4

# Accelerating Crop Domestication in the Era of Gene Editing

*Angel Del Valle-Echevarria*
Hawaii Agricultural Research Center, Kunia, HI, USA

*Nathan Fumia*
University of Hawaii at Manoa, Honolulu, HI, USA

*Michael A. Gore*
Plant Breeding and Genetics Section, School of Integrative Plant Science, Cornell University, Ithaca, NY, USA

*Michael Kantar*
University of Hawaii at Manoa, Honolulu, HI, USA



## ABSTRACT

Through plant agriculture, humans have modified their natural environment to produce food, fiber, fuel, and medicine. These products are the result of artificial selection that favors the accumulation of desirable phenotypes over time. This conscious, or subconscious, selection is initiated by the phenomenon of domestication, which is a well-described process with distinct stages on a continuum. While desirable phenotypes are seemingly constant across wide phylogenetic distances, the molecular basis of the phenotypic changes is often not. As the amount of genomic information has increased, the knowledge regarding the genetic basis of domestication in multiple plant species has become more accessible. The development of better phenotypic measurement tools, more sophisticated crossing schemes, and new statistical methods has helped to link genes to domestication traits. New molecular technologies such

*Plant Breeding Reviews, Volume 45*, First Edition. Edited by Irwin Goldman.

as the application of CRISPR-Cas on plants have greatly facilitated the ability to recreate domestication phenotypes faster than traditional breeding. Breeders can use new selection techniques to rapidly bring new species into use for ever-changing human needs.

**KEYWORDS:** Machine learning, CRISPR, domestication syndrome, lost crop, de novo domestication

## OUTLINE



## ABBREVIATIONS

| | |
|---|---|
| CRISPR-Cas | *C*lustered *R*egularly *I*nterspaced *S*hort *P*alindromic *R*epeats-CRISPR-*as*sociated nucleases |
| gRNA | guide RNA |
| crRNA | CRISPR RNA |
| tracrRNA | *trans*-CRISPR RNA |
| pegRNA | prime editing guide RNA |
| RNP | ribonucleoprotein |
| DSB | double-strand breaks |
| NHEJ | non-homologous end-joining |
| HR | homologous recombination |
| GBLUP | Genomic best linear unbiased prediction |

## I. INTRODUCTION

Plants provide a key means for people to relate to their environment, and the origins of domesticated plants provide insights into cultural traditions of human society. Domesticated plants (e.g. crops) have been incorporated in diverse production systems, and these systems have been customized to exploit a type of crop (e.g., cereals) or a specific crop (e.g., apple), in defined geographies for specific outcomes (Meyer et al. 2012; Kurashima et al. 2019). Exploring crop origins has a rich history (De Candolle 1890; Vavilov 1926; Harlan 1971) that has remained of interest for both understanding evolution and use in breeding (Hufford et al. 2012; Baute et al. 2015; Chen et al. 2020). Domestication is a well-known process with observable steps that proceed at different rates in different organisms (Kantar et al. 2017). Key phenotypic traits termed the "domestication syndrome" (Harlan et al. 1973) have been characterized across a wide range of species, and while the syndrome is widely generalizable, the genetic basis and consequences can be different (Meyer and Purugganan 2013; Allaby et al. 2019). Human migrations have helped extend domestic crop ranges and expand the crop ecological niche (Fuller et al. 2014; Purugganan 2019). Technological packages (e.g., chemical fertilizers, mechanization, draft animals, and improved seed) made the adoption of some crops more prevalent in different eras and led to their greater expansion (e.g., maize).

Modern breeding has further refined the domestication syndrome (Vaughan et al. 2007), as during improvement, it became clear that most traits of importance to humans are complex, varying in a continuous rather than discrete fashion. In response, breeders have changed the approach to selection by focusing on statistical approaches and more recently, prediction approaches (Wallace et al. 2018; Ramstein et al. 2019; Bernardo 2020) for the agronomically and economically important complex traits that often have a strong interaction with the environment. As the number of phenotypes has been restricted, so has the number of genotypes (Ross-Ibarra et al. 2007). The concepts of selection, genetic drift, and linkage disequilibrium (Table 4.1) help define the genomic changes that occur during domestication (Zohary 1999; Wright et al. 2005; Morrell and Clegg 2007; Allaby et al. 2008; Yamasaki et al. 2008; Brown 2010). Much work has been done exploring the convergence of domestication phenotypes and the genetic basis of these phenotypes across species (Meyer and Purugganan 2013; Kantar et al. 2017). Since the advent of widespread genomic data, a new theme has emerged: exploring the genomic impacts of domestication and mutations that are related to syndrome traits. These efforts have

**Table 4.1** Definitions of terms important to domestication.

| *Term* | *Definition* |
|---|---|
| Domestication | Domestication is the process by which humans select desirable qualities in plants or animals to make them more useful to humans and dependent on human intervention for persistence. |
| Domestication syndrome | The "domestication syndrome" consists of agriculturally important traits altered to achieve a domestic form (Harlan 1992). These traits do not always have a similar genetic basis, but they do have phenotypic similarities (Vaughn et al. 2007; Weeden 2007). The domestication syndrome includes increased seed size, changes in timing of flowering, greater flowering synchrony, decreased inflorescences per plant, decreased stature, loss of defensive structures (barbs, thorns, chemical production), loss of seed dispersal mechanisms, increased fruit size, and more predictable germination (loss of seed dormancy) (Harlan 1992; Koinange et al. 1996; Sang 2009). |
| Wild progenitor | Species from which a domesticated species was selected. |
| Landrace | Domesticate that has most domestication syndrome traits fixed but has not undergone modern breeding. |
| Cultural diffusion | The human dispersal of technology across a wide geographic area. |
| Natural selection | Process by which favorable traits that increase reproductive success increase in frequency in populations. This is accompanied by a change in allele frequency. |
| Artificial selection | Process in which individual or small groups of organisms containing favorable characteristics are chosen by humans to establish subsequent generations (Falconer and Mackay 1996). |
| Genetic drift | Variation in allele frequencies due to random sampling from small populations. |
| Linkage disequilibrium | Nonrandom association of alleles at two or more loci (Falconer and Mackay 1996). |
| Genetic bottleneck | Bottlenecks occur when rare alleles are lost from populations due to decreases in population size (Nei et al. 1975; Tanksley and McCouch 1997). |
| Gene flow | Transfer of genes between populations or species. |
| Effective population size (Ne) | The number of breeding individuals contributing to the diversity in subsequent generations. |
| Introgression | Transmission of genes between genetic backgrounds or species. |
| Quantitative trait loci | QTL are regions of the genome that control a continuously varying phenotype (Doerge 2002). |
| Monophyletic phylogeny | Where two groups in a relationship tree contain a single most recent common ancestor, implying a single common history. |
| Polyphyletic phylogeny | Where two groups in a relationship tree do not contain the most recent common ancestor, implying multiple histories. |
| Selective sweep | The process of a mutation going from rare to fixation within a population; soft sweeps are from standing variation. |

explored convergent evolution across a wide range of species (Glemin and Bataillon 2009; Purugganan and Fuller 2011) to understand how knowledge of domestication can be used in rapid development of new crops (Chopra et al. 2020). Targeting these genes through both traditional (Chopra et al. 2020) and targeted mutagenesis, both transgenic and transgene free (Chen et al. 2019a), has become a popular method for creating new domesticates. This review aims to describe how plant breeders can leverage these known genomic changes that occurred during domestication in combination with molecular, statistical, and computational approaches to improve species for human needs.

## II. MOLECULAR BIOLOGY IN DOMESTICATING AND IMPROVING NOVEL CROPS

Understanding the molecular basis of domestication is of major interest in plant breeding and genetics since it can aid in speeding-up the breeding process of orphan crops, landraces, and/or wild-crop relatives. Most well-known domesticated traits are conditioned by loss-of-function mutations in single genes that have a major regulatory role in the organism (see Meyer and Purugganan 2013). The increasing number of plant genomes available has made the use of molecular biology tools much more tractable (Chen et al. 2019b). Mutant analysis identified loci and specific mutations responsible for domestication traits (Li et al. 2020). These mutations can be readily introduced to a plant by targeting the domestication-associated gene with the easy-to-use gene editing technique CRISPR-Cas (Wiedenheft et al. 2009; Barrangou et al. 2007).

