forget the time we spent together chasing for delicious foods and beautiful sceneries. And I am looking forward to our next trip! Ai 艾姐, you've been a guide to me, and you are the one I could always trust. Thank you for leading me to an outer world and reminding me to calibrate myself! Chen 琛姐, I'm very happy to have you as a sport buddy. Wish you can always release your talent and go up a storey higher! Jia 小胡佳, I feel very lucky to have you walk in front of me, and I can obtain a lot of experience from you. Wish you and Kai 凯哥 and little baby happiness forever! Hui 慧姐, not only a dinner partner, but also a friend in life, thank you for always being there when I need help! Xinyi 馨艺, I'm glad to meet with you and your family. I really enjoy the time playing Ukulele with Ingrid 小樱, and switch games with Haiyang 海洋! Biyue 小玥师姐, although we didn't meet a lot, you are a reliable and experienced senior to me. Feng 小凤, thank you for being in my late PhD life. This winter is not so cold for me with your accompany. Ming 明哥, I was surprised by this such small world when I noticed your ex-superior is my mom's college roommate. Hope this small world can make us work together again! He 赫哥, my fellow-townsman, I'm very lucky to have you stand behind me. You are always welcome to come to me with bugs (on bioinformatics). Shuai 帅哥, I admire your passion to science and enjoy the discussions with you. Hope we will have the chance to work together again! My colleagues and former colleagues, maybe we didn't have much interaction, but you are more than colleagues to me. People in the volleyball team 二排, we made a lot of memories together, and they will become my sweet dreams, with everyone in the team. I won't forget the blue sky of summer night, and my promise. I bid you all a very fond farewell.

To my friends in China, you guys were also the strength keeping me up during the dark days, but I missed many important times of you, graduations, weddings… It's time to turn our facetime meetings into physical ones.

To my mom, I left home for studying since I was 18 years old, but this is the first time I haven't being at home for 4 consecutive years. Both of us had a hard time, when the time you were sick but I wasn't able to hug you, but finally we made it through. To my dad, thank you for being a supporter, with me, and mom. I know you, even if without a word, I know you are there. My sails are set, and I’m coming home.

I’ve been thinking of my future. With insufficient passion on scientific works, will I become a good scientist? I still don’t have an answer to this question, but I decide to move on. Take a step, and step again, and time will finally lead me to the truth.

Acta Universitatis Agriculturae Sueciae

Doctoral Thesis No. 2024:25

Plants exhibit plenty of adaptations that promote cross-pollination and reduce the occurrence of self-fertilization, and understanding the mechanisms behind these phenomena is a key goal of evolutionary biology. We explored the genetic architecture of adaptations in two closely related species, common bean and runner bean. The work in this thesis presents how mating system has influenced patterns of adaptation and gene expression variation.

Le Yu received her graduate education at Shandong University, Jinan, China. She acquired a M.Sc. degree in Bioinformatics and Molecular Biology at Lanzhou University, China.

Acta Universitatis Agriculturae Sueciae presents doctoral theses from the Swedish University of Agricultural Sciences (SLU).

SLU generates knowledge for the sustainable use of biological natural resources. Research, education, extension, as well as environmental monitoring and assessment are used to achieve this goal.

ISSN 1652-6880
ISBN (print version) 978-91-8046-314-0
ISBN (electronic version) 978-91-8046-315-7