



Genomic, Transcriptomic and Epigenomic Tools to Study the Domestication of Plants and Animals: A Field Guide for Beginners

Josué Barrera-Redondo, Daniel Piñero and Luis E. Eguiarte*

Departamento de Ecología Evolutiva, Instituto de Ecología, Universidad Nacional Autónoma de México, Mexico City, Mexico

OPEN ACCESS

Edited by:

TingFung Chan,
The Chinese University of Hong Kong,
China

Reviewed by:

Eric Von Wettberg,
The University of Vermont,
United States
Martin Mascher,
Leibniz Institute for Plant Genetics
and Cultural Plant Research (IPK),
Germany
David Irwin,
University of Toronto, Canada

*Correspondence:

Luis E. Eguiarte
fruns@unam.mx

Specialty section:

This article was submitted to
Evolutionary and Population Genetics,
a section of the journal
Frontiers in Genetics

Received: 25 April 2020

Accepted: 22 June 2020

Published: 15 July 2020

Citation:

Barrera-Redondo J, Piñero D and
Eguiarte LE (2020) Genomic,
Transcriptomic and Epigenomic Tools
to Study the Domestication of Plants
and Animals: A Field Guide
for Beginners. *Front. Genet.* 11:742.
doi: 10.3389/fgene.2020.00742

In the last decade, genomics and the related fields of transcriptomics and epigenomics have revolutionized the study of the domestication process in plants and animals, leading to new discoveries and new unresolved questions. Given that some domesticated taxa have been more studied than others, the extent of genomic data can range from vast to nonexistent, depending on the domesticated taxon of interest. This review is meant as a rough guide for students and academics that want to start a domestication research project using modern genomic tools, as well as for researchers already conducting domestication studies that are interested in following a genomic approach and looking for alternate strategies (cheaper or more efficient) and future directions. We summarize the theoretical and technical background needed to carry out domestication genomics, starting from the acquisition of a reference genome and genome assembly, to the sampling design for population genomics, paleogenomics, transcriptomics, epigenomics and experimental validation of domestication-related genes. We also describe some examples of the aforementioned approaches and the relevant discoveries they made to understand the domestication of the studied taxa.

Keywords: population genomics, pangenomics, ancient DNA, differential expression analysis, epialleles, genome editing

INTRODUCTION

The modern study of domestication of plants and animals is multidisciplinary, and relevant contributions come from botany, zoology, archeology, genetics, ethnobiology, biogeography, and linguistics (Larson et al., 2014). Modern domestication studies seek to understand the dates of domestication, the places where domestication started and number of times that domestication took place, as well as the details of the evolutionary and ecological forces that led to the divergence between the domesticated taxa and their wild relatives and ancestors (Zeder, 2006; Larson et al., 2014).

Given that domestication is an evolutionary process, genetics emerged as a powerful tool to understand the domestication of plants and animals, revealing the demographic history of the domesticated taxa and the genetic variants that underlie their domesticated phenotypes (Zeder et al., 2006; Gepts, 2014). The advent of high-throughput sequencing technologies sparked the use of genomic studies to understand the domestication of crops and animals in a much deeper

level than previously imagined, as researchers can now pinpoint the genetic changes that allowed domestication to happen (Ross-Ibarra et al., 2007; Gepts, 2014).

WHY AND HOW TO USE A GENOMIC APPROACH IN DOMESTICATION STUDIES? TOP-DOWN AND BOTTOM-UP APPROACHES FOR THE STUDY OF DOMESTICATION

In genetics, we refer to *top* or *up* when referring to a specific phenotype, while we refer to *bottom* or *down* when referring to the underlying genotype responsible for that trait. Thus, top-down approaches start by studying a particular phenotype and searching for its genetic basis. Huge advances in the genetic study of domestication traits have been made using classic top-down approaches (e.g., Sax, 1923; Paterson et al., 1988; Doebley and Stec, 1991; Doebley et al., 1995), which are performed by analyzing the phenotypic traits of interest between wild and domesticated taxa, and then finding the genetic variant or variants that correlate with the phenotypic traits through the mapping of quantitative trait loci and linkage disequilibrium (Ross-Ibarra et al., 2007; Kantar et al., 2017). These top-down approaches are precise in finding causal variants involved in the evolution of specific traits, but usually they are very labor-intensive and are biased towards *a priori* selected phenotypes to be compared between wild and domesticated taxa (Ross-Ibarra et al., 2007; Kantar et al., 2017).

In contrast to top-down approaches, bottom-up approaches start by analyzing the genetic variation within genomes in order to detect potential signals of selection related to the domestication process and finally associate such evolutionary signals to important loci and domestication phenotypes (Ross-Ibarra et al., 2007; Kantar et al., 2017). In the last decade, high-throughput sequencing technologies allowed us to analyze entire genomes of one or several individuals of domesticated taxa, and to compare them to different varieties or to their wild relatives (e.g., Hufford et al., 2012; Yang et al., 2012; Li et al., 2013; Wang et al., 2019; Zeng et al., 2019).

Bottom-up approaches do not need an *a priori* phenotypic target, enabling a genome-wide search of domestication-related loci without previous background of possible candidates, revealing important traits that can hardly be studied using a top-down approach (Ross-Ibarra et al., 2007; Kantar et al., 2017). Nevertheless, the results of bottom-up approaches can be limited by the sampling scheme, the density of genetic markers, and the detection of false positives (Tiffin and Ross-Ibarra, 2014), so these genomic approaches have to be properly and carefully designed in order to obtain satisfying results (De Mita et al., 2013).

Genomic data facilitated the widespread and reliable use of bottom-up approaches to study plant and animal domestication, but top-down strategies were also aided by genomics, allowing a more efficient search of genotype-phenotype correlations through genome-wide association studies (GWAS; Wang G.-D. et al., 2014), which can be defined as experimental

designs that are used to detect the association between genetic variation in a population and phenotypic traits of interest (Visscher et al., 2017).

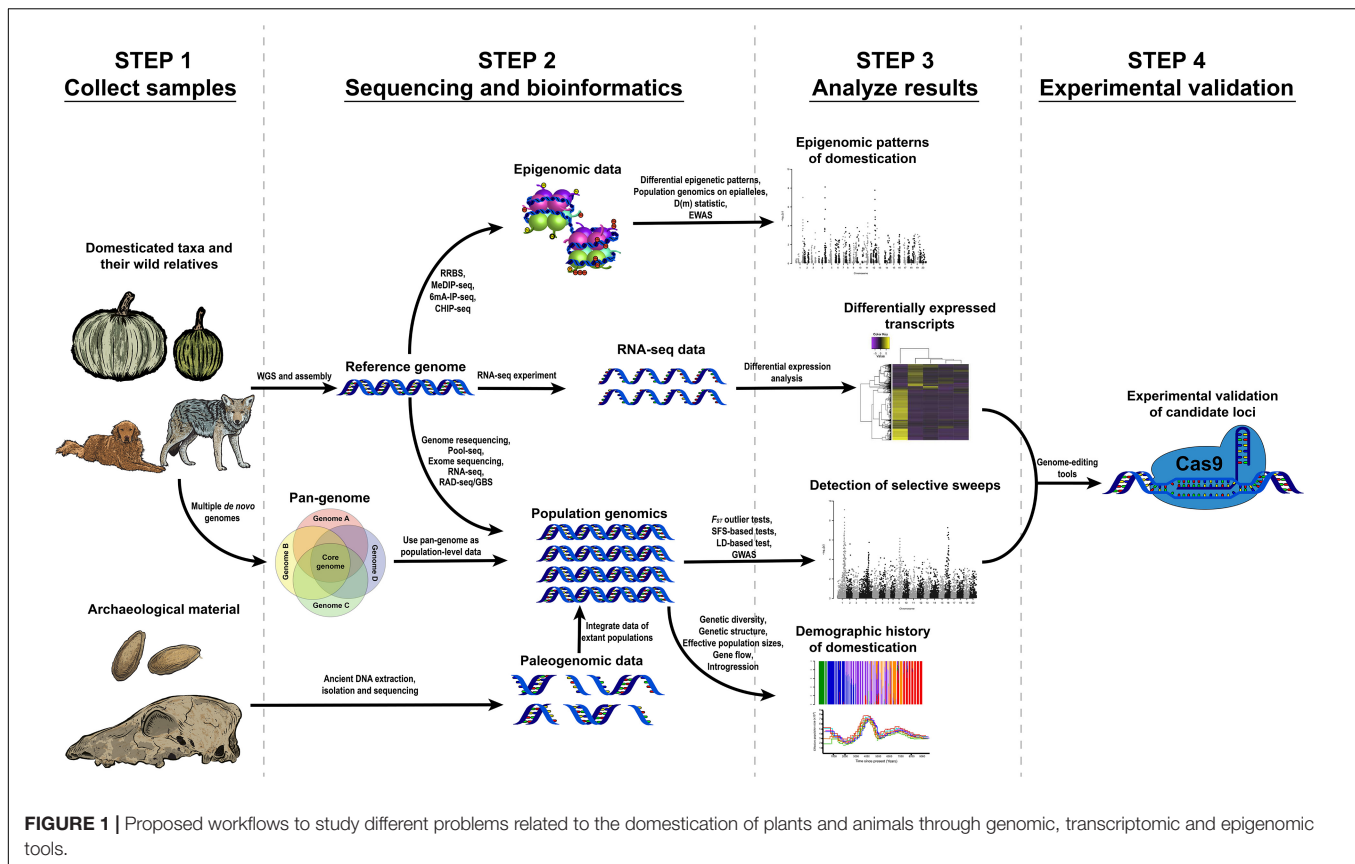
Genome-wide genetic markers allows to differentiate between global and local evolutionary signals occurring throughout the genome (Diao and Chen, 2012), discerning the signals of selection during domestication (Vitti et al., 2013) from other fine-scale signals of demographic events that occurred during the domestication process (Meyer and Purugganan, 2013; Guerra García and Piñero, 2017).

The use of modern genomic tools is not limited to population genetics, as other interesting approaches can reveal important aspects of the domestication process. For instance, one can analyze changes in the transcriptional activity of genes related to domestication (Hekman et al., 2015), demonstrate the phenotypic effects of certain alleles through the use of genomic editing tools (Zhou J. et al., 2019), search for epigenetic patterns that changed between domesticated and wild taxa (Janowitz Koch et al., 2016) or analyze the genetic makeup of archeological samples (Irving-Pease et al., 2019).

This review describes the necessary steps and data to start a genomic research project towards understanding domestication, the questions that can be approached using genomic data and the main results obtained from previous studies using these methods (Figure 1).

WHOLE-GENOME ASSEMBLY AND REFERENCE GENOMES

Whole-genome assembly is one of the first steps in modern domestication studies, since it generates a reference genome that is useful for downstream analyses. Whole-genome assembly projects require the use of high-throughput sequencing technologies such as Illumina (e.g., Sun et al., 2017), PacBio (e.g., Badouin et al., 2017; VanBuren et al., 2018), Oxford Nanopore (e.g., Belser et al., 2018) or a combination of these (e.g., Bickhart et al., 2017; Zhou Y. et al., 2019) to sequence the genome of interest of a single individual. Before starting a genome assembly project, a rough estimate of the haploid genome size must be known as well as the ploidy of the organism, since the assembly difficulty and sequencing cost are determined by both factors (Sims et al., 2014). In order to successfully assemble eukaryotic genomes, where repetitive elements usually comprise a significant portion of its content [ranging from 3% in tiny genomes such as *Utricularia gibba* (Ibarra-Laclette et al., 2013) up to 65.5% in huge genomes such as *Ambystoma mexicanum* (Nowoshilow et al., 2018)], it is necessary to generate sequencing libraries with large insert sizes – called mate-pair libraries – or use long-read sequencing technologies such as PacBio or Oxford Nanopore (Levy and Myers, 2016; Sohn and Nam, 2016). Additionally, the use of chromosome conformation capture (Mascher et al., 2017), optical mapping (Dong et al., 2013) or linkage maps obtained from crosses (Fierst, 2015) will help achieve chromosome-level assemblies that are highly desirable to adequately assess haplotypes, linkage disequilibrium, putative genomic rearrangements and the



genomic location of candidate loci (Sohn and Nam, 2016; see **Table 1**).

After sequencing and assembling the genome of at least one individual, it must be properly annotated before it can be of any use. Since eukaryotic genes are structurally complex, genome assemblies require the additional sequencing of RNA data from the same species to be used as transcriptomic evidence, alongside homology evidence from other curated genomes and *ab initio* predictions based on the underlying structure of genes, in order to be successfully annotated (Yandell and Ence, 2012; see **Table 1**). Even though whole-genome assembly projects were previously restricted to large research groups (e.g., Schnable et al., 2009; Tomato Genome Consortium, 2012), the sequencing cost per nucleotide is declining constantly in all the aforementioned technologies, making genome analyses accessible for a large part of the research community (Muir et al., 2016). The current bottleneck for small research groups is usually not the cost of sequencing itself, but rather the availability of computational resources capable of storing and analyzing huge amounts of data (Muir et al., 2016).

The main purpose of assembling a genome in a domestication study is to use it as a reference for high-quality population data to infer the selection, introgression and recombination processes, and to design posterior studies for experimental validation of candidate loci. Even though several population-level analyses based on reduced-representation genome sequencing can be performed in the absence of a reference genome (De

Wit et al., 2012; Mastretta-Yanes et al., 2015), the use of a reference genome alongside population data enables the correct identification of otherwise anonymous loci into specific genes or regions within the genome and it makes possible the identification and the proper handling of linkage between loci (Fitz-Gibbon et al., 2017). Also, it can help to discriminate between orthologous and paralogous loci, which is critical given the large size of many genomes and the frequent genome duplication processes experienced during the evolution of plant and animal lineages (Clark and Donoghue, 2018; Zadesenets and Rubtsov, 2018).

Thus, the availability of a reference genome is desired for genomic analyses concerning domestication. Luckily, domesticated taxa are usually economically relevant, drawing the attention of several research groups worldwide and in some cases helping to fund the projects. Therefore, reference genomes are usually available for domesticated species, since such data is also relevant for other research areas, such as crop improvement and breeding programs (Ellegren, 2014). However, it should be noted that using a single reference genome can lead to reference bias, where sequenced individuals that are more distantly related to the reference will tend to have fewer predicted variants due to mismatches while mapping the reads (Günther and Nettelblad, 2019).

Besides its use as a reference genome for population-level data, the analysis of several whole-genome assemblies between

TABLE 1 | List of key publications with other reviews that are focused on specific topics, as well as some notable examples of research articles using some of the methods described in this review with reliable results.

Topic	Citation	Usefulness/importance
Genome assembly and reference genomes	Sohn and Nam, 2016	In-depth review on genome assembly. Includes compelling explanations behind the genome assembly algorithms and an extensive list of genome assembly strategies.
	Yandell and Ence, 2012	In-depth review on eukaryotic genome annotation, a description of the available tools to predict genes and best practices when predicting genes.
Sequencing strategies	Meirmans, 2015	Classic review concerning common pitfalls that should be avoided in a population genomic study. A compulsory review for any newcomer to population genomics.
	Dorant et al., 2019	Empirical study that compares the efficiency of Pool-seq, RADseq and Rapture to detect weak signals of genetic structure in lobsters.
	Inbar et al., 2020	Empirical study that compares the efficiency of whole-genome sequencing, Pool-seq and RADseq for GWAS in ants.
Pan-genomics	Golicz et al., 2016a	In-depth review about pan-genomics in plant species, its advantages over the use of reference genomes, a guide on how to generate pan-genomes and the importance of studying structural variants. The article is dedicated to plants, but the rationale and methods can also be applied to other eukaryotes.
	Khan et al., 2020	Opinion article detailing the relevance of pan-genomes as a necessary next step from reference genomes. The authors also highlight the importance of including wild taxa into pan-genomics and propose the idea of genus-level super-pan-genomes.
	Gao et al., 2019	Landmark study of the tomato pan-genome. The authors sequenced 725 accessions from the domesticated tomato and its wild relatives. They found 4,873 additional genes, including several well-characterized genes that were absent from the reference-genome. They also evaluated the presence-absence variants between the wild and domesticated tomatoes, which were enriched in disease-resistance genes.
Demographic analyses	Linck and Battey, 2019	Research study focused on the effects of minor allele-frequency filters to detect genetic structure in populations. Gives a good explanation on the rationale behind the clustering-based methods to detect structure.
	Mather et al., 2020	Review dedicated to the theoretical background and technical requirements of PSMS and MSMC to infer changes in effective population sizes and coalescent times.
	Gerbault et al., 2014	Excellent review on how to use Bayesian approaches to test different demographic models of domestication.
	Frantz et al., 2015	Landmark study on pig domestication. The authors make use of Approximate Bayesian computation to compare domestication scenarios, they use clustering-based methods to detect genetic structure and used a graph-based method to infer the genetic relationship between pig and wild boar populations.
Selection scans	Vitti et al., 2013	Good review focused on explaining the rationale behind many of the bottom-up tests to detect selection and the genomic signals they are sensitive to.
	De Mita et al., 2013; Lotterhos and Whitlock, 2015	Classic simulation-based studies that compare different scenarios to evaluate the best sampling strategies and the most powerful methods to detect selection throughout the genome, according to the reproductive nature of the organism under study.
	Gibson, 2018	A primer dedicated to understanding the principles behind GWAS and its ability to detect polygenic effects on quantitative traits.
	Hufford et al., 2012	A landmark paper that illustrates how to perform genome scans to detect domestication-related loci in domesticated taxa, and the importance of these loci for crop improvement. The paper studies the domestication of maize, but a similar study design can be applied to domesticated animals.
Paleogenomics	Irving-Pease et al., 2019	Exhaustive book chapter dedicated to the study of ancient DNA to understand domestication.
	Allaby et al., 2019	Research study that casts into doubt the long-lasting idea that domestication processes lead to strong population bottlenecks by re-analyzing data based on ancient DNA samples.
	Daly et al., 2018	Remarkable study that sequenced and analyzed 83 mitochondrial genomes and 51 nuclear genomes from ancient goat samples. The authors found signals of ancient introgression events, as well as ancient selective signals related to several traits that are shared with modern goats.
Transcriptomics	Fang and Cui, 2011	General guideline on how to adequately design an RNA-seq experiment to avoid technical mistakes and generate meaningful results.
	Yang and Kim, 2015	General guideline on how to analyze RNA-seq data to assess differential expression.
	Hekman et al., 2015	In-depth review dedicated to study the domestication process through transcriptomics, including methodological strategies and challenges.
	Hradilová et al., 2017	An excellent study that combines transcriptomic data with metabolomic data and morphological data between domesticated and wild peas. The analysis of multi-omic data allowed them to get a better understanding behind seed dormancy and pod dehiscence in domesticated peas.
Epigenomics	Guerrero-Bosagna, 2012; Heard and Martienssen, 2014; Burggren, 2016	Contrasting views on the role of transgenerational epigenetic inheritance in evolution. The topic is still debated and should be viewed critically.
	Jensen, 2015	In-depth review on the rationale and advances of epigenetic studies to understand domestication. The manuscript is focused on animal behavior, but many of the ideas can also be applied to domesticated plants.