### A. Gene Editing Using the CRISPR-Cas System

The CRISPR-Cas (*C*lustered *R*egularly *I*nterspaced *S*hort *P*alindromic *R*epeats and the CRISPR-*as*sociated nucleases) system is a prokaryotic adaptive immune system that conveys resistance against viruses and has been well described in the bacterium *Streptococcus pyogenes* (Barrangou et al. 2007; Wiedenheft et al. 2009). In practice, two main features of the CRISPR-Cas system are engineered to edit genomes of interest: target RNA sequence (CRISPR RNA and *trans*-CRISPR RNA duplex (crRNA:tracrRNA) or crRNA only) and Cas nuclease (e.g., Cas9 or Cas12a) (Figure 4.1). Jinek et al. (2012) determined that the crRNA:tracrRNA duplex can be synthesized in a single recombinant molecule termed guide RNA (gRNA). Depending on the CRISPR-Cas

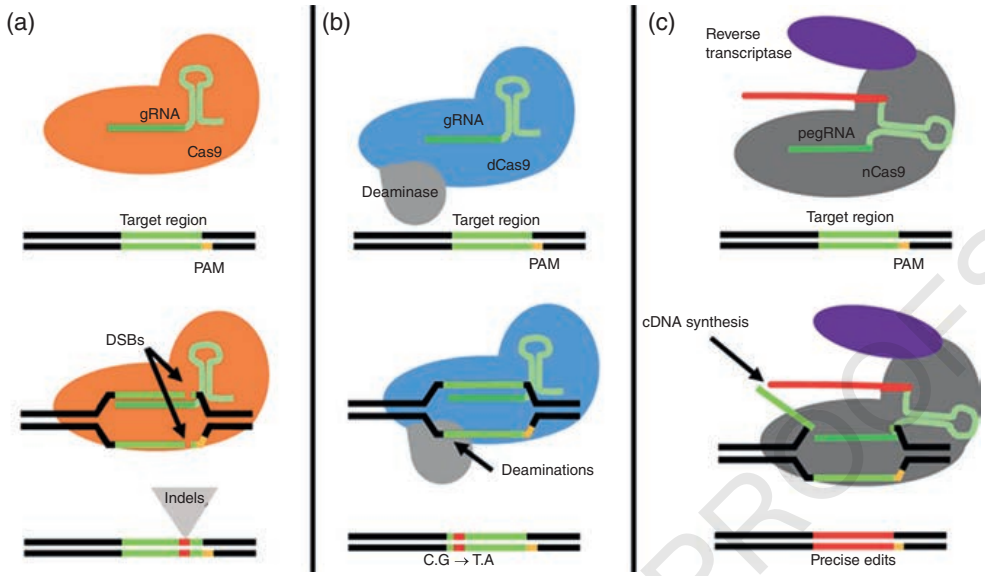


**Fig. 4.1.** The CRISPR-Cas tool for genetic engineering. (Left-A) The conventional CRISPR-Cas9. CRISPR-Cas9 is an RNA-mediated DNA endonuclease that produces double-strand breaks (DSB) in DNA. The ribonucleoprotein (RNP) complex of a gRNA (green) and wild-type Cas9 (orange) survey the DNA for the target sequence that is upstream of a protospacer adjacent motif (PAM; yellow). After cleavage, the cell's DNA repair system fixes the double-strand break resulting in small insertions/deletions. (Middle-B) The base editor system. The base editor system is based on a loss-of-function Cas9 that does not make cuts but binds to DNA with a gRNA (known as dead Cas9 = dCas9; blue) and is fused with a deaminase (gray). Similar to the wild-type Cas9, dCas9 surveys the genome for the target sequence complementary to the gRNA but it does not make any cuts. The deaminase enzyme deaminates cytosines converting them into uracils, which in turn the DNA repair system incorporates thymines. This strategy is used to create specific point mutations. (Right-C) The prime editor system. The prime editor system is based on a mutated Cas9 that cuts only one strand (known as nickase Cas9 = nCas9; dark gray) and is fused with a reverse transcriptase (purple). The gRNA is further modified to include the conventional target (in green) and a template that has the desired mutations (red line) called "prime editing guide RNA" (pegRNA). After one strand is nicked, it pairs with the template side of the pegRNA by complementation, and the reverse transcriptase synthesizes from the nicked strand using the pegRNA as the template (complementary DNA [cDNA]). The resulting molecule has a section replaced with multiple desired mutations (precise edits).



system, the crRNA or the crRNA:tracrRNA (also known as gRNA) bind to Cas12a and Cas9 nucleases, respectively (Jinek et al. 2012; Jao et al. 2013). The Cas nuclease is a RNA-guided DNA nuclease that binds to the crRNA (i.e., Cas12a) or crRNA:tracrRNA, as well as gRNA (i.e., Cas9), forming a ribonucleoprotein (RNP) complex that surveys the genome for short motifs that allow it to bind to the DNA and determine

if there is base complementation between DNA and crRNA (Jinek et al. 2012). These short motifs are called *p*rotospacer *a*djacent *m*otifs (PAM) and are critical for the usage of this tool in eukaryotic organisms (Anzalone et al. 2020). There are two major Cas systems used in plant biology: Cas9 and Cas12a (Jinek et al. 2012; Jao et al. 2013). Cas9 and Cas12a make *d*ouble-*s*trand *b*reaks (DSBs) in their targets by RNA–DNA complementation and have two major differences: PAM sequence and the target RNA to be used (Anzalone et al. 2020). Cas9 recognizes a 5′-NGG-3′ PAM, where 20–24 nucleotides immediately upstream need to be complementary to the crRNA (Jinek et al. 2012). Cas12a, also known as Cpf1, recognizes a 5′-TTTV-3′ (V = C, G, A) PAM, where 22–24 nucleotides immediately downstream of the PAM need to be complementary to the crRNA (Jao et al. 2013). When Cas nucleases catalyze DSBs, it triggers the cell's DNA repair pathway: non-homologous end-joining (NHEJ) or homologous recombination (HR) (reviewed by Anzalone et al. 2020). NHEJ is more common than HR since HR only happens in the presence of a sister chromatid, which is only present in actively dividing cells, or when supplying the cell with single-stranded donor DNA (Cubbon et al. 2018). As stated earlier, most well-known domesticated traits are conditioned by loss-of-function mutations in genes that have major regulatory roles, where the mistake-prone NHEJ DNA repair mechanism suffices in producing a loss-of-function mutation.

## B. Using CRISPR-Cas in Plants

The CRISPR-Cas system is introduced to the plant as an expression plasmid via *Agrobacterium*-mediated transformation, particle bombardment, or protoplast transfection (Belhaj et al. 2015). The CRISPR-Cas transformation system has proven successful in generating edited plants. More importantly, because gene editing occurs outside of the transgene locus, plants that are self-compatible can be self-pollinated to segregate the transgene out while keeping the edited gene in the subsequent generations (Rodriguez-Leal et al. 2017). Fortunately, the CRISPR-Cas system can be also delivered in a DNA-free strategy by synthesizing in vitro the crRNA, tracrRNA, gRNA, and Cas nuclease encoding transcript, as well as by purifying recombinant Cas nucleases from *Escherichia coli* (Woo et al. 2015, Liang et al. 2017; Liang et al. 2018). This latest strategy was developed to address the concerns posed in government regulations regarding the introduction of recombinant DNA in plants (Woo et al. 2015). The DNA-free strategy has yielded desired mutation without any trace of the CRISPR-Cas gene editing module (Woo et al. 2015). Regardless of methodology, most

studies using CRISPR-Cas on plants have focused on proof-of-concept experiments to determine the efficiency of the system in their target plant by disrupting *phytoene desaturase* (*PDS*) and *chlorophyll A oxygenase* (*CAO1*) (Reviewed in Rojas-Vasquez and Gatica-Arias 2020). Loss-of-function mutations on *PDS* and *CAO1* lead to albinism, making it an easy phenotype for quickly scoring editing efficiency and mosaicism (Miao et al. 2013). This strategy has been a standard in protocol development of new crops entering the gene editing realm.

AQ1

CRISPR-Cas has been used in plants to convey resistance to disease by having the CRISPR-Cas module recognize the pathogen and neutralize it (Tripathi et al. 2019), disrupting disease susceptibility genes (Chandrasekaran et al. 2016; Jia et al. 2016; Wang et al. 2016; Peng et al. 2017; Fister et al. 2018; Tashkandi et al. 2018; Bari et al. 2019; Gomez et al. 2019; Mehta et al. 2019; Navet and Tian 2020), modifying crop architecture for productivity enhancement (Zsögön et al. 2018; Shi et al. 2017; Shao et al. 2020), herbicide resistance (Xu et al. 2014; Sun et al. 2016; Hummel et al. 2018; Tian et al. 2018), nutrition enhancement (Sun et al. 2017; Li et al. 2018a–d), and introducing haploids (Che et al. 2018; Wang et al. 2020a, b). Furthermore, CRISPR-Cas is capable of accelerating the breeding process by introducing known causative polymorphisms in elite lines, as well as in wild species and orphan crops to speed up domestication.

## C. Plant Genome Domestication and Breeding Using the CRISPR-Cas System

The best example that clearly shows the application of CRISPR-Cas, as a means to domesticate and improve crops, is the work on tomato (*Solanum lycopersicum L.*) and its relatives (Soyk et al. 2017; Rodriguez-Leal et al. 2017; Lemmon et al. 2018; Kwon et al. 2020). Genes associated with key traits, such as fruit size/shape, inflorescence, and overall plant architecture, are known and have been shown to improve productivity in crops (Meyer and Purugganan 2013). However, in some instances, mixing desired genotypes may result in undesired phenotypes due to negative epistasis and may hinder obtaining the desired output (Chae et al. 2014; Soyk et al. 2017).

In tomato, there are two major loci involved in inflorescence that were selected independently at different times of its domestication: *enhancer of jointless 2* (*ej2*) and *jointless-2* (*j2*). Both loci are attributed to bigger fruit and higher productivity, since in their recessive state both cause less branching and increased fertility independently, but together the opposite phenotype occurs (i.e., more branching, smaller fruits, and infertility). In an elegant series of genetic and genomic

analyses, Soyk et al. (2017) determined the genetic basis of *ej2* and *j2* as two MADS-box genes, both homologs of *Arabidopsis thaliana SEPALLATA4* (*SEP4*), that are involved in tomato meristem maturation. Soyk et al. (2017) identified that the natural alleles had a Copia transposable element in the first intron of *j2* and a 564 nucleotide insertion in the fifth intron of *ej2*, where these insertions produced aberrant splicing transcript variants causing a loss-of-function and weak-function, respectively. To validate these findings, Soyk et al. (2017) used CRISPR-Cas9 targeting both *SEP4* homologs and produced the *j2* and *ej2* phenotypes, confirming their genetic identity. Soyk et al. (2017) evaluated the agronomic potential of weakly branched genotypes for improving flower production and yield by crossing lines that segregated for *j2 ej2*. The *j2 ej2*/+ hybrid lines produced inflorescences with more branches and flowers compared to *j2* control hybrids, resulting in an increased yield by 41–71% and individual fruit weight increased by 19–22%, while sugar content (Brix) remained unchanged (Soyk et al. 2017). This work shows promise that targeted mutagenesis using CRISPR-Cas allows for faster identification of causative genes of domestication traits, flexibility in mixing desired genotypes, and exploiting dosage effects (i.e., homozygous for one locus and heterozygous for another) from selected alleles thus improving inflorescence architecture and yield.

