



Current approaches and methods in plant domestication studies



AZALEA GUERRA-GARCÍA^{1,2*} AND DANIEL PIÑERO¹

Botanical Sciences
95 (3): 345-362, 2017

DOI: 10.17129/botsci.1209

Copyright: © 2017 Guerra-García & Piñero. This is an open access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Abstract

Background: The domestication process has left signatures in the genomes of domesticated species. Before the existence of molecular markers, only phenotypic traits could be used in domestication studies and breeding programs, but these approaches required long time and effort. In the last decades, the use of molecular markers dramatically increased, and the development of massive sequencing tools have enabled to obtain thousands or even millions of molecular markers.

Questions: The main goal of this review is to bring a general and an integrative perspective of the data, approaches and questions that can be answered using new sequencing tools.

Species study and data description: This work focuses on domesticated plants, comparing genetic and genomic data.

Results: The use of molecular markers in the last decades has increased the efficiency and accuracy of plant breeding, allowing to access information about domestication history and to identify genes affected by domestication. Some patterns have been identified in domesticated species: (1) genetic diversity reductions due to demographic bottlenecks and artificial selection; (2) frequently, mutations related with domestication syndrome preexisted at low frequency in natural populations; (3) accumulation of deleterious mutation; (4) gene flow between wild and cultivated populations; (5) phenotypic convergence usually do not result from molecular convergence. There are several approaches that can be used in massive sequencing tools: de novo genome sequencing, whole genome resequencing, reduction of genome complexity using restriction enzymes, transcriptome analysis and epigenetic studies.

Conclusions: Despite the progress made, enormous challenges remain: storage of large databases; development of fast and accurate methods to evaluate phenotypes; identification of paralogous genes in polyploid species; and the analysis of large and highly diverse genomes.

Keywords: artificial selection, genomics, plant breeding, molecular markers

Métodos y enfoques modernos del estudio de la domesticación

Resumen

Antecedentes: La domesticación es un proceso que ha dejado huellas en los genomas de las especies domesticadas. Anterior al surgimiento de los marcadores moleculares, únicamente se utilizaban atributos fenotípicos para estudiar la domesticación y para seleccionar individuos para el mejoramiento de cultivos. En las últimas décadas, el uso de marcadores moleculares se ha incrementado dramáticamente, y las técnicas de secuenciación masiva han permitido obtener miles, e incluso millones de marcadores.

Preguntas: El objetivo de esta revisión es brindar una visión general e integrativa de los distintos tipos de datos, enfoques y preguntas que se pueden contestar usando nuevas herramientas de secuenciación.

Especies de estudio y descripción de datos: Especies de plantas cultivadas, contrastando datos genéticos con datos genómicos.

Resultados: El uso de marcadores moleculares ha incrementado la eficiencia del mejoramiento de cultivos y ha permitido hacer inferencias sobre la historia de la domesticación y los genes involucrados. Se han encontrado algunos patrones en las especies domesticadas: (1) reducción de la diversidad genética, debido a cuellos de botella y selección artificial; (2) frecuentemente, las mutaciones relacionadas al síndrome de domesticación preexistían en poblaciones silvestres; (3) acumulación de alelos deletéreos; (4) flujo genético entre formas cultivadas y silvestres; (5) la convergencia fenotípica comúnmente no resulta de convergencia a nivel molecular. Hay una amplia gama de estrategias que se pueden usar con las herramientas de secuenciación masiva que van desde secuenciar genomas completos, usar enzimas de restricción para reducir la complejidad del genoma, hasta el análisis de transcriptomas y estudios epigenómicos.

Conclusiones: Pese a estos avances, aún quedan grandes retos como el análisis de grandes bases de datos; desarrollo de estrategias para obtener información fenotípica de forma rápida y precisa; identificación de genes parálogos en especies poliploidales; análisis de genomas grandes y muy diversos.

Palabras clave: selección, genómica, mejoramiento, marcadores moleculares

Author Contributions

Azalea Guerra-García wrote the paper.

Daniel Piñero reviewed the drafts of the manuscript.

Running head: Approaches and methods in plant domestication studies

¹ Departamento de Ecología Evolutiva, Instituto de Ecología, Universidad Nacional Autónoma de México

² Posgrado en Ciencias Biológicas, Universidad Nacional Autónoma de México

* Corresponding author:
azalea.guerra.g@gmail.com

T

he domestication process and its effect on genomes. During the Neolithic, a major change occurred in the way humans obtained food and resources. Human transitioned from hunters-gatherers to practice agriculture, a sedentary lifestyle, settlement of villages and the creation of ceramic, eventually leaded to the development of hierarchical civilizations (Gepts 2014, Larson *et al.* 2014). During this process, domestication was a key technological tool (Gepts 2014).

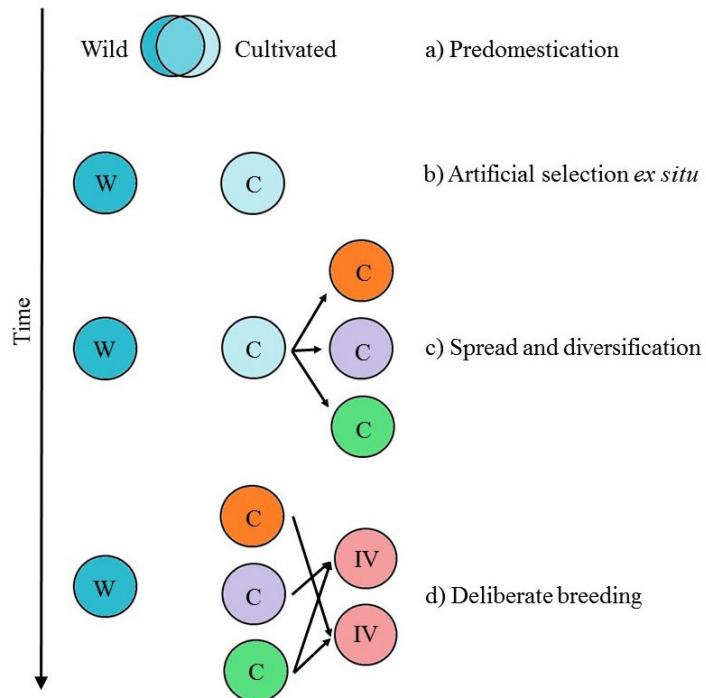
Domestication is an evolutionary process, in which humans promote the adaptation of wild species to agroecological niches and to local preferences (Figure 1. Casas *et al.* 1996, Larson *et al.* 2014). The evolutionary trajectory from wild to domesticated population is a complex multi-stage process, that involves several steps (Figure 2. Meyer & Purugganan 2013):

- Predomestication: humans begin to purposely plant and look after wild plants with favorable traits.
- Artificial selection in environments created by humans: at this the stage, human selective pressures promote alleles related with domestication traits.
- Spread and diversification: propagation and local adaptation to different agroecological and cultural environments. In this stage, cultivated and wild populations diverge and domesticated traits diversify.



Figure 1. Examples of wild and domesticated forms of crops. The first image of each row is the wild relative. **a)** teosinte and maize (*Zea mays*); **b)** chilli pepper (*Capsicum annuum*); **c)** common bean (*Phaseolus vulgaris*); **d)** cotton (*Gossypium hirsutum*). Images taken from CONABIO and CIAT and CIAT.

Figure 2. Evolutionary stages of domestication. W represents wild populations, C cultivated populations and IM improved lines. **a)** Predomestication; **b)** Management in environments created and controlled by humans; **c)** Spread and adaptation of cultivated populations; **d)** Deliberate breeding. It is important to notice that these stages are not mutually exclusive and can occur in presence of wild progenitors, increasing the probability of gene flow. Modified from Meyer & Purugganan (2013).



d) Deliberate breeding: this last stage has been practiced intensively in the last century. In classical breeding, controlled crosses between inbred lines are made and then individuals with desirable traits are selected. In the last decades, molecular and genetic tools have been integrated to breeding practices.

These stages are not mutually exclusive and can occur simultaneously, even today. For example, Nahuas communities in Mexico consume wild progenitors of crops like common bean, chili pepper, agave and tomato (in total 68 wild species) using techniques and tools that are available since the Archaic period (Zizumbo-Villarreal *et al.* 2012).

The domestication process affects the genomes of cultivated species and leaves a signature that can be identified. Domestication results in a pronounced alteration in the diversity and differentiation of few genes, while most of the genome reduces its variation as a result of population bottlenecks that occur in the initial predomestication stage (Charlesworth *et al.* 1997, Gepts 2004).

What do domesticated plants have in common? While the domestication process has varied in different plant groups depending on the mating system, growth form, harvested plant parts and economic importance there are several aspects that have been identified to be common to most domestication processes. These include loss of genetic diversity, the genes associated with the domestication process, an accumulation of deleterious variants, the relation between the cultivar and its wild relatives and the possibility of phenotype convergence and its molecular basis in different domestication processes.

Genetic diversity reduction. One of the most important determinants in crop evolution is the level of genetic diversity contained in the domesticated accessions, especially with reference to the wild ancestral gene pool. Genetic diversity is important as a necessary condition for further evolution in response to selection pressures, not only in the wild but also in breeding programmes (Gepts 2002). Genetic diversity reduction has been widely described in some of the most important crops. It is due to genetic drift resulted from population bottlenecks, and to artificial selection and the consequent selective sweep (Charlesworth *et al.* 1997, Gepts & Papa 2003, Gepts 2014). The domestication process itself, the spread from the center of origin and modern breeding practices, involve population bottlenecks because only some genotypes are selected, promoting a genetic diversity reduction (Gepts & Papa 2003). This pattern has been

reported in numerous species, for example, Mesoamerican common bean (~20 % reduction; Schmutz *et al.* 2014); sunflower (~30 %; Renaut & Rieseberg 2015); soybean (~30 %; Li *et al.* 2013); maize (~17 %; Hufford *et al.* 2012); and in cultivated *Agave* species (from 21 to 66 %; Eguiarte *et al.* 2013). In sunflower, maize and soybean, the largest diversity decrease occurs between wild populations and landraces; on the contrary, the difference in genetic variation between landraces and improved varieties is low, indicating that breeding practices, at least in these cases, do not have major effects in terms of genetic diversity, at least in these species (Li *et al.* 2013, Hufford *et al.* 2012, Renaut & Rieseberg 2015).