(Continued)

TABLE 1 | Continued

Topic	Citation	Usefulness/importance
Genome-editing tools	Janowitz Koch et al., 2016	A landmark paper showing the importance of epigenetic marks on dog domestication and its association with behavioral traits. The study doesn't just compare the methylation marks between wolves and dogs, but also assess the heritability of the methylation marks and proposes a formal test to detect selection on epialleles.
	Boettcher and McManus, 2015).	Review on novel genome-editing techniques and RNA interference. Useful to compare and choose the best tool to validate candidate loci.
	Shan et al., 2020 Soyk et al., 2017	A general guide on how to develop a CRISPR/Cas9 system on a non-model plant species. Landmark paper that uses genome-editing to validate two candidate genes related to fruit size and reduced fruit dropping in tomato. The authors also detect the emergence of undesirable traits in domesticated tomatoes due to an epistatic effect between both domesticated loci and introduce wild alleles to generate new tomato phenotypes with reduced degrees of the undesirable traits.
Perspectives	Piperno, 2017	Review centered on the potential application of an extended synthesis framework to understand domestication. Centered around the concepts of niche construction, transgenerational epigenetic inheritance and developmental plasticity.

domesticated and wild taxa will help us reveal structural differences between the genome of a domesticated taxon and its closest wild relatives, such as duplications, chromosome rearrangements or presence/absence of entire genes and genomic regions (Yang et al., 2012; Wang W. et al., 2014; Xie et al., 2019). Since selection and bottlenecks during domestication often leads to the fixation of mutations that involve a loss of function (Renaut and Rieseberg, 2015; Moyers et al., 2018), comparative analyses using genome assemblies of wild ancestors may also reveal these changes in genes that could not be properly predicted within the domesticated genome (Moyers et al., 2018). In this sense, further efforts should be made to assemble high-quality genomes of wild relatives alongside the domesticated taxon of interest (Brozynska et al., 2016; Xie et al., 2019).

STRATEGIES TO GATHER ADEQUATE POPULATION GENOMICS DATA

Genome assemblies alone give us a limited view on domestication, unless several genomes of wild relatives (if known and available) and domesticated individuals are sequenced, because evolution is a population-level process, and in consequence population data is necessary to address most of the evolutionary questions in domestication (Wang G.-D. et al., 2014; Guerra García and Piñero, 2017). Population genomics examines the genetic variation within and between populations that is scattered across the entire genome to assess the demographic history, phylogenetic relations and selective pressures of a species (Jorde, 2001). Several types of genomic data can be evaluated at the population level, including single nucleotide polymorphisms (SNPs), indels and copy number variations; but SNPs are the most commonly analyzed of the three (Seal et al., 2014).

All population-level sequencing techniques share common pitfalls that should be known and avoided before investing any money on sequencing. Population sampling should be planned carefully, as the sampling scheme has a stronger impact over sequencing to obtain reliable results in any analysis (Meirns, 2015). Also, different populations should be mixed, rather than

being sequenced on separate libraries or sequencing lanes, as failing to do so will generate sequencing biases that can be confused with biological patterns (Meirns, 2015; see Table 1).

Once adequate genomic population data is gathered, we need to analyze the demographic processes that shaped the genetic variation and the population structure of contemporary populations during the domestication process. This data is necessary to perform tests to detect natural and artificial selection, which are required to understand the genetic base of domestication syndromes (Ross-Ibarra et al., 2007). There are several approaches to obtain population data at a genomic scale, which differ in the fraction of the genome that is sequenced, therefore determining the sequencing cost of each sample (Schreiber et al., 2018).

Whole-Genome Sequencing of Populations

After assembling a reference genome, one of the next possible strategies to understand domestication is to sequence the complete genome of several individuals. This approach requires the alignment of the sequencing reads back to a reference genome, in order to infer the variable sites between individuals and know the genetic elements (e.g., genes, upstream regulators, repetitive elements, non-coding RNAs) associated to those sites. The main benefit of this approach is its potential to retrieve all the variant sites within an individual's genome that are structurally represented in the reference genome. Whole-genome sequencing can be used in almost any population-level test of interest (Schreiber et al., 2018). Common practices recommend a sequencing depth around 30× per individual, but empirical studies in pigs suggests that even 10x is enough to cover up to 99% of a genome with accurate detection of variant sites (Jiang L.G. et al., 2019). The main drawback of this approach is the sequencing cost of each sample, which is significantly higher compared with other approaches, especially for organisms with large genomes such as polyploid crops or mammals (Schreiber et al., 2018). This can lead researchers to evaluate a trade-off between sequencing depth and number of sampled individuals to optimize their resources. Simulation studies suggest that sequencing more individuals is more convenient to obtain

reliable results, even at the expense of lower sequencing depths per individual (Fumagalli, 2013).

Alternatives to Whole-Genome Sequencing

Other approaches aim to reduce the sequencing cost per samples by pooling the DNA of several individuals into a single sequencing library (Futschik and Schlötterer, 2010) or by reducing the portion of the genome that is sequenced (often named as reduced-representation sequencing), either by sequencing arbitrary defined segments scattered across the genome, by targeting the desired portions of the genome or by sequencing the transcriptionally active portions of the genome (Schreiber et al., 2018). These techniques are especially helpful for organisms with very large genomes, and some of these methods can even be used in the absence of a reference genome (Mastretta-Yanes et al., 2015; Schreiber et al., 2018). Furthermore, the reduced representation of the genome means that those fewer regions that are targeted can have a high sequencing depth, leading to higher accuracy of the observed genetic variation and better heterozygosity estimations (Schreiber et al., 2018). Additionally, the reduced sequencing cost per sample allows for a large number of sequenced individuals and populations that, with a proper sampling strategy, can lead to robust results (De Mita et al., 2013; Lotterhos and Whitlock, 2015). Due to the fragmented nature of these sequencing techniques, reduced representation data alone may be insufficient to pinpoint all or even the most important possible causal genetic variants associated to the domestication syndromes (Lowry et al., 2017), but they are still useful to infer basic genetic statistics, infer demographic properties and past demographic scenarios, detect some signatures of selective sweeps across the genome and even perform GWAS for domestication traits of interest (Andrews et al., 2016; Schreiber et al., 2018).

Pool Sequencing

Pool sequencing (Pool-seq) is a promising alternative to whole-genome sequencing with a much lower cost (Futschik and Schlötterer, 2010). As the name suggests, Pool-seq consists of sequencing a large pool of individuals for a given population into a single high-throughput sequencing library, instead of sequencing each individual separately, allowing an accurate estimation of allele frequencies and other parameters of population genetics at the expense of losing individual-level information (Futschik and Schlötterer, 2010). This method requires to map the reads against a reference genome of the same species in order to work (Schlötterer et al., 2014). It is intended for sequencing large pools of individuals (>40 individuals per population is recommended, but >100 is optimal), otherwise the allele frequencies will not be estimated accurately (Schlötterer et al., 2014). The relative amount of pooled DNA of each individual in a Pool-seq study should be similar in order to avoid overrepresentation of individual alleles, a task that is often challenging (Schlötterer et al., 2014).

Pool-seq has several limitations that should be considered based on the objectives of the research project. It is difficult to

discard a low-frequency allele from a sequencing error, but this problem is potentially fixed by either establishing a minor allele frequency threshold for SNP calling or by using pool replicates (Schlötterer et al., 2014). One important limitation is the inability of Pool-seq data to estimate linkage disequilibrium and haplotype phasing, which is particularly important to evaluate the non-independence of genetic signals in demographic studies and selective scans (Schlötterer et al., 2014). Finally, assessing genetic structure can be difficult and sometimes misleading when using Pool-seq, due to potential biases in individual allele representations within the pool (Dorant et al., 2019). This makes Pool-seq an adequate method for GWAS, selective sweeps and some methods based on allele frequencies when resources are limited (Luu et al., 2017; Inbar et al., 2020), but the loss of individual-level information makes many of the demographic inferences difficult, as populations need to be predefined before sequencing (Dorant et al., 2019), and the bioinformatic tools that handle Pool-seq data are scarce.

Exome Capture and Sequencing

Exome sequencing is another lower-cost alternative to whole-genome sequencing which targets the protein-coding regions of the genome (Warr et al., 2015; Kaur and Gaikwad, 2017). Protein-coding genes represent a small fraction of eukaryotic genomes, which is particularly useful for most population genomic studies, since it represents mostly functional elements within genomes (Kaur and Gaikwad, 2017). This technique is usually performed using hybridization probes, which requires previous knowledge of the genome content as well as a *priori* selection of regions of interest in order to design probes (Kaur and Gaikwad, 2017). Fortunately, hybridization probes are already available for several domesticated plants and animals (Warr et al., 2015; Kaur and Gaikwad, 2017).

Despite its advantages, exome sequencing can generate an uneven sequencing depth in certain genomic positions, unlike whole-genome sequencing that shows a uniform distribution of reads throughout the genome (Lelieveld et al., 2015). Another important limitation of exome sequencing is its bias towards the protein-coding portion of the genome, since increasing evidence shows that many of the genetic changes that have been directly associated to domestication traits are located within *cis*-regulatory elements, noncoding RNAs and other *trans*-regulatory elements, rather than within the open reading frame of the genes (Swinnen et al., 2016). Despite its limitations, demographic history and selective sweeps can still be detected using this sequencing method (Pankin et al., 2018).

RNA Sequencing of Populations

Transcriptome sequencing (also known as RNA-seq) is another useful approach to obtain population-level data from the transcriptionally active elements within genomes (De Wit et al., 2012). RNA-seq can be mapped against a reference genome to detect genetic variants and determine the genomic regions of interest, but it can also be analyzed in the absence of a reference genome (De Wit et al., 2012), since transcriptomes can be assembled *de novo* (Haas et al., 2013) and the functional

annotation of the assembled transcripts is relatively easy (Bryant et al., 2017).

However, transcription profiles are dependent on the sequenced tissues and organs, the development stage of the organism, and the influence of external stimuli, capturing just the transcripts that are active at the moment of RNA extraction (Hekman et al., 2015). This complexity can generate important biases in the relative abundance of certain transcripts over others and overlook potential adaptive genes whose expression are context dependent (Hekman et al., 2015; Kaur and Gaikwad, 2017). Nonetheless, RNA-seq is still a good option for species with large genomes that are hard to assemble (De Wit et al., 2012). Similarly to exome-sequencing, RNA-seq data can be used to evaluate demographic history and selective sweeps, but the selective signals are restricted to the transcriptionally active part of the genome, and cannot be used to evaluate structural variants (Schreiber et al., 2018).

Restriction Site-Associated DNA Sequencing

Restriction site-associated DNA sequencing (RAD-seq), which may also be referred to as genotyping by sequencing (GBS), has been one of the most popular options for cost-affordable population genomics in the last decade (Davey and Blaxter, 2010). The technique consists in using restriction enzymes to digest the DNA and sequence the regions adjacent to the restriction sites that are scattered across the genome (Davey and Blaxter, 2010). It can also be combined with sequence capture techniques to target specific loci of interest (Ali et al., 2016). RADseq data can either be mapped against a reference genome or it can be assembled *de novo* (Catchen et al., 2013; Mastretta-Yanes et al., 2015), making it a versatile technique for species with scant genomic resources.

However, empirical studies show that using certain *de novo* approaches for RAD-seq data can lead to fewer predicted SNPs due to errors in the definition of loci and treatment of sequencing errors (Shafer et al., 2017), all which may subsequently alter downstream analyses, especially those based on the distribution of allele frequencies within the genome of a population, also known as the site frequency spectrum (SFS) (Shafer et al., 2017). For this reason, a reference-based approach is highly recommended as long as the reference genome is closely related to the population dataset (Shafer et al., 2017). Furthermore, RADseq data could involve errors when a polymorphism resides within a restriction site, which prevents the enzyme to cut in individuals carrying such polymorphism, leading to failures in sequencing that region in homozygous individuals (null alleles) and makes heterozygous individuals to look like homozygotes (allele dropout) (Andrews et al., 2016). Finally, the capacity of RADseq libraries to adequately perform selective scans has been casted into serious question (Lowry et al., 2017). Its potential capacity to detect selective sweeps is dependent on the genome size, the density of variants detected for a given genomic region and specially the length of the extent of linkage disequilibrium in the genome (Lowry et al., 2017). Thus, when a species genome has short regions in linkage disequilibrium (due to high recombination rates) and the SNP density is low (particularly in

large genomes), odds are that the selective scans will likely miss a significant portion of selective sweeps associated to domestication (Lowry et al., 2017).

PAN-GENOME ANALYSES IN DOMESTICATED AND WILD TAXA

An increasing number of studies are revealing that structural variants (copy-number variation, presence/absence of genomic regions, inversions, transversions, translocations) are common within plant and animal populations (Khan et al., 2020). Thus, the use of a single reference genome hampers our ability to study the full repertoire of genetic variation within a species (Golicz et al., 2016a; Zhao et al., 2018). Structural variants such copy-number variation can contain functional genomic elements that are usually under relaxed selective pressures and can serve as the basis of adaptation given specific environments and selective regimes (Lye and Purugganan, 2019). Coincidentally, copy-number variation and other structural variants play an important role in the emergence of domestication traits, as well as diversification traits in landrace varieties (Lye and Purugganan, 2019). Some studies estimate that at least one third of the known domestication loci are structural variants, and up to one in seven genes can be hemizygous (*i.e.*, with one copy) in grapevine individuals (Zhou Y. et al., 2019). Despite its importance, structural variants cannot be properly analyzed using any of the aforementioned techniques. This led the research community to adopt the concept of the pan-genome, an idea that first appeared in microbiology (Tettelin et al., 2005), into the study of plant and animal genomes (Golicz et al., 2016a).

The concept of pan-genome rests on the idea that the genomes of individuals within a population or species share a core set of genes that unifies them (*i.e.*, the core genome), but also contain a fraction of genes that are absent from one or more individuals (*i.e.*, the accessory or dispensable genome), which altogether give rise to the pan-genome of such population or species (Tettelin et al., 2005).

There are three main methods to generate a pan-genome: the alignment and comparison of multiple *de novo* genome assemblies, the iterative assembly of several genomes from an initial reference or the use of *de Bruijn* graph assemblers to jointly assemble several genomes (Golicz et al., 2016a; see **Table 1**). Since domestication reduces the genetic diversity of a taxon, often eliminating portions of the dispensable genome that contain genes involved in local adaptation, the use of wild relatives is crucial to generate a representative pan-genome for a species (Khan et al., 2020). Once a pan-genome is generated, it can be used alongside whole-genome sequencing data to analyze the structural variants between and within populations, revealing novel loci involved in the development of domestication-related traits that would have stayed hidden when using a single reference genome (Li et al., 2014; Zhao et al., 2018). Besides, the use of a pan-genome alleviates the inherent reference biases of a single reference genome (Günther and Nettelblad, 2019).

Pan-genome studies have revealed additional selective sweeps and structural variants associated to the domestication process,

which were not identified using sequencing data with a single reference genome (Li et al., 2014; Zhao et al., 2018). Pan-genomes are already available for several species (Figure 2) such as maize (Brohammer et al., 2018), wheat (Montenegro et al., 2017), *Brassica oleracea* (Golicz et al., 2016b) or *Brassica napus* (Hurgobin et al., 2018); and pan-genome analyses to study domestication have already been performed in soybean (Li et al., 2014), rice (Zhao et al., 2018), sunflower (Hübner et al., 2019) and tomato (Gao et al., 2019). While current eukaryote pan-genome analyses are focused on plant species (Golicz et al., 2016a, see Table 1) and goats (Li et al., 2019), other livestock researchers may soon venture into this field. As sequencing technologies become cheaper, multiple pan-genomes from different species of the same genus should eventually be combined to create a super-pan-genome that represents the entire genetic content available in a genus with one or more domesticated taxa, as it would include the diversity of all their wild relatives (Khan et al., 2020).

POPULATION GENETICS AND DEMOGRAPHIC ANALYSES OF THE DOMESTICATION PROCESS

Demography and population size changes during the domestication process is tightly related to unraveling some of the most fundamental questions of the domestication process. These analyses can help answer questions such as possible centers of origin and diversification, patterns of migration and expansion throughout these centers, gene flow between domesticated and wild taxa, number of domestication events, the extent of genetic erosion in the domesticated taxon, levels of global genetic differentiation between wild and domesticated taxa, the patterns of adaptive and neutral introgression among them, and in some cases even the number of generations that have elapsed since domestication and other processes such as differentiation and local adaptation of domesticated taxa (Meyer and Purugganan, 2013; Guerra García and Piñero, 2017).

Genetic Diversity in Populations

A first necessary step for the SNP data is to extract and compare the summary statistics of population genetics within and between populations (Andrews et al., 2016). This information describes the genetic diversity in populations, including the estimate of allele frequencies (usually denoted as p or the frequency of the most abundant allele), observed heterozygosity (H_O), expected heterozygosity (H_E), nucleotide diversity (π), number of segregating sites (S) and number of private alleles (*i.e.*, alleles only found in one population). These summary statistics can reveal the level of genetic erosion in domesticated plants and animals when compared to the ancestral wild population, which is expected due to severe bottlenecks, selective sweeps and inbreeding (Groeneveld et al., 2010; Gepts, 2014). One should be aware that reference bias can influence the relative genetic variation observed between the wild and domesticated populations, which could be alleviated using more than one reference or using a pan-genome (Günther and Nettelblad, 2019).

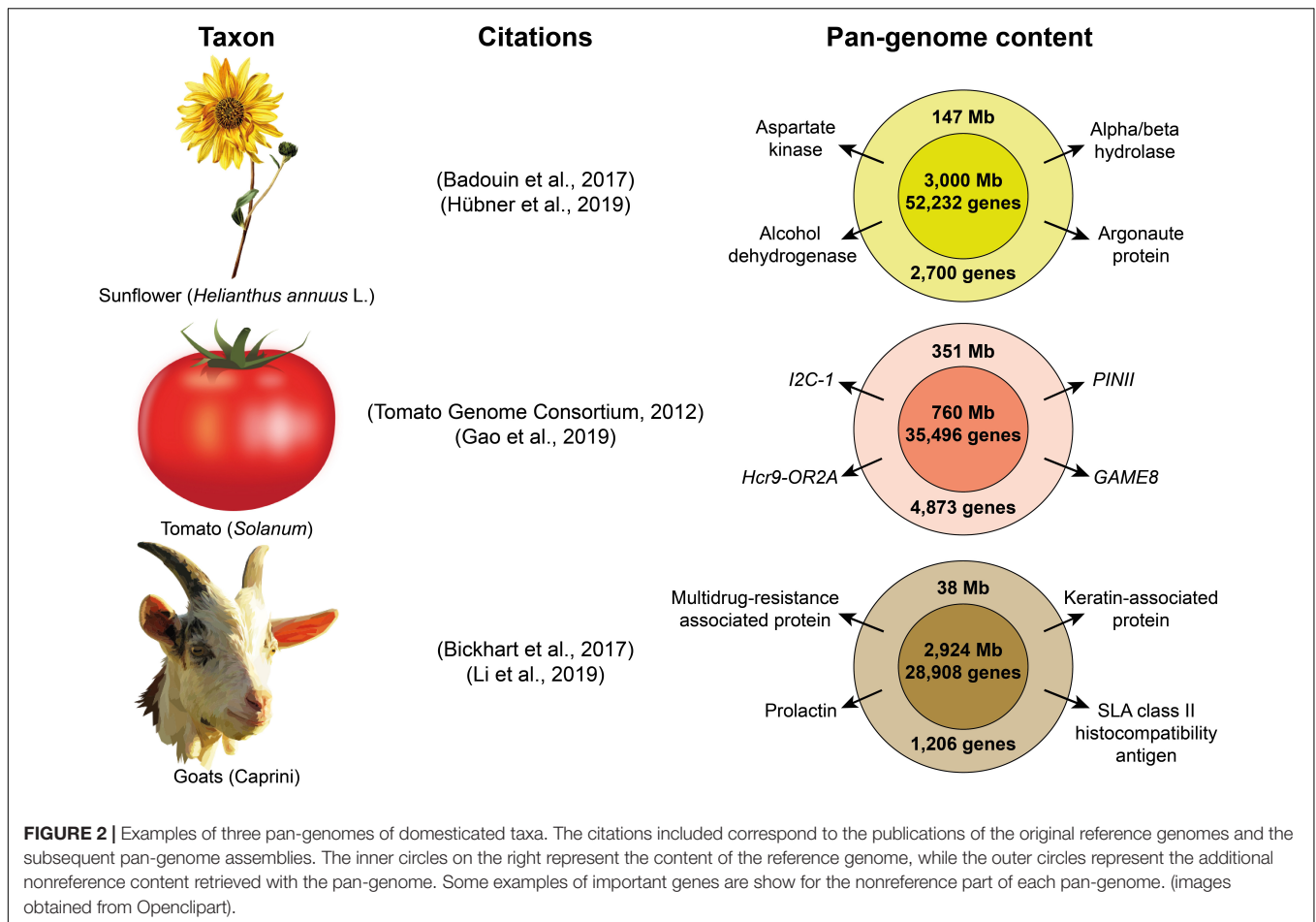
Population Structure

It is also important to describe the population structure (*i.e.*, the genetic differentiation among populations) of domesticated taxa and of their wild relatives, as it can reveal the influence of historical events that shaped the genetic diversity of the organisms (Linck and Battey, 2019). The level of population structure between wild and domesticated taxa can be determined by several factors, such as the number of generations since domestication started, the intensity of the selective pressures imposed to the domesticated taxon, the intensity of the bottlenecks suffered through the domestication process, and the frequency of gene flow between the domesticated taxon and its wild relative (Meyer and Purugganan, 2013).