Rodriguez-Leal et al. (2017) is another great example of using the CRISPR-Cas system that exploits the knowledge of domestication genes to generate variation in breeding programs. Unlike the Soyk et al. (2017) study, where natural and CRISPR-mediated mutations altered the protein product and the combination of loss-of-function and weak-function alleles gives rise to variation, the Rodriguez-Leal et al. (2017) CRISPR-Cas strategy focused on altering the regulation of known domestication genes that confer desired traits. The main trait explored in Rodriguez-Leal et al. (2017) was fruit size that is conditioned by the *CLAVATA-WUSHEL* (*CLV-WUS)* complex whose function is to determine meristem size (Somssich et al. 2016). The two major loci involved in fruit size are *fasciated (fas, CLV3)* and *locule number* (*lc, WUS*), where their weak loss-of-function and weak gain-of-function result in bigger fruits (Van der Knaap, et al. 2014). This study targeted the promoter regions of *cis*-regulatory elements (CRE) and other regulatory regions that have been identified by quantitative trait loci (QTL) and genome-wide-association studies (GWAS) analyses. Using CRISPR-Cas9, Rodriguez-Leal et al. (2017) targeted the CArG element upstream of *SlWUS* and 2 kb promoter region immediately upstream of *SlCLV3*, producing the expected *lc* and *fas* phenotypes, respectively. The generation of edited regulatory regions was analyzed further to assess for variation of fruit size traits

(Rodriguez-Leal et al. 2017). Furthermore, the promoter alleles generated via CRISPR-Cas9 of *SlCLV3* generated novel genetic and phenotypic variation (Rodriguez-Leal et al. 2017).

Taking advantage of the *trans*-effect and *trans*-generational inheritance of CRISPR-Cas9, Rodriguez-Leal et al. (2017) made crosses between transgenic hemizygous Cas9 and wild-type (WT) plants. About 24% of plants generated that had novel promoter regions due to CRISPR-Cas9 showed more floral organs than WT, most plants showed weaker effects, and fewer were similar to *fas* or stronger (Rodriguez-Leal et al. 2017). As the authors state in this study (Rodriguez-Leal et al. 2017): "These findings demonstrate the power of combining meiotically heritable Cas9-gRNA activity with a sensitized background to efficiently engineer numerous cis-regulatory alleles with readily observable phenotypic consequences." This approach may help breeders increase the germplasm diversity in elite lines without the need of introducing exotic germplasm to acquire desirable traits.

### D. Transgene-free Generation of Edited Plants via the CRISPR-Cas System

As mentioned previously, the CRISPR-Cas system can be delivered to the plant in the traditional DNA format (i.e., transgenic) or DNA-free. Transgenic gene edited crops that are self-compatible can easily be selected to maintain the desired edited genotype while removing the transgene by simple self-crossing due to the *trans*-effect of Cas9. However, "one size does not fit all" if a crop is desired to be gene edited and transgene-free. First, not all crops are self-compatible, which makes developing transgene-free gene edited plants using this strategy quite laborious. Second, not all crops have short life cycles, which makes the generation time to obtain a transgene-free gene edited plant unfeasible. Third, not all crops are propagated by seed, as some are propagated vegetatively (i.e., cassava) or the variety is sterile (i.e., banana). Lastly, in regard to US agriculture exports, not all products will be regulated the same way, potentially hindering trade. For example, the United States regulates the product while the European Union regulates the process; therefore, a transgene-free gene edited plant would be considered a genetically modified organism (GMO) in the European Union market and would be subjected to a costly and lengthy process of regulation (Pammer and Heller 2010; Callaway 2018; Ledford 2019) but would be able to market in the United States without issues, as long as it does not have the transgene. Taking into consideration these issues, several protocols have been released that do not rely on transgene constructs

to produce gene edited plants, but the same components are produced in vitro and delivered to the plant cells (Woo et al. 2015; Liang et al. 2017; Liang et al. 2018). These protocols provide opportunities for both plant breeding and domestication. One approach that has been suggested is to remove mildly deleterious mutations by creating multiplexed CRISPR constructs in a transgene-free protocol.

### E. Engineered Variants of the CRISPR-Cas System

Several modifications have been implemented on *S. pyogenes* Cas9, as well as the discovery of CRISPR-Cas systems in other prokaryotes, to produce specific mutations in the target sequence and/or increase the chance of modification (Klompe et al. 2019). For example, recombinant loss-of-function Cas9 (referred to as "dead Cas9" or "dCas9") with a deaminase is commonly used to make specific nucleotide changes in the sequence, without DNA breaks, referred to as "base editors" (Komor et al. 2016; Gaudelli et al. 2017). The usage of base editors in crop improvement is reviewed in detail in Mishra et al. (2020). An example of the recent use of base editors to introduce a desirable trait was herbicide resistance in maize and rice (Li et al. 2020; Kuang et al. 2020). Sulfonylurea herbicides are widely used in fields to control weeds (Brown 1990) because it inhibits a key enzyme in amino acid biosynthesis, *ACETOLACTATE SYNTHASE* (*ALS*) (Bernasconi et al. 1995). In maize, *ZmALS1* and *ZmALS2* were targeted by cytosine base editors (C to T) and quantified resistance against sulfonylurea herbicides (Li et al. 2020). Homozygous *ZmALS1* mutants and *ZmALS1 ZmALS2* double mutants showed herbicide resistance up to 15X the recommended dosage (Li et al. 2020). A similar strategy was done in rice, using a gRNA library targeting different regions in *OsALS1* and transformed cells went through sulfonylurea selection (Kuang et al. 2020). Rice lines that showed resistance against sulfonylurea were sequenced to determine the causative mutation in *OsALS1* and in parallel identify the gRNA that produced the desired mutation (Kuang et al. 2020). The elite commercial rice cultivar "Nangeng 46" was subjected to base editing using the gRNA identified in the initial experiment, and the resulting regenerated plants had the desired resistance against sulfonylurea herbicide, avoiding the need of extensive backcrossing (Kuang et al. 2020). It is worth noting that due to self-compatibility, maize and rice can be self-pollinated and select progeny that are homozygous for the edited allele while not having the base editor construct that deems it transgenic (Li et al. 2020; Kuang et al. 2020).

Although base editors have key advantages, such as producing gain-of-function and loss-of-function mutations, they have limitations.

Target sites and off-target editing is inherent to Cas9, where target sites are only present upstream of the PAM sequence and these regions might have homology with other unrelated regions in the genome (Mishra et al. 2020). Specific to base editors, the catalytic window where the base edit can be performed is limited (Mishra et al. 2020). To address these issues, the prime editor system has revolutionized precise editing by overcoming the aforementioned limitations (Anzalone et al. 2019). The prime editor is a recombinant nickase Cas9, which only cuts one strand, with a reverse transcriptase (RT) enzyme that binds to a prime editing gRNA (pegRNA), surveying the genome for the target sequence (Anzalone et al. 2019). The novelty of this strategy is that multiple edits can be embedded in the pegRNA (Anzalone et al. 2019). This strategy has been shown to be effective in rice and wheat protoplasts targeting an array of genes that have been edited in the past (Lin et al. 2020). To date, there has not been a crop evaluation after prime editing but the ease of the technique shows promise in crop domestication and improvement. However, prime editors have the potential of being used by plant geneticists and breeders to recapitulate an exact causal variant that had been under selection during the domestication process, allowing it to be introduced into a wild or semiwild population where it did not exist. Furthermore, prime editors may present a more accurate strategy to correct deleterious alleles to a less deleterious state, especially when less deleterious alleles are not available for certain genes in breeding populations (Ramu et al. 2017). This scenario might be analogous to human gene therapy in plants when purging of deleterious alleles is not possible for all loci.

## III. BRINGING IN GENES FROM THE WILD INTO DOMESTICATED CROPS

Plant breeding has two pillars, domestication and improvement (Figure 4.2). Domestication is Mendelian/oligogenic centric with many traits under the control of at most a few large effect genes, while improvement is quantitative genetics centric with traits genetically controlled by many genes of small effect. While the use of CRISPR-Cas has led to the rapid recovery of domestication phenotypes very quickly (e.g., loss of function), quantitative expression can be altered by creating mutations in domestication gene regulatory regions (Zsögön et al. 2017; Lemmon et al. 2018; Zsögön et al. 2018). The biggest breakthrough in rapid plant domestication involved the use of CRISPR-Cas9 in the solanaceous orphan crop groundcherry (*Physalis pruinosa*),