However, exceptions to this pattern have been reported. For example, in carrots, high levels of gene flow between wild and domesticated populations, besides a strong inbreeding depression, have prevented from genetic variation reduction (Iorizzo *et al.* 2013). In the Andean region, cultivated common bean is more genetically diverse than wild populations (Schmutz *et al.* 2014). In central Mexico, managed populations of the cacti *Stenocereus stellatus* (Pfeiff.) Riccob., presents higher levels of genetic variation than unmanaged populations, showing that the management made by humans can potentially alter the genetic diversity of non-domesticated species (Cruse-Sanders *et al.* 2013).

Mutations in genes under domestication-. Identification and isolation of genes that underlie domestication traits provide the opportunity to identify patterns present in these loci. These genes show a wide range of functions, from transcription factors to metabolic enzymes, although many encode similar enzymes or are involved in the same metabolic pathways (Meyer & Purugganan 2013).

According to Meyer & Purugganan (2013), the most frequent functional changes occur in alleles involved in function loss and gene expression. These mutations promote major effects in phenotypes, which usually distinguish domesticated and wild populations. For example, in woodland strawberry (*Fragaria vesca* L.), a 2-bp deletion in the coding region of the *KSN* gene –which is a transcriptional factor– introduces a frameshift, resulting in continuous flowering (Iwata *et al.* 2012). Also, the function loss of *Vrs1* gene in barley (*Hordeum vulgare* L.), changed the inflorescence architecture from two-rowed to six-rowed type (Komatsuda *et al.* 2007). Mutations in genes under domestication include single nucleotide polymorphisms (SNPs), insertions and deletions (indels), transposon insertions, gene duplication (including gene copy number) and chromosomal rearrangements. The most frequent alterations are SNPs, followed by indels (Meyer & Purugganan 2013).

An important question is whether mutations that lead to domestication phenotypes are new or preexisted in wild populations. This has implications in the nature of selective sweeps and crop species evolution dynamics; for example, a selective sweep caused by a preexisting mutation leaves weaker signatures of selection and allows rapid evolution, because it has a higher frequency than a new mutation (Gepts 2014).

There are some examples of alleles that underlie domesticated traits that appeared recently and are not present in wild populations. That is the case of *LGI* gene in rice, associated with closed panicle (Huang *et al.* 2012) or the *SUN* gene duplication in tomato, which regulates fruit shape (Rodriguez *et al.* 2011). However, there are cases of alleles that preexisted in wild populations at low frequency that were subjected to human selective pressures. For example, the *Brassica oleracea* *CAL* gene encodes a transcription factor that regulates floral meristem development, and a nonsense mutation leads to the proliferation of floral meristem in domesticated cauliflower and broccoli (Purugganan *et al.* 2000). This mutation is either fixed or present at a high frequency in cultivars, but it is also present at low frequencies in wild populations (Purugganan *et al.* 2000). Another example is the single-stem phenotype of domesticated maize, which is controlled by the dominant *Tb-1* allele. This allele arose before domestication by the insertion of a *Hopscotch* transposon leading to overexpression. The insertion occurred 28,000–23,000 years BP, predating domestication (Studer *et al.* 2011). This suggests that many domesticated traits arise not from new mutations but rather from mutations that segregate in wild population of crop species.

Accumulation of deleterious mutations-. Mutations can have several effects on fitness that range from lethal to neutral and advantageous, but most of these new variants in coding regions are

expected to be deleterious, because they may alter phylogenetically conserved sites or can cause loss of protein function (Kono *et al.* 2016). Under mutation-selection balance and in sexually reproducing species, the accumulation of deleterious alleles -- particularly strongly deleterious and with dominant effects -- is infrequent, because recombination brings together deleterious variants, resulting in unfit genotypes that are eliminated from the population, purging deleterious mutations by purifying selection. Weakly deleterious variants on the other hand, can be potentially maintained under some circumstances, like population size reduction and/or inbreeding, reducing the effective recombination rate and promoting the accumulation of deleterious mutations (Morrell *et al.* 2011, Renaut & Rieseberg 2015).

Besides, as a species expands into new environments (either natural or artificial), genetic drift could increase, due to the reduction of effective population size and a posterior fast growth rate. During the process of domestication and improvement, populations undergo multiple bottlenecks, accompanied by strong artificial selection and the relaxation of selective pressures on traits important in the wild (Morrell *et al.* 2011). These bottlenecks increase the probability that a deleterious variant increases its frequency and gets fixed (Mezmouk & Ross-Ibarra 2014). Also, the linkage between desirable beneficial and deleterious mutations may reduce the efficiency of selection to eliminate the latter, contributing cumulatively to fitness reduction and probably constraining crop yield (Renaut & Rieseberg 2015).

The selective and demographic processes of domestication have led to three hypotheses about the patterns of deleterious mutations (Kono *et al.* 2016):

- a. There will be more deleterious alleles in domesticated than in wild relatives.
- b. Deleterious variants will be enriched near loci that have been under artificial selection.
- c. There will be less deleterious mutations in elite cultivars than in landraces due to strong selection for yield.

There is evidence that supports these hypotheses, as it is estimated that 20 % of non-synonymous variants in rice (Lu *et al.* 2006) and in maize (Mezmouk & Ross-Ibarra 2014) present deleterious effects. In sunflower, landraces and elite lines possess more non-synonymous SNPs compared to wild relatives, particularly regions with low recombination rates and less deleterious mutation, are present in elite cultivars compared to landraces (Renaut & Rieseberg 2015). Also, hundreds of deleterious mutations have been identified in barley and soybean cultivars, being non-sense mutations the least frequent (Kono *et al.* 2016).

Recent advances in sequencing technologies permit to use bioinformatics tools to examine the patterns of deleterious variants in populations. Pure bioinformatic approaches use sequence conservation to infer variants with a significant probability of being deleterious. As a consequence, the identification and elimination of these alleles can potentially provide a complementary approach to breeding practices if identified variants are truly deleterious (Kono *et al.* 2016).

Gene flow between domesticated plants and their wild relatives.- Gene flow between crops and their wild relatives is common, since these two types of population coexist in sympatry, and in most cases, crops and their wild progenitors belong to the same biological species (Gepts 2014). The potential consequences of gene flow are diverse. For example, gene flow between wild and domesticated populations can result in diversity recovery in cultivated forms, as has been described in maize due to continuous introgression with teosinte (Hufford *et al.* 2012). This is also the case in grapevine during cultivars spread through Europe (Myles *et al.* 2011). Also, emergence of weeds has been reported, as occurred in hybrid rice populations from China (Jiang *et al.* 2011, Xia *et al.* 2011) and the United States (Olsen *et al.* 2007).

Despite gene flow, wild and domesticated plants remain phenotypically distinct, at least with respect to domesticated syndrome traits, probably because of human selective pressures on cultivated populations, and natural selection of wild forms. For example, gene flow among wild and cultivated squash taxa have been detected in Mexico (Montes-Hernández & Eguiarte 2002). But farmers do not select seeds from the individuals that present intermediate morphological characters for seed stock (Montes-Hernández *et al.* 2005).

Following the definition in its broad sense, introgression is the transfer of genes between genetically distinguishable populations (Rieseberg & Carney 1998). Introgression can take place between populations of wild species (wild-wild) that are related and between a cultivated spe-

cies and its close wild relatives (crop–wild; Rieseberg & Carney 1998). Nevertheless, introgression is not uniform across the genome and it is unlikely to occur on regions close to genes related with domestication (Gepts 2014). Gene flow may also be asymmetric, although the direction is not consistent: for example, in common bean it mainly occurs from domesticated to wild types (Papa & Gepts 2003), but the opposite happens in maize (from *Zea mays* L. subspecies *mexicana* teosinte to cultivated maize; Hufford *et al.* 2013).

Introgression may produce profound effects on domestication evolutionary trajectories. The evolutionary history of rice (*Oryza sativa* L.) is an example: Molina *et al.* (2011) suggest that rice domestication occurred in China, where *japonica* variety arose. Later, *indica* variety was originated by a hybridization event between *japonica* and a putative *indica* protoform or with *O. ryfipogon* from South Asia. This resulted in the introgression of *japonica* domesticated genes.

Another case that illustrates the impact of gene flow occurs in domesticated citrus species. Wu *et al.* (2014) sequenced and compared the genomes of some *Citrus* species and showed that cultivated types derive from two progenitor species: introgression from *C. maxima* (Burm.) Merrill to ancestral mandarin species *C. reticulata* Blanco originated tangerine; sweet orange is the offspring of previously admixed individuals of these two species; and sour orange is an *F1* hybrid of pure *C. maxima* and *C. reticulata* parents. The exception is pomelo, which is the result of selection on *C. maxima* (Wu *et al.* 2014).

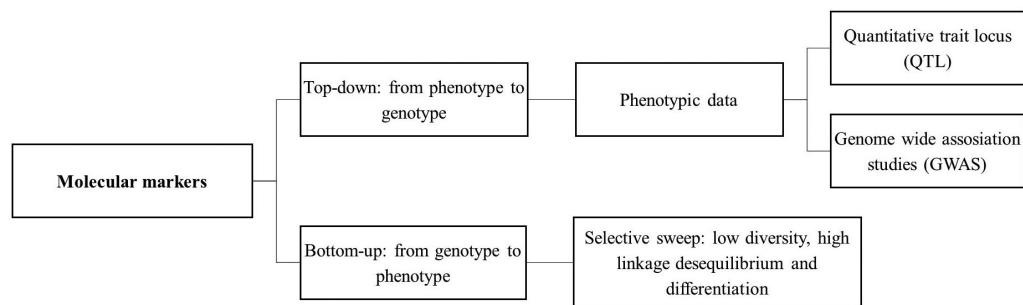
Phenotypic convergence.— The Law of Homologous Series, formulated by Vavilov (1922), and ‘analogous variation’ term used by Darwin, are based on the observation that similar phenotypes were selected during domestication. This leads to the question: phenotypic parallelisms result from molecular convergences? In other words, selection on particular traits affects the same genes in different species? If this is true, parallelisms potentially can be used in breeding programs focusing on regions previously identified to be related with important domestication traits (Pickersgill 2009). However, contradictory data has been found and there is evidence for and against this. For example, Grube *et al.* (2000) found five clusters of resistance genes common to potato, tomato and chili pepper, and seven clusters common to two of these three crops; they concluded that the position of resistance genes, is conserved in Solanaceae. However, they found only two examples in which these homologous genes controlled resistance to the same disease. Therefore, knowing the sequence of alleles and the structure of their products can be useful in the development of new pathogens or pest control methods (Pickersgill 2009). Another example is the *Shattering gene (Sh1)*, which avoids dehiscence in maize, rice, sorghum and common bean; furthermore, there is evidence that *Sh1* independently emerged three times in sorghum (Lin *et al.* 2012, Schmutz *et al.* 2014).