The F -statistics are classic estimates of population genetics that are based on the heterozygosity values within and among populations, which can reveal patterns of inbreeding, gene flow and differentiation between and within populations (Andrews et al., 2016). Of these, the F_{ST} statistic is of particular interest, since it can be used to detect population structure between wild and domesticated populations, or between different domesticated varieties (Andrews et al., 2016). These estimates are relatively simple to calculate, but they require *a priori* assignment of individuals to discrete populations, which may be wrongly assigned, may not reflect natural populations or may simply be unknown (Linck and Battey, 2019).

Methods based on population clustering have become popular for describing genetic structure, as they do not require *a priori* population assignment. These clustering methods can be classified into parametric and non-parametric methods (Linck and Battey, 2019). Parametric methods, also known as model-based methods, assign individuals into a predefined number of K populations based on their genotypes and the allele frequency of each locus (Pritchard et al., 2000). Several parametric methods have been described that successfully analyze genomic datasets to infer population structure (*e.g.*, Tang et al., 2006; Alexander et al., 2009; Raj et al., 2014), but one has to be careful when using them, as they assume linkage equilibrium and Hardy-Weinberg equilibrium in the dataset (Linck and Battey, 2019), so SNPs should be filtered accordingly before these methods can be confidently used (Wigginton et al., 2005; Mathew et al., 2018). Furthermore, parametric methods have been found to be susceptible to changes in the SFS generated by minor allele frequency thresholds that are commonly used to filter population genomics data because low-frequency polymorphisms are expected to contain information about recent events, which adds uncertainty to the assignment of individuals in populations that reflect ancient demographic events (Linck and Battey, 2019).

Non-parametric methods include principal component analyses, discriminant analyses of principal components and K -means clustering. These methods define populations and genetic structure by transforming the genetic data into uncorrelated variables – named eigenvectors or principal components – to identify groups within the dataset (Patterson et al., 2006; Jombart et al., 2010; Linck and Battey, 2019). Non-parametric methods were designed to work with large amounts



of genomic data (Patterson et al., 2006; Jombart et al., 2010) and they are more robust to changes in the SFS than the parametric methods, so it is recommended to run both types of methods and compare their results before making further inferences (Linck and Battey, 2019).

Inferences in Changes of Population Sizes Throughout Time

An important aspect of the demographic history of domesticated taxa is the analysis of the change in the effective population size (N_e) in the populations throughout time (Chen J. et al., 2018). The concept of N_e reflects the estimated populations size in a Wright-Fisher model given an observed genetic variation, so these estimations hardly reflect the census population size of real populations (Charlesworth, 2009), and can also be affected by reference biases and allele dropouts. Changes in N_e can reveal or at least hint on the demographic history of taxa throughout the domestication process, such as expansions or bottlenecks. These changes can help to understand other evolutionary aspects of domestication concerning natural and artificial selection, such as the efficiency of selection and the accumulation of deleterious mutations in domesticated taxa (Chen J. et al., 2018; Allaby et al., 2019).

The domestication process is expected to include a bottleneck as a consequence of subsampling the genetic diversity in the wild ancestor, followed by a population expansion as domesticated taxa diversify (Meyer and Purugganan, 2013), although this idea has been recently challenged by paleogenomic studies (Allaby et al., 2019). Many methods exist to explore the changes in N_e throughout time, whose approach sometimes depends on the type of data available. It should be noted that all the methods to infer historical changes in N_e are susceptible to predicting false bottlenecks when populations are structured, so as indicated above, genetic structure should be evaluated and properly accounted for (Nielsen and Beaumont, 2009).

Studies with few individuals and high sequencing depth may use the Pairwise Sequentially Markovian Coalescent model (PSMC; Li and Durbin, 2011) or the Multiple Sequential Markovian Coalescent model (MSMC; Schiffels and Durbin, 2014) to analyze the demographic history of domesticated and wild taxa. The PSMC and MSMC models can infer changes in N_e throughout time (bottlenecks and expansions) by calculating the distribution of the time of coalescence between all the heterozygous loci in complete diploid genomes (Li and Durbin, 2011; Schiffels and Durbin, 2014). These models can also calculate the time of coalescence (*i.e.*, separation, and in some cases the domestication time) between two genomes given a

specified mutation rate, recombination rate and generation time (Li and Durbin, 2011).

However, the genomes used in PSMC or MSMC must be of very good quality, having an average sequencing depth of the very least 18x, at least 10 reads per site, and less than 25% of missing data (Nadachowska-Brzyska et al., 2016). Besides, PSMC has several limitations compared to other estimators of N_e and is particularly susceptible to predicting false bottlenecks when populations are structured (Mazet et al., 2015). Nevertheless, this can be properly handled by comparing models of instantaneous N_e size change against models of classical symmetric islands using a maximum-likelihood approach (Mazet et al., 2015).

Multiple Sequential Markovian Coalescent can infer more recent changes in N_e compared to PSMC (Schiffels and Durbin, 2014), so it may be convenient to explore recent demographic expansions in diversified domesticated taxa (Allaby et al., 2019). For example, MSMC was used to infer population bottlenecks in East Asian and Western Eurasian dogs, as well as divergence times between wolves and dogs around 60,000–20,000 years ago (Frantz et al., 2016), while PSMC was used to determine a severe bottleneck in African rice around 15,000–13,000 years ago (Meyer et al., 2016).

Other methods rely on population data at a genomic scale from many (sometimes hundreds) individuals (as obtained from exome sequencing or RAD-seq), namely the extended Bayesian skyline plots (Heled and Drummond, 2008; Trucchi et al., 2014) and the stairway plots (Liu and Fu, 2015). Since N_e is a crucial concept in coalescent theory, extended Bayesian skyline plots and stairway plots rely on the SFS calculated from the population data to estimate N_e (Heled and Drummond, 2008; Liu and Fu, 2015). The inferences made from these two methods are comparable to those obtained from PSMC and MSMC, although they rely on different kinds of datasets (Liu and Fu, 2015). Furthermore, stairway plots are more efficient in inferring recent demographic history, whereas PSMC is more reliable for ancient demographic events (Liu and Fu, 2015).

Estimating Gene Flow and Introgression Between Populations

Ancient gene flow and local ancestry (*i.e.*, the genetic ancestry of an individual for an specific chromosomal position; Thornton and Bermejo, 2014) are also important aspects of plant and animal domestication that need to be addressed, since they can describe the genetic contribution of different ancestral populations in the genomic architecture of extant populations, such as wild and domesticated taxa (Price et al., 2009; Pickrell and Pritchard, 2012).

One approach to assess ancient gene flow are graph-based methods that incorporate the possibility of ancient gene flow between distantly related populations (Pickrell and Pritchard, 2012). This type of methods represents the relationships between populations as a bifurcating tree, where internal nodes can also be interconnected forming a graph that represents ancient gene flow that contributed to modern genetic variation (Pickrell and Pritchard, 2012). For example, graph-based analyses have

revealed constant gene flow between sympatric populations of domesticated and wild pearl millet (Burgarella et al., 2018), constant gene flow between domesticated and wild pigs (Frantz et al., 2015) but lack of hybridization events between wild and domesticated populations of goats and sheep (Alberto et al., 2018).

Another popular test to infer ancient admixture is the ABBA-BABA test, also known as the D -statistic, which evaluates the allelic patterns of three taxa and compares them to an outgroup to identify genomic regions with an excess of shared derived variants that are not concordant to the species tree (*i.e.*, ABBA-BABA patterns), which suggest introgression events (Durand et al., 2011). The f_d^A test, which is derived from the D -statistic, can help discriminate between introgression events and nonrandom mating in ancestral structured populations (Martin et al., 2015). The D -statistic is sensitive to both introgression and incomplete lineage sorting, so both signals can be separated by testing deviations in the symmetry of branch lengths between the gene trees and the species tree (Edelman et al., 2019). By the same logic, the D_3 test can also infer introgression events by analyzing the symmetry in branch lengths, without the need for an outgroup (Hahn and Hibbins, 2019). The D -statistic has been used to infer several introgression events between species of the *Bos* genus during domestication (Wu et al., 2018).

On the other hand, local ancestry methods can reveal which chromosomal segments in the genome were inherited from different ancestral source populations (Price et al., 2009). These methods use the data obtained from linkage disequilibrium between loci to assign ancestry in each portion of the genome in comparison to reference populations that depict ancestral source populations, requiring an *a priori* assignation of unadmixed reference populations in order to assign local ancestry to the populations of interest (Price et al., 2009). The analysis reveals chromosomal blocks that can be assigned to either a wild or a domesticated ancestry in hybrid populations, which may reveal historical processes of introgression and local adaptation in modern domesticated populations, as well as potential targets for selective breeding (Janzen et al., 2019).

Many methods exist that can infer local ancestry using genome-wide population data, and all of them require a high-quality reference genome (preferably assembled at a chromosome-level) in order to detect the ancestry of chromosomal segments (*e.g.*, Price et al., 2009; Baran et al., 2012; Maples et al., 2013; Dias-Alves et al., 2018). For example, a local ancestry analysis of East Asian domestic cattle revealed introgressed blocks inherited from ancient banteng and yak populations that contained genes enriched in sensory perception of smell, transmembrane transport and antigen processing (Chen N. et al., 2018).

Using Demographic Simulations to Infer Domestication Scenarios

The previous descriptive tools can help us explore possible evolutionary and demographic scenarios in the absence of *a priori* hypotheses (Liu and Fu, 2015). However, for domesticated taxa we usually have additional classic botanical, zoological,

morphological, paleoclimatic, archeological, ethnobiological and biogeographical data that may suggest some likely scenarios (Gerbault et al., 2014). Thus, demographic modeling can be used to test explicit demographic scenarios by comparing simulations of SFS in such scenarios to the observed data (Gerbault et al., 2014; Liu and Fu, 2015). There are many methods available for demographic modeling, which can be more suitable depending on the type of scenarios that need to be tested (Anderson et al., 2005; Gutenkunst et al., 2009; Excoffier and Foll, 2011; Cornuet et al., 2014). All these methods rely on some basic tenets of coalescent theory (Liu and Fu, 2015), so they are also susceptible to possible biases in the observed genetic variation in the populations.

For example, the approximate Bayesian computation (ABC) method compares the summary statistics of several simulated scenarios against the observed data to accept or reject certain demographic hypotheses (Cornuet et al., 2014; Gerbault et al., 2014). This method can help us determine certain parameters of our models and can be used with genome-wide datasets (Cornuet et al., 2014).

Other methods based on diffusion approximation can help us infer the demographic history of multiple populations and their interaction through migration and admixture using biallelic SNP data (Gutenkunst et al., 2009). Demographic modeling has helped test the number of domestication events as well as intercontinental migratory events in cattle (Pitt et al., 2019). Coalescent simulations have supported a common origin for all the domesticated varieties of pearl millet (Burgarella et al., 2018), while the ABC method has revealed that the most likely scenario in the domestication of the scarlet runner bean consists of a single domestication event around 21,000 years ago with a mild bottleneck effect (Guerra-García et al., 2017).

IDENTIFYING GENES UNDER SELECTION DURING DOMESTICATION

Demographic processes are important to understand the general history that led to the domestication of plant and animal taxa, but many studies are specially interested in finding the selected genes that explain the phenotypic differences between domesticated taxa and their wild counterparts (Wang G.-D. et al., 2014; Kantar et al., 2017). Indeed, the detection of these genes under selection during domestication is critical to understand the genetic basis of domestication syndromes, especially for detecting genetic variation relevant for future improvement and selective breeding (Hufford et al., 2012).

When a genetic variant increases its frequency due to positive selection (*i.e.*, selection favoring the fixation of a new allele), the adjacent alleles (*i.e.*, physically connected in the same chromosomal region) also increase their frequency in a process known as hitchhiking (Smith and Haigh, 2007). Once the genetic variant under selection reaches a high frequency or fixation, the hitchhiking effect reduces or even eliminates the genetic variation around the selected locus, producing what is known as a selective sweep (Vitti et al., 2013; Pavlidis and Alachiotis, 2017). The size and intensity of a selective sweep

depends on the rate of recombination in the genome, and on the intensity of the selective pressure (Smith and Haigh, 2007), which may be weaker in conscious selection compared to some cases of natural selection (Fugère and Hendry, 2018; Yang et al., 2019). Luckily, the signals of a selective sweep can be detected when the selection event occurred “recently” in an evolutionary timescale, as it is the case for domestication (Vitti et al., 2013).

Different bottom-up methods using population genomics data have been developed to detect the regions in the genome that were selected for during domestication, which we will refer to as candidate loci. We can mention methods for detecting regions with higher population differentiation compared to the rest of the genome, methods for detecting local changes in the SFS throughout the genome, and methods that detect extended regions with strong linkage disequilibrium compared to other haplotypes in the genome (see **Supplementary Table S1** for a summary of methods to detect selective sweeps). Alternatively, a GWAS can be performed to detect the association of a genetic variant to a specific phenotype of interest (Wang G.-D. et al., 2014).

F_{ST} Outlier Tests to Detect Candidate Genes

Besides the standard use of F_{ST} to detect global population structure, the F_{ST} statistic can also be used to detect signals of selective sweeps between populations, namely between wild and domesticated taxa (Gepts, 2014). While a global F_{ST} statistic (involving all the analyzed loci or SNPs) can reveal the overall genetic structure between populations, a local F_{ST} statistic calculated for each locus or SNP along the genome can evaluate whether particular regions of the genome are more differentiated from what is expected due to demographic processes, which can be interpreted as signals of a selective sweep (Nei and Maruyama, 1975). Many different methods exist that are based on the F_{ST} statistic, which are collectively known as F_{ST} outlier tests (Foll and Gaggiotti, 2008; Excoffier et al., 2009; Bonhomme et al., 2010; de Villemereuil and Gaggiotti, 2015; Lotterhos and Whitlock, 2015), that differ mainly on the underlying model used to calculate the null distribution of the F_{ST} values, and thus its ability to detect outliers (**Supplementary Table S1**).

F_{ST} outlier tests are able to detect selective pressures following a bottom-up approach, but their efficiency is determined by a multitude of factors that should be carefully accounted for before using them, such as the sampling scheme used to obtain the population data, the total size of the dataset (*i.e.*, number of populations, of individual per population and of SNPs analyzed), the intensity of the selective pressure, the selfing or allogamous nature of its sexual reproduction, and the migration patterns and genetic structure among populations (De Mita et al., 2013; Lotterhos and Whitlock, 2014, 2015).

Some successful examples in the use of F_{ST} outlier tests include the detection of domestication candidate genes in apple involved in fruit development, size, acidity and sugar metabolism (Khan et al., 2014), the finding of candidate domestication genes involved in metabolism and oil biosynthesis in sunflower

(Baute et al., 2015), the description of candidate diversification genes between pig breeds associated to the shape of the skull (Wilkinson et al., 2013), and the identification of candidate loci between wild and domesticated salmon strains involved in body weight, condition factor, male maturation and a brain related protein (Vasemägi et al., 2012).

Site Frequency Spectrum Based Tests to Detect Selective Sweeps

Selective sweeps alter the SFS that would be expected under neutral evolution processes because of the reduction in the genetic diversity around the loci under selection (Vitti et al., 2013). The genomic region under selection skews the SFS into an excess of high frequency derived alleles when the selective sweep was recent, since the alleles that were linked to the favored selected locus also reach high frequencies (Fay and Wu, 2000). However, after all the high-frequency alleles reached fixation, the genomic region under the selective sweep will have little to no variation, while mutations will slowly generate new allelic variants, skewing the SFS into an excess of low frequency variants (Zeng et al., 2006). Several tests have been developed to detect skews in the SFS, each of them capable of detecting changes in different parts of the SFS (**Supplementary Table S1**), making them complementary to one another (Zeng et al., 2006; Vitti et al., 2013).

Even though SFS based tests are powerful tools to detect selection, it is important to remember that the SFS at the global genomic scale is also altered by demographic events such as bottlenecks that produces an excess of low frequency variants, and expansions that generates an excess of intermediate frequency variants (Vitti et al., 2013). Thus, it is mandatory to have a previous prediction of the demographic history of the populations in order to properly adjust the null hypothesis in each test (Ross-Ibarra et al., 2007).

The well-known summary statistic called Tajima's D is sensitive to changes in low-frequency variants, making it particularly useful to detect selective sweeps before and after the selected locus reaches fixation, although low-frequency variants can also be observed in loci under purifying selection (Tajima, 1989; Zeng et al., 2006). Tajima's D is also sensitive to intermediate-frequency alleles, making it useful to detect balancing selection (Tajima, 1989) or even some forms of soft selective sweeps generated by standing genetic variation (Prezeworski et al., 2005).

Conversely, Fay and Wu's H is sensitive to changes in high-frequency variants, which are only altered by positive selection, making it very useful when used alongside Tajima's D (Fay and Wu, 2000). Unlike Tajima's D , Fay and Wu's H needs an outgroup species in order to differentiate ancestral alleles from derived alleles and thereby to know whether the derived alleles are at high or low frequencies (Fay and Wu, 2000).

Zeng et al. (2006)'s E is sensitive to both low and high frequency variants, making it particularly powerful to detect selective sweeps before or after the selected locus reached fixation, also needing an outgroup in order to differentiate derived alleles from ancestral alleles).

There are some tools available to implement SFS based tests using genome-wide data, that can perform all the above tests (*i.e.*, Korneliussen et al., 2013, 2014; Rozas et al., 2017). For example, Tajima's D test was used alongside other methods to detect selective sweeps associated to the domestication of yaks (Qiu et al., 2015), Zeng's E test helped discover 125 selective sweeps associated to the domestication of horses (Librado et al., 2016), and the complementary implementation of Tajima's D , Fay and Wu's H and Zeng's E revealed several candidate genes that share similar functions between peach and almond (Velasco et al., 2016).

The reduction of diversity (ROD) test is another popular SFS-based method to detect selective sweeps that has been particularly useful for the study of domestication (**Supplementary Table S1**). ROD compares local π values of domesticated taxa against the local π values of its wild relatives, using sliding windows alongside the genome (Guo et al., 2012; Huang et al., 2012; Qi et al., 2013; Schmutz et al., 2014). The ROD method has been used to successfully detect candidate domestication genes in rice (Huang et al., 2012), watermelon (Guo et al., 2012), cucumber (Qi et al., 2013), common bean (Schmutz et al., 2014), and chickpea (Varshney et al., 2019), to name a few.