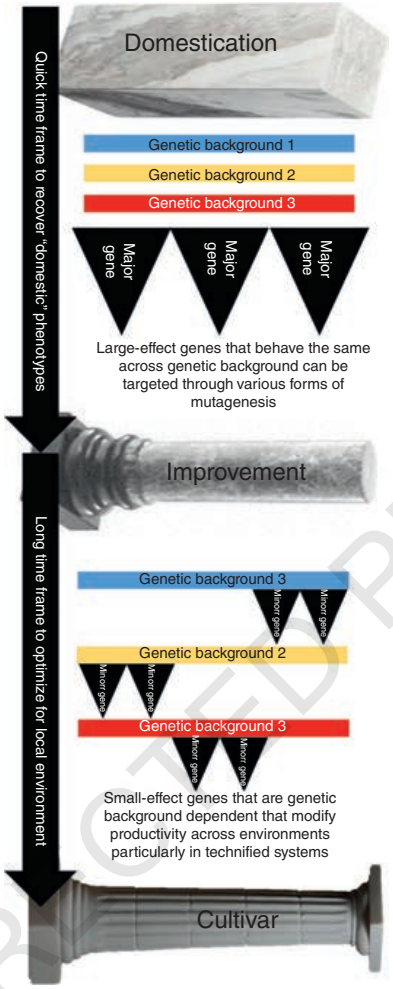


**Fig. 4.2.** The pillars of genetic change in plant breeding are domestication and improvement. These pillars have different leverage points in the breeding process and are amenable to different technologies. In general, the goal during this process is to fix large effect genes to dramatically shift phenotypic means within the species followed by using polygenic standing variation to sculpt form and function to fit local environments and farming practices. Domestication phenotypes have been recovered by modifications in few genes that have large effects that are largely genetic background independent, while improvement is generally controlled by many genes that differ in effect between different environments that are genetic background dependent. Breeding for domestication has been done by creating variation in wild species through random mutagenesis (chemical, radiation, ultraviolet light) and targeted mutagenesis (CRISPR-Cas), by selecting from standing variation using high-throughput phenotyping, or by screening large germplasm collections for variants in domestication genes.

where the disruption of three genes (associated with tomato domestication) led to a phenotypically domesticated groundcherry (Lemmon et al. 2018). In groundcherry they were able to modify domestication syndrome traits (plant architecture, flower number, and fruit size) to rapidly achieve a plant that was more amenable to cultivation in years rather than decades or centuries. The work in groundcherry supports long-term hypotheses that proposed that a subset of genes was under human selection for centuries and are the basis of the present domesticated cultivars.

Crop improvement cannot be done as quickly, as a major factor in improvement is the need to test for genotype-by-environment interaction (Ewing et al. 2019). Crop improvement is in effect genome centric where fine-tuning of the crop to the environments is key to widespread cultivation. New breeding techniques such as marker-assisted selection and genomic selection have helped to improve the speed of crop improvement and the ability to understand and exploit genotype-by-environment interactions (Bernardo 2008; Bernardo 2020). New high-throughput phenotyping methods are being developed in order to gain more data with more precision, both spatially and temporally (Cobb et al. 2013; Rahaman et al. 2015; Pauli et al. 2016).

Linking genomic variation to the environment helps in understanding the history of domestication and provides targets for selection. There are two major ways that are used to identify this variation. The first is bottom-up approaches looking for signals of selection (Tiffin and Ross-Ibarra 2014) and environmental association analysis (Bragg et al. 2015). Bottom-up approaches do not explore the phenotype explicitly but rather look for signals of selection or use climate data as a phenotype and can be done on germplasm collections (e.g., Anderson et al. 2016). This has advantages for identifying targets for breeding, as existing material can be used. The second is top-down approaches such as genetic mapping approaches to connect genotypes to phenotypes (Ross-Ibarra et al. 2007). Top-down approaches have the advantage of being able to explore interactions more easily and estimates of phenotypic variance are more accurate. If collections have been georeferenced, there are many different potential data sources for bioclimatic and soil data for the present day, the past, and the future; some of the most common include Worldclim database (Fick and Hijmans 2017), CHELSA climate data (Karger et al. 2017), and ISRIC soil data (Hengl et al. 2017).

The precise strategy to use once an important trait has been identified as important to either domestication or improvement depends on many factors, including generation time, ploidy, ability to use tissue culture

methods, transformation potential, relationship to elite material, and whether the trait is controlled by a major gene (e.g., marker-assisted backcrossing or CRISPR-Cas) or is quantitative in nature (e.g., genomic selection). New transformation methods that do not rely on tissue culture may provide an opportunity to increase the speed of trait introgression in a wider range of species (Atkins and Voytas 2020). Additionally, it will be possible to increase breeding efficiency through using techniques such as speed breeding (Watson et al. 2018). As these techniques are becoming more trackable in all species, it will be possible to compress breeding timelines in terms of the number of generations of an organism to achieve an elite variety.

## IV. GOING INTO THE UNKNOWN: CAN WE REDOMESTICATE IN A MORE SPECIFIC WAY TO CREATE BETTER CROPS?

Globalization, which has been connecting global food systems (Khoury et al. 2020) and changing societal preferences (Khoury et al. 2015), has led to a greater interest in adapting ancient or underutilized crops to new localities. This has led to the desire to develop methodologies that can quickly adapt semi-domestic crops to different environments and bring undomesticated or abandoned crops into the food system. When using CRISPR for breeding, we are assuming that the locus accounts for all of the variation of the genotype, omitting possible epistasis (genotype-by-genotype interactions) accounting for traits, which need to be obtained by breeding. A potential approach is first selecting for large effect genes (i.e., classic domestication syndrome traits), and then improve these relationships by selecting on small effect genes in different genetic backgrounds (Figure 4.2). This implies coupling many different breeding strategies and being very aware of breeding goals for specific markets. There are many potential ways to use different plant material at different stages in the process of domestication.

There have been many ways proposed to use crop wild relatives. To explore how to use semi-domestic and wild-crafted species in breeding, it is beneficial to use potato (*Solanum tuberosum*) as a case study. The variation among species phylogenetically related to our domesticated crops is large, and the utilization of this vast biodiversity is an established method of crop improvement (Jansky et al. 2013). For example, potato is the fourth most important crop internationally (Castañeda-Álvarez et al. 2016) and has 199 known wild relatives with a large environmental range (Hijmans and Spooner 2001; Hijmans et al. 2002); yet, similar to many crops, cultivated potato (*S. tuberosum*)

lacks genetic diversity (Khoury et al. 2015). However, prioritizing the selection of parental species for use in introgression for crop improvement or hybridization for a new domestication program is not clear. With the ever-changing climate, it is important to develop crop varieties tailored toward forecasted environmental niches. In doing so, large positive genotype-by-environment interactions can be generated, benefiting producers in each environmental niche by harnessing the adaptation to climatic criterion evolved over millions of years.

One potential approach to leverage this variation is the use of an estimated breeding value measure. The estimated breeding value measure could be altered for the use of prioritizing parental species as it does for parental lines in crop improvement programs around the globe (Pironon et al. 2020). This approach would incorporate pertinent biological, evolutionary, and ecological factors into the decision-making process for an optimal parental species for a specific production climate through the use of phylogenetic distance from the crop, biological factors (e.g., endosperm balance number), ecological divergence or conservatism relative to the crop, and discrete climate classification of wild relatives' occurrence. With the ever-growing availability of genomic information, prediction tools can also be used. For example, genomic best linear unbiased prediction (GBLUP) where the random effect of climate dynamics could be predicted (Bernardo 2020) could be used as part of a process to identify the best parent species and the best accessions within species. Recent increase in computing capacity provides the means to utilize these data analytics techniques. Optimizing these approaches would accelerate the use of semi-domesticated plants and operationalize their diverse value to aid in the selection process prior to introgression for crop improvement.

## V. DO CROP MODELS OFFER OPPORTUNITIES FOR ASSISTING IN *DE NOVO* DOMESTICATION OF WILD SPECIES?

There are a number of candidate plant species for *de novo* domestication that can be sourced from wild populations for which there exists opportunities to collect genotypic, phenotypic, and environmental data. These species include legumes (Schlautman et al. 2018; Herron et al. 2020), silphium (Van Tassel et al. 2017), and intermediate wheatgrass (Zhang et al. 2016). Collectively, these candidate species can be prioritized based on criteria related to the consumer and market landscape, but there is still difficulty in identifying which accessions among many of a wild species should be the focus. Should it be those

adapted to a specific locale or having a novel phenotype conferring tolerance to an environmental stress? A possible path forward is the use of process-based crop models that connect environmental variation and plant performance via a mechanistic approach to simulate the potential of traits under different environment and management regimes (Peng et al. 2020). In this context, the implementation of crop models could involve their use as a decision support tool to select candidate accession/species combinations with a certain level of adaptability to enter the *de novo* domestication pipeline. Such a model-assisted selection approach for tapping new domesticates could be especially optimal for perennial plant species with long-generation times. An example of this could be modeling soil and plant hydraulic properties to identify wild accessions of a woody species with favorable drought-adaptive phenotypes for shifting agroecological zones (Wang et al. 2020a). The selected wild accessions of annual or perennial species could then serve as founders for constructing breeding populations, followed by the application of genomic selection combined with crop modeling for the optimization of adaptive traits (Cooper et al. 2016; Technow et al. 2015).

## VI. CAN WE REVIVE LOST DOMESTICATES AND HOW WOULD WE BREED THESE?

Not only are lost crops a void in the archaeological record, but their disappearance could also have implications on food and nutrition security. As an example, even before the adaptation of maize to eastern North America, several domesticates including sumpweed, goosefoot, maygrass, erect knotweed, and little barley comprised the native crop complex (Reviewed in Mueller et al. 2017). All that exists of these five crops now are their remains collected from archaeological sites in eastern North America (Asch and Asch 1985; Powell 2000; Simon and Parker 2006). There lies the opportunity for the use of paleogenomics to reconstruct the evolutionary history of a species through a comparative effort that involves ancient domesticate and extant wild DNA samples (Pont et al. 2019). This is slightly different from the typical analysis of comparing DNA from archaeological early domesticate remains to that of modern domesticates (Ramos-Madrigal et al. 2016; Scott et al. 2019), because in our proposed scenario there are only wild populations within native ranges, and in the worst cases the species is extinct. Such an analysis could provide retrospective insights into candidate domestication loci, serving as targets for CRISPR, especially if considered orthologous to any of the major effect loci considered critical for

rapid crop domestication (Lenser and Theißen 2013). Such abandoned domesticates could also be complimented by the addition of traits that were not present during the first time these plants were domesticated. Another parallel avenue to explore is the utilization of genomic prediction models to help better understand the extent to which the lost domesticates were adapted to their agricultural production systems. Through an approach inspired by Swarts et al. (2017), constructed training and test sets consisting of individuals from extant wild (or semiwild, if needed to facilitate phenotyping) populations would be used to train genomic prediction models for phenotypes related to productivity and local adaptation to generate genomic estimated breeding values (GEBVs) for unobserved phenotypes in the genotyped ancient DNA sample population. This would potentially improve our understanding of ancient phenotypes at the time of domestication and identify phenotypic constraints and targets for the modern-day genomics-assisted breeding process.