On the other hand, evidence against molecular convergence in crops includes two cotton species (*Gossypium hirsutum* L. and *G. barbadense* L.), which were independently domesticated, but selection by humans acted in different genetic components that control fiber development (Hovav *et al.* 2008). Another case where parallelism in phenotypes does not correspond at the gene level occurs between common bean and pea: *p* and *v* genes are partially responsible of dehiscence inhibition in common bean, whereas *dpo1* and *dpo2* genes control this trait in pea, despite *p* and *v* genes are also present (Weeden 2007). Also, in common bean—which was independently domesticated in South America and Mesoamerica—Schmutz *et al.* (2014) identified 1,835 candidate genes under artificial selection in Mexican populations, and 748 in Andean ones, however only 59 were shared genes between the two gene pools. Probably, in some cases, genetic architecture behind domestication is so diverse that similar phenotypic changes are due to selection on different genes.

Using molecular markers to study domestication and advance breeding programs. Before the development of molecular markers, farmers and breeders only had phenotypic traits to select desirable individuals to interbreed. Nevertheless, relative long periods of time and several generations were required to evaluate and select useful genotypes. Some decades ago, molecular markers started to be used in breeding programs and in the development of new cultivars or varieties (Kim *et al.* 2015).

Following the first molecular markers tools (He *et al.* 2014), Sanger DNA sequencing al-

Figure 3. Overview of massive parallel sequencing (NGS) applications in crop genetics and breeding.



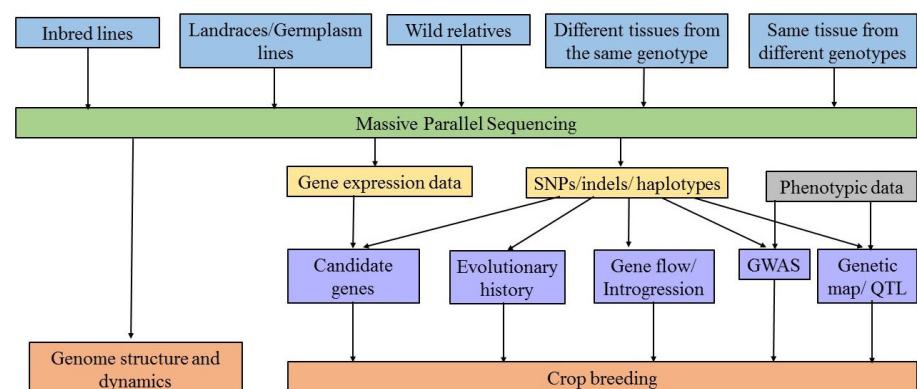
lowed the identification of variants at the single base pair resolution. However, sequencing small genomes or large regions from complex genomes using Sanger technique is expensive and laborious. The Human Genome Project encouraged the development of techniques for sequence whole genomes at low cost and less time-consuming. These sequencing tools are known as massive parallel sequencing (MPS), also called Next Generation Sequencing (NGS) and have allowed the routine use of single nucleotide polymorphisms (SNPs) in the last decade, even in no model species (Escalante *et al.* 2014). These massive sequencing tools are used by breeders to sequence large populations, study genetic composition of crop varieties, understand evolutionary relationships between cultivars and wild relatives, identify genotype-genotype ($G \times G$) and genotypes-environment ($G \times A$) interactions, and increase the resolution of gene and quantitative trait locus (QTL) discovery, providing the basis for modeling complex genotype-phenotype relationships at the whole-genome level (Figure 3. Cobb *et al.* 2013, Varshney *et al.* 2014).

Several methods are used to identify relationships between genes or genomic regions and domestication syndrome traits. These can be classified in *bottom-up* and *top-down* approaches (Figure 4).

Top-down domestication studies. Classical studies to infer the genetic basis of domestication traits use a top-down approach, especially when used in mapping QTLs. Traditional QTL mapping studies begin with two inbred parental lines that differ in their trait values. After creating an F_1 generation by crossing the parental lines, F_1 individuals are crossed to create an F_2 mapping population. These individuals are measured for the trait of interest and genotyped at markers that segregate between parental populations. QTLs are then identified by correlations between trait value and genotype at a locus. The ability to identify a QTL using this method depends on the scale of recombination in the mapping population. For this reason, recombinant inbred lines (RILs) are often used to increase mapping resolution (Mackay *et al.* 2009).

On the other hand, genome-wide association studies (GWAS) utilize association mapping, also known as linkage disequilibrium (LD) mapping, to map QTLs by taking advantage of historic LD to identify statistically significant phenotype-genotype associations (Varshney *et al.* 2014). Unlike classical methods to identify QTLs, in GWAS population mapping is usually

Figure 4. Approaches used to identify genes or genomic regions under selection: Top-down *vs.* bottom-up.



an outbred population and the markers used are SNPs. In general, the results are better than the ones using RILs. However, the power to detect a QTL depends of the phenotypic variance that is explained by the QTL that is determined by its effect and allele frequency. Because of this, GWAS generally has low power to detect QTL variants with small effect or that are rare in the population (Mackay *et al.* 2009). Nevertheless, GWAS have been successfully performed in several crops, and the use of massive sequencing makes possible to genotype larger populations with a higher density of markers (tens to hundreds of thousands of markers are needed in GWAS), increasing mapping resolution (Varshney *et al.* 2014, Andrews *et al.* 2016).

Results of the top-down approach have shown that many traits are controlled at least in part by a few major genes, that genetic effects predominate over environmental ones, and that some genes are linked. It is likely that this approach underestimates the number of genomic regions involved in domestication syndrome. For example, Hufford *et al.* (2012) estimated that only ~3 % of maize genome have been affected by domestication using top-down approach.

The disadvantages of the top-down approach include that phenotypic information is required, and that only few traits can be considered at the same time (Tang *et al.* 2010). A recent study using a top-down approach was performed in rice, in which 203 varieties with well-characterized phenotypes were sequenced, and three traits were mapped (Wang *et al.* 2016). For amylose content and seed length traits, this approach leads to direct identification of the previously identified causal SNPs as the major-effect loci. For pericarp color, they identified a new major-effect locus. Although previously known loci can explain color variation in the varieties of two main subspecies of Asian domesticated rice, *japonica* and *indica*, the new locus identified is unique to another domesticated rice subgroup, aus, and together with existing loci, can fully explain the major variation in pericarp color (Wang *et al.* 2016).

Bottom-up domestication studies. Compared to classic genetic studies, in which relatively few markers are used, genomic approaches provide additional information about the diversity and LD patterns at the genome level. Therefore, genomics provides a broader view across taxa and across individual genomes (Gepts 2014). The bottom-up approach consists in population genetic screenings of genome-wide diversity, including sequence diversity departures from neutrality, scans for selective sweeps or highly divergent regions (F_{ST}). Further research, involving functional genetics, is then needed to directly establish a causal effect between candidate genes or alleles and phenotypes.

Comparison of sequence diversity combined with neutrality tests in wild and domesticated populations can identify regions that were affected by domestication. Furthermore, comparisons between landraces and wild forms focus specifically on the effect of initial domestication, whereas comparisons between landraces and improved cultivars measure the effect of subsequent selection, including modern breeding (Tang *et al.* 2010, Gepts 2014). The bottom-up approach does not require previous phenotypic data, allowing the identification of a larger amount of candidate genes or genomic regions under selective pressures, unlike the top-down approach which can only consider few genes or regions (Tang *et al.* 2010). For example, with a top-down approach, in which only ~3 % of the maize genome was estimated to be affected by domestication, using a bottom-up approach 1,764 genes (representing ~6.6 % of its genome) were inferred to be related to domestication and 1,508 genes (~5.6 %) related to breeding (Hufford *et al.* 2012).

An example of selective sweep detection using a bottom-up approach includes sorghum, where selective sweeps that encompass genes for starch synthesis enzymes, seed shattering, plant height and time to maturity, were detected using 1,000 accessions and ~265,000 SNPs (Morris *et al.* 2013). Another example was found in rice, where by resequencing 1,529 wild and domesticated accessions, 55 selective sweeps were detected, including *Bh4* (hull color), *PROG1* (tiller angle), *sh4* (seed shattering), *qSW5* (grain width) and *OsC1* (leaf sheath and apiculus color; Huang *et al.* 2012).

Challenges to overcome using a bottom-up approach include that the phenotypic traits affected by the genes or regions encompassed in selective sweeps may be and remain unknown, and that demographic events may hide selective signatures because they alter diversity patterns, and standard test typically assume that populations evolve according to the idealized Wright-Fisher model, with panmictic populations of constant size (Ross-Ibarra *et al.* 2007). When these

assumptions are inaccurate, because domestication involves genetic bottlenecks followed by demographic expansion, tests to detect selection can be inaccurate (Ross-Ibarra *et al.* 2007).

Advantages of using molecular markers.- There are two main types of genomic-assisted breeding: marker-assisted selection (MAS) and genomic selection (GS). MAS, which includes marker-assisted back-crossing (MABC), uses molecular markers that map within specific genes or QTLs known to be associated with target traits or phenotypes to select individuals that carry favorable alleles and/or discard those that do not (Mackay *et al.* 2009).

GS uses all available marker data for a population as predictor of breeding value. GS integrates marker data from a training population with phenotypic and when available, pedigree data to generate a prediction model. The model output genomic estimated breeding values (GEBVs) for all genotyped individual, which are used as a predictor of how well a plant will perform as a parent for crossing (Varshney *et al.* 2014).

The advantage of using molecular markers in breeding programs is that genotypic data obtained from a seed or seedling can be used to predict the phenotypic performance of mature individuals without the need for extensive phenotypic evaluation over long periods of time, allowing for more selection cycles and greater genetic gain per unit of time (Varshney *et al.* 2014). Previously, marker data were expensive per data point, and laborious to generate, and MAS were constrained by the number of available markers. As a result, only markers in important genome regions were utilized to predict the presence or absence of agriculturally valuable traits. By contrast, the use of massive parallel sequencing technologies provides genome-wide markers coverage at a low cost per data point, allowing to assess the inheritance of the entire genome with nucleotide-level precision (Cobb *et al.* 2013, Varshney *et al.* 2014).