Linkage Disequilibrium (LD) Based Methods to Detect Selection

Given that selective sweeps remove the variation in regions adjacent to the locus under selection, they can form haplotype blocks that extend in strong LD compared to other haplotypes in the same locus because they reached a medium-to-high frequency in the population swift enough so they are not yet disrupted by recombination (Sabeti et al., 2002; Vitti et al., 2013). This pattern has been exploited to develop several methods based on LD to detect selective sweeps of recent origin (Vitti et al., 2013). Interestingly, LD-based methods are sensitive enough to detect both strong and soft selective sweeps (Garud et al., 2015), as well as partial or incomplete selective sweeps (Vitti et al., 2013), making them excellent tools to study recent and ongoing selection events, such as those occurring during domestication and the subsequent diversification of landraces (**Supplementary Table S1**).

Since the above rationale relies on LD decay due to recombination, any method based on LD requires to control for local variation in recombination rates in order to reduce false positives (Sabeti et al., 2002). The extended haplotype homozygosity (EHH) is a widely used statistic in LD-based methods that is defined as the probability that two orthologous genomic regions carrying a "core" haplotype of interest (*i.e.*, the part of the haplotype that is shared by all the individuals carrying it, such as the allele under positive selection) in the population are identical by descent (*i.e.*, they were inherited by the same ancestor), as one looks to a specified distance farther away from the core region (Sabeti et al., 2002).

Among the LD based methods that uses the EHH, we can mention the long-range haplotype (LRH) test, sometimes named the relative EHH (rEHH) test, which controls for local

recombination rates by comparing the EHH of several haplotypes localized within the same locus (Sabeti et al., 2002). Other EHH based methods include the whole-genome long-range haplotype (WGLRH) test that uses sliding windows to perform the LRH test (Zhang et al., 2006), the long-range haplotype similarity (LRHs) test (Hanchard et al., 2006), the integrated haplotype score (iHS) which is particularly sensitive to incomplete selective sweeps and soft sweeps (Voight et al., 2006) and the cross-population extended haplotype homozygosity (XP-EHH) statistic that is able to detect selective sweeps after the selected allele reached fixation (Sabeti et al., 2007). The iHS and the XP-EHH statistics can be regarded as complementary to each other, enabling the detection of incomplete and complete selective sweeps in the target population (Vitti et al., 2013).

All the LD-based tests that make use of the EHH statistic require the previous phasing of the chromosomes in order to work (*i.e.*, assignation of alleles in an individual to their corresponding maternal and paternal haplotypes), which may or may not be possible depending on the sequencing depth and type of data available for the analysis (Delaneau et al., 2013). For instance, a reference genome is usually needed in order to phase genotypes, since most methods rely on the information of proximity between alleles and their distribution within individuals in a population to assign haplotypes (Delaneau et al., 2013) although new methods are emerging that can phase genotypes without a reference genome (Money et al., 2017).

There are other LD-based methods that do not make use of the EHH statistic, such as the LD decay (LDD) test, which rely on individuals that are homozygous for any given SNPs to look for LD differences between alleles in a population (Wang et al., 2006) or the ω statistic that scans for high SNP correlation coefficients around a site under selection (Kim and Nielsen, 2004; Alachiotis et al., 2012). Another method that do not require chromosome phasing is the regression-based test, which relies on the reduction of heterozygosity as one approaches the locus under selection in a genome to infer selective sweeps (Wiener and Pong-Wong, 2011). Other LD-based methods exploit the estimation of identity-by-descent using genome-wide data to detect haplotypes that are shared between several unrelated individual (> 10 generations) to infer selective sweeps without previous knowledge of the pedigree of individual (Han and Abney, 2013), so they might prove useful to study recent domestication processes.

Some examples of LD-based methods used to explore the domestication process includes an analysis using LRH to detect signatures of selection associated to dairy and beef cattle breeds (Bomba et al., 2015), a study using the XP-EHH statistic to find signals of selective sweeps in Jinhua pigs (Li et al., 2016), and a paper focused on the diversification of goat landraces that calculated the iHS and the XP-EHH statistics alongside other tests to detect selective sweeps between goat breeds (Bertolini et al., 2018).

Other important tests include the XP-CLR test (Chen et al., 2010) and the μ statistic (Alachiotis and Pavlidis, 2018) which implement multiple signatures to detect selective sweeps (**Supplementary Table S1**) and have been used to detect candidate loci in maize and African rice, respectively (Hufford et al., 2012; Ndjiondjop et al., 2019).

Using GWAS to Detect Domestication-Associated Loci

Genome-wide association studies have been used extensively to uncover the genetic variants that underlie domestication traits (Shi and Lai, 2015). The domestication traits that can be analyzed through a GWAS can encompass any biological characteristic from simple morphological traits (Jiao et al., 2012) to the production of certain metabolites (Shang et al., 2014), tame behavior in animals (Ilska et al., 2017), resistance or susceptibility to certain diseases (Wang et al., 2012), or adaptation to certain environmental conditions (Song et al., 2018).

An important advantage of the GWAS over the bottom-up approaches is its ability to detect polygenic effects on single traits of interest, which is commonplace considering that genes interact between them and the environment to generate phenotypes (Gibson, 2018).

A prerequisite before performing a GWAS is to have large sample sizes in both the number of sequenced genetic variants and the number of individuals included in the study, as they are necessary to obtain the statistical power to detect variants with small effects and to reduce the risk of false positives (Wang G.-D. et al., 2014).

Some recent examples include the use of a GWAS to identify candidate genes with unknown functions involved in several agronomic traits, including drought and heat tolerance in chickpea (Varshney et al., 2019); a GWAS that revealed loci associated to fruit size and quality in peach (Cao et al., 2019); and a GWAS that uncovered the genetic variants involved in the absence of anthocyanin in domesticated rice compared to its wild relative (Zheng et al., 2019).

ANCIENT DNA AND PALEOGENOMICS OF DOMESTICATED TAXA

Extant domesticated taxa lack the information of ancient genetic diversity that was lost through bottlenecks, selection and genetic drift (Ramos-Madriral et al., 2016). However, the analysis of ancient DNA can allow the research community to overcome some of these limitations (Irving-Pease et al., 2019). Ancient DNA retrieved from archeological sites allows the study of the rate at which domestication happened, as well as revealing which genes were important at the beginning of this process (Vallebuena-Estrada et al., 2016; Irving-Pease et al., 2019). Thus, paleogenomics is becoming a novel research area for understanding the process of plant and animal domestication (Irving-Pease et al., 2019).

Extraction and Sequencing of Ancient DNA

An important limitation of paleogenomic analyses is the level of preservation of the ancient DNA itself, as well as the total yield of extracted DNA (Sawyer et al., 2012). The DNA molecules that are extracted from tissues that are not conserved on permafrost and are older than 100 years are usually shorter than 100 bp (Sawyer et al., 2012). The strand breaks of

such fragments are also non-random, as purines are enriched before the strand breaks (Sawyer et al., 2012). Additionally, these fragments incorporate cytosine-to-uracil mutations on their ends, further hindering the analysis of the sequenced fragments (Sawyer et al., 2012). Even though these characteristics hamper the sequencing and analysis of ancient DNA, they are also useful to differentiate between real ancient DNA and extant DNA contamination (Sawyer et al., 2012). Furthermore, due to the scarce ancient material located throughout few archeological sites worldwide, sample sizes in paleogenomic studies are very small, usually one or few individuals per location and sometimes only one locality (e.g., Wales et al., 2016; Ramos-Madrigo et al., 2016).

Given the above difficulties and the uniqueness of the biological material retrieved from archeological sites, it is crucial to extract and sequence as much ancient DNA as possible while avoiding DNA contamination (Gamba et al., 2016). Major efforts have been made to develop efficient protocols for ancient DNA extraction (Gamba et al., 2016) and single-strand library preparation for high-throughput sequencing (e.g., Gansauge et al., 2017). Organelle genomes were usually the target for ancient DNA sequencing because multiple copies of these can be found within each plant and animal cell and can reveal several demographic processes (Wales et al., 2016; Irving-Pease et al., 2019). Nonetheless, more evolutionary information can be retrieved from nuclear DNA, which is the main target for modern paleogenomic studies (Wales et al., 2016; Irving-Pease et al., 2019).

Insights of Paleogenomic Data in Domestication

Paleogenomic studies are challenging some of our previous ideas of the domestication process, such as the occurrence of ancient domestication bottlenecks, which appear to be absent in several archeological plant genomes, suggesting that the reduced diversity in domesticated taxa may be a more gradual process from what was expected using DNA of extant populations (Allaby et al., 2019). For example, several archeological samples of *Sorghum bicolor* from different time periods (ranging from 1800 to 100 years ago) were compared to extant individuals of the species, revealing that this crop did not suffer an initial domestication bottleneck, but rather that the reduction in genetic diversity, and its associated mutational load, occurred gradually throughout time (Smith et al., 2019).

Paleogenomics is also revealing important aspects of plant and animal domestication, such as the first genetic steps towards domestication syndromes as well as the overall graduality of the process (Ramos-Madrigo et al., 2016; Vallebuena-Estrada et al., 2016; Daly et al., 2018). For example, archeological remains of goat populations have revealed multiple domestication processes in ancient wild goats, possible dispersal routes of ancient goat populations and signs of early selective pressures towards candidate genes involved in pigmentation, milk production, size, reproduction and changes in diet (Daly et al., 2018). Likewise, several archeological maize samples retrieved from the Tehuacán

Valley in Mexico have revealed that early domesticates already presented signals of selective sweeps on important candidate genes, such as *teosinte branched1* and *brittle endosperm2*, but lacked selective sweep signals in other important candidate genes present on modern maize populations, even though these ancient maize populations were already endogamous and more closely related to modern maize than to wild teosinte, revealing that maize domestication was a gradual process ranging thousands of years (Ramos-Madrigo et al., 2016; Vallebuena-Estrada et al., 2016).

Other examples demonstrate the importance of paleogenomic studies in domesticated taxa, including grapevine (Wales et al., 2016), barley (Mascher et al., 2016), sunflower (Wales et al., 2019), horses (Schubert et al., 2014), dogs (Frantz et al., 2016) and cats (Otonari et al., 2017).

RNA SEQUENCING TO DETECT DIFFERENTIALLY EXPRESSED GENES ASSOCIATED TO DOMESTICATION

Besides the use of RNA-seq to obtain population-level data, comparative transcriptomics is a good way to find or support the validity of candidate genes (Hekman et al., 2015). Transcriptomic analyses between domesticated and wild taxa can reveal important changes in gene expression associated to domestication (Koenig et al., 2013; Hekman et al., 2015; Hradilová et al., 2017). Likewise, the analysis of hybrids between domesticated and wild individuals can reveal important patterns of allele-specific regulation and the role of *cis/trans* regulatory elements in the emergence of domestication traits (Bell et al., 2013; Lemmon et al., 2014).

The Experimental Design of Differential Expression Analyses

Transcriptomic profiles are tissue-specific and time-dependent (Hekman et al., 2015). Thus, a good experimental design can reveal important loci involved in the phenotypic differences associated to domestication syndromes, such as suppression of secondary metabolites, changes in form, size, taste, absence of defense mechanisms, seed dormancy, docile behavior, among other traits (Hekman et al., 2015). This can be done by comparing the total RNA expression of the tissue or organ of interest (Koenig et al., 2013), as well as comparing RNA expression throughout the developmental stages of such tissue or organ (Hradilová et al., 2017).

Since transcriptomic analyses are experimental by nature, experimental designs require biological replicates for each treatment, condition or organ to assess the variability in the data; as well as controlled environmental conditions to reduce possible biases and sources of error (Fang and Cui, 2011; Schurch et al., 2016). Empirical studies recommend using at least six biological replicates for each condition in the experiment, even though the use of three replicates is common, but discouraged (Burden et al., 2014; Schurch et al., 2016). Additionally, it is important to avoid committing

errors in the experimental design that can bias the results of the RNA-seq experiment, such as using different sequencing technologies for each sample, using different methods for library preparations throughout the samples, sequencing each treatment in a different sequencing flowcell or different lanes within a flowcell (Fang and Cui, 2011). Other technical biases associated to adapter ligation and within-lane variation can be properly assessed when using biological replicates (Fang and Cui, 2011; see **Table 1**).

RNA-seq data can also be complemented with metabolomic data to infer the association between the differential expression of genes and the presence/absence of metabolites between wild and domesticated taxa (Hradilová et al., 2017).

After obtaining high-quality data with an appropriate experimental design, RNA-seq analyses usually follow a similar workflow, which should culminate in the detection of differentially expressed genes between a wild plant and its domesticated counterpart (Yang and Kim, 2015; see **Table 1**). These differentially expressed genes are most likely candidates that may explain to some degree the changes associated to domestication (Koenig et al., 2013; Hradilová et al., 2017). Nonetheless, one must be careful while interpreting the results of these studies, as some differentially expressed genes between wild and domesticated taxa may be a consequence, rather than a cause, of the domestication traits under study (Albert et al., 2012).

Successful Examples of RNA-seq Experiments to Understand Domestication

RNA-seq analysis has been successfully employed to discover differentially expressed genes involved in the domestication of several plant species. For example, RNA-seq analyses between maize and teosinte found 600 differentially expressed genes and 1,100 genes with altered patterns of co-expression, mainly involved in biotic stress responses, and many of which were previously found as candidate genes using selective scans (Myers et al., 2012). Similar results have been found in tomato (Koenig et al., 2013), pea (Hradilová et al., 2017), common bean (Singh et al., 2018), and carrot (Machaj et al., 2018). This approach has also led to the discovery of differentially expressed genes between dogs and wolves associated to tameness (Li et al., 2013), as well as changes related to the immune system and aerobic capacity (Yang et al., 2018). Another study found differential isoform expression between wild and domesticated sorghum accessions, revealing that domestication can alter the patterns of alternative splicing (Ranwez et al., 2017). Hybrid studies have been performed between maize and teosinte, suggesting potential selection on *cis* regulatory elements associated with changes in ear tissue and previously reported candidate genes (Lemmon et al., 2014). Another hybrid study in *Capsicum annuum* using network analyses revealed that loss of function in *cis* regulatory sequences lead to transcriptional changes in *trans* elements that are associated with fruit morphology (Díaz-Valenzuela et al., 2020).

MODERN EPIGENOMICS AND METHODOLOGICAL STRATEGIES TO EXPLORE DOMESTICATION

Epigenetics is classically defined as the heritable mechanisms that regulate gene expression without direct modifications to the DNA sequence, namely DNA methylation, RNA methylation, covalent histone modifications and chromatin assembly states (Sakurada, 2010; Zhao et al., 2017). Epigenetic variants, sometimes called epialleles, are local differences in these epigenetic marks between individuals in a population, which can have similar dynamics to genetic variants (Weigel and Colot, 2012; Guo et al., 2015). Since epigenetic mechanisms underly the ability of organisms to respond to changing environmental conditions, some epigenetic marks associated to these responses are more susceptible to change due to environmental input, while other marks involved in cell differentiation, embryonic development and core cellular functions might be more stable (Turner, 2009).

Most of the domestication studies that explain phenotypic differences between wild and domesticated taxa focus on genetic variation. However, the study of epigenomics may explain some of the missing heritability in domestication traits (i.e., the gap between the heritability of a trait estimated by classic genetics and GWAS), the patterns of differentially expressed genes that do not have clear signs of selective sweeps, or even connect the causality between the genetic variation that was selected for during domestication and the resulting phenotypes (Schmitz et al., 2013; Trerotola et al., 2015; Janowitz Koch et al., 2016; Bélteky et al., 2018).

Epigenetic variation can be inherited from one generation to the next in a process known as trans-generational epigenetic inheritance, which has been documented in plants and animals (Heard and Martienssen, 2014), even though the overall importance of this trans-generational epigenetic inheritance in plant and animal evolution is still debated (see **Table 1**). Nevertheless, we consider that studying epigenetic patterns associated to transcriptional activity and phenotypic traits should help understand the emergence of domestication phenotypes (Bélteky et al., 2018). If epigenetic variants such as single methylation polymorphisms (SMPs) show complete transgenerational inheritance, they can even be analyzed using the theoretical tools of population genetics to detect selective sweeps (Schmitz et al., 2013; Janowitz Koch et al., 2016).

In a similar fashion to GWAS, the use of epigenome-wide association studies (EWAS) can also reveal the association of an epigenetic variant to a trait of interest in domesticated taxa (Feeney et al., 2014). The same precautions taken in transcriptomic data should also be taken for epigenomic data, since the patterns of epigenetic marks in organisms are tissue-specific, time-dependent and sensitive to environmental input, meaning that epigenomic data should be analyzed for specific organs or tissues of interest in a controlled environment (Jensen, 2015). This is particularly important for the epigenetic marks that respond to environmental input, since domesticated taxa and their wild relatives live under different environmental

conditions. Growing both taxa under controlled conditions will alter the natural state of these marks, but will also help differentiate the heritable epialleles associated to domestication traits (Turner, 2009).

Obtaining Population Data From Epigenetic Marks

The most studied epigenetic mark is DNA 5-methylcytosine, which refers to the DNA methylation in cytosines which are usually associated to transcriptional gene silencing (He et al., 2011). Cytosine methylome data can be obtained using high-throughput sequencing technologies alongside bisulfite sequencing (Meissner, 2005). Bisulfite sequencing consists in the deamination of unmethylated cytosines through a bisulfite reaction, converting them into uracil, which are encoded as thymine by sequencing technologies (Frommer et al., 1992). The comparison of sequenced DNA that was treated with bisulfite alongside sequenced DNA without treatment can discriminate between methylated and unmethylated cytosines in an organ, tissue or cell-type of interest (Frommer et al., 1992).

Reduced representation bisulfite sequencing (RRBS) is a high-throughput technique with a similar rationale to RAD-seq that enriches the sequencing of CG rich regions of the genome after the digestion of restriction enzymes (Meissner, 2005). This makes the RRBS technique a cost-effective option to analyze cytosine methylation patterns in mammals, since its cytosine DNA methylation happens at CG sites (Meissner, 2005; He et al., 2011). Plant cytosine methylomes should instead be analyzed through MethylC-seq, which consists of whole-genome sequencing and bisulfite treatment (Urich et al., 2015), as cytosine methylation can also happen in CHG and CHH sites in plant genomes (He et al., 2011). Cytosine methylation can also be detected using methylated DNA immunoprecipitation sequencing (MeDIP-seq), which consists in shearing the genomic DNA into small pieces followed by the immunoprecipitation of the methylated cytosines using antibodies that recognizes 5-methylcytosine and finally sequencing the DNA sequences with the methylated sites using standard high-throughput sequencing technologies (Weber et al., 2005).

Besides cytosine methylation, adenine has also been shown to be methylated in both plants and animals (N6-methyldeoxyadenosine), which cannot be detected using bisulfite sequencing (Luo et al., 2015). However, genomic regions with methylated adenines can be detected using N6-methyldeoxyadenosine immunoprecipitation sequencing (6mA-IP-seq), which uses the same rationale as MeDIP-seq but requires antibodies that specifically targets N6-methyldeoxyadenosine (Fu et al., 2015). PacBio and Nanopore sequencing technologies are known to be sensitive to DNA methylation, regardless of it being on a cytosine or adenine, so they are currently being used as powerful, albeit expensive tools to evaluate DNA methylation patterns in genomes (Gouil and Keniry, 2019).

Histone modifications refers to either posttranslational covalent modifications in histones (methylations, acetylations, phosphorylations, ubiquitylations, ADP-ribosylations,

sumoylations, crotonylations, malonylations, succinylations) or the substitution of canonical histones by histone variants with different amino acid composition (Bowman and Poirier, 2015). These histone modifications determine the functionality of local genomic regions by changing the state of the chromatin either through its direct effects on the chemical interactions between DNA and histones or through the recruitment of chromatin remodeling complexes (Bowman and Poirier, 2015).