## VII. CAN MACHINE LEARNING BE USED TO DETECT DOMESTICATION LOCI?

Statistical inference approaches founded on population genetic models have been the classical workhorses for detecting a selective sweep, which is a genomic signature of low nucleotide diversity produced by a rapid increase in the frequency of an advantageous allele that ultimately becomes fixed due to positive directional selection (Stephan 2019). The wealth of DNA sequence data at the population level has allowed for the identification of many selective sweeps, presumably the consequence of domestication across a number of crop species (Shi and Lai 2015). The most frequently used tests for detecting sweeps, however, can produce inaccurate results when the assumptions of population genetics models are violated (Stephan 2019). Furthermore, simulation methods that use joint likelihood approximation of multiple population genetics summary statistics for detecting positive selection are vulnerable to the "curse of dimensionality" (Lin et al. 2011), which is a data mining problem that results from sparsity of data as the number of dimensions increases. This is precisely the environment where machine learning (ML) classifiers such as decision trees and artificial neural networks (deep learning) would thrive to accurately distinguish selection from neutrality (Schrider and Kern 2018; Koropoulis et al. 2020). Model architecture, hyperparameter selection, and overfitting are some other important considerations when using ML models that impact efficacy (Wang et al. 2020b).

Supervised ML approaches (e.g., support vector machine and random forest) have the flexibility to be trained on simulated data from a range of population genetic models to improve power for selective sweep detection (Lin et al. 2011; Pavlidis et al. 2010; Ronen et al. 2013), with a variant of random forest (supervised learning algorithm for identifying mean regression of many individual decision trees) shown to detect and differentiate hard and soft sweeps in the presence of non-equilibrium demography (Schrider and Kern 2016). Simulated data from population genetic models covering a range of evolutionary and demographic scenarios will continue to be needed for training ML models, as many of the population genetic parameters underlying selective sweeps are not estimated with high precision (Schrider and Kern 2018). Although an ML model for either categorical (classification) or continuous (regression) variables can be initially trained for a single crop species, there is potential to apply the same model to closely related species that share a similar population history. Also, it could be used as a pre-trained model in a transfer learning approach to limit the amount of training data needed for implementation in other crop species more distant at the evolutionary and population genetic levels (Wang et al. 2020b).

## VIII. CONCLUSION AND FUTURE DIRECTIONS

As the field of domestication moves forward, more nuance within the process is continuing to come to light. This offers insight into the way artificial selection has shaped genomes and the ways individual genes have been changed. This knowledge creates many selection targets in many different types of plants. Combining different domestication phenotypic syndromes uses a combination of plant breeding approaches. When the phenotypic goals (e.g., agronomic, production, and nutritional) have been defined, many tools are available to put these traits into the plants that people use. We outlined a model for *de novo* domestication (using both genome editing and ML techniques) of fixing large effect genes in order to dramatically change plant form and function, followed by selecting for polygenic standing variation present in appropriate genetic backgrounds to further sculpt phenotypes in order to refine the plant form during improvement. Domestication has occurred across a multitude of species in many different geographies, a process that has led to many different outcomes. When new domesticates that are viewed as superior arrive, other plants often are abandoned. By using advanced phenotyping and genomics, it is possible

to rapidly advance breeding populations to make ancient domesticates more amenable to modern practices and to breed for as yet unimagined agroecosystems.

## IX. ACKNOWLEDGMENTS

We would like to thank the Hawaii Agriculture Research Center (HARC) Sustainable Agriculture Fellowship for supporting Nathan Fumia, and we would like to thank Cornell University for supporting the sabbatical of Dr. Michael A. Gore to contribute to this manuscript. We thank Savanah Dale for comments on an earlier version of the manuscript.

## LITERATURE CITED

Allaby, R.G., D.Q. Fuller, and T.A. Brown. 2008. The genetic expectations of a protracted model for the origins of domesticated crops. *Proc. Natl. Acad. Sci.* 105:13982–13986.

Allaby, R. G., R.L. Ware, and L. Kistler. 2019. A re-evaluation of the domestication bottleneck from archaeogenomic evidence. *Evol. Appl. 12*(1):29–37.

Anderson, J.E., T.J. Kono, R.M. Stupar, M.B. Kantar, and P. Morrell. 2016. Environmental association analyses identify candidates for abiotic stress tolerance in Glycine soja, the wild progenitor of cultivated soybeans. *G3: Genes. Genomes, Genetics* 6:835–843.

Anzalone, A.V., L.W. Koblan, and D.R. Liu. 2020. Genome editing with CRISPR-Cas nucleases, base editors, transposases and prime editors. *Nat. Biotechnol.* 38:824–844.

Anzalone, A.V., P.B. Randolph, J.R. Davis, A.A. Sousa, L.W. Koblan, J.M. Levy, P.J. Chen, C. Wilson, G.A. Newby, A. Raguram, and D.R. Liu. 2019. Search-and-replace genome editing without double-strand breaks or donor DNA. *Nature* 576:149–157.

Asch, D.L. and N.B. Asch. (1985). Prehistoric plant cultivation in west-central Illinois. p. 149–203. In Ford, R.I. (ed.), *Prehistoric Food Production in North America*, University of Michigan Museum, Ann Arbor.

Atkins, P.A. and D.F. Voytas. 2020. Overcoming bottlenecks in plant gene editing. *Curr. Opin. Plant Biol.* 54:79–84.

Bari, V.K., J.A. Nassar, S.M. Kheredin, A. Gal-On, M. Ron, A. Britt, D. Steele, J. Yoder, and R. Aly. 2019. CRISPR/Cas9-mediated mutagenesis of CAROTENOID CLEAVAGE DIOXYGENASE 8 in tomato provides resistance against the parasitic weed Phelipanche aegyptiaca. *Sci. Rep.* 9(1):1–12.

Barrangou, R., C. Fremaux, H. Deveau, M. Richards, P. Boyaval, S. Moineau, D.A. Romero, and P. Horvath. 2007. CRISPR provides acquired resistance against viruses in prokaryotes. *Science* 315:1709–1712.

Baute G.J., Kane, N.C., Grassa, C., Lai, Z. and L.H. Rieseberg. 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

Belhaj, K., A. Chaparro-Garcia, S. Kamoun, N.J. Patron, and V. Nekrasov. 2015. Editing plant genomes with CRISPR/Cas9. *Curr. Opin. Biotechnol.* 32:76–84.

Bernasconi, P., A.R. Woodworth, B.A. Rosen, M.V. Subramanian, and D.L. Siehl. 1995. A naturally occurring point mutation confers broad range tolerance to herbicides that target acetolactate synthase. *J. Biol. Chem.* 270:17381–17385.

Bernardo, R. 2008. Molecular markers and selection for complex traits in plants: learning from the last 20 years. *Crop Sci.* 48:1649–1664.

Bernardo, R. 2020. Reinventing quantitative genetics for plant breeding: something old, something new, something borrowed, something BLUE. *Heredity:*1–11.

Bragg, J.G., M.A. Supple, R.L. Andrew, and J.O. Borevitz. 2015. Genomic variation across landscapes: insights and applications. *New Phytol.* 207:953–967.

Brown, H.M. 1990. Mode of action, crop selectivity, and soil relations of the sulfonylurea herbicides. *Pestic. Sci.* 29:263–281.

Brown A.H.D. 2010. Variation under domestication in plants: 1859 and today. *Philos. Trans. R. Soc. B* 365:2523–2530.

Callaway, E. (2018). CRISPR plants now subject to tough GM laws in European Union. *Nature* 560(7716):16–17.

Castañeda-Álvarez, N. P., C.K. Khoury, H. A. Achicanoy, V. Bernau, H. Dempewolf, R.J. Eastwood, L. Guarino, R.H. Harker, A. Jarvis, N. Maxted, J.V. Müller, J. Ramirez-Villegas, C.C. Sosa, P.C. Struik, H. Vincent, and J. Toll 2016. Global conservation priorities for crop wild relatives. *Nat. Plants* 2:1–6.

Chae, E., K. Bomblies, S.T. Kim, D. Karelina, M. Zaidem, S. Ossowski, C. Martín-Pizarro, R.A. Laitinen, B.A. Rowan, H. Tenenboim, and S. Lechner. 2014. Species-wide genetic incompatibility analysis identifies immune genes as hot spots of deleterious epistasis. *Cell* 159:1341–1351.

Chandrasekaran, J., M. Brumin, D. Wolf, D. Leibman, C. Klap, M. Pearlsman, A. Sherman, T. Arazi, A. Gal-On (2016). Development of broad virus resistance in non-transgenic cucumber using CRISPR/Cas9 technology. *Mol. Plant Pathol. 17*(7):1140–1153.

Che, P. A. Anand, E. Wu, J.D. Sander, M.K. Simon, W. Zhu, A.L. Sigmund, G. Zastrow-Hayes, M. Miller, D. Liu, S.J. Lawit, Z.-Y. Zhao, M.C. Albertsen, T.J. Jones. (2018). Developing a flexible, high-efficiency Agrobacterium-mediated sorghum transformation system with broad application. *Plant Biotechnol. J.* 16(7):1388–1395.