Today, studies are no longer limited by our ability to genotype large populations or by the number of markers, but rather by the high cost and low throughput of phenotypic strategies for traits of interest and in environments relevant to plant breeding (Cobb *et al.* 2013).

How genomic markers help understand the history of domesticated species?- Molecular markers that are not affected by selection bring information about demographic histories, including the history of the domestication process (Tang *et al.* 2010). Earlier models of domestication proposed a single domestication event and suggested that domestication occurred through strong selection and severe genetic bottleneck, which resulted in reproductive isolation between wild and cultivated populations (Meyer & Purugganan 2013). Then, models integrated gene flow and introgression between cultivated and wild relatives, as occurred in grape (*Vitis vinifera* L. subsp. *vinifera*), that emerged in Near East region, but evidence of introgression from the wild progenitor (*Vitis vinifera* subsp. *sylvestris*) during Europe introduction was found (Myles *et al.* 2011). High levels of genetic diversity and fast decay of LD in *vinifera* suggest a weak bottleneck at the beginning of domestication process, followed by thousands of years of vegetative propagation (Myles *et al.* 2011).

There are some cases in which multiple domestication events occurred. For example, resequencing common bean genomes -- including wild and domesticated forms -- confirmed two independent domestication events, one occurred in Mesoamerica and the other one in the Andean region. Besides, these two gene pools diverged 165,000 years ago, long before domestication started (Schmutz *et al.* 2014). Finally, an alternative domestication model proposes that crops are domesticated from interspecific hybridization, sometimes followed by clonal propagation. This is common in tree crops, as is the case of species of the genus *Citrus*, previously discussed. New groups of bananas and plantains developed when diploid domesticated bananas (genome AA) spread into the range of wild *Musa balbisiana* Colla (genome BB), producing the AAB and ABB triploids (Heslop-Harrison & Schwarzacher 2007).

How is genomic data obtained? Several massive parallel sequencing platforms are commercially available, and while the specific methods vary, they obtain millions of short DNA sequence reads from random locations in the genome (Escalante *et al.* 2014). Massive sequencing technologies have been widely used for *de novo* sequencing and several crop and other angiosperm genomes have been sequenced, which are used as reference in no model species studies

(Table 1). Some of the genome assemblies are in draft stage and extensive work is ongoing in the direction of closing the gaps and re-sequencing. In addition to the genome sequence, transcriptomes and expressions profiles are also available for many crops. The large genome size and polyploidy exhibited by many crop species make more difficult the sequencing and further analysis. In some allopolyploid crops, the genomes of progenitors species were first sequenced and then were used to assemble the polyplid genomes of the domesticated forms. This strategy was used in cotton (Wang *et al.* 2012, Li *et al.* 2014, Li *et al.* 2015), strawberry (Hirakawa *et al.* 2014) and peanut (Bertioli *et al.* 2016).

Several strategies to construct genomic libraries (collection of DNA that will be sequenced) have been developed. When the genomic library consists in fragments from whole genomes, it is called whole genome resequencing (WGRS). This approach has been used in several crop species, as common bean (Schmutz *et al.* 2014), rice (Huang *et al.* 2012, Wang *et al.* 2016), maize (Hufford *et al.* 2012, Xu *et al.* 2014); and soybean (Li *et al.* 2013) to identify selection signatures and to infer the domestication history. Nevertheless, WGRS is relatively expensive, particularly in large genomes or polyploid species, reducing the number of samples that can be multiplexed (Kim *et al.* 2015).

In the last few years, different strategies to genotype a larger number of samples at reduced costs have been developed, including microarrays and methods to construct reduced representation libraries (RRLs). The cost per SNP using microarrays is relatively low but it is necessary to know the sequence of the variants (SNPs) that will be used (Bolger *et al.* 2014). Construction of RRLs, also known as restriction-associated DNA sequencing (RAD-seq), is a method for sequencing loci adjacent to restriction cut sites across the genome. RAD-seq targets a subset of the genome, therefore allow multiplex a higher number of samples without prior genomic information (Elshire *et al.* 2011, Andrews *et al.* 2016). The use of these

Table 1. Examples of published sequenced crop genomes. The statistics for each genome are taken from the publication, despite several model plant genomes have had significant updates to genome assemblies and gene counts.

Species	Common name	Genome size (Mb)	Assembly size	Ploidy and chromosome number	# predicted genes	Reference
Domesticated in Mexico						
<i>Amaranthus hypochondriacus</i>	Amaranth	466	377	2n = 32	23,059	Clouse <i>et al.</i> 2016
<i>Capsicum annuum</i>	Hot pepper	3,480	3,060	2n = 24	34,903	Kim <i>et al.</i> (2014)
<i>Carica papaya</i>	Papaya	372	271	2n = 18	24,746	Ming <i>et al.</i> (2008)
<i>Gossypium hirsutum</i>	Cotton	2,250-2,430	2,173	2n = 4x = 52	76,943	Li <i>et al.</i> (2015)
<i>Phaseolus vulgaris</i>	Common bean	587	473	2n = 22	27,197	Schmutz <i>et al.</i> (2014)
<i>Theobroma cacao</i>	Cocoa	430	327	2n = 20	28,798	Argout <i>et al.</i> (2011)
<i>Zea mays</i>	Maize	2,300	2,048	2n = 20	32,540	Schnable <i>et al.</i> (2009)
Domesticated in America						
<i>Anana comosus</i>	Pineapple	526	382	2n = 50	27,024	Ming <i>et al.</i> (2015)
<i>Chenopodium quinoa</i>	Quinoa	1,450-1,500	1,390	2n = 4x = 36	44,776	Jarvis <i>et al.</i> (2017)
<i>Gossypium barbadense</i>	Cotton	2,470	1,395 A subgenome 776 B subgenome	2n = 4x = 52	77,526	Liu <i>et al.</i> (2015)
<i>Manihot esculenta</i>	Cassava	770	532	2n = 36	30,666	Prochnik <i>et al.</i> (2012)
<i>Nicotina tabacum</i>	Tobacco	4,500	3,700	2n = 4x = 48	90,000	Sierro <i>et al.</i> (2014)
<i>Solanum lycopersicum</i>	Tomato	900	760	2n = 24	34,727	The Tomato Genome Consortium (2012)
<i>Solanum tuberosum</i>	Potato	844	727	2n = 12	39,031	The Potato Genome Sequencing Consortium (2011)
<i>Vaccinium macrocarpon</i>	Cranberry	470	420	2n = 12	36,364	Polashock <i>et al.</i> (2014)
Other important economic crops						
<i>Glycine max</i>	Soybean	1,115	950	2n = 20	46,430	Schmutz <i>et al.</i> (2010)
<i>Oriza sativa indica</i>	Rice	466	429	2n = 24	46,022-55,615	Yu <i>et al.</i> (2002)
<i>Oriza sativa japonica</i>	Rice	420	390	2n = 24	37,544	Goff <i>et al.</i> (2002)
<i>Sorghum bicolor</i>	Sorghum	730	698	2n = 20	27,640	Paterson <i>et al.</i> (2009)
<i>Triticum aestivum</i>	Bread wheat	17,000	3,800	2n = 6x = 42	90,000-94,000	Brenchley <i>et al.</i> (2012)

tools has been intensified to genotype large populations, for both top-down and bottom up approaches. For example, the 539 inbred lines of maize found in the International Maize and Wheat Improvement Center (*Centro Internacional de Mejoramiento de Maíz y Trigo*, CIMMYT) have been genotyped using genotyping by sequencing (GBS) to know the genetic diversity and structure (Wu *et al.* 2016).

To design a genomic study it is necessary to consider the size and complexity of the genome, and if a reference genome is available. Besides, the amount of markers that will be identified depends on the approach used to construct the library (WGRS, RAD-seq or microarray). This must be considered within the context of the biological question. For instance, to infer neutral processes, as demographic histories, it only requires from hundreds to a few thousands of molecular markers. If the goal is to identify selection signatures or functionally characterize regions, dozens of thousands to hundreds of thousands of markers are required; for example, for GWAS or gene mapping, at least hundreds of thousands of SNPs are needed (Andrews *et al.* 2016).

The functional aspect of genomics in domestication studies. The goal of functional genomics is to understand the complex relationships between the genome and the phenotype. This involves dynamic processes as gene transcription, translation, regulation of gene expression and protein interactions. These processes comprise a number of *-omics* approaches, such as transcriptomics (gene expression), proteomics (protein production), and metabolomics (characterization of metabolic products; Huang *et al.* 2016). Recently, some studies have added the functional genomics aspect of domestication to assess how many and which genes show differences in expression when wild and domesticated forms are compared.

From the *-omics*, transcriptomics has received more attention (Huang *et al.* 2016). The transcriptome is the set of messenger RNA molecules in one cell or a population of cells, and provides information about expression and genic regulation. RNA-seq (RNA sequencing), also called whole transcriptome shotgun sequencing (WTSS), uses massive sequencing to reveal the presence and quantity of RNA in a sample at a given moment in time. RNA-seq is an alternative of genome reduction representation, that can be used to identify differential expressed genes, alternative splicing (mechanism by which different forms of mature mRNAs are generated from the same gene), and genetic regulatory networks, as well as for annotation, and gene and markers discovery (Andrews *et al.* 2016, Huang *et al.* 2016). Transcriptomics may complement genomic studies, because RNAseq focused on coding regions, while genomics approaches, as RADseq, include coding and noncoding regions (Andrews *et al.* 2016).

For example, in developing cotton fiber, ~15 % of 1,300 proteins are significant up or down-regulated (Hu *et al.* 2013). Most of the changes took place in the early developmental stages, which is consistent with human selection for earlier activation of fiber elongation in domesticated types. Nevertheless, there were a few changes that overlapped between transcripts and proteins, probably because protein abundances depend on multiple factors, as translation, post-translational modifications and degradation processes (Hu *et al.* 2013).

In maize, Swanson-Wagner *et al.* (2012) detected hundreds of genes whose expression patterns were altered during domestication, some of them are involved in biotic and abiotic stress responses. More recently, Huang *et al.* (2016) studied the transcriptomes of six teosinte accessions and found that approximately 75 % of the genes were highly conserved between maize and teosinte. Moreover, they also found 1,516 unigenes (set of transcripts located at the same loci) that were specifically expressed in teosinte, and identified 99 unigenes with strong selection signals, of which 57 might be under strong selection during maize domestication and improvement. This kind of functional studies allows an integrative understanding of the genomic effects that domestications has on species.