Chromatin immunoprecipitation sequencing (ChIP-seq) can be used to assess the genome-wide association between DNA regions and specific histone modifications (Schmidt et al., 2009). ChIP-seq consists in the initial fixation of DNA-protein interactions using formaldehyde followed by DNA fragmentation and subsequent enrichment of the target histone modification using magnetic beads coupled to antibodies in order to sequence the genomic regions where the histone modification is present (Schmidt et al., 2009). ChIP-seq can also be used to assess the interaction between any DNA-binding protein such as transcriptional factors and specific genomic regions (Schmidt et al., 2009).

Epigenomic Studies Applied to Understand Domestication

The current epigenomic analyses regarding domestication have focused on DNA methylation patterns (Jensen, 2015; Ding and Chen, 2018), but some studies have also ventured into histone modification patterns (He et al., 2014). Recent efforts are trying to connect the discoveries of genomics and epigenetics to understand the evolution of tameness in domesticated animals (Jensen, 2015). A study using RRBS that compared the DNA methylation patterns between wolves and dogs revealed signals of natural selection acting on SMPs which are enriched in transposons and genes involved in the regulation of neurotransmitters, suggesting a dog-specific silencing of genes involved in behavior (Janowitz Koch et al., 2016). Similarly, a recent study using MeDIP-seq in red junglefowl populations that were bred to have either high or low fear to humans discovered genomic region that were differentially methylated in genes that were previously related to tameness (Bélteky et al., 2018).

Other studies focused on plant domestication have found differentially methylated sites associated to domestication syndromes (Song et al., 2017; Shen et al., 2018). A study using MethylC-seq found 519 differentially methylated genes between domesticated and wild cotton from which some of them are associated with the observed differences in flowering time and seed dormancy between the wild and domesticated taxa (Song et al., 2017). Another study using MethylC-seq found 4,248 differentially methylated regions between wild and domesticated soybean and 1,164 differentially methylated regions between domesticated and improved soybean (Shen et al., 2018). As expected, the differentially methylated regions in soybean had higher genetic diversity compared to the regions with evidence of selective sweeps that were previously found, and interestingly, 22.5% of the differentially methylated sites could be associated to a causal genetic variant (suggesting that these genetic variants were responsible for the observed epigenetic

patterns), whereas the rest of the differentially methylated regions could be interpreted as genuine epialleles located within genes involved in carbohydrate metabolism (Shen et al., 2018).

EXPERIMENTAL VALIDATION OF CANDIDATE GENES

Once we have evidence of candidate genes involved in the domestication syndrome, the necessary next step to understand the genetic basis of domestication is to design *in vitro* systems, knock-out, knock-down or knock-in experiments that validate the involvement of such genes in the observed phenotypes (Zhang et al., 2017). This can be performed either by direct alteration of the genome in the organism of interest, by using RNA interference or by designing heterologous systems in a model organism (Boettcher and McManus, 2015). As an example, a knock-out experiment with backcrosses between domesticated and wild mice elucidated the role of some genes involved in behavioral changes associated to mouse domestication (Chalfin et al., 2014).

Previous knock-out and knock-in experiments were restricted to model organisms, but nowadays experimental validation of candidate genes can be supported via knock-out and knock-in experiments, using novel genome editing tools (e.g., Shalem et al., 2014; Hahn et al., 2017; Ueta et al., 2017). Genome-editing tools are already available for a broad range of taxa, including dozens of crop species, but developing a working system in non-model organisms can still be a difficult task that can take several months or even years to accomplish (Shan et al., 2020), so doing collaborative studies alongside experimental researchers is recommended. In this moment, the leading toolset to perform genome editing is the Clustered Regulatory Interspaced Short Palindromic Repeats (CRISPR) system alongside the CRISPR associated protein 9 (Cas9), commonly known as CRISPR/Cas9, which can be used to eliminate, introduce or replace specific segments of DNA within a targeted site in a genome (Cong et al., 2013). Another useful tool for genome editing is the Transcription Activator-Like Effector Nuclease (TALEN) technology, which has its own advantages in comparison to CRISPR/CAS9 (Zhang et al., 2017). RNA interference can also help in validating the function of candidate genes, although it is limited to knock-down experiments (Boettcher and McManus, 2015). Heterologous expression in model organisms is a cost-effective alternative to validate candidate genes (e.g., Schweiger et al., 2010), although this method overlooks the interaction networks that exist *in vivo* which are accountable for the emergence of phenotypes (Rodríguez-Mega et al., 2015).

Regardless the genome-editing tool of choice (Boettcher and McManus, 2015; Zhang et al., 2017), genome edition is proving its usefulness to validate the effect of candidate genes involved in domestication through the introduction of domesticated alleles on wild relatives and vice-versa (Zhou J. et al., 2019), which can prove that the gene is indeed involved in the appearance of the domesticated phenotype (Zhou J. et al., 2019). This can be performed in the same way as a usual knock-out or knock-in experiment, where the edited locus must be validated

through PCR and Sanger sequencing, a PCR-RFLP analysis or using Western-blot in case of a protein knock-out (e.g., Ueta et al., 2017). The expected result of these type of studies is to find a modified phenotype after editing a candidate locus, either a wild individual with a domesticated-like phenotypic trait or a domesticated individual with a wild-like phenotypic trait (Zhou J. et al., 2019).

Of course, the above studies will hardly reproduce a complete domesticated or wild phenotype, since genetic elements interact in complex regulatory networks, including other elements within the genome as well as epigenetic and environmental components (Rodríguez-Mega et al., 2015), but nonetheless will be useful to understand the role of those genes in the emergence of domesticated phenotypes.

Once the candidate genes are validated, genome-editing tools can also become useful to introduce desirable traits from wild relatives to its domesticated counterparts, a goal of great interest for crop improvement (Zhou J. et al., 2019) and currently used to accelerate plant breeding and to fine-tune desirable traits (Wolter et al., 2019). Furthermore, recent efforts are trying to domesticate plant crops *de novo* by inserting the desired domestication alleles into their wild relatives, generating crops with the desired domestication phenotypes but without the problems of low genetic variation and accumulation of deleterious mutations that are an inevitable consequence of regular domestication processes (Fernie and Yan, 2019).

CONCLUSION AND PERSPECTIVES

Plant and animal domestication can be studied using genomic, transcriptomic and epigenomic strategies, revealing the action of evolutionary, ecological and anthropogenic processes (Kantar et al., 2017). These tools can lead us beyond the description of the possible historical scenarios that shaped the domesticated species, since we can explore the effects of domestication on the transcriptomic activity of a species (Hekman et al., 2015), test the validity of candidate genes associated to domestication phenotypes (Zhou J. et al., 2019) and analyze epigenetic patterns associated to domestication traits (Jensen, 2015). Many domesticated taxa remain genetically unexplored, and as sequencing technologies become cheaper and more efficient, domestication genomics will soon be available for polyploids and species with huge genomes (e.g., Edger et al., 2019).

Nonetheless, the modern study of domestication of plants and animals should still be multidisciplinary, since genetics only tells us part of the story (Larson et al., 2014). An extended synthesis framework should also be considered to understand domestication, as these new studies are helping us understand niche construction and the emergence of domesticated phenotypes (Piperno, 2017). Other potential lines of work remain to be addressed in domestication studies, such as the changes in the chromatin architecture (e.g., Concia et al., 2020), the use of comparative proteomic atlases (e.g., Jiang Y. et al., 2019) and the analysis of cell-type divergences during development using single-cell RNA-seq data (Arendt et al., 2016). The use of this multi-omic approaches will help us create and

compare developmental atlases (e.g., Walley et al., 2016) between wild and domesticated taxa to understand how morphology diverged during domestication.

AUTHOR CONTRIBUTIONS

JB-R, DP, and LE wrote the manuscript. All authors contributed to the article and approved the submitted version.

FUNDING

This study was funded by Comisión Nacional para el Conocimiento y Uso de la Biodiversidad (CONABIO) KE004 “Diversidad genética de las especies de *Cucurbita* en México e hibridación entre plantas genéticamente modificadas y especies silvestres de *Cucurbita*” and CONABIO PE001 “Diversidad genética de las especies de *Cucurbita* en México. Fase II. Genómica evolutiva y de poblaciones, recursos genéticos y domesticación, both awarded to Rafael Lira-Saade and LE.

REFERENCES

- Alachiotis, N., and Pavlidis, P. (2018). RAiSD detects positive selection based on multiple signatures of a selective sweep and SNP vectors. *Commun. Biol.* 1, 1–11. doi: 10.1038/s42003-018-0085-8
- Alachiotis, N., Stamatakis, A., and Pavlidis, P. (2012). OmegaPlus: a scalable tool for rapid detection of selective sweeps in whole-genome datasets. *Bioinformatics* 28, 2274–2275. doi: 10.1093/bioinformatics/bts419
- Albert, F. W., Somel, M., Carneiro, M., Aximu-Petri, A., Halbwax, M., Thalmann, O., et al. (2012). A comparison of brain gene expression levels in domesticated and wild animals. *PLoS Genet.* 8:e1002962. doi: 10.1371/journal.pgen.1002962
- Alberto, F. J., Boyer, F., Orozco-terWengel, P., Streeter, I., Servin, B., de Villemereuil, P., et al. (2018). Convergent genomic signatures of domestication in sheep and goats. *Nat. Commun.* 9:813. doi: 10.1038/s41467-018-03206-y
- Alexander, D. H., Novembre, J., and Lange, K. (2009). Fast model-based estimation of ancestry in unrelated individuals. *Genome Res.* 19, 1655–1664. doi: 10.1101/gr.094052.109
- Ali, O. A., O'Rourke, S. M., Amish, S. J., Meek, M. H., Luikart, G., Jeffres, C., et al. (2016). RAD capture (Rapture): flexible and efficient sequence-based genotyping. *Genetics* 202, 389–400. doi: 10.1534/genetics.115.183665
- Allaby, R. G., Ware, R. L., and Kistler, L. (2019). A re-evaluation of the domestication bottleneck from archaeogenomic evidence. *Evol. Appl.* 12, 29–37. doi: 10.1111/eva.12680
- Anderson, C. N. K., Ramakrishnan, U., Chan, Y. L., and Hadly, E. A. (2005). Serial SimCoal: a population genetics model for data from multiple populations and points in time. *Bioinformatics* 21, 1733–1734. doi: 10.1093/bioinformatics/btl154
- Andrews, K. R., Good, J. M., Miller, M. R., Luikart, G., and Hohenlohe, P. A. (2016). Harnessing the power of RADseq for ecological and evolutionary genomics. *Nat. Rev. Genet.* 17, 81–92. doi: 10.1038/nrg.2015.28
- Arendt, D., Musser, J. M., Baker, C. V., Bergman, A., Cepko, C., Erwin, D. H., et al. (2016). The origin and evolution of cell types. *Nat. Rev. Genet.* 17, 744–757. doi: 10.1038/nrg.2016.127
- Badouin, H., Gouzy, J., Grassa, C. J., Murat, F., Staton, S. E., Cottret, L., et al. (2017). The sunflower genome provides insights into oil metabolism, flowering and Asterid evolution. *Nature* 546, 148–152. doi: 10.1038/nature22380
- Baran, Y., Pasaniuc, B., Sankararaman, S., Torgerson, D. G., Gignoux, C., Eng, C., et al. (2012). Fast and accurate inference of local ancestry in Latino populations. *Bioinformatics* 28, 1359–1367. doi: 10.1093/bioinformatics/bts144
- Baute, G. J., Kane, N. C., Grassa, C. J., Lai, Z., and Rieseberg, L. H. (2015). Genome scans reveal candidate domestication and improvement genes in cultivated sunflower, as well as post-domestication introgression with wild relatives. *New Phytol.* 206, 830–838. doi: 10.1111/nph.13255
- Bell, G. D., Kane, N. C., Rieseberg, L. H., and Adams, K. L. (2013). RNA-seq analysis of allele-specific expression, hybrid effects, and regulatory divergence in hybrids compared with their parents from natural populations. *Genome Biol. Evol.* 5, 1309–1323. doi: 10.1093/gbe/evt072
- Belser, C., Istace, B., Denis, E., Dubarry, M., Baurens, F. C., Falentin, C., et al. (2018). Chromosome-scale assemblies of plant genomes using nanopore long reads and optical maps. *Nat. Plants* 4, 879–887. doi: 10.1038/s41477-018-0289-4
- Bélteky, J., Agnval, B., Bektic, L., Höglund, A., Jensen, P., and Guerrero-Bosagna, C. (2018). Epigenetics and early domestication: differences in hypothalamic DNA methylation between red junglefowl divergently selected for high or low fear of humans. *Genet. Sel. Evol.* 50:13. doi: 10.1186/s12711-018-0384-z
- Bertolini, F., Servin, B., Talenti, A., Rochat, E., Kim, E. S., Oget, C., et al. (2018). Signatures of selection and environmental adaptation across the goat genome post-domestication. *Genet. Sel. Evol.* 50:57. doi: 10.1186/s12711-018-0421-y
- Bickhart, D. M., Rosen, B. D., Koren, S., Sayre, B. L., Hastie, A. R., Chan, S., et al. (2017). Single-molecule sequencing and chromatin conformation capture enable *de novo* reference assembly of the domestic goat genome. *Nat. Genet.* 49, 643–650. doi: 10.1038/ng.3802
- Boettcher, M., and McManus, M. T. (2015). Choosing the right tool for the job: RNAi, TALEN, or CRISPR. *Mol. Cell* 58, 575–585. doi: 10.1016/j.molcel.2015.04.028
- Bomba, L., Nicolazzi, E. L., Milanese, M., Negrini, R., Mancini, G., Biscarini, F., et al. (2015). Relative extended haplotype homozygosity signals across breeds reveal dairy and beef specific signatures of selection. *Genet. Sel. Evol.* 47:25. doi: 10.1186/s12711-015-0113-9
- Bonhomme, M., Chevalet, C., Servin, B., Boitard, S., Abdallah, J., Blott, S., et al. (2010). Detecting selection in population trees: the lewontin and krakauer test extended. *Genetics* 186, 241–262. doi: 10.1534/genetics.104.117275
- Bowman, G. D., and Poirier, M. G. (2015). Post-translational modifications of histones that influence nucleosome dynamics. *Chem. Rev.* 115, 2274–2295. doi: 10.1021/cr500350x
- Brohammer, A. B., Kono, T. J. Y., and Hirsch, C. N. (2018). “The maize pan-genome,” in *The Maize Genome. Compendium of Plant Genomes*, eds J. Bennetzen, S. Flint-Garcia, C. Hirsch, and R. Tuberosa (Cham: Springer), 13–29. doi: 10.1007/978-3-319-97427-9_2
- Brozynska, M., Furtado, A., and Henry, R. J. (2016). Genomics of crop wild relatives: expanding the gene pool for crop improvement. *Plant Biotechnol. J.* 14, 1070–1085. doi: 10.1111/pbi.12454

ACKNOWLEDGMENTS

We acknowledge the Doctorado en Ciencias Biomédicas for the academic support provided to JB-R during the development of this project. We thank Alejandra Moreno Letelier for her insights, which greatly improved the quality of the manuscript.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fgene.2020.00742/full#supplementary-material>