Chen, Q., L.F. Samayoa, C.J. Yang, P.J. Bradbury, B.A. Olukolu, M.A. Neumeyer, M.C. Romay, Q. Sun, A. Lorant, E.S. Buckler, and J. Ross-Ibarra. 2020. The genetic architecture of the maize progenitor, teosinte, and how it was altered during maize domestication. *PLoS Genet.* 16:e1008791.

Chen, K., Y. Wang, R. Zhang, H. Zhang, and C. Gao. 2019a. CRISPR/Cas genome editing and precision plant breeding in agriculture. *Annu. Rev. Plant Biol.* 70:667–697.

Chen, F., Y. Song, X. Li, J. Chen, L. Mo, X. Zhang, Z. Lin, and L. Zhang. 2019b. Genome sequences of horticultural plants: past, present, and future. *Hortic. Res.* 6:1–23.

Chopra, R., E.B. Johnson, R. Emenecker, E.B. Cahoon, J. Lyons, D.J. Kliebenstein, E. Daniels, K.M. Dorn, M. Esfahanian, N. Nicole Folstad, K. Frels, M. McGinn, M. Ott, C. Gallaher, K. Altendorf, A. Berroyer, B. Ismail, J.A. Anderson, D.L. Wyse, T. Ulmasov, J.C. Sedbrook, and M.D. Marks. 2020. Identification and stacking of crucial traits required for the domestication of pennycress. *Nat. Food* 1:84–91.

Cobb, J.N., G. DeClerck, A. Greenberg, R. Clark, and S. McCouch. 2013. Next-generation phenotyping: requirements and strategies for enhancing our understanding of genotype–phenotype relationships and its relevance to crop improvement. *Theor. Appl. Genet.* 126:867–887.

Cooper, M., F. Technow, C. Messina, C. Gho, and L.R. Totir. 2016. Use of crop growth models with whole-genome prediction: application to a maize multienvironment trial. *Crop Sci.* 56:2141–2156.

Cubbon, A., I. Ivancic-Bace, and E.L. Bolt. 2018. CRISPR-Cas immunity, DNA repair and genome stability. *Biosci. Rep.* 38(5).

De Candolle A. 1890. *Origin of Cultivated Plants*. Appleton.

Doerge R.W. 2002. Mapping and analysis of quantitative trait loci in experimental populations. *Nat. Rev. Genet.* 3:43–52.

Ewing, P.M., B.C. Runck, T.Y. Kono, and M.B. Kantar. 2019. The home field advantage of modern plant breeding. *PLoS One* 14. https://doi.org/10.1371/journal.pone.0227079.

Falconer D.S. and T.F.C. Mackay. 1996. *Introduction to Quantitative Genetics, Ed 4*. Longmans Green, Harlow, Essex, UK

Fick, S.E. and R.J. Hijmans. 2017. WorldClim 2: new 1-km spatial resolution climate surfaces for global land areas. *Int. J. Climatol.* 37:4302–4315.

Fister, A.S., L. Landherr, S.N. Maximova, and M.J. Guiltinan. 2018. Transient expression of CRISPR/Cas9 machinery targeting TcNPR3 enhances defense response in theobroma cacao. *Front. Plant Sci.* 9:268.

Fuller, D.Q., T. Denham, M. Arroyo-Kalin, L. Lucas, C.J. Stevens, L. Qin, R.G. Allaby, and M.D. Purugganan. 2014. Convergent evolution and parallelism in plant domestication revealed by an expanding archaeological record. *Proc. Natl. Acad. Sci.* 111:6147–6152.

Gaudelli, N.M., A.C. Komor, H.A. Rees, M.S. Packer, A.H. Badran, D.I. Bryson, and D.R. Liu. 2017. Programmable base editing of A• T to G• C in genomic DNA without DNA cleavage. *Nature* 551(7681):464–471.

Glémin, S. and T. Bataillon. 2009. A comparative view of the evolution of grasses under domestication. *New Phytol.* 183:273–290.

Gomez, M.A., Z.D. Lin, T. Moll, R.D. Chauhan, L. Hayden, K. Renninger, G. Beyene, N.J. Taylor, J.C. Carrington, B.J. Staskawicz, and R.S. Bart. 2019. Simultaneous CRISPR/Cas9-mediated editing of cassava eIF4E isoforms nCBP-1 and nCBP-2 reduces cassava brown streak disease symptom severity and incidence. *Plant Biotechnol. J.* 17(2):421–434.

Harlan J.R. 1971. Agricultural origins: centers and noncenters. *Science* 174:468–474.

Harlan J.R. 1992. *Crops and Man*. American Society of Agronomy, Madison, WI.

Harlan J.R, J.M.J De Wet, and E.G. Price. 1973. Comparative evolution of cereals. *Evolution* 27:311–325.

Hengl, T, J. Mendes de Jesus, G.B. Heuvelink, M. Ruiperez Gonzalez, M. Kilibarda, A. Blagotić, W. Shangguan, M.N. Wright, X. Geng, B. Bauer-Marschallinger, and M.A. Guevara. 2017. SoilGrids250m: global gridded soil information based on machine learning. *PLoS One* 12:e0169748.

Herron, S.A., M.J. Rubin, C. Ciotir, T.E. Crews, D.L. Van Tassel, and A.J. Miller, 2020. Comparative analysis of early life stage traits in annual and perennial Phaseolus crops and their wild relatives. *Front. Plant Sci.* 11:34.

Hijmans, R. and D. Spooner. 2001. Geographic distribution of wild potato species. *Am. J. Bot.* 88:2101–2112.

Hijmans, R.J., D.M. Spooner, A.R. Salas, L. Guarino, J. de la Cruz (2002). *Atlas of Wild Potatoes*, p. 130. ISBN: 978-92-9043-518-1, ISBN: 92-9043-518-6

Hufford, M. B., X. Xu, J. van Heerwaarden, T. Pyhäjärvi, J.-M. Chia, R. A. Cartwright, R. J. Elshire,J. C. Glaubitz, K. E. Guill, S. Kaeppler, J. Lai, P. L. Morrell, L. M. Shannon, C. Song, N. M. Springer, R. A. Swanson-Wagner, P. Tiffin, J. Wang, G. Zhang, J. Doebley, M. D. McMullen, E. S. Buckler, D. Ware, S. Yang, and J. Ross-Ibarra. 2012. Comparative population genomics of maize domestication and improvement. *Nat. Genet.* 44:808–811.

Hummel, A.W., R.D. Chauhan, T. Cermak, A.M. Mutka, A. Vijayaraghavan, A. Boyher, C.G. Starker, R. Bart, D.F. Voytas, and N.J. Taylor. 2018. Allele exchange at the EPSPS locus confers glyphosate tolerance in cassava. *Plant Biotechnol. J.* 16(7):1275–1282.

Jia, H., V. Orbovic, J.B. Jones, and N. Wang. 2016. Modification of the PthA4 effector binding elements in Type I CsLOB1 promoter using Cas9/sgRNA to produce transgenic

Duncan grapefruit alleviating XccΔpthA4: DCsLOB1.3 infection. *Plant Biotechnol. J.* 14(5):1291–1301.

Jao, L.E., S.R. Wente, and W. Chen. 2013. Efficient multiplex biallelic zebrafish genome editing using a CRISPR nuclease system. *Proc. Natl. Acad. Sci.* 110:13904–13909.

Jinek, M., K. Chylinski, I. Fonfara, M. Hauer, J. Doudna, and E. Charpentier. 2012. A programmable dual-RNA-guided DNA endonuclease in adaptive bacterial immunity. *Science* 337:816–821

Kantar, M.B., A.R. Nashoba, J. Anderson, B. Blackman, and L.H. Reiseberg. 2017. The genetics and genomics of plant domestication. *Bioscience* 67:971–982.

Karger, D.N., O. Conrad, J. Böhner, T. Kawohl, H. Kreft, R.W. Soria-Auza, N.E. Zimmermann, H.P. Linder, and M. Kessler. 2017. Climatologies at high resolution for the earth's land surface areas. *Sci. Data* 4:170122.

Khoury, C. K., D. Carver, S.L. Greene, K.A. Williams, H.A. Achicanoy, M. Schori, B. Leon, J. Wiersema, A. Frances. (2020). Crop wild relatives of the United States require urgent conservation action. *Proc. Natl. Acad. Sci.* 117:33351–33357.

Khoury, C.K., N.P. Castañeda-Alvarez, H.A. Achicanoy, C.C. Sosa, V. Bernau, M.T. Kassa, S.L. Nortonf, L. Jos, G. van der Maeseng, H.D. Upadhyaya, J. Ramírez-Villegas, A. Jarvis, P.C. Struik. 2015. Crop wild relatives of pigeonpea [Cajanus cajan (L.) Millsp.]: distributions, ex situ conservation status, and potential genetic resources for abiotic stress tolerance. *Biol. Conserv.* 184:259–270.

Klompe, S.E., P.L. Vo, T.S. Halpin-Healy, and S.H. Sternberg. 2019. Transposon-encoded CRISPR–Cas systems direct RNA-guided DNA integration. *Nature* 571:219–225.

Koinange E.M.K., S.P. Singh, and P. Gepts. 1996. Genetic Control of the domestication syndrome in common-bean. *Crop Sci.* 36:1037–1145.

Komor, A., Y. Kim, M. Packer, J. Zuris, and R. Liu. 2016. Programmable editing of a target base in genomic DNA without double-stranded DNA cleavage. *Nature* 533:420–424. https://doi.org/10.1038/nature17946. https://www.nature.com/articles/nature17946#citeas.