Conclusions

The domestication process and breeding have modified the genomes of the plant species that we consume on a daily basis. The recent development of massive parallel sequencing tools allows us to access to information from whole genomes, accelerating identification of genes affected by domestication and to correlate domestication syndrome traits with their molecular basis.

Because many crop species represent a good model for evolutionary studies and have high economic value, there are detailed genetic studies of the most important crops and their wild relatives. From these studies, general patterns have been inferred and in some species, at least partially, the dynamics of the evolutionary process associated to domestication and diversification have been elucidated. Nevertheless, it is necessary to expand our knowledge to no model species, particularly to perennial plants, including tree crops, which have received far less attention. Besides, more functional genomics studies are needed to have an integrative knowledge of the changes that have affected crop species.

Despite major advances of the last years in genomics, important challenges still remain. One important challenge is the analysis of big databases that are generated because computer clusters are frequently needed. Also, the analysis of large, complex and diverse genomes as is the case of conifers is still a relevant problem. We also need to develop efficient approaches to include and analyze repetitive regions. In addition, the identification of paralogues, and the study of polyploid species, remain as some of the most critical problems when analyzing massive sequencing data. Besides, in GWAS, phenotyping is a major operational bottleneck that limits the power and resolution of genetic analysis. There is an urgent need for high throughput, cost-effective, and precise phenotyping methods.

Given the importance of wild progenitors to analyze changes derived from domestication, more attention should be paid to the genetic diversity and adaptation of the wild forms. Besides, in the context of climate change, this approach facilitates the development of climate-tolerant cultivars. In order to achieve these goals, it is necessary to integrate training and research across scientific fields, including genetics, plant breeding, molecular biology, evolution, statistics and bioinformatics.

Acknowledgements

We thank Luis E. Eguiarte for encourage us to make this review, and Alicia Mastretta for reviewing the manuscript. AGG acknowledge the scholarship provided by Consejo Nacional de Ciencia y Tecnología (CONACyT) and thanks the Posgrado en Ciencias Biológicas, from Universidad Nacional Autónoma de México (UNAM). This work is supported by CONACyT PD-CPN2014-01, project 247730.

Literature cited

- Andrews KR, Good JM, Miller MR, Luikart G, Hohenlohe PA. 2016. Harnessing the power of RADseq for ecological and evolutionary genomics. *Nature Reviews Genetics* **17**: 81-92. DOI:10.1038/nrg.2015.28
- Argout X, Salse J, Aury JM, Guiltinan MJ, Droc G, Gouzy J, Allegre M, Chaparro C, Legavre T, Maximova SN, Abrouk M, Murat F, Fouet O, Poulain J, Ruiz M, Roguet Y, Rodier-Goud M, Barbosa-Neto JF, Sabot F, Kudrna D, Ammiraju JS, Schuster SC, Carlson JE, Sallet E, Schiex T, Dievart A, Kramer M, Gelle L, Shi Z, Bérard A, Viot C, Boccardo M, Risterucci AM, Guignon V, Sabau X, Axtell MJ, Ma Z, Zhang Y, Brown S, Bourge M, Golser W, Song X, Clement D, Rivallan R, Tahi M, Akaza JM, Pitollat B, Gramacho K, D'Hont A, Brunel D, Infante D, Kebe I, Costet P, Wing R, McCombie WR, Guiderdoni E, Quétier F, Panaud O, Wincker P, Bocs S, Lanaud C. 2011. The genome of *Theobroma cacao*. *Nature Genetics* **43**: 101-108. DOI: 10.1038/ng.736
- Bertioli DJ, Cannon SB, Froenicke L, Huang G, Farmer AD, Cannon EK, Liu X, Gao D, Clevenger J, Dash S, Ren L, Moretzsohn MC, Shirasawa K, Huang W, Vidigal B, Abernathy B, Chu Y, Niederhuth CE, Umale P, Araújo AC, Kozik A, Kim KD, Burow MD, Varshney RK, Wang X, Zhang X, Barkley N, Guimaraes PM, Isobe S, Guo B, Liao B, Stalker HT, Schmitz RJ, Scheffler BE, Leal-Bertioli SC, Xun X, Jackson SA, Michelmore R, Ozias-Akins P. 2016. The genome sequences of *Arachis duranensis* and *Arachis ipaensis*, the diploid ancestors of cultivated peanut. *Nature Genetics* **48**: 438-446. DOI: 10.1038/ng.3517
- Bolger ME, Weisshaar B, Scholz U, Stein N, Usadel B, Mayer KF. 2014. Plant genome sequencing – application for crop improvement. *Current opinion in Biotechnology* **26**: 31-37. DOI: 10.1016/j.copbio.2013.08.019
- Brenchley R, Spannagl M, Pfeifer M, Barker GLA, D'Amore R, Allen AM, McKenzie N, Kramer M, Kerhornou A, Bolser D, Kay S, Waite D, Trick M, Bancroft I, Gu Y, Huo N, Luo MC, Sehgal S, Gill B, Kianian S, Anderson O, Kersey P, Dvorak J, McCombie WR, Hall A, Mayer KF, Edwards KJ, Bevan

- MW, Hall N. 2012. Analysis of the bread wheat genome using whole-genome shotgun sequencing. *Nature* **491**: 705–710. DOI: 10.1038/nature11650
- Casas A, Vázquez MC, Viveros JL, Caballero J. 1996. Plant management among the Nahua and the Mixtec in the Balsas River Basin, Mexico: an Ethnobotanical approach to the study of plant domestication. *Human Ecology* **24**: 455–478. DOI: 10.1007/BF02168862
- Charlesworth B, Nordborg M, Charlesworth D. 1997. The effects of local selection, balanced polymorphism and background selection on equilibrium patterns of genetic diversity in subdivided populations. *Genetics Research* **70**: 155–174.
- CIAT. Colección de frijol. <<http://genebank.ciat.cgiar.org/genebank/beancollection.do>> (accessed march, 2017).
- Clouse JW, Adhikary D, Page JT, Ramaraj T, Deyholos MK, Udall JA, Fairbanks DJ, Jellen EN, Maughan PJ. 2016. The amaranth genome: genome, transcriptome, and physical map assembly. *The Plant Genome* **9**: 1–14. DOI: 10.3835/plantgenome2015.07.0062
- Cobb JN, DeClerck G, Greenberg A, Clark R, McCouch S. 2013. Next-generation phenotyping: requirements and strategies for enhancing our understanding of genotype–phenotype relationships and its relevance to crop improvement. *Theoretical and Applied Genetics* **126**: 867–887. DOI: 10.1007/s00122-013-2066-0
- CONABIO. Centros de origen y diversificación. <<http://www.biodiversidad.gob.mx/genes/centrosOrigen/centrosOrig.html>> (accessed march, 2017).
- Cruse-Sanders JM, Parker KC, Friar EA, Huang DI, Mashayekhi S, Prince LM, Otero-Arnaiz A, Casas A. 2013. Managing diversity: Domestication and gene flow in *Stenocereus stellatus* Riccob. (Cactaceae) in Mexico. *Ecology and Evolution* **3**: 1340–1355. DOI: 10.1002/eee3.524
- Eguiarte LE, Aguirre-Planter E, Aguirre X, Colín R, González A, Rocha M, Scheinvar E, Trejo L, Souza V. 2013. From isozymes to genomics: Population genetics and conservation of *Agave* in Mexico. *The Botanical Review* **79**: 483–506. DOI: 10.1007/s12229-013-9123-x
- Elshire RJ, Glaubitz JC, Sun Q, Poland JA, Kawamoto K, Buckler ES, Mitchell SE. 2011. A robust, simple genotyping-by-sequencing (GBS) approach for high diversity species. *PLoS ONE* **6**: 19379. DOI: 10.1371/journal.pone.0019379
- Escalante AE, Jardón-Barbolla L, Ramírez-Barahona S, Eguiarte LE. 2014. The study of biodiversity in the era of massive sequencing. *Revista Mexicana de la Biodiversidad* **85**: 1249–1264. DOI: 10.7550/rmb.43498
- Gepts P, Papa R. 2003. Evolution during Domestication. Encyclopedia of Life Sciences DOI: 10.1038/npg.els.0003071.
- Gepts P. 2002. A Comparison between Crop Domestication, Classical Plant Breeding, and Genetic Engineering. *Crop Science* **42**: 1780–1790. DOI: 10.2135/cropsci2002.1780.
- Gepts P. 2004. Crop domestication as a long-term selection experiment. *Plant Breeding Reviews* **24**: 1–44. DOI: 10.1002/9780470650288.ch1
- Gepts P. 2014. The contribution of genetic and genomic approaches to plant domestication studies. *Current Opinion in Plant Biology* **18**: 51–59. DOI: 10.1016/j.pbi.2014.02.001
- Goff SA, Ricke D, Lan T, Presting G, Wang R, Dunn M, Glazebrook J, Sessions A, Oeller P, Varma H, Hadley D, Hutchison D, Martin C, Katagiri F, Lange BM, Moughamer T, Xia Y, Budworth P, Zhong J, Miguel T, Paszkowski U, Zhang S, Colbert M, Sun WL, Chen L, Cooper B, Park S, Wood TC, Mao L, Quail P, Wing R, Dean R, Yu Y, Zharkikh A, Shen R, Sahasrabudhe S, Thomas A, Cannings R, Gutin A, Pruss D, Reid J, Tavtigian S, Mitchell J, Eldredge G, Scholl T, Miller RM, Bhatnagar S, Adey N, Rubano T, Tusneem N, Robinson R, Feldhaus J, Macalma T, Olliphant A, Briggs S. 2002. A draft sequence of the rice genome (*Oryza sativa* L. ssp. *japonica*). *Science* **296**: 92–100. DOI: 10.1126/science.1068275
- Grube RC, Radwanski ER, Jahn M. 2000. Comparative genetics of disease resistance within the Solanaeae. *Genetics* **155**: 873–887.
- He J, Zhao X, Laroche A, Lu Z-X, Li Z. 2014. Genotyping by sequencing (GBS), an ultimate marker-assisted selection (MAS) tool to accelerate plant breeding. *Frontiers in Plant Science* **5**: 1–8. DOI: 10.3389/fpls.2014.00484
- Heslop-Harrison JS, Schwarzacher T. 2007. Domestication, genomics and the future for banana. *Annals of Botany* **100**: 1073–1084. DOI: 10.1093/aob/mcm191
- Hirakawa H, Shirasawa K, Kosugi S, Tashiro K, Nakayama S, Yamada M, Kohara M, Watanabe A, Kishida Y, Fujishiro T, Tsuruoka H, Minami C, Sasamoto S, Kato M, Nanri K, Komaki A, Yanagi T, Guoxin Q, Maeda F, Ishikawa M, Kuhara S, Sato S, Tabata S, Isobe SN. 2014. Dissection of the octoploid strawberry genome by deep sequencing of the genomes of *Fragaria* species. *DNA Research* **21**: 169–181. DOI: 10.1093/dnare/dst049
- Hovav R, Chaudhary B, Udall JA, Flagel L, Wendel JF. 2008. Parallel domestication, convergent evolution and duplicated gene recruitment in allopolyploid cotton. *Genetics* **179**: 1725–1733. DOI: 10.1534/genetics.108.089656