- Bryant, D. M., Johnson, K., DiTommaso, T., Tickle, T., Couger, M. B., Payzin-Dogru, D., et al. (2017). A tissue-mapped axolotl de novo transcriptome enables identification of limb regeneration factors. *Cell Rep.* 18, 762–776. doi: 10.1016/j.celrep.2016.12.063
- Burden, C. J., Qureshi, S. E., and Wilson, S. R. (2014). Error estimates for the analysis of differential expression from RNA-seq count data. *PeerJ* 2:e576. doi: 10.7717/peerj.576
- Burgarella, C., Cubry, P., Kane, N. A., Varshney, R. K., Mariac, C., Liu, X., et al. (2018). A western Sahara centre of domestication inferred from pearl millet genomes. *Nat. Ecol. Evol.* 2, 1377–1380. doi: 10.1038/s41559-018-0643-y
- Burggren, W. (2016). Epigenetic inheritance and its role in evolutionary biology: re-evaluation and new perspectives. *Biology (Basel)*. 5:24. doi: 10.3390/biology5020024
- Cao, K., Li, Y., Deng, C. H., Gardiner, S. E., Zhu, G., Fang, W., et al. (2019). Comparative population genomics identified genomic regions and candidate genes associated with fruit domestication traits in peach. *Plant Biotechnol. J.* 17, 1954–1970. doi: 10.1111/pbi.13112
- Catchen, J., Hohenlohe, P., Bassham, S., Amores, A., and Cresko, W. (2013). Stacks: an analysis tool set for population genomics. *Mol. Ecol.* 22, 3124–3140. doi: 10.1016/j.biotechadv.2011.08.021.Secreted
- Chalfin, L., Dayan, M., Levy, D. R., Austad, S. N., Miller, R. A., Iraqi, F. A., et al. (2014). Mapping ecologically relevant social behaviours by gene knockout in wild mice. *Nat. Commun.* 5:4569. doi: 10.1038/ncomms5569
- Charlesworth, B. (2009). Effective population size and patterns of molecular evolution and variation. *Nat. Rev. Genet.* 10, 195–205. doi: 10.1038/nrg2526
- Chen, H., Patterson, N., and Reich, D. (2010). Population differentiation as a test for selective sweeps. *Genome Res.* 20, 393–402. doi: 10.1101/gr.100545.109
- Chen, J., Ni, P., Li, X., Han, J., Jakovljić, I., Zhang, C., et al. (2018). Population size may shape the accumulation of functional mutations following domestication. *BMC Evol. Biol.* 18:4. doi: 10.1186/s12862-018-1120-6
- Chen, N., Cai, Y., Chen, Q., Li, R., Wang, K., Huang, Y., et al. (2018). Whole-genome resequencing reveals world-wide ancestry and adaptive introgression events of domesticated cattle in East Asia. *Nat. Commun.* 9:2337. doi: 10.1038/s41467-018-04737-0
- Clark, J. W., and Donoghue, P. C. J. (2018). Whole-genome duplication and plant macroevolution. *Trends Plant Sci.* 23, 933–945. doi: 10.1016/j.tplants.2018.07.006
- Concia, L., Veluchamy, A., Ramirez-Prado, J. S., Martin-Ramirez, A., Huang, Y., Perez, M., et al. (2020). Wheat chromatin architecture is organized in genome territories and transcription factories. *Genome Biol.* 21, 1–20. doi: 10.1186/s13059-020-01998-1
- Cong, L., Ran, F. A., Cox, D., Lin, S., Barretto, R., Habib, N., et al. (2013). Multiplex genome engineering using CRISPR/Cas systems. *Science* 339, 819–823. doi: 10.1126/science.1231143
- Cornuet, J.-M., Pudlo, P., Veyssier, J., Dehne-Garcia, A., Gautier, M., Leblois, R., et al. (2014). DIYABC v2.0: a software to make approximate Bayesian computation inferences about population history using single nucleotide polymorphism, DNA sequence and microsatellite data. *Bioinformatics* 30, 1187–1189. doi: 10.1093/bioinformatics/btt763
- Daly, K. G., Maisano Delsler, P., Mullin, V. E., Scheu, A., Mattiangeli, V., Teasdale, M. D., et al. (2018). Ancient goat genomes reveal mosaic domestication in the Fertile Crescent. *Science* 361, 85–88. doi: 10.1126/science.aas9411
- Davey, J. W., and Blaxter, M. L. (2010). RADSeq?: next-generation population genetics. *Brief. Funct. Genomics* 9, 416–423. doi: 10.1093/bfgp/elq031
- De Mita, S., Thuillet, A.-C., Gay, L., Ahmadi, N., Manel, S., Ronfort, J., et al. (2013). Detecting selection along environmental gradients: analysis of eight methods and their effectiveness for outbreeding and selfing populations. *Mol. Ecol.* 22, 1383–1399. doi: 10.1111/mec.12182
- de Villemereuil, P., and Gaggiotti, O. E. (2015). A new F_{ST}-based method to uncover local adaptation using environmental variables. *Methods Ecol. Evol.* 6, 1248–1258. doi: 10.1111/2041-210X.12418
- De Wit, P., Pespeni, M. H., Ladner, J. T., Barshir, D. J., Seneca, F., Jaris, H., et al. (2012). The simple fool's guide to population genomics via RNA-Seq: an introduction to high-throughput sequencing data analysis. *Mol. Ecol. Resour.* 12, 1058–1067. doi: 10.1111/1755-0998.12003
- Delaneau, O., Howie, B., Cox, A. J., Zagury, J.-F., and Marchini, J. (2013). Haplotype estimation using sequencing reads. *Am. J. Hum. Genet.* 93, 687–696. doi: 10.1016/j.ajhg.2013.09.002
- Diao, L., and Chen, K. C. (2012). Local ancestry corrects for population structure in *Saccharomyces cerevisiae* genome-wide association studies. *Genetics* 192, 1503–1511. doi: 10.1534/genetics.112.144790
- Dias-Alves, T., Mairal, J., and Blum, M. G. B. (2018). Loter: a software package to infer local ancestry for a wide range of species. *Mol. Biol. Evol.* 35, 2318–2326. doi: 10.1093/molbev/msy126
- Diaz-Valenzuela, E., Sawers, R. H., and Cibrián-Jaramillo, A. (2020). Cis- and trans-regulatory variations in the domestication of the chili pepper fruit. *Mol. Biol. Evol.* 37, 1593–1603. doi: 10.1093/molbev/msaa027
- Ding, M., and Chen, Z. J. (2018). Epigenetic perspectives on the evolution and domestication of polyploid plant and crops. *Curr. Opin. Plant Biol.* 42, 37–48. doi: 10.1016/j.pbi.2018.02.003
- Doebley, J., and Stec, A. (1991). Genetic analysis of the morphological differences between maize and teosinte. *Genetics* 129, 285–295.
- Doebley, J., Stec, A., and Gustus, C. (1995). teosinte branched1 and the origin of maize: evidence for epistasis and the evolution of dominance. *Genetics* 141, 333–346.
- Dong, Y., Xie, M., Jiang, Y., Xiao, N., Du, X., Zhang, W., et al. (2013). Sequencing and automated whole-genome optical mapping of the genome of a domestic goat (*Capra hircus*). *Nat. Biotechnol.* 31, 135–141. doi: 10.1038/nbt.2478
- Dorant, Y., Benestan, L., Rougemont, Q., Normandeau, E., Boyle, B., Rochette, R., et al. (2019). Comparing Pool-seq, Rapture, and GBS genotyping for inferring weak population structure: the American lobster (*Homarus americanus*) as a case study. *Ecol. Evol.* 9, 6606–6623. doi: 10.1002/ece3.5240
- Durand, E. Y., Patterson, N., Reich, D., and Slatkin, M. (2011). Testing for ancient admixture between closely related populations. *Mol. Biol. Evol.* 28, 2239–2252. doi: 10.1093/molbev/msr048
- Edelman, N. B., Frandsen, P. B., Miyagi, M., Clavijo, B., Davey, J., Dikow, R. B., et al. (2019). Genomic architecture and introgression shape a butterfly radiation. *Science* 366, 594–599. doi: 10.1126/science.aaw2090
- Edger, P. P., Poorten, T. J., VanBuren, R., Hardigan, M. A., Colle, M., McKain, M. R., et al. (2019). Origin and evolution of the octoploid strawberry genome. *Nat. Genet.* 51, 541–547. doi: 10.1038/s41588-019-0356-4
- Ellegren, H. (2014). Genome sequencing and population genomics in non-model organisms. *Trends Ecol. Evol.* 29, 51–63. doi: 10.1016/j.tree.2013.09.008
- Excoffier, L., and Foll, M. (2011). fastsimcoal: a continuous-time coalescent simulator of genomic diversity under arbitrarily complex evolutionary scenarios. *Bioinformatics* 27, 1332–1334. doi: 10.1093/bioinformatics/btr124
- Excoffier, L., Hofer, T., and Foll, M. (2009). Detecting loci under selection in a hierarchically structured population. *Heredity (Edinb)*. 103:285. doi: 10.1038/hdy.2009.74
- Fang, Z., and Cui, X. (2011). Design and validation issues in RNA-seq experiments. *Brief. Bioinform.* 12, 280–287. doi: 10.1093/bib/bbr004
- Fay, J. C., and Wu, C. I. (2000). Hitchhiking under positive Darwinian selection. *Genetics* 155, 1405–1413.
- Feeney, A., Nilsson, E., and Skinner, M. K. (2014). Epigenetics and transgenerational inheritance in domesticated farm animals. *J. Anim. Sci. Biotechnol.* 5:48. doi: 10.1186/2049-1891-5-48
- Fernie, A. R., and Yan, J. (2019). De novo domestication: an alternative route toward new crops for the future. *Mol. Plant* 12, 615–631. doi: 10.1016/j.molp.2019.03.016
- Fierst, J. L. (2015). Using linkage maps to correct and scaffold de novo genome assemblies: methods, challenges, and computational tools. *Front. Genet.* 6:220. doi: 10.3389/fgene.2015.00220
- Fitz-Gibbon, S., Hipp, A. L., Pham, K. K., Manos, P. S., and Sork, V. L. (2017). Phylogenomic inferences from reference-mapped and de novo assembled short-read sequence data using RADseq sequencing of California white oaks (*Quercus* section *Quercus*). *Genome* 60, 743–755. doi: 10.1139/gen-2016-0202
- Foll, M., and Gaggiotti, O. (2008). A genome-scan method to identify selected loci appropriate for both dominant and codominant markers: a Bayesian perspective. *Genetics* 180, 977–993. doi: 10.1534/genetics.108.092221
- Frantz, L. A. F., Mullin, V. E., Pionnier-Capitan, M., Lebrasseur, O., Ollivier, M., Perri, A., et al. (2016). Genomic and archaeological evidence suggest a dual origin of domestic dogs. *Science* 352, 1228–1231. doi: 10.1126/science.aaf3161
- Frantz, L. A. F., Schraiber, J. G., Madsen, O., Megens, H.-J., Cagan, A., Bosse, M., et al. (2015). Evidence of long-term gene flow and selection during

- domestication from analyses of Eurasian wild and domestic pig genomes. *Nat. Genet.* 47, 1141–1148. doi: 10.1038/ng.3394
- Frommer, M., McDonald, L. E., Millar, D. S., Collis, C. M., Watt, F., Grigg, G. W., et al. (1992). A genomic sequencing protocol that yields a positive display of 5-methylcytosine residues in individual DNA strands. *Proc. Natl. Acad. Sci. U.S.A.* 89, 1827–1831. doi: 10.1073/pnas.89.5.1827
- Fu, Y., Luo, G.-Z., Chen, K., Deng, X., Yu, M., Han, D., et al. (2015). N6-methyldeoxyadenosine marks active transcription start sites in chlamydomonas. *Cell* 161, 879–892. doi: 10.1016/j.cell.2015.04.010
- Fugère, V., and Hendry, A. P. (2018). Human influences on the strength of phenotypic selection. *Proc. Natl. Acad. Sci. U.S.A.* 115, 10070–10075. doi: 10.1073/pnas.1806013115
- Fumagalli, M. (2013). Assessing the effect of sequencing depth and sample size in population genetics inferences. *PLoS One* 8:e79667. doi: 10.1371/journal.pone.0079667
- Futschik, A., and Schlötterer, C. (2010). The next generation of molecular markers from massively parallel sequencing of pooled DNA samples. *Genetics* 186, 207–218. doi: 10.1534/genetics.110.114397
- Gamba, C., Hanghøj, K., Gaunitz, C., Alfarhan, A. H., Alquraishi, S. A., Al-Rasheid, K. A. S., et al. (2016). Comparing the performance of three ancient DNA extraction methods for high-throughput sequencing. *Mol. Ecol. Resour.* 16, 459–469. doi: 10.1111/1755-0998.12470
- Gansauge, M.-T., Gerber, T., Glocke, I., Korleviæ, P., Lippik, L., Nagel, S., et al. (2017). Single-stranded DNA library preparation from highly degraded DNA using T4 DNA ligase. *Nucleic Acids Res.* 45:gkx033. doi: 10.1093/nar/gkx033
- Gao, L., Gonda, L., Sun, H., Ma, Q., Bao, K., Tieman, D. M., et al. (2019). The tomato pan-genome uncovers new genes and a rare allele regulating fruit flavor. *Nat. Genet.* 51, 1044–1051. doi: 10.1038/s41588-019-0410-2
- Garud, N. R., Messer, P. W., Buzbas, E. O., and Petrov, D. A. (2015). Recent selective sweeps in North American *Drosophila melanogaster* show signatures of soft sweeps. *PLoS Genet.* 11:e1005004. doi: 10.1371/journal.pgen.1005004
- Gepts, P. (2014). The contribution of genetic and genomic approaches to plant domestication studies. *Curr. Opin. Plant Biol.* 18, 51–59. doi: 10.1016/j.pbi.2014.02.001
- Gerbault, P., Allaby, R. G., Boivin, N., Ruzdinski, A., Grimaldi, I. M., Pires, J. C., et al. (2014). Storytelling and story testing in domestication. *Proc. Natl. Acad. Sci. U.S.A.* 111, 6159–6164. doi: 10.1073/pnas.1400425111
- Gibson, G. (2018). Population genetics and GWAS: a primer. *PLoS Biol.* 16:e2005485. doi: 10.1371/journal.pbio.2005485
- Golicz, A. A., Batley, J., and Edwards, D. (2016a). Towards plant pangenomics. *Plant Biotechnol. J.* 14, 1099–1105. doi: 10.1111/pbi.12499
- Golicz, A. A., Bayer, P. E., Barker, G. C., Edger, P. P., Kim, H., Martinez, P. A., et al. (2016b). The pangenome of an agronomically important crop plant *Brassica oleracea*. *Nat. Commun.* 7, 1–8. doi: 10.1038/ncomms13390
- Gouil, Q., and Keniry, A. (2019). Latest techniques to study DNA methylation. *Essays Biochem.* 63, 639–648. doi: 10.1042/EBC20190027
- Groeneveld, L. F., Lenstra, J. A., Eding, H., Toro, M. A., Scherf, B., Pilling, D., et al. (2010). Genetic diversity in farm animals – a review. *Anim. Genet.* 41, 6–31. doi: 10.1111/j.1365-2052.2010.02038.x
- Guerra García, A., and Piñero, D. (2017). Current approaches and methods in plant domestication studies. *Bot. Sci.* 95:345. doi: 10.17129/botsci.1209
- Guerra-García, A., Suárez-Atilano, M., Mastretta-Yanes, A., Delgado-Salinas, A., and Piñero, D. (2017). Domestication genomics of the open-pollinated scarlet runner bean (*Phaseolus coccineus* L.). *Front. Plant Sci.* 8:1891. doi: 10.3389/fpls.2017.01891
- Guerrero-Bosagna, C. (2012). Finalism in darwinian and lamarckian evolution: lessons from epigenetics and developmental biology. *Evol. Biol.* 39, 283–300. doi: 10.1007/s11692-012-9163-x
- Günther, T., and Nettelblad, C. (2019). The presence and impact of reference bias on population genomic studies of prehistoric human populations. *PLoS Genet.* 15:e1008302. doi: 10.1371/journal.pgen.1008302
- Guo, S., Zhang, J., Sun, H., Salse, J., Lucas, W. J., Zhang, H., et al. (2012). The draft genome of watermelon (*Citrullus lanatus*) and resequencing of 20 diverse accessions. *Nat. Genet.* 45, 51–58. doi: 10.1038/ng.2470
- Guo, Z., Song, G., Liu, Z., Qu, X., Chen, R., Jiang, D., et al. (2015). Global epigenomic analysis indicates that Epialleles contribute to Allele-specific expression via Allele-specific histone modifications in hybrid rice. *BMC Genomics* 16:232. doi: 10.1186/s12864-015-1454-z
- Gutenkunst, R. N., Hernandez, R. D., Williamson, S. H., and Bustamante, C. D. (2009). Inferring the joint demographic history of multiple populations from multidimensional SNP frequency data. *PLoS Genet.* 5:e1000695. doi: 10.1371/journal.pgen.1000695
- Haas, B. J., Papanicolaou, A., Yassour, M., Grabherr, M., Blood, P. D., Bowden, J., et al. (2013). De novo transcript sequence reconstruction from RNA-seq using the Trinity platform for reference generation and analysis. *Nat. Protoc.* 8, 1494–1512. doi: 10.1038/nprot.2013.084
- Hahn, F., Eisenhut, M., Mantegazza, O., and Weber, A. (2017). Generation of targeted knockout mutants in *Arabidopsis thaliana* using CRISPR/Cas9. *Bio Protoc.* 7, 1–20. doi: 10.21769/BioProtoc.2384
- Hahn, M. W., and Hibbins, M. S. (2019). A three-sample test for introgression. *Mol. Biol. Evol.* 36, 2878–2882. doi: 10.1093/molbev/msz178
- Han, L., and Abney, M. (2013). Using identity by descent estimation with dense genotype data to detect positive selection. *Eur. J. Hum. Genet.* 21, 205–211. doi: 10.1038/ejhg.2012.148
- Hanchard, N. A., Rockett, K. A., Spencer, C., Coop, G., Pinder, M., Jallow, M., et al. (2006). Screening for recently selected alleles by analysis of human haplotype similarity. *Am. J. Hum. Genet.* 78, 153–159. doi: 10.1086/499252
- He, S., Yan, S., Wang, P., Zhu, W., Wang, X., Shen, Y., et al. (2014). Comparative analysis of genome-wide chromosomal histone modification patterns in maize cultivars and their wild relatives. *PLoS One* 9:e97364. doi: 10.1371/journal.pone.0097364
- He, X.-J., Chen, T., and Zhu, J.-K. (2011). Regulation and function of DNA methylation in plants and animals. *Cell Res.* 21, 442–465. doi: 10.1038/cr.2011.23
- Heard, E., and Martienssen, R. A. (2014). Transgenerational epigenetic inheritance: myths and mechanisms. *Cell* 157, 95–109. doi: 10.1016/j.cell.2014.02.045
- Hekman, J. P., Johnson, J. L., and Kukekova, A. V. (2015). Transcriptome analysis in domesticated species: challenges and strategies. *Bioinform. Biol. Insights* 9(Suppl. 4), 21–31. doi: 10.4137/BBI.S29334
- Heled, J., and Drummond, A. J. (2008). Bayesian inference of population size history from multiple loci. *BMC Evol. Biol.* 8:289. doi: 10.1186/1471-2148-8-289
- Hradilová, I., Trnina, O., Váľková, M., Cechová, M., Janská, A., Prokešová, L., et al. (2017). A combined comparative transcriptomic, metabolomic, and anatomical analyses of two key domestication traits: pod dehiscence and seed dormancy in pea (*Pisum* sp.). *Front. Plant Sci.* 8:542. doi: 10.3389/fpls.2017.00542
- Huang, X., Kurata, N., Wei, X., Wang, Z.-X., Wang, A., Zhao, Q., et al. (2012). A map of rice genome variation reveals the origin of cultivated rice. *Nature* 490, 497–501. doi: 10.1038/nature11532
- Hübner, S., Bercovich, N., Todesco, M., Mandel, J. R., Odenheimer, J., Ziegler, E., et al. (2019). Sunflower pan-genome analysis shows that hybridization altered gene content and disease resistance. *Nat. Plants* 5, 54–62. doi: 10.1038/s41477-018-0329-0
- Hufford, M. B., Xu, X., van Heerwaarden, J., Pyhäjärvi, T., Chia, J.-M., Cartwright, R. A., et al. (2012). Comparative population genomics of maize domestication and improvement. *Nat. Genet.* 44, 808–811. doi: 10.1038/ng.2309
- Hurgobin, B., Golicz, A. A., Bayer, P. E., Chan, C. K. K., Tirnaz, S., Dolatabadian, A., et al. (2018). Homoeologous exchange is a major cause of gene presence/absence variation in the amphidiploid *Brassica napus*. *Plant Biotechnol. J.* 16, 1265–1274. doi: 10.1111/pbi.12867
- Ibarra-Laclette, E., Lyons, E., Hernández-Guzmán, G., Pérez-Torres, C. A., Carretero-Paulet, L., Chang, T. H., et al. (2013). Architecture and evolution of a minute plant genome. *Nature* 498, 94–98. doi: 10.1038/nature12132
- Ilska, J., Haskell, M. J., Blott, S. C., Sánchez-Molano, E., Polgar, Z., Lofgren, S. E., et al. (2017). Genetic characterization of dog personality traits. *Genetics* 206, 1101–1111. doi: 10.1534/genetics.116.192674
- Inbar, S., Cohen, P., Yahav, T., and Privman, E. (2020). Comparative study of population genomic approaches for mapping colony-level traits. *PLoS Comput. Biol.* 16:e1007653. doi: 10.1371/journal.pcbi.1007653
- Irving-Pease, E. K., Ryan, H., Jamieson, A., Dimopoulos, E. A., Larson, G., and Frantz, L. A. F. (2019). “Paleogenomics of animal domestication,” in *Paleogenomics: Genome-Scale Analysis of Ancient DNA*, eds C. Lindqvist and O. P. Rajora (Cham: Springer International Publishing), 225–272. doi: 10.1007/13836_2018_55
- Janowitz Koch, I., Clark, M. M., Thompson, M. J., Deere-Machemer, K. A., Wang, J., Duarte, L., et al. (2016). The concerted impact of domestication and