Koropoulis, A., N. Alachiotis, and P. Pavlidis. 2020. Detecting positive selection in populations using genetic data. P. 87–123. In: Dutheil J.Y. (ed.) *Statistical Population Genomics*, Humana, New York. https://doi.org/10.1007/978-1-0716-0199-0_5

Kurashima, N., L. Fortini, and T. Ticktin. 2019. The potential of indigenous agricultural food production under climate change in Hawai'i. *Nat. Sustain.* 2:191.

Kuang Y., S. Li, B. Ren, F. Yan, C. Spetz, X. Li, X. Zhou, and H. Zhou. 2020. Base-editing-mediated artificial evolution of OsALS1 in planta to develop novel herbicide-tolerant rice germplasms. *Mol. Plant* 13:565–572

Kwon, C.T., J. Heo, Z.H. Lemmon. Y. Capua, S.F. Hutton, J. Van Eck, S.J. Park, Z.B. Lippman 2020. Rapid customization of Solanaceae fruit crops for urban agriculture. *Nat. Biotechnol.* 38:182–188.

Ledford, H. 2019. CRISPR conundrum: strict European court ruling leaves food-testing labs without a plan. *Nature* 572(7767):15–16.

Lemmon, Z.H., N.T. Reem, J. Dalrymple, S. Soyk, K.E. Swartwood, D. Rodriguez-Leal, J. Van Eck, and Z.B. Lippman. 2018. Rapid improvement of domestication traits in an orphan crop by genome editing. *Nat. Plants*, 4:766–770.

Lenser, T. and G. Theißen. 2013. Molecular mechanisms involved in convergent crop domestication. *Trends Plant Sci.* 18:704–714.

Li, Y., J. Zhu, H. Wu, C. Liu, C. Huang, J. Lan, Y. Zhao, and C. Xie. 2020. Precise base editing of non-allelic acetolactate synthase genes confers sulfonylurea herbicide resistance in maize. *Crop J.* 8:449–456.

Li, A., S. Jia, A. Yobi, Z. Ge, S.J. Sato, C. Zhang, R. Angelovici, T.E. Clemente, and D.R. Holding. 2018a. Editing of an alpha-kafirin gene family increases, digestibility and protein quality in sorghum. *Plant Physiol.* 177(4):1425–1438.

Li, C., Y. Yue, H. Chen, W. Qi, and R. Song. 2018b. The ZmbZIP22 transcription factor regulates 27-kD γ-zein gene transcription during maize endosperm development. *Plant Cell* 30(10):2402–2424.

Li, C., Y. Zong, Y. Wang, S. Jin, D. Zhang, Q. Song, R. Zhang, and C. Gao 2018c. Expanded base editing in rice and wheat using a Cas9-adenosine deaminase fusion. *Genome Biol.*, 19(1):1–9.

Li, R., R. Li, X. Li, D. Fu, B. Zhu, H. Tian, Y. Luo, and H. Zhu. 2018d. Multiplexed CRISPR/Cas9-mediated metabolic engineering of γ-aminobutyric acid levels in Solanum lycopersicum. *Plant Biotechnol. J.*, 16(2), 415–427.

Liang, Z., K. Chen, T. Li, Y. Zhang, Y. Wang, Q. Zhao, J. Liu, H. Zhang, C. Liu, Y. Ran, and C. Gao. 2017. Efficient DNA-free genome editing of bread wheat using CRISPR/Cas9 ribonucleoprotein complexes. *Nat. Commun.* 8:1–5.

Liang, Z., K. Chen, Y. Zhang, J. Liu, K. Yin, J.L. Qiu, and C. Gao. 2018. Genome editing of bread wheat using biolistic delivery of CRISPR/Cas9 in vitro transcripts or ribonucleoproteins. *Nat. Protoc.* 13:413.

Lin, K., H. Li, C. Schlötterer, and A. Futschik. 2011. Distinguishing positive selection from neutral evolution: boosting the performance of summary statistics. *Genetics* 187:229–244.

Lin, Q. Y. Zong, C. Xue, S. Wang, S. Jin, Z. Zhu, Y. Wang, A.V. Anzalone, A. Raguram, J.L. Doman, and D.R. Liu. 2020. Prime genome editing in rice and wheat. *Nat. Biotechnol.* 38:582–585.

Mehta, D., A. Stürchler, R.B. Anjanappa, S.S.E.A. Zaidi, M. Hirsch-Hoffmann, W. Gruissem, and H. Vanderschuren. 2019. Linking CRISPR-Cas9 interference in cassava to the evolution of editing-resistant geminiviruses. *Genome Biol.* 20(1):1–10.

Miao, J., D. Guo, J. Zhang, Q. Huang, G. Qin, X. Zhang, J. Wan, H. Gu, and L. Qu. 2013. Targeted mutagenesis in rice using CRISPR-Cas system. *Cell Res. 23(*10):1233–1236.

Mishra, R., R.K. Joshi, and K. Zhao. 2020. Base editing in crops: current advances, limitations and future implications. *Plant Biotechnol. J.* 18(1):20–31.

Meyer, R.S., and M.D. Purugganan. 2013. Evolution of crop species: genetics of domestication and diversification. *Nat. Rev. Genetics 14*(12): 840–852.

Meyer, R.S., A.E. DuVal, and H.R. Jensen. 2012. Patterns and processes in crop domestication: an historical review and quantitative analysis of 203 global food crops. *New Phytol.* 196:29–48.

Mueller, N.G., G.J. Fritz, P. Patton, S. Carmody, and E.T. Horton. 2017. Growing the lost crops of eastern North America's original agricultural system. *Nat. Plants* 3:1–5.

Morrell, P.L. and M.T. Clegg. 2007. Genetic evidence for a second domestication of barley (Hordeum vulgare) east of the Fertile Crescent. *Proc. Natl. Acad. Sci.* 104:3289–3294.

Navet, N. and M. Tian. 2020. Efficient targeted mutagenesis in allotetraploid sweet basil by CRISPR/Cas9. *Plant Direct* 00:1–13

Nei, M., T. Maruyama, and R. Chakraborty. 1975. The bottleneck effect and genetic variability in populations. *Evolution* 29:1–10.

Pammer, P. and O. Heller. 2010. Organisms obtained by mutagenesis are GMOs and are, in principle, subject to the obligations laid down by the GMO Directive. Regulation: 1–2. https://curia.europa.eu/jcms/upload/docs/application/pdf/2018-07/cp180111en.pdf.

Pavlidis, P., J.D. Jensen, and W. Stephan. 2010. Searching for footprints of positive selection in whole-genome SNP data from nonequilibrium populations. *Genetics* 185:907–922.

Pauli, D., S.C. Chapman, R. Bart, C.N. Topp, C.J. Lawrence-Dill, J. Poland, and M.A. Gore. 2016. The quest for understanding phenotypic variation via integrated approaches in the field environment. *Plant Physiol.* 172:622–634.

Peng, A., S. Chen, T. Lei, L. Xu, Y. He, L. Wu, L. Yao, and X. Zou. 2017. Engineering canker-resistant plants through CRISPR/Cas9-targeted editing of the susceptibility gene Cs LOB 1 promoter in citrus. *Plant Biotechnol. J.* 15(12):1509–1519.

Peng, B., K. Guan, J. Tang, E.A. Ainsworth, S. Asseng, C.J. Bernacchi, et al. 2020. Towards a multiscale crop modelling framework for climate change adaptation assessment. *Nat. Plants* 6(4):338–348.

Pironon, S., J.S. Borrell, I. Ondo, R. Douglas, C. Phillips, C.K. Khoury, M.B. Kantar, N. Fumia, M. Soto Gomez, J. Viruel, R. Govaerts, F. Forest, and A. Antonelli. 2020. Toward unifying global hotspots of wild and domesticated biodiversity. *Plants* 9:1128.

Pont, C., S. Wagner, A. Kremer, L. Orlando, C. Plomion, and J. Salse. 2019. Paleogenomics: reconstruction of plant evolutionary trajectories from modern and ancient DNA. *Genome Biol.*, 20:29.

Purugganan, M.D. 2019. Evolutionary insights into the nature of plant domestication. *Curr. Biol.*, 29:R705–R714.

Purugganan, M.D. and D.Q. Fuller. 2011. Archaeological data reveal slow rates of evolution during plant domestication. *Int. J. Org. Evol.*, 65:171–183.

Rahaman, M., D. Chen, Z. Gillani, C. Klukas, and M. Chen. 2015. Advanced phenotyping and phenotype data analysis for the study of plant growth and development. *Front. Plant Sci.* 6:619.

Ramos-Madrigal, J., B.D. Smith, J.V. Moreno-Mayar, S. Gopalakrishnan, J. Ross-Ibarra, M.T.P. Gilbert, and N. Wales. 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.

Ramstein, G.P., S.E. Jensen, and E.S. Buckler. 2019. Breaking the curse of dimensionality to identify causal variants in Breeding 4. *Theor. Appl. Genet.* 132 559–567.

Ramu, P., W. Esuma, R. Kawuki, I.Y. Rabbi, C. Egesi, J.V. Bredeson, R.S. Bart, J. Verma, E.S. Buckler, and F. Lu. 2017. Cassava haplotype map highlights fixation of deleterious mutations during clonal propagation. *Nat. Genet.* 49:959–963

Rodríguez-Leal, D., Z.H. Lemmon, J. Man, M.E. Bartlett, and Z.B. Lippman. 2017. Engineering quantitative trait variation for crop improvement by genome editing. *Cell* 171:470–480.

Rojas-Vásquez, R. and A. Gatica-Arias. 2020. Use of genome editing technologies for genetic improvement of crops of tropical origin. *Plant Cell Tissue Organ Culture (PCTOC) 140*(1):215–244.