- Hu G, Koh J, Yoo M, Grupp K, Chen S, Wendel JF. 2013. Proteomic profiling of developing cotton fibers from wild and domesticated *Gossypium barbadense*. *New Phytologist* **200**: 570-582. DOI: 10.1111/nph.12381
- Huang J, Gau Y, Jia H, Zhang Z. 2016. Characterization of the teosinte transcriptome reveals adaptive sequence divergence during maize domestication. *Molecular Ecology Resources*. DOI: 10.1111/1755-0998.12526
- Huang X, Kurata N, Wei X, Wang Z-X, Wang A, Zhao Q, Zhao Y, Lui K, Lu H, Li W, Lu Y, Zhou C, Fan D, Weng Q, Zhu C, Huang T, Zhang L, Wang Y, Feng L, Furuumi H, Kubo T, Miyabayashi T, Yuan X, Xu Q, Dong G, Zhan Q, Li C, Fujiyama A, Toyoda A, Lu T, Feng Q, Qian Q, Li J, Han B. 2012. A map of rice genome variation reveals the origin of cultivated rice. *Nature* **490**: 497-501. DOI: 10.1038/nature11532
- Hufford MB, Lubinsky P, Pyhäjärvi T, Devengenzo MT, Ellstrand NC, Ross-Ibarra J. 2013. The genomic signatures of crop-wild introgression in maize. *PLOS Genetics* **9**: e1003477. DOI: 10.1371/journal.pgen.1003477
- Hufford MB, Xu X, van Heerwaarden J, Pyhäjärvi T, Chia J-M, Cartwright RA, Elshire RJ, Glaubitz JC, Guill KE, Kaepller SM, Lai J, Morrell PL, Shannon LM, Song C, Springer NM, Swanson-Wagner RA, Tiffin P, Wang J, Zhang G, Doebley J, McMullen MD, Ware D, Buckler ES, Yang S, Ross-Ibarra J. 2012. Comparative population genomics of maize domestication and improvement. *Nature Genetics* **44**: 808-811. DOI: 10.1038/ng.2309
- Iorizzo M, Senalik DA, Ellison SL, Grzelbelus D, Cavagnaro PF, Allender C, Brunet J, Spooner DM, Van Deynze A, Simon PW. 2013. Genetic structure and domestication of carrot (*Daucus carota* subsp. *sativus*) (Apiaceae). *American Journal of Botany* **100**: 930-938. DOI: 10.3732/ajb.1300055
- Iwata H, Gaston A, Remay A, Thouroude T, Jeauffre J, Kawamura K, Oyant LH, Araki T, Denoyes B, Foucher F. 2012. The *TFL1* homologue *KSN* is a regulator of continuous flowering in rose and strawberry. *The Plant Journal* **69**: 116-125. DOI: 10.1111/j.1365-313X.2011.04776.x
- Jarvis DE, Ho YS, Lightfoot DJ, Schmöckel SM, Li B, Borm TJA, Ohyanagi H, Mineta K, Michell CT, Saber N, Kharbatia NM, Rupper RR, Sharp AR, Dally N, Boughton BA, Woo YH, Gao G, Schijlen EGWM, Guo X, Momin AA, Negrão S, Al-Babili S, Gehring C, Roessner U, Jung C, Murphy K, Arold ST, Gojobori T, van der Linden CG, van der Loo EN, Jellen EN, Maughan PJ, Tester M. 2017. The genome of *Chenopodium quinoa*. *Nature* **542**: 304-312. DOI: 10.1038/nature21370
- Jiang Z, Xia H, Basso B, Lu B-R. 2011. Introgression from cultivated rice influences genetic differentiation of weedy rice population at local spatial scale. *Theoretical and Applied Genetics* **124**: 309-322. DOI: 10.1007/s00122-011-1706-5
- Kim C, Guo H, Kong W, Chandnani R, Shuang LS, Paterson AH. 2015. Application of genotyping by sequencing technology to a variety of crop breeding programs. *Plant Science* **242**: 14-22. DOI: 10.1016/j.plantsci.2015.04.016
- Kim S, Park M, Yeom SI, Kim YM, Lee JM, Lee HA, Seo E, Choi J, Cheong K, Kim KT, Jung K, Lee GW, Oh SK, Bae C, Kim SB, Lee HY, Kim SY, Kim MS, Kang BC, Jo YD, Yang HB, Jeong HJ, Kang WH, Kwon JK, Shin C, Lim JY, Park JH, Huh JH, Kim JS, Kim BD, Cohen O, Paran I, Suh MC, Lee SB, Kim YK, Shin Y, Noh SJ, Park J, Seo YS, Kwon SY, Kim HA, Park JM, Kim HJ, Choi SB, Bosland PW, Reeves G, Jo SH, Lee BW, Cho HT, Choi HS, Lee MS, Yu Y, Do Choi Y, Park BS, van Deynze A, Ashrafi H, Hill T, Kim WT, Pai HS, Ahn HK, Yeam I, Giovannoni JJ, Rose JK, Sørensen I, Lee SJ, Kim RW, Choi IY, Choi BS, Lim JS, Lee YH, Choi D. 2014. Genome sequence of the hot pepper provides insights into the evolution of pungency in *Capsicum* species. *Nature Genetics* **46**: 270-278. DOI: 10.1038/ng.2877
- Komatsuda T, Pourkheirandish M, He C, Azhagavel P, Kanamori H, Perovic D, Stein N, Graner A, Wicker T, Tagiri A, Lundqvist U, Fujimura T, Matsuoka M, Matsumoto T, Yano M. 2007. Six-rowed barley originated from a mutation in a homeodomain-leucine zipper I-class homeobox gene. *Proceedings of the National Academy of Sciences of the United States of America* **104**: 1424-1429. DOI: 10.1073/pnas.0608580104
- Kono TJK, Fu F, Mohammadi M, Hoffman PJ, Liu C, Stupar RM, Smith KP, Tiffin P, Fay JC, Morrel PL. 2016. The role of deleterious substitutions in crop genomes. *Molecular Biology and Evolution* **33**: 2307-2317. DOI: 10.1093/molbev/msw102
- Larson G, Piperno DR, Allaby RG, Purugganan MD, Andersson L, Arollo-Kalin M, Barton L, Vigueira CC, Denham T, Dobney K, Doust AN, Gepts P, Gilbert MT, Gremillion KJ, Lucas L, Lukens L, Marshall FB, Olsen KM, Pires JC, Richerson PJ, Rubio de Casas R, Sanjur OI, Thomas MG, Fuller DQ. 2014. Current perspectives and the future of domestication studies. *Proceedings of the National Academy of Sciences of the United States of America* **111**: 6139-6146. DOI: 10.1073/pnas.1323964111
- Li F, Fan G, Lu C, Xiao G, Zou C, Kohel RJ, Ma Z, Shang H, Ma X, Wu J, Liang X, Huang G, Percy RG, Liu K, Yang W, Chen W, Du X, Shi C, Yuan Y, Ye W, Liu X, Zhang X, Liu W, Wei H, Wei S, Huang G, Zhang X, Zhu S, Zhang H, Sun F, Wang X, Liang J, Wang J, He Q, Huang L, Wang J, Cui J, Song G,