- transposon insertions on methylation patterns between dogs and grey wolves. *Mol. Ecol.* 25, 1838–1855. doi: 10.1111/mec.13480
- Janzen, G. M., Wang, L., and Hufford, M. B. (2019). The extent of adaptive wild introgression in crops. *New Phytol.* 221, 1279–1288. doi: 10.1111/nph.15457
- Jensen, P. (2015). Adding “epi-” to behaviour genetics: implications for animal domestication. *J. Exp. Biol.* 218, 32–40. doi: 10.1242/jeb.106799
- Jiang, L. G., Li, B., Liu, S. X., Wang, H. W., Li, C. P., Song, S. H., et al. (2019). Characterization of proteome variation during modern maize breeding. *Mol. Cell. Proteomics* 18, 263–276. doi: 10.1074/mcp.RA118.001021
- Jiang, Y., Jiang, Y., Wang, S., Zhang, Q., and Ding, X. (2019). Optimal sequencing depth design for whole genome re-sequencing in pigs. *BMC Bioinform.* 20:556. doi: 10.1186/s12859-019-3164-z
- Jiao, Y., Zhao, H., Ren, L., Song, W., Zeng, B., Guo, J., et al. (2012). Genome-wide genetic changes during modern breeding of maize. *Nat. Genet.* 44, 812–815. doi: 10.1038/ng.2312
- Jombart, T., Devillard, S., and Balloux, F. (2010). Discriminant analysis of principal components: a new method for the analysis of genetically structured populations. *BMC Genet.* 11:94. doi: 10.1186/1471-2156-11-94
- Jorde, L. B. (2001). Population genomics: a bridge from evolutionary history to genetic medicine. *Hum. Mol. Genet.* 10, 2199–2207. doi: 10.1093/hmg/10.20.2199
- Kantar, M. B., Nashoba, A. R., Anderson, J. E., Blackman, B. K., and Rieseberg, L. H. (2017). The genetics and genomics of plant domestication. *Bioscience* 67, 971–982. doi: 10.1093/biosci/bix114
- Kaur, P., and Gaikwad, K. (2017). From genomes to GENE-omes: exome sequencing concept and applications in crop improvement. *Front. Plant Sci.* 8:2164. doi: 10.3389/fpls.2017.02164
- Khan, A. W., Garg, V., Roorikiwal, M., Golicz, A. A., Edwards, D., and Varshney, R. K. (2020). Super-pangenome by integrating the wild side of a species for accelerated crop improvement. *Trends. Plant Sci.* 25, 148–158. doi: 10.1016/j.tplants.2019.10.012
- Khan, M. A., Olsen, K. M., Sovero, V., Kushad, M. M., and Korban, S. S. (2014). Fruit quality traits have played critical roles in domestication of the apple. *Plant Genome* 7, 1–18. doi: 10.3835/plantgenome2014.04.0018
- Kim, Y., and Nielsen, R. (2004). Linkage disequilibrium as a signature of selective sweeps. *Genetics* 167, 1513–1524. doi: 10.1534/genetics.103.025387
- Koenig, D., Jimenez-Gomez, J. M., Kimura, S., Fulop, D., Chitwood, D. H., Headland, L. R., et al. (2013). Comparative transcriptomics reveals patterns of selection in domesticated and wild tomato. *Proc. Natl. Acad. Sci. U.S.A.* 110, E2655–E2662. doi: 10.1073/pnas.1309606110
- Korneliussen, T. S., Albrechtsen, A., and Nielsen, R. (2014). ANGSD: analysis of next generation sequencing data. *BMC Bioinform.* 15:356. doi: 10.1186/s12859-014-0356-4
- Korneliussen, T. S., Moltke, I., Albrechtsen, A., and Nielsen, R. (2013). Calculation of Tajima's D and other neutrality test statistics from low depth next-generation sequencing data. *BMC Bioinform.* 14:289. doi: 10.1186/1471-2105-14-289
- Larson, G., Piperno, D. R., Allaby, R. G., Purugganan, M. D., Andersson, L., Arroyo-Kalin, M., et al. (2014). Current perspectives and the future of domestication studies. *Proc. Natl. Acad. Sci. U.S.A.* 111, 6139–6146. doi: 10.1073/pnas.1323964111
- Lelieveld, S. H., Spielmann, M., Mundlos, S., Veltman, J. A., and Gilissen, C. (2015). Comparison of exome and genome sequencing technologies for the complete capture of protein-coding regions. *Hum. Mutat.* 36, 815–822. doi: 10.1002/humu.22813
- Lemmon, Z. H., Bukowski, R., Sun, Q., and Doebley, J. F. (2014). The role of *Cis* regulatory evolution in maize domestication. *PLoS Genet.* 10:e1004745. doi: 10.1371/journal.pgen.1004745
- Levy, S. E., and Myers, R. M. (2016). Advancements in next-generation sequencing. *Annu. Rev. Genomics Hum. Genet.* 17, 95–115. doi: 10.1146/annurev-genom-083115-022413
- Li, H., and Durbin, R. (2011). Inference of human population history from individual whole-genome sequences. *Nature* 475, 493–496. doi: 10.1038/nature10231
- Li, R., Fu, W., Su, R., Tian, X., Du, D., Zhao, Y., et al. (2019). Towards the complete goat pan-genome by recovering missing genomic segments from the reference genome. *Front. Genet.* 10:1169. doi: 10.3389/fgene.2019.01169
- Li, Y., Von Holdt, B. M., Reynolds, A., Boyko, A. R., Wayne, R. K., Wu, D. D., et al. (2013). Artificial selection on brain-expressed genes during the domestication of dog. *Mol. Biol. Evol.* 30, 1867–1876. doi: 10.1093/molbev/mst088
- Li, Y., Zhou, G., Ma, J., Jiang, W., Jin, L., Zhang, Z., et al. (2014). De novo assembly of soybean wild relatives for pan-genome analysis of diversity and agronomic traits. *Nat. Biotechnol.* 32, 1045–1052. doi: 10.1038/nbt.2979
- Li, Z., Chen, J., Wang, Z., Pan, Y., Wang, Q., Xu, N., et al. (2016). Detection of selection signatures of population-specific genomic regions selected during domestication process in Jinhua pigs. *Anim. Genet.* 47, 672–681. doi: 10.1111/age.12475
- Librado, P., Fages, A., Gaunitz, C., Leonardi, M., Wagner, S., Khan, N., et al. (2016). The evolutionary origin and genetic makeup of domestic horses. *Genetics* 204, 423–434. doi: 10.1534/genetics.116.194860
- Linck, E., and Battey, C. J. (2019). Minor allele frequency thresholds strongly affect population structure inference with genomic data sets. *Mol. Ecol. Resour.* 19, 639–647. doi: 10.1111/1755-0998.12995
- Liu, X., and Fu, Y.-X. (2015). Exploring population size changes using SNP frequency spectra. *Nat. Genet.* 47, 555–559. doi: 10.1038/ng.3254
- Lotterhos, K. E., and Whitlock, M. C. (2014). Evaluation of demographic history and neutral parameterization on the performance of F_{ST} outlier tests. *Mol. Ecol.* 23, 2178–2192. doi: 10.1111/mec.12725
- Lotterhos, K. E., and Whitlock, M. C. (2015). The relative power of genome scans to detect local adaptation depends on sampling design and statistical method. *Mol. Ecol.* 24, 1031–1046. doi: 10.1111/mec.13100
- Lowry, D. B., Hoban, S., Kelley, J. L., Lotterhos, K. E., Reed, L. K., Antolin, M. F., et al. (2017). Breaking RAD: an evaluation of the utility of restriction site-associated DNA sequencing for genome scans of adaptation. *Mol. Ecol. Resour.* 17, 142–152. doi: 10.1111/1755-0998.12635
- Luo, G.-Z., Blanco, M. A., Greer, E. L., He, C., and Shi, Y. (2015). DNA N6-methyladenine: a new epigenetic mark in eukaryotes? *Nat. Rev. Mol. Cell Biol.* 16, 705–710. doi: 10.1038/nrm4076
- Luu, K., Bazin, E., and Blum, M. G. B. (2017). pcadapt?: an R package to perform genome scans for selection based on principal component analysis. *Mol. Ecol. Resour.* 17, 67–77. doi: 10.1111/1755-0998.12592
- Lye, Z. N., and Purugganan, M. D. (2019). Copy number variation in domestication. *Trends Plant Sci.* 24, 352–365. doi: 10.1016/j.tplants.2019.01.003
- Machaj, G., Bostan, H., Macko-Podgórn, A., Iorizzo, M., and Grzebelus, D. (2018). Comparative transcriptomics of root development in wild and cultivated carrots. *Genes (Basel)*. 9:431. doi: 10.3390/genes9090431
- Maples, B. K., Gravel, S., Kenny, E. E., and Bustamante, C. D. (2013). RFMix: a discriminative modeling approach for rapid and robust local-ancestry inference. *Am. J. Hum. Genet.* 93, 278–288. doi: 10.1016/j.ajhg.2013.06.020
- Martin, S. H., Davey, J. W., and Jiggins, C. D. (2015). Evaluating the use of ABBA-BABA statistics to locate introgressed loci. *Mol. Biol. Evol.* 32, 244–257. doi: 10.1093/molbev/msu269
- Mascher, M., Gundlach, H., Himmelbach, A., Beier, S., Twardziok, S. O., Wicker, T., et al. (2017). A chromosome conformation capture ordered sequence of the barley genome. *Nature* 544, 427–433. doi: 10.1038/nature22043
- Mascher, M., Schuenemann, V. J., Davidovich, U., Marom, N., Himmelbach, A., Hübner, S., et al. (2016). Genomic analysis of 6,000-year-old cultivated grain illuminates the domestication history of barley. *Nat. Genet.* 48, 1089–1093. doi: 10.1038/ng.3611
- Mastretta-Yanes, A., Arrigo, N., Alvarez, N., Jorgensen, T. H., Piñero, D., and Emerson, B. C. (2015). Restriction site-associated DNA sequencing, genotyping error estimation and de novo assembly optimization for population genetic inference. *Mol. Ecol. Resour.* 15, 28–41. doi: 10.1111/1755-0998.12291
- Mather, N., Traves, S. M., and Ho, S. Y. (2020). A practical introduction to sequentially Markovian coalescent methods for estimating demographic history from genomic data. *Ecol. Evol.* 10, 579–589. doi: 10.1002/ece3.5888
- Mathew, B., Léon, J., and Sillanpää, M. J. (2018). A novel linkage-disequilibrium corrected genomic relationship matrix for SNP-heritability estimation and genomic prediction. *Heredity (Edinb)*. 120, 356–368. doi: 10.1038/s41437-017-0023-4
- Mazet, O., Rodríguez, W., and Chikhi, L. (2015). Demographic inference using genetic data from a single individual: separating population size variation from population structure. *Theor. Popul. Biol.* 104, 46–58. doi: 10.1016/j.tpb.2015.06.003

- Meirmans, P. G. (2015). Seven common mistakes in population genetics and how to avoid them. *Mol. Ecol.* 24, 3223–3231. doi: 10.1111/mec.13243
- Meissner, A. (2005). Reduced representation bisulfite sequencing for comparative high-resolution DNA methylation analysis. *Nucleic Acids Res.* 33, 5868–5877. doi: 10.1093/nar/gki901
- Meyer, R. S., Choi, J. Y., Sanches, M., Plessis, A., Flowers, J. M., Amas, J., et al. (2016). Domestication history and geographical adaptation inferred from a SNP map of African rice. *Nat. Genet.* 48, 1083–1088. doi: 10.1038/ng.3633
- Meyer, R. S., and Purugganan, M. D. (2013). Evolution of crop species: genetics of domestication and diversification. *Nat. Rev. Genet.* 14, 840–852. doi: 10.1038/nrg3605
- Money, D., Migicovsky, Z., Gardner, K., and Myles, S. (2017). LinkImputeR: user-guided genotype calling and imputation for non-model organisms. *BMC Genomics* 18:523. doi: 10.1186/s12864-017-3873-5
- Montenegro, J. D., Golicz, A. A., Bayer, P. E., Hurgobin, B., Lee, H., Chan, C. K. K., et al. (2017). The pangenome of hexaploid bread wheat. *Plant Journal* 90, 1007–1013. doi: 10.1111/tpj.13515
- Moyers, B. T., Morrell, P. L., and McKay, J. K. (2018). Genetic costs of domestication and improvement. *J. Hered.* 109, 103–116. doi: 10.1093/jhered/esx069
- Muir, P., Li, S., Lou, S., Wang, D., Spakowicz, D. J., Salichos, L., et al. (2016). The real cost of sequencing: scaling computation to keep pace with data generation. *Genome Biol.* 17:53. doi: 10.1186/s13059-016-0917-0
- Myers, C. L., Springer, N. M., Schaefer, R., Ross-Ibarra, J., Swanson-Wagner, R., Tiffin, P., et al. (2012). Reshaping of the maize transcriptome by domestication. *Proc. Natl. Acad. Sci. U.S.A.* 109, 11878–11883. doi: 10.1073/pnas.1201961109
- Nadachowska-Brzyska, K., Burri, R., Smeds, L., and Ellegren, H. (2016). PSMC analysis of effective population sizes in molecular ecology and its application to black-and-white *Ficedula* flycatchers. *Mol. Ecol.* 25, 1058–1072. doi: 10.1111/mec.13540
- Ndjiondjop, M. N., Alachiotis, N., Pavlidis, P., Goungoulou, A., Kpeki, S. B., Zhao, D., et al. (2019). Comparisons of molecular diversity indices, selective sweeps and population structure of African rice with its wild progenitor and Asian rice. *Theor. Appl. Genet.* 132, 1145–1158. doi: 10.1007/s00122-018-3268-2
- Nei, M., and Maruyama, T. (1975). Lewontin-Krakauer test for neutral genes. *Genetics* 80:395.
- Nielsen, R., and Beaumont, M. A. (2009). Statistical inferences in phylogeography. *Mol. Ecol.* 18, 1034–1047. doi: 10.1111/j.1365-294X.2008.04059.x
- Nowoshilow, S., Schloissnig, S., Fei, J. F., Dahl, A., Pang, A. W., Pippel, M., et al. (2018). The axolotl genome and the evolution of key tissue formation regulators. *Nature* 554, 50–55. doi: 10.1038/nature25458
- Ottoni, C., Van Neer, W., De Cupere, B., Daligault, J., Guimaraes, S., Peters, J., et al. (2017). The palaeogenetics of cat dispersal in the ancient world. *Nat. Ecol. Evol.* 1:0139. doi: 10.1038/s41559-017-0139
- Pankin, A., Altmüller, J., Becker, C., and von Korff, M. (2018). Targeted resequencing reveals genomic signatures of barley domestication. *New Phytol.* 218, 1247–1259. doi: 10.1111/nph.15077
- Paterson, A. H., Lander, E. S., Hewitt, J. D., Peterson, S., Lincoln, S. E., and Tanksley, S. D. (1988). Resolution of quantitative traits into Mendelian factors by using a complete linkage map of restriction fragment length polymorphisms. *Nature* 335, 721–726. doi: 10.1038/335721a0
- Patterson, N., Price, A. L., and Reich, D. (2006). Population structure and eigenanalysis. *PLoS Genet.* 2:e190. doi: 10.1371/journal.pgen.0020190
- Pavlidis, P., and Alachiotis, N. (2017). A survey of methods and tools to detect recent and strong positive selection. *J. Biol. Res.* 24:7. doi: 10.1186/s40709-017-0064-0
- Pickrell, J. K., and Pritchard, J. K. (2012). Inference of population splits and mixtures from genome-wide allele frequency data. *PLoS Genet.* 8:e1002967. doi: 10.1371/journal.pgen.1002967
- Piperno, D. R. (2017). Assessing elements of an extended evolutionary synthesis for plant domestication and agricultural origin research. *Proc. Natl. Acad. Sci. U.S.A.* 114, 6429–6437. doi: 10.1073/pnas.1703658114
- Pitt, D., Sevana, N., Nicolazzi, E. L., MacHugh, D. E., Park, S. D. E., Colli, L., et al. (2019). Domestication of cattle: two or three events? *Evol. Appl.* 12, 123–136. doi: 10.1111/eva.12674
- Prezeworski, M., Coop, G., and Wall, J. D. (2005). The signature of positive selection on standing genetic variation. *Evolution* 59, 2312–2323. doi: 10.1111/j.0014-3820.2005.tb00941.x
- Price, A. L., Tandon, A., Patterson, N., Barnes, K. C., Rafaels, N., Ruczinski, I., et al. (2009). Sensitive detection of chromosomal segments of distinct ancestry in admixed populations. *PLoS Genet.* 5:e1000519. doi: 10.1371/journal.pgen.1000519
- Pritchard, J. K., Stephens, M., and Donnelly, P. (2000). Inference of population structure using multilocus genotype data. *Genetics* 155, 945–959. doi: 10.1111/j.1471-8286.2007.01758.x
- Qi, J., Liu, X., Shen, D., Miao, H., Xie, B., Li, X., et al. (2013). A genomic variation map provides insights into the genetic basis of cucumber domestication and diversity. *Nat. Genet.* 45, 1510–1515. doi: 10.1038/ng.2801
- Qiu, Q., Wang, L., Wang, K., Yang, Y., Ma, T., Wang, Z., et al. (2015). Yak whole-genome resequencing reveals domestication signatures and prehistoric population expansions. *Nat. Commun.* 6:10283. doi: 10.1038/ncomms10283
- Raj, A., Stephens, M., and Pritchard, J. K. (2014). fastSTRUCTURE: variational inference of population structure in large SNP data sets. *Genetics* 197, 573–589. doi: 10.1534/genetics.114.164350
- Ramos-Madrugal, J., Smith, B. D., Moreno-Mayar, J. V., Gopalakrishnan, S., Ross-Ibarra, J., Gilbert, M. T. P., et al. (2016). Genome sequence of a 5,310-year-old maize cob provides insights into the early stages of maize domestication. *Curr. Biol.* 26, 3195–3201. doi: 10.1016/j.cub.2016.09.036
- Ranwez, V., Serra, A., Pot, D., and Chantret, N. (2017). Domestication reduces alternative splicing expression variations in sorghum. *PLoS One* 12:e0183454. doi: 10.1371/journal.pone.0183454
- Renaut, S., and Rieseberg, L. H. (2015). The accumulation of deleterious mutations as a consequence of domestication and improvement in sunflowers and other compositae crops. *Mol. Biol. Evol.* 32, 2273–2283. doi: 10.1093/molbev/msv106
- Rodríguez-Mega, E., Piñeyro-Nelson, A., Gutierrez, C., García-Ponce, B., Sánchez, M. D. L. P., Zluhan-Martínez, E., et al. (2015). Role of transcriptional regulation in the evolution of plant phenotype: a dynamic systems approach. *Dev. Dyn.* 244, 1074–1095. doi: 10.1002/dvdy.24268
- Ross-Ibarra, J., Morrell, P. L., and Gaut, B. S. (2007). Plant domestication, a unique opportunity to identify the genetic basis of adaptation. *Proc. Natl. Acad. Sci. U.S.A.* 104, 8641–8648. doi: 10.1073/pnas.0700643104
- Rozas, J., Ferrer-Mata, A., Sánchez-DelBarrio, J. C., Guirao-Rico, S., Librado, P., Ramos-Onsins, S. E., et al. (2017). DnaSP 6: DNA sequence polymorphism analysis of large data sets. *Mol. Biol. Evol.* 34, 3299–3302. doi: 10.1093/molbev/msx248
- Sabeti, P. C., Reich, D. E., Higgins, J. M., Levine, H. Z. P., Richter, D. J., Schaffner, S. F., et al. (2002). Detecting recent positive selection in the human genome from haplotype structure. *Nature* 419, 832–837. doi: 10.1038/nature01140
- Sabeti, P. C., Vargill, P., Fry, B., Lohmueller, J., Hostetter, E., Cotsapas, C., et al. (2007). Genome-wide detection and characterization of positive selection in human populations. *Nature* 449, 913–918. doi: 10.1038/nature06250
- Sakurada, K. (2010). Environmental epigenetic modifications and reprogramming-recalcitrant genes. *Stem Cell Res.* 4, 157–164. doi: 10.1016/j.scr.2010.01.001
- Sawyer, S., Krause, J., Guschanski, K., Savolainen, V., and Pääbo, S. (2012). Temporal patterns of nucleotide misincorporations and DNA fragmentation in ancient DNA. *PLoS One* 7:e34131. doi: 10.1371/journal.pone.0034131
- Sax, K. (1923). The association of size differences with seed-coat pattern and pigmentation in *Phaseolus vulgaris*. *Genetics* 8, 552–560.
- Schiffels, S., and Durbin, R. (2014). Inferring human population size and separation history from multiple genome sequences. *Nat. Genet.* 46, 919–925. doi: 10.1038/ng.3015
- Schlötterer, C., Tobler, R., Kofler, R., and Nolte, V. (2014). Sequencing pools of individuals – mining genome-wide polymorphism data without big funding. *Nat. Rev. Genet.* 15, 749–763. doi: 10.1038/nrg3803
- Schmidt, D., Wilson, M. D., Spyrou, C., Brown, G. D., Hadfield, J., and Odom, D. T. (2009). ChIP-seq: using high-throughput sequencing to discover protein–DNA interactions. *Methods* 48, 240–248. doi: 10.1016/j.ymeth.2009.03.001
- Schmitz, R. J., Schultz, M. D., Urich, M. A., Nery, J. R., Pelizzola, M., Libiger, O., et al. (2013). Patterns of population epigenomic diversity. *Nature* 495, 193–198. doi: 10.1038/nature11968
- Schmutz, J., McClean, P. E., Mamidi, S., Wu, G. A., Cannon, S. B., Grimwood, J., et al. (2014). A reference genome for common bean and genome-wide analysis of dual domestications. *Nat. Genet.* 46, 707–713. doi: 10.1038/ng.3008
- Schnable, P. S., Ware, D., Fulton, R. S., Stein, J. C., Wei, F., Pasternak, S., et al. (2009). The B73 maize genome: complexity, diversity, and dynamics. *Science* 326, 1112–1115. doi: 10.1126/science.1178534