- Wang K, Xu X, Yu JZ, Zhu Y, Yu S. 2015. Genome sequence of cultivated Upland cotton (*Gossypium hirsutum* TM-1) provides insights into genome evolution. *Nature Biotechnology* **33**: 524-530. DOI: 10.1038/nbt.3208
- Li F, Fan G, Wang K, Sun F, Yuan Y, Song G, Li Q, Ma Z, Lu C, Zou C, Chen W, Liang X, Shang H, Liu W, Shi C, Xiao G, Gou C, Ye W, Xu X, Zhang X, Wei H, Li Z, Zhang G, Wang J, Liu K, Kohel RJ, Percy RG, Yu JZ, Zhu YX, Wang J, Yu S. 2014. Genome sequence of the cultivated cotton *Gossypium arboreum*. *Nature Genomics* **46**: 467-472. DOI: 10.1038/ng.2987
- Li YH, Zhao SC, Ma JX, Li D, Yan L, Li J, Qi XT, Guo XS, Zhang L, He WM, Chang RZ, Liang QS, Guo Y, Ye C, Wang XB, Tao Y, Guan RX, Wang JY, Liu YL, Jin LG, Zhang XQ, Liu ZX, Zhang LJ, Chen J, Wang KJ, Nielsen R, Li RQ, Chen PY, Li WB, Reif JC, Purugganan M, Wang J, Zhang MC, Wang J, Qiu LJ. 2013. Molecular footprints of domestication and improvement in soybean revealed by whole genome re-sequencing. *BMC Genomics* **14**: 1-12. DOI: 10.1186/1471-2164-14-579
- Lin Z, Li X, Shannon LM, Yeh C-T, Wang ML, Bai G, Peng Z, Li J, Trick HN, Clemente TE, Doebley J, Schnable PS, Tuinstra MR, Tesso TT, White F, Yu J. 2012. Parallel domestication of the *Shattering1* genes in cereals. *Nature Genetics* **44**: 720-724. DOI: doi:10.1038/ng.2281
- Liu X, Zhao B, Zheng HJ, Hu Y, Lu G, Yang CQ, Chen JD, Chen JJ, Chen DY, Zhang L, Zhou Y, Wang LJ, Guo WZ, Bai YL, Ruan JX, Shangguan XX, Mao YB, Shan CM, Jiang JP, Zhu YQ, Jin L, Kang H, Chen ST, He XL, Wang R, Wang YZ, Chen J, Wang LJ, Yu ST, Wang BY, Wei J, Song SC, Lu XY, Gao ZC, Gu WY, Deng X, Ma D, Wang S, Liang WH, Fang L, Cai CP, Zhu XF, Zhou BL, Jeffrey Chen Z, Xu SH, Zhang YG, Wang SY, Zhang TZ, Zhao GP, Chen XY. 2015. *Gossypium barbadense* genome sequence provides insight into evolution of extra-long staple fiber and specialized metabolites. *Scientific Reports* **5**: 1-14. DOI: 10.1038/srep14139
- Lu J, Tang T, Tang H, Huang J, Shi S, Wu CI. 2006. The accumulation of deleterious mutations in rice genomes: a hypothesis on the cost of domestication. *Trends in Genetics* **22**: 126-131. DOI: 10.1016/j.tig.2006.01.004
- Mackay TF, Stone E, Ayroles JF. 2009. The genetics of quantitative traits: challenges and prospects. *Nature Reviews Genetics* **10**: 565-577. DOI: 10.1038/nrg2612
- Meyer RS, Purugganan MD. 2013. Evolution of crop species: genetics of domestication and diversification. *Nature Reviews Genetics* **14**: 840-852. DOI: 10.1038/nrg3605
- Mezmouk S, Roos-Ibarra J. 2014. The pattern and distribution of deleterious mutations in maize. *G3: Genes, Genomes, Genetics* **4**: 163-171. DOI: 10.1534/g3.113.008870
- Ming R, Hou S, Feng Y, Yu Q, Dionne-Laporte A, Saw JH, Senin P, Wang W, Ly BV, Lewis KL, Salzberg SL, Feng L, Jones MR, Skelton RL, Murray JE, Chen C, Qian W, Shen J, Du P, Eustice M, Tong E, Tang H, Lyons E, Paull RE, Michael TP, Wall K, Rice DW, Albert H, Wang ML, Zhu YJ, Schatz M, Nagarajan N, Acob RA, Guan P, Blas A, Wai CM, Ackerman CM, Ren Y, Liu C, Wang J, Wang J, Na JK, Shakirov EV, Haas B, Thimmapuram J, Nelson D, Wang X, Bowers JE, Gschwend AR, Delcher AL, Singh R, Suzuki JY, Tripathi S, Neupane K, Wei H, Irikura B, Paidi M, Jiang N, Zhang W, Presting G, Windsor A, Navajas-Pérez R, Torres MJ, Feltus FA, Porter B, Li Y, Burroughs AM, Luo MC, Liu L, Christopher DA, Mount SM, Moore PH, Sugimura T, Jiang J, Schuler MA, Friedman V, Mitchell-Olds T, Shippen DE, dePamphilis CW, Palmer JD, Freeling M, Paterson AH, Gonsalves D, Wang L, Alam M. 2008. The draft genome of the transgenic tropical fruit papaya (*Carica papaya* Linnaeus). *Nature* **452**: 991-996. DOI: 10.1038/nature06856
- Ming R, VanBuren R, Wai CM, Tang H, Schatz MC, Bowers JE, Lyons E, Wang ML, Chen J, Biggers E, Zhang J, Huang L, Zhang L, Miao W, Zhang J, Ye Z, Miao C, Lin Z, Wang H, Zhou H, Yim WC, Priest HD, Zheng C, Woodhouse M, Edger PP, Guyot R, Guo HB, Guo H, Zheng G, Singh R, Sharma A, Min X, Zheng Y, Lee H, Gurtowski J, Sedlacek FJ, Harkess A, McKain MR, Liao Z, Fang J, Liu J, Zhang X, Zhang Q, Hu W, Qin Y, Wang K, Chen LY, Shirley N, Lin YR, Liu LY, Hernandez AG, Wright CL, Bulone V, Tuskan GA, Heath K, Zee F, Moore PH, Sunkar R, Leebens-Mack JH, Mockler T, Bennetzen JL, Freeling M, Sankoff D, Paterson AH, Zhu X, Yang X, Smith JA, Cushman JC, Paull RE, Yu Q. 2015. The pineapple genome and the evolution of CAM photosynthesis. *Nature genetics* **47**: 1435-1442. DOI: 10.1038/ng.3435
- Molina J, Sikora M, Garud N, Flowers JM, Rubinstein S, Reynolds A, Huang P, Jackson S, Schaal BA, Bustamante CD, Boyko AR, Purugganan MD. 2011. Molecular evidence for a single evolutionary origin of domesticated rice. *Proceedings of the National Academy of Sciences of the United States of America* **7**: 8351-8356. DOI: 10.1073/pnas.1104686108
- Montes-Hernández S, Eguiarte LE. 2002. Genetic structure indirect estimates of gene flow in three taxa of *Cucurbita* (Cucurbitaceae) in western Mexico. *American Journal of Botany* **89**: 1156-1163. DOI: 10.3732/ajb.89.7.1156
- Montes-Hernández S, Merrick LC, Eguiarte LE. 2005. Maintenance of squash (*Cucurbita* spp.) landrace diversity by farmers' activities in Mexico. *Genetic Resources and Crop Evolution* **52**: 697-707. DOI:10.1007/s10722-003-6018-4

- Morrell PL, Bucker ES, Ross-Ibarra J. 2011. Crop genomics: advances and applications. *Nature Reviews Genetics* **13**: 85-96. DOI: 10.1038/nrg3097
- Morris GP, Ramu P, Deshpande SP, Hash CT, Shah T, Upadhyaya HD, Riera-Lizarazu O, Brown PJ, Acharya CB, Mitchell SE *et al.* 2013. Population genomic and genome-wide association studies of agroclimatic traits in sorghum. *Proceedings of the National Academy of Sciences of the United States of America* **110**: 453-458. DOI: 10.1073/pnas.1215985110
- Myles S, Boyko AR, Owens CL, Brown PJ, Grassi F, Aradhya MK, Prins B, Reynolds A, Chia J-M, Ware D, Bustamente CD, Buckler ES. 2011. Genetic domestication history of the grape. *Proceedings of the National Academy of Sciences of the United States of America* **108**: 3530-3535. DOI: 10.1073/pnas.1009363108
- Olsen KM, Caicedo AL, Jia Y. 2007. Evolutionary of genomics of weedy rice in the USA. *Journal of Integrative Plant Biology* **49**: 811-816. DOI: 10.1111/j.1744-7909.2007.00506.x
- Papa R, Gepts P. 2003. Asymmetry of gene flow and differential geographical structure of molecular diversity in wild and domesticated common bean (*Phaseolus vulgaris* L.) from Mesoamerica. *Theoretical and Applied Genetics* **106**: 239-250. DOI: 10.1007/s00122-002-1085-z
- Paterson AH, Bowers JE, Bruggmann R, Dubchak I, Grimwood J, Gundlach H, Haberer G, Hellsten U, Mitros T, Poliakov A, Schmutz J, Spannagl M, Tang H, Wang X, Wicker T, Bharti AK, Chapman J, Feltus FA, Gowik U, Grigoriev IV, Lyons E, Maher CA, Martis M, Narechania A, Otillar RP, Penning BW, Salamov AA, Wang Y, Zhang L, Carpita NC, Freeling M, Gingle AR, Hash CT, Keller B, Klein P, Kresovich S, McCann MC, Ming R, Peterson DG, Mehboob-ur-Rahman, Ware D, Westhoff P, Mayer KF, Messing J, Rokhsar DS. 2009. The Sorghum bicolor genome and the diversification of grasses. *Nature* **457**: 551-556. DOI: 10.1038/nature07723
- Pickersgill B. 2009. Domestication of plants revisited – Darwin to the present day. *Botanical Journal of the Linnean Society* **161**: 203-212. DOI: 10.1111/j.1095-8339.2009.01007.x
- Polashock J, Zelzion E, Fajardo D, Zalapa J, Georgi L, Bhattacharya D, Vorsa N. 2014. The American cranberry: first insights into the whole genome of a species adapted to bog habitat. *BMC Plant Biology* **14**: 165-181. DOI: 10.1186/1471-2229-14-165
- Prochnik S, Marri PR, Desany B, Rabinowicz PD, Kodira C, Mohiuddin M, Rodriguez F, Fauquet C, Tohme J, Harkins T, Rokhsar DS, Rounsley S. 2012. The cassava genome: current progress, future directions. *Tropical Plant Biology* **5**: 88-94. DOI: 10.1007/s12042-011-9088-z
- Purugganan MD, Boyles AL, Suddith J. 2000. Variation and selection at the CAULIFLOWER floral homeotic gene accompanying the evolution of domesticated *Brassica oleracea*. *Genetics* **155**: 855-862.
- Renaut S, Rieseberg LH. 2015. The accumulation of deleterious mutation as a consequence of domestication and improvement in sunflowers and other Compositae crops. *Molecular Biology and Evolution* **32**: 2273-2283. DOI: 10.1093/molbev/msv106
- Rieseberg LH, Carney SE. 1998. Plant hybridization. *New Phytologist* **140**: 599-624. DOI: 10.1046/j.1469-8137.1998.00315.x
- Rodriguez GR, Muños S, Anderson C, Sim SC, Michel A, Causse M, Gardener BB, Francis D, van der Knaap E. 2011. Distribution of SUN, OVATE, LC and FAS in the tomato germplasm and the relationships to fruit shape diversity. *Plant physiology* **156**: 275-285. DOI: 10.1104/pp.110.167577
- Ross-Ibarra J, Morrell PL, Gaut BS. 2007. Plant domestication, a unique opportunity to identify the genetic basis of adaptation. *Proceedings of the National Academy of Sciences of the United States of America* **104**: 8641-8648. DOI: 10.1073/pnas.0700643104
- Schmutz J, Cannon SB, Schlueter J, Ma J, Mitros T, Nelson W, Hyten DL, Song Q, Thelen JJ, Cheng J, Xu D, Hellsten U, May GD, Yu Y, Sakurai T, Umezawa T, Bhattacharyya MK, Sandhu D, Valliyodan B, Lindquist E, Peto M, Grant D, Shu S, Goodstein D, Barry K, Futrell-Griggs M, Abernathy B, Du J, Tian Z, Zhu L, Gill N, Joshi T, Libault M, Sethuraman A, Zhang XC, Shinozaki K, Nguyen HT, Wing RA, Cregan P, Specht J, Grimwood J, Rokhsar D, Stacey G, Shoemaker RC, Jackson SA. 2010. Genome sequence of the palaeopolyploid soybean. *Nature* **463**: 178-183. DOI: 10.1038/nature08670
- Schmutz J, McClean PE, Mamidi S, Wu GA, Cannon SB, Grimwood J, Jenkins J, Shu S, Song Q, Chavarro C, Torres-Torres M, Geffroy V, Moghaddam SM, Gao D, Abernathy B, Barry K, Blair M, Brick MA, Chovatia M, Gepts P, Goodstein DM, Gonzales M, Hellsten U, Hyten DL, Jia G, Kelly JD, Kudrna D, Lee R, Richard MM, Miklas PN, Osorno JM, Rodrigues J, Thareau V, Urrea CA, Wang M, Yu Y, Zhang M, Wing RA, Cregan PB, Rokhsar DS, Jackson SA. 2014. A reference genome for common bean and genome-wide analysis of dual domestication. *Nature genetics* **46**: 707-713. DOI: 10.1038/ng.3008
- Schnable PS, Ware D, Fulton RS, Stein JC, Wei F, Pasternak S, Liang C, Zhang J, Fulton L, Graves TA, Minx P, Reily AD, Courtney L, Kruchowski SS, Tomlinson C, Strong C, Delehaunty K, Fronick C, Courtney B, Rock SM, Belter E, Du F, Kim K, Abbott RM, Cotton M, Levy A, Marchetto P, Ochoa K, Jackson SM, Gillam B, Chen W, Yan L, Higginbotham J, Cardenas M, Waligorski J, Applebaum E, Phelps L, Falcone J, Kanchi K, Thane T, Scimone A, Thane N, Henke J, Wang T, Ruppert J, Shah N, Rotter K, Hodges J, Ingenthron E, Cordes M, Kohlberg S, Sgro J, Delgado B, Mead K, Chinwalla A,