- Schreiber, M., Stein, N., and Mascher, M. (2018). Genomic approaches for studying crop evolution. *Genome Biol.* 19:140. doi: 10.1186/s13059-018-1528-8
- Schubert, M., Jónsson, H., Chang, D., Der Sarkissian, C., Ermini, L., Ginolhac, A., et al. (2014). Prehistoric genomes reveal the genetic foundation and cost of horse domestication. *Proc. Natl. Acad. Sci. U.S.A.* 111, E5661–E5669. doi: 10.1073/pnas.1416991111
- Schurch, N. J., Schofield, P., Gierliński, M., Cole, C., Sherstnev, A., Singh, V., et al. (2016). How many biological replicates are needed in an RNA-seq experiment and which differential expression tool should you use? *RNA* 22, 839–851. doi: 10.1261/rna.053959.115
- Schweiger, W., Boddu, J., Shin, S., Poppenberger, B., Berthiller, F., Lemmens, M., et al. (2010). Validation of a candidate deoxynivalenol-inactivating UDP-glucosyltransferase from barley by heterologous expression in yeast. *Mol. Plant Microbe Interact.* 23, 977–986. doi: 10.1094/MPMI-23-7-0977
- Seal, A., Gupta, A., Mahalaxmi, M., Aykhal, R., Singh, T. R., and Arunachalam, V. (2014). Tools, resources and databases for SNPs and indels in sequences: a review. *Int. J. Bioinform. Res. Appl.* 10:264. doi: 10.1504/IJBRA.2014.060762
- Shafer, A. B. A., Peart, C. R., Tusso, S., Maayan, I., Brelsford, A., Wheat, C. W., et al. (2017). Bioinformatic processing of RAD-seq data dramatically impacts downstream population genetic inference. *Methods Ecol. Evol.* 8, 907–917. doi: 10.1111/2041-210X.12700
- Shalem, O., Sanjana, N. E., Hartenian, E., Shi, X., Scott, D. A., Mikkelsen, T. S., et al. (2014). Genome-scale CRISPR-Cas9 knockout screening in human cells. *Science* 343, 84–87. doi: 10.1126/science.1247005
- Shan, S., Soltis, P. S., Soltis, D. E., and Yang, B. (2020). Considerations in adapting CRISPR/Cas9 in nongenetic model plant systems. *Appl. Plant Sci.* 8:e11314. doi: 10.1002/aps3.11314
- Shang, Y., Ma, Y., Zhou, Y., Zhang, H., Duan, L., Chen, H., et al. (2014). Biosynthesis, regulation, and domestication of bitterness in cucumber. *Science* 346, 1084–1088. doi: 10.1126/science.1259215
- Shen, Y., Zhang, J., Liu, Y., Liu, S., Liu, Z., Duan, Z., et al. (2018). DNA methylation footprints during soybean domestication and improvement. *Genome Biol.* 19:128. doi: 10.1186/s13059-018-1516-z
- Shi, J., and Lai, J. (2015). Patterns of genomic changes with crop domestication and breeding. *Curr. Opin. Plant Biol.* 24C, 47–53. doi: 10.1016/j.pbi.2015.01.008
- Sims, D., Sudbery, I., Illott, N. E., Heger, A., and Ponting, C. P. (2014). Sequencing depth and coverage: key considerations in genomic analyses. *Nat. Rev. Genet.* 15, 121–132. doi: 10.1038/nrg3642
- Singh, J., Zhao, J., and Vallejos, C. E. (2018). Differential transcriptome patterns associated with early seedling development in a wild and a domesticated common bean (*Phaseolus vulgaris* L.) accession. *Plant Sci.* 274, 153–162. doi: 10.1016/j.plantsci.2018.05.024
- Smith, J. M., and Haigh, J. (2007). The hitch-hiking effect of a favourable gene. *Genet. Res. (Camb)* 89, 391–403. doi: 10.1017/S0016672308009579
- Smith, O., Nicholson, W. V., Kistler, L., Mace, E., Clapham, A., Rose, P., et al. (2019). A domestication history of dynamic adaptation and genomic deterioration in Sorghum. *Nat. Plants* 5, 369–379. doi: 10.1038/s41477-019-0397-9
- Sohn, J., and Nam, J.-W. (2016). The present and future of de novo whole-genome assembly. *Brief. Bioinform.* 19:bbw096. doi: 10.1093/bib/bbw096
- Song, J., Li, J., Sun, J., Hu, T., Wu, A., Liu, S., et al. (2018). Genome-wide association mapping for cold tolerance in a core collection of rice (*Oryza sativa* L.) landraces by using high-density single nucleotide polymorphism markers from specific-locus amplified fragment sequencing. *Front. Plant Sci.* 9:875. doi: 10.3389/fpls.2018.00875
- Song, Q., Zhang, T., Stelly, D. M., and Chen, Z. J. (2017). Epigenomic and functional analyses reveal roles of epialleles in the loss of photoperiod sensitivity during domestication of allotetraploid cottons. *Genome Biol.* 18:99. doi: 10.1186/s13059-017-1229-8
- Soyk, S., Lemmon, Z. H., Oved, M., Fisher, J., Liberatore, K. L., Park, S. J., et al. (2017). Bypassing negative epistasis on yield in tomato imposed by a domestication gene. *Cell* 169, 1142–1155. doi: 10.1016/j.cell.2017.04.032
- Sun, H., Wu, S., Zhang, G., Jiao, C., Guo, S., Ren, Y., et al. (2017). Karyotype stability and unbiased fractionation in the paleo-allotetraploid cucurbita genomes. *Mol. Plant* 10, 1293–1306. doi: 10.1016/j.molp.2017.09.003
- Swinnen, G., Goossens, A., and Pauwels, L. (2016). Lessons from domestication: targeting Cis-regulatory elements for crop improvement. *Trends Plant Sci.* 21, 506–515. doi: 10.1016/j.tplants.2016.01.014
- Tajima, F. (1989). Statistical method for testing the neutral mutation hypothesis by DNA polymorphism. *Genetics* 123, 585–595.
- Tang, H., Coram, M., Wang, P., Zhu, X., and Risch, N. (2006). Reconstructing genetic ancestry blocks in admixed individuals. *Am. J. Hum. Genet.* 79, 1–12. doi: 10.1086/504302
- Tettelin, H., Masignani, V., Cieslewicz, M. J., Donati, C., Medini, D., Ward, N. L., et al. (2005). Genome analysis of multiple pathogenic isolates of *Streptococcus agalactiae*: implications for the microbial “pan-genome.”. *Proc. Natl. Acad. Sci. U.S.A.* 102, 13950–13955. doi: 10.1073/pnas.0506758102
- Thornton, T. A., and Bermejo, J. L. (2014). Local and global ancestry inference and applications to genetic association analysis for admixed populations. *Genet. Epidemiol.* 38, S5–S12. doi: 10.1002/gepi.21819
- Tiffin, P., and Ross-Ibarra, J. (2014). Advances and limits of using population genetics to understand local adaptation. *Trends Ecol. Evol.* 29, 673–680. doi: 10.1016/j.tree.2014.10.004
- Tomato Genome Consortium (2012). The tomato genome sequence provides insights into fleshy fruit evolution. *Nature* 485:635. doi: 10.1038/nature11119
- Trerotola, M., Relli, V., Simeone, P., and Alberti, S. (2015). Epigenetic inheritance and the missing heritability. *Hum. Genomics* 9:17. doi: 10.1186/s40246-015-0041-3
- Trucchi, E., Gratton, P., Whittington, J. D., Cristofari, R., Le Maho, Y., Stenseth, N. C., et al. (2014). King penguin demography since the last glaciation inferred from genome-wide data. *Proc. R. Soc. B Biol. Sci.* 281:20140528. doi: 10.1098/rspb.2014.0528
- Turner, B. M. (2009). Epigenetic responses to environmental change and their evolutionary implications. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* 364, 3403–3418. doi: 10.1098/rstb.2009.0125
- Ueta, R., Abe, C., Watanabe, T., Sugano, S. S., Ishihara, R., Ezura, H., et al. (2017). Rapid breeding of parthenocarpic tomato plants using CRISPR/Cas9. *Sci. Rep.* 7:507. doi: 10.1038/s41598-017-00501-4
- Urich, M. A., Nery, J. R., Lister, R., Schmitz, R. J., and Ecker, J. R. (2015). MethyC-seq library preparation for base-resolution whole-genome bisulfite sequencing. *Nat. Protoc.* 10:475. doi: 10.1038/nprot.2014.114
- Vallebuena-Estrada, M., Rodríguez-Arévalo, I., Rougon-Cardoso, A., Martínez González, J., García Cook, A., Montiel, R., et al. (2016). The earliest maize from San Marcos Tehuacán is a partial domesticate with genomic evidence of inbreeding. *Proc. Natl. Acad. Sci. U.S.A.* 113, 14151–14156. doi: 10.1073/pnas.1609701113
- VanBuren, R., Wai, C. M., Colle, M., Wang, J., Sullivan, S., Bushakra, J. M., et al. (2018). A near complete, chromosome-scale assembly of the black raspberry (*Rubus occidentalis*) genome. *Gigascience* 7, 1–9. doi: 10.1093/gigascience/giy094
- Varshney, R. K., Thudi, M., Roorkiwal, M., He, W., Upadhyaya, H. D., Yang, W., et al. (2019). Resequencing of 429 chickpea accessions from 45 countries provides insights into genome diversity, domestication and agronomic traits. *Nat. Genet.* 51, 857–864. doi: 10.1038/s41588-019-0401-3
- Vasemägi, A., Nilsson, J., McGinnity, P., Cross, T., O'Reilly, P., Glebe, B., et al. (2012). Screen for footprints of selection during domestication/captive breeding of atlantic salmon. *Comp. Funct. Genomics* 2012, 1–14. doi: 10.1155/2012/628204
- Velasco, D., Hough, J., Aradhya, M., and Ross-Ibarra, J. (2016). Evolutionary genomics of peach and almond domestication. *G3* 6, 3985–3993. doi: 10.1534/g3.116.032672
- Visscher, P. M., Wray, N. R., Zhang, Q., Sklar, P., McCarthy, M. I., Brown, M. A., et al. (2017). 10 years of GWAS discovery: biology, function, and translation. *Am. J. Hum. Genet.* 101, 5–22. doi: 10.1016/j.ajhg.2017.06.005
- Vitti, J. J., Grossman, S. R., and Sabeti, P. C. (2013). Detecting natural selection in genomic data. *Annu. Rev. Genet.* 47, 97–120. doi: 10.1146/annurev-genet-111212-133526
- Voight, B. F., Kudaravalli, S., Wen, X., and Pritchard, J. K. (2006). A map of recent positive selection in the human genome. *PLoS Biol.* 4:e72. doi: 10.1371/journal.pbio.0040072
- Wales, N., Akman, M., Watson, R. H. B., Sánchez Barreiro, F., Smith, B. D., Gremillion, K. J., et al. (2019). Ancient DNA reveals the timing and persistence of organellar genetic bottlenecks over 3,000 years of sunflower domestication and improvement. *Evol. Appl.* 12, 38–53. doi: 10.1111/eva.12594
- Wales, N., Ramos Madrigal, J., Cappellini, E., Carmona Baez, A., Samaniego Castruita, J. A., Romero-Navarro, J. A., et al. (2016). The limits and potential

- of paleogenomic techniques for reconstructing grapevine domestication. *J. Archaeol. Sci.* 72, 57–70. doi: 10.1016/j.jas.2016.05.014
- Walley, J. W., Sartor, R. C., Shen, Z., Schmitz, R. J., Wu, K. J., Urlich, M. A., et al. (2016). Integration of omic networks in a developmental atlas of maize. *Science* 353, 814–818. doi: 10.1126/science.aag1125
- Wang, E. T., Kodama, G., Baldi, P., and Moyzis, R. K. (2006). Global landscape of recent inferred Darwinian selection for *Homo sapiens*. *Proc. Natl. Acad. Sci. U.S.A.* 103, 135–140. doi: 10.1073/pnas.0509691102
- Wang, G.-D., Xie, H.-B., Peng, M.-S., Irwin, D., and Zhang, Y.-P. (2014). Domestication genomics: evidence from animals. *Annu. Rev. Anim. Biosci.* 2, 65–84. doi: 10.1146/annurev-animal-022513-114129
- Wang, M., Yan, J., Zhao, J., Song, W., Zhang, X., Xiao, Y., et al. (2012). Genome-wide association study (GWAS) of resistance to head smut in maize. *Plant Sci.* 196, 125–131. doi: 10.1016/j.plantsci.2012.08.004
- Wang, W., Feng, B., Xiao, J., Xia, Z., Zhou, X., Li, P., et al. (2014). Cassava genome from a wild ancestor to cultivated varieties. *Nat. Commun.* 5:5110. doi: 10.1038/ncomms6110
- Wang, W., Zhang, X., Zhou, X., Zhang, Y., La, Y., Zhang, Y., et al. (2019). Deep genome resequencing reveals artificial and natural selection for visual deterioration, plateau adaptability and high prolificacy in chinese domestic sheep. *Front. Genet.* 10:300. doi: 10.3389/fgene.2019.00300
- Warr, A., Robert, C., Hume, D., Archibald, A., Deeb, N., and Watson, M. (2015). Exome sequencing: current and future perspectives. *G3* 5, 1543–1550. doi: 10.1534/g3.115.018564
- Weber, M., Davies, J. J., Wittig, D., Oakeley, E. J., Haase, M., Lam, W. L., et al. (2005). Chromosome-wide and promoter-specific analyses identify sites of differential DNA methylation in normal and transformed human cells. *Nat. Genet.* 37, 853–862. doi: 10.1038/ng1598
- Weigel, D., and Colot, V. (2012). Epialleles in plant evolution. *Genome Biol.* 13, 249. doi: 10.1186/gb-2012-13-10-249
- Wiener, P., and Pong-Wong, R. (2011). A regression-based approach to selection mapping. *J. Hered.* 102, 294–305. doi: 10.1093/jhered/esr014
- Wigginton, J. E., Cutler, D. J., and Abecasis, G. R. (2005). A note on exact tests of hardy-weinberg equilibrium. *Am. J. Hum. Genet.* 76, 887–893. doi: 10.1086/429864
- Wilkinson, S., Lu, Z. H., Megens, H.-J., Archibald, A. L., Haley, C., Jackson, I. J., et al. (2013). Signatures of diversifying selection in european pig breeds. *PLoS Genet.* 9:e1003453. doi: 10.1371/journal.pgen.1003453
- Wolter, F., Schindele, P., and Puchta, H. (2019). Plant breeding at the speed of light: the power of CRISPR/Cas to generate directed genetic diversity at multiple sites. *BMC Plant Biol.* 19:176. doi: 10.1186/s12870-019-1775-1
- Wu, D.-D., Ding, X.-D., Wang, S., Wójcik, J. M., Zhang, Y., Tokarska, M., et al. (2018). Pervasive introgression facilitated domestication and adaptation in the *Bos* species complex. *Nat. Ecol. Evol.* 2, 1139–1145. doi: 10.1038/s41559-018-0562-y
- Xie, M., Chung, C. Y.-L., Li, M., Wong, F.-L., Wang, X., Liu, A., et al. (2019). A reference-grade wild soybean genome. *Nat. Commun.* 10:1216. doi: 10.1038/s41467-019-09142-9
- Yandell, M., and Ence, D. (2012). A beginner's guide to eukaryotic genome annotation. *Nat. Rev. Genet.* 13, 329–342. doi: 10.1038/nrg3174
- Yang, C. J., Samayoa, L. F., Bradbury, P. J., Olukolu, B. A., Xue, W., York, A. M., et al. (2019). The genetic architecture of teosinte catalyzed and constrained maize domestication. *Proc. Natl. Acad. Sci. U.S.A.* 116, 5643–5652. doi: 10.1073/pnas.1820997116
- Yang, I. S., and Kim, S. (2015). Analysis of whole transcriptome sequencing data: workflow and software. *Genomics Inform.* 13:119. doi: 10.5808/GL.2015.13.4.119
- Yang, L., Koo, D. H., Li, Y., Zhang, X., Luan, F., Havey, M. J., et al. (2012). Chromosome rearrangements during domestication of cucumber as revealed by high-density genetic mapping and draft genome assembly. *Plant J.* 71, 895–906. doi: 10.1111/j.1365-3113X.2012.05017.x
- Yang, X., Zhang, H., Shang, J., Liu, G., Xia, T., Zhao, C., et al. (2018). Comparative analysis of the blood transcriptomes between wolves and dogs. *Anim. Genet.* 49, 291–302. doi: 10.1111/age.12675
- Zadesenets, K. S., and Rubtsov, N. B. (2018). Genome duplication in animal evolution. *Russ. J. Genet.* 54, 1125–1136. doi: 10.1134/S1022795418090168
- Zeder, M. A. (2006). Central questions in the domestication of plants and animals. *Evol. Anthropol. Issues News Rev.* 15, 105–117. doi: 10.1002/evan.20101
- Zeder, M. A., Emshwiller, E., Smith, B. D., and Bradley, D. G. (2006). Documenting domestication: the intersection of genetics and archaeology. *Trends Genet.* 22, 139–155. doi: 10.1016/j.tig.2006.01.007
- Zeng, K., Fu, Y.-X., Shi, S., and Wu, C.-I. (2006). Statistical tests for detecting positive selection by utilizing high-frequency variants. *Genetics* 174, 1431–1439. doi: 10.1534/genetics.106.061432
- Zeng, L., Tu, X.-L., Dai, H., Han, F.-M., Lu, B.-S., Wang, M.-S., et al. (2019). Whole genomes and transcriptomes reveal adaptation and domestication of pistachio. *Genome Biol.* 20:79. doi: 10.1186/s13059-019-1686-3
- Zhang, C., Bailey, D. K., Awad, T., Liu, G., Xing, G., Cao, M., et al. (2006). A whole genome long-range haplotype (WGLRH) test for detecting imprints of positive selection in human populations. *Bioinformatics* 22, 2122–2128. doi: 10.1093/bioinformatics/btl365
- Zhang, H., Zhang, J., Lang, Z., Botella, J. R., and Zhu, J.-K. (2017). Genome editing—principles and applications for functional genomics research and crop improvement. *CRC. Crit. Rev. Plant Sci.* 36, 291–309. doi: 10.1080/07352689.2017.1402989
- Zhao, B. S., Roundtree, I. A., and He, C. (2017). Post-transcriptional gene regulation by mRNA modifications. *Nat. Rev. Mol. Cell. Biol.* 18, 31–42. doi: 10.1038/nrm.2016.132
- Zhao, Q., Feng, Q., Lu, H., Li, Y., Wang, A., Tian, Q., et al. (2018). Pan-genome analysis highlights the extent of genomic variation in cultivated and wild rice. *Nat. Genet.* 50, 278–284. doi: 10.1038/s41588-018-0041-z
- Zheng, J., Wu, H., Zhu, H., Huang, C., Liu, C., Chang, Y., et al. (2019). Determining factors, regulation system, and domestication of anthocyanin biosynthesis in rice leaves. *New Phytol.* 223, 705–721. doi: 10.1111/nph.15807
- Zhou, J., Li, D., Wang, G., Wang, F., Kunjal, M., Joldersma, D., et al. (2019). Application and future perspective of CRISPR/Cas9 genome editing in fruit crops. *J. Integr. Plant Biol.* 62, 269–286. doi: 10.1111/jipb.12793
- Zhou, Y., Minio, A., Massonnet, M., Solares, E., Lv, Y., Beridze, T., et al. (2019). The population genetics of structural variants in grapevine domestication. *Nat. Plants* 5, 965–979. doi: 10.1038/s41477-019-0507-8

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2020 Barrera-Redondo, Piñero and Eguarte. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.