- Leonard S, Crouse K, Collura K, Kudrna D, Currie J, He R, Angelova A, Rajasekar S, Mueller T, Lomeli R, Scara G, Ko A, Delaney K, Wissotski M, Lopez G, Campos D, Braidotti M, Ashley E, Golser W, Kim H, Lee S, Lin J, Dujmic Z, Kim W, Talag J, Zuccolo A, Fan C, Sebastian A, Kramer M, Spiegel L, Nascimento L, Zutavern T, Miller B, Ambroise C, Muller S, Spooner W, Narechania A, Ren L, Wei S, Kumari S, Faga B, Levy MJ, McMahan L, Van Buren P, Vaughn MW, Ying K, Yeh CT, Emrich SJ, Jia Y, Kalyanaraman A, Hsia AP, Barbazuk WB, Baucom RS, Brutnell TP, Carpita NC, Chaparro C, Chia JM, Derragon JM, Estill JC, Fu Y, Jeddelloh JA, Han Y, Lee H, Li P, Lisch DR, Liu S, Liu Z, Nagel DH, McCann MC, SanMiguel P, Myers AM, Nettleton D, Nguyen J, Penning BW, Ponnala L, Schneider KL, Schwartz DC, Sharma A, Soderlund C, Springer NM, Sun Q, Wang H, Waterman M, Westerman R, Wolfgruber TK, Yang L, Yu Y, Zhang L, Zhou S, Zhu Q, Bennetzen JL, Dawe RK, Jiang J, Jiang N, Presting GG, Wessler SR, Aluru S, Martienssen RA, Clifton SW, McCombie WR, Wing RA, Wilson RK. 2009. The B73 maize genome: Complexity, diversity and dynamics. *Science* **326**: 1112–1115. DOI: 10.1126/science.1178534
- Sierro N, Battey JN, Ouadi S, Bakaher N, Bovet L, Willig A, Geopfert S, Peitsch MC, Ivanov NV. 2014. The tobacco genome sequence and its comparison with those of tomato and potato. *Nature Communications* **5**: 1–9. DOI: 10.1038/ncomms4833.
- Studer A, Zhao Q, Ross-Ibarra J, Doebley J. 2011. Identification of a functional transposon insertion in the maize domestication gene tb1. *Nature Genetics* **43**: 1160–1163. DOI: 10.1038/ng.942
- Swanson-Wagner R, Briskine R, Schaefer R, Schaefer R, Hufford MB, Ross-Ibarra J, Myers CL, Tiffin P, Springer NM. 2012. Reshaping of the maize transcriptome by domestication. *Proceedings of the National Academy of Sciences of the United States of America* **109**: 11878–11883. DOI: 10.1073/pnas.1201961109
- Tang H, Sezen U, Paterson AH. 2010. Domestication and plant genomes. *Current Opinion in Plant Biology* **13**: 160–166. DOI: 10.1016/j.pbi.2009.10.008
- The Potato Genome Sequencing Consortium. Genome sequence and analysis of the tuber crop potato. *Nature* **475**: 189–195. DOI: 10.1038/nature10158
- The Tomato Genome Consortium. 2012. The tomato genome sequence provides insights into fleshy fruit evolution. *Nature* **485**: 635–641. DOI: 10.1038/nature11119
- Varshney RK, Terauchi R, McCouch SR. 2014. Harvesting the promising fruits of genomics: Applying genome sequencing technologies to crop breeding. *PLOS Biology* **12**: 1–8. DOI: 10.1371/journal.pbio.1001883
- Vavilov NI. 1922. The law of homologous series in variation. *Journal of Genetics* **12**: 47–89
- Wang H, Xu X, Vieira FG, Xiao Y, Li Z, Wang J, Neilsen R, Chu C. 2016. The power of inbreeding: NGS based GWAS of rice reveals convergent evolution during rice domestication. *Molecular Plant* **9**: 975–85. DOI: 10.1016/j.molp.2016.04.018
- Wang K, Wang Z, Li F, Ye W, Wang J, Song G, Yue Z, Cong L, Shang H, Zhu S, Zou C, Li Q, Yuan Y, Lu C, Wei H, Gou C, Zheng Z, Yin Y, Zhang X, Liu K, Wang B, Song C, Shi N, Kohel RJ, Percy RG, Yu JZ, Zhu YX, Wang J, Yu S. 2012. The draft genome of a diploid cotton *Gossypium raimondii*. *Nature Genetics* **44**: 1098–1103. DOI: 10.1038/ng.2371
- Weeden NF. 2007. Genetic changes accompanying the domestication of *Pisum sativum*: is there a common basis to the domestication syndrome for legumes?. *Annals of Botany* **100**: 1017–1025. DOI: 10.1093/aob/mcm122
- Wu GA, Prochnik S, Jenkins J, Salse J, Hellsten U, Murat F, Perrier X, Ruiz M, Scalabrin S, Terol J, Takita MA, Labadie K, Poulain J, Couloux A, Jabbari K, Cattonaro F, Del Fabbro C, Pinosio S, Zuccolo A, Chapman J, Grimwood J, Tadeo FR, Estornell LH, Muñoz-Sanz JV, Ibanez V, Herrero-Ortega A, Aleza P, Pérez-Pérez J, Ramón D, Brunel D, Luro F, Chen C, Farmerie WG, Desany B, Kodira C, Mohiuddin M, Harkins T, Fredrikson K, Burns P, Lomsadze A, Borodovsky M, Reforgiato G, Freitas-Astúa J, Quetier F, Navarro L, Roose M, Wincker P, Schmutz J, Morgante M, Machado MA, Talon M, Jaillon O, Ollitrault P, Gmitter F, Rokhsar D. 2014. Sequencing of diverse mandarin, pummelo and orange genomes reveals complex history of admixture during citrus domestication. *Nature Biotechnology* **32**: 656–662. DOI: 10.1038/nbt.2906
- Wu Y, San Vicente F, Huang K, Dhliwayo T, Costich DE, Semagn K, Sudha N, Olsen M, Prasanna BM, Zhang X, Babu R. 2016. Molecular characterization of CIMMYT maize inbred lines with genotyping-by-sequencing SNPs. *Theoretical and Applied Genetics* **129**: 753–765. DOI: 10.1007/s00122-016-2664-8
- Xia HB, Wang W, Xia H, Zhao W, Lu BR. 2011. Conspecific crop-weed introgression influences evolution of weedy rice (*Oryza sativa* f. *spontanea*) across a geographical range. *PLoS ONE* **6**: 1–11. DOI: 10.1371/journal.pone.0016189
- Xu J, Yuan Y, Xu Y, Zhang G, Guo X, Wu F, Wang Q, Rong T, Pan G, Cao M, Tang Q, Gao S, Liu Y, Wang J, Lan H, Lu Y. 2014. Identification of candidate genes for drought tolerance by whole-genome resequencing in maize. *BMC Plant Biology* **14**: 1–15. DOI: 10.1186/1471-2229-14-83.

Received:
January 17th, 2017

Accepted:
May 12th, 2017

- Yu J, Hu S, Wang J, Wong KS, Li S, Liu B, Deng Y, Dai L, Zhou Y, Zhang X, Cao M, Liu J, Sun J, Tang J, Chen Y, Huang X, Lin W, Ye C, Tong W, Cong L, Geng J, Han Y, Li L, Li W, Hu G, Huang X, Li W, Li J, Liu Z, Li L, Liu J, Qi Q, Liu J, Li L, Li T, Wang X, Lu H, Wu T, Zhu M, Ni P, Han H, Dong W, Ren X, Feng X, Cui P, Li X, Wang H, Xu X, Zhai W, Xu Z, Zhang J, He S, Zhang J, Xu J, Zhang K, Zheng X, Dong J, Zeng W, Tao L, Ye J, Tan J, Ren X, Chen X, He J, Liu D, Tian W, Tian C, Xia H, Bao Q, Li G, Gao H, Cao T, Wang J, Zhao W, Li P, Chen W, Wang X, Zhang Y, Hu J, Wang J, Liu S, Yang J, Zhang G, Xiong Y, Li Z, Mao L, Zhou C, Zhu Z, Chen R, Hao B, Zheng W, Chen S, Guo W, Li G, Liu S, Tao M, Wang J, Zhu L, Yuan L, Yang H. 2002. A draft sequence of the rice genome (*Oryza sativa* L. ssp. *indica*). *Science* **296**: 79–92. DOI: 10.1126/science.1068037
- Zizumbo-Villarreal D, Flores-Silva A, Colunga-García Marín P. 2012. The archaic diet in Mesoamerica: Incentive for milpa development and species domestication. *Economic Botany* **66**: 328–343. DOI: 10.1007/s12231-012-9212-5