Ronen, R., N. Udpa, E. Halperin, and V. Bafna. 2013. Learning natural selection from the site frequency spectrum. *Genetics* 195:181–193.

Ross-Ibarra, J., P.L. Morrell, and B.S. Gaut. 2007. Plant domestication, a unique opportunity to identify the genetic basis of adaptation. *Proc. Natl. Acad. Sci.* 104:8641–8648.

Sang T. 2009. Genes and mutations underlying domestication transitions in grasses. *Plant Physiol.* 149:63–70.

Schlautman, B., S. Barriball, C. Ciotir, S. Herron, and Miller, A.J. 2018. Perennial grain legume domestication phase I: criteria for candidate species selection. *Sustainability* 10:730.

Schrider, D.R. and A.D. Kern. 2016. S/HIC: robust Identification of soft and hard sweeps using machine learning. *PLoS Genet.* 12:e1005928.

Schreiber, M., N. Stein, and M. Mascher. 2018. Genomic approaches for studying crop evolution. *Genome Biol.* 19:140.

Schrider, D.R. and A.D. Kern. 2018. Supervised machine learning for population genetics: a new paradigm. *Trends Genet.* 34:301–312.

Scott, M.F., L.R. Botigué, S. Brace, C.J. Stevens, V.E. Mullin, A. Stevenson, M.G. Thomas, D.Q. Fuller, and R. Mott. 2019. A 3,000-year-old Egyptian emmer wheat genome reveals dispersal and domestication history. *Nature Plants* 5:1120–1128.

Shao, X. S. Wu, T. Dou, H. Zhu, C. Hu, H. Huo, W. He, G. Deng, O. Sheng, F. Bi, H. Gao, T. Dong C. Li, Q. Yang, G. Yi. 2020. Using CRISPR/Cas9 genome editing system to create MaGA20ox2 gene-modified semi-dwarf banana. *Plant Biotechnol. J.* 18(1):17–19.

Shi, J., H. Gao, H. Wang, H.R. Lafitte, R.L. Archibald, M. Yang, S.M. Hakimi, H. Mo, and J.E. Habben. 2017. ARGOS 8 variants generated by CRISPR-Cas9 improve maize grain yield under field drought stress conditions. *Plant Biotechnol. J.* 15(2):207–216.

Shi, J. and J. Lai. 2015. Patterns of genomic changes with crop domestication and breeding. *Curr. Opin. Plant Biol.* 24:47–53.

Simon, M.L. and K.E. Parker. 2006. Prehistoric plant use in the American Bottom: new thoughts and interpretations. *South. Archaeol.* 25:212–257.

Stephan, W. 2019. Selective sweeps. *Genetics*, 211:5–13.

Soyk, S., Z.H. Lemmon, M. Oved, J. Fisher, K.L. Liberatore, S.J. Park, A. Goren, K. Jiang, A. Ramos, E. Van Der Knaap, J. Van Eck, D. Zamir, Y. Eshed, and Z.B. Lippman 2017. Bypassing negative epistasis on yield in tomato imposed by a domestication gene. *Cell* 169:1142–1155.

Somssich, M., B.I. Je, R. Simon, and Jackson, D. 2016. Clavata-Wuschel signaling in the shoot meristem. *Development* 143:3238–3248.

Sun, Y., X. Zhang, C. Wu, Y. He, Y. Ma, H. Hou, X. Guo, W. Du, Y. Zhao, and L. Xia. 2016. Engineering herbicide-resistant rice plants through CRISPR/Cas9-mediated homologous recombination of acetolactate synthase. *Molecular Plant* 9(4):628–631.

Sun, Y., G. Jiao, Z. Liu, X. Zhang, J. Li, X. Guo, W. Du, J. Du, F. Francis, Y. Zhao, and L. Xia. 2017. Generation of high-amylose rice through CRISPR/Cas9-mediated targeted mutagenesis of starch branching enzymes. *Front. Plant Sci.* 8:298.

Swarts, K., R.M. Gutaker, B. Benz, M. Blake, R. Bukowski, J. Holland, and H.A. Burbano. 2017. Genomic estimation of complex traits reveals ancient maize adaptation to temperate North America. *Science* 357(6350):512–515.

Tashkandi, M., Z. Ali, F. Aljedaani, A. Shami, and M.M. Mahfouz. 2018. Engineering resistance against Tomato yellow leaf curl virus via the CRISPR/Cas9 system in tomato. *Plant Signaling Behav.* 13(10):e1525996.

Tanksley, S.D. and S.R. McCouch. 1997. Seed banks and molecular maps: unlocking genetic potential from the wild. *Science* 277:1063–1066. doi: 10.1126/science.277.5329.1063.

Technow, F., C.D. Messina, L.R. Totir, and M. Cooper. 2015 Integrating crop growth models with whole genome prediction through approximate bayesian computation. *PLoS One* 10:e0130855.

Tian, S., L. Jiang, X. Cui, X. Cui, J. Zhang, S. Guo, M. Li, H. Zhang, Y. Ren, G. Gong, M. Zong, F. Liu, Q. Chen, Y. Xu. 2018. Engineering herbicide-resistant watermelon variety through CRISPR/Cas9-mediated base-editing. *Plant Cell Rep.* 37:1353–1356. https://doi.org/10.1007/s00299-018-2299-0.

Tiffin, P. and J. Ross-Ibarra. 2014. Advances and limits of using population genetics to understand local adaptation. *Trends Ecol. Evol.*, 29:673–680.

Tripathi, J. N., V.O. Ntui, M. Ron, S.K. Muiruri, A. Britt, and L. Tripathi. 2019. CRISPR/Cas9 editing of endogenous banana streak virus in the B genome of Musa spp. overcomes a major challenge in banana breeding. *Commun. Biol. 2*(1):1–11.

van der Knaap, E., M. Chakrabarti, Y.H. Chu, J.P. Clevenger, E. Illa-Berenguer, Z. Huang, N. Keyhaninejad, Q. Mu, L. Sun, Y. Wang, and Wu, S. 2014. What lies beyond the eye: the molecular mechanisms regulating tomato fruit weight and shape. *Front. Plant Sci.* 5:227.

Van Tassel, D.L., K.A. Albrecht, J.D. Bever, A.A. Boe, Y. Brandvain, T.E. Crews, M. Gansberger, P. Gerstberger, L. González-Paleo, B.S. Hulke, and Kane, N.C., 2017. Accelerating Silphium domestication: an opportunity to develop new crop ideotypes and breeding strategies informed by multiple disciplines. *Crop Sci.* 57:1274–1284.

Vaughan, D.A., E. Balazs, and J.S. Heslop-Harrison. 2007. From crop domestication to super-domestication. *Ann. Bot.* 100:893–901.

Vavilov N.I. 1926. *Origin and geography of cultivated plants*, Cambridge University Press, Cambridge.

Wallace, J.G., E. Rodgers-Melnick, and E.S. Buckler. 2018. On the road to breeding 4.0: unraveling the good, the bad, and the boring of crop quantitative genomics. *Annu. Rev. Genet. 52:*421–444.

Wang, D.R., M.D. Venturas, D.S. Mackay, D.J. Hunsaker, K.R. Thorp, M.A. Gore, and D. Pauli. 2020a. Use of hydraulic traits for modeling genotype-specific acclimation in cotton under drought. *New Phytol.* 228:898–909

Wang, H., E. Cimen, N. Singh, and E. Buckler. 2020b. Deep learning for plant genomics and crop improvement. *Curr. Opin. Plant Biol.* 54:34–41.

Wang F., C. Wang, P. Liu, C. Lei, W. Hao, Y. Gao, Y.-G. Liu, and K. Zhao. 2016. Enhanced rice blast resistance by CRISPR/Cas9-targeted mutagenesis of the ERF transcription factor gene OsERF922. *PLoS One* 11(4):e0154027. https://doi.org/10.1371/journal.pone.0154027

Watson, A., S. Ghosh, M.J. Williams, W.S. Cuddy, J. Simmonds, M.D. Rey, M.A.M. Hatta, A. Hinchliffe, A. Steed, D. Reynolds, and N.M. Adamski. 2018. Speed breeding is a powerful tool to accelerate crop research and breeding. *Nature plants* 4:23–29.

Weeden, N.F. 2007. Genetic changes accompanying the domestication of Pisum sativum: is there a common genetic basis to the 'domestication syndrome' for legumes? *Ann. Bot.* 100:1017–1025.

Wiedenheft, B., K. Zhou, M. Jinek, S.M. Coyle, W. Ma, and J.A. Doudna. 2009. Structural basis for DNase activity of a conserved protein implicated in CRISPR-mediated genome defense. *Structure* 17:904–912.

Woo, J.W., J. Kim, S.I. Kwon, C. Corvalán, S.W. Cho, H. Kim, S.G. Kim, S.T. Kim, S. Choe, and Kim, J.S., 2015. DNA-free genome editing in plants with preassembled CRISPR-Cas9 ribonucleoproteins. *Nat. Biotechnol.* 33:1162–1164.

Wright, S.I., I.V. Bi, S.G. Schroeder, M. Yamasaki, J.F. Doebley, M.D. McMullen, and Gaut, B.S. 2005. The effects of artificial selection on the maize genome. *Science* 308:1310–1314.

Xu, R., H. Li, R. Qin, L. Wang, L. Li, P. Wei, and J. Yang, 2014. Gene targeting using the Agrobacterium tumefaciens-mediated CRISPR-Cas system in rice. *Rice* 7:5. https://doi.org/10.1186/s12284-014-0005-6.

Yamasaki, M., S. G. Schroeder, H. Sanchez-Villeda, B.S. Gaut, and M.D. McMullen. 2008. Empirical analysis of selection screens for domestication and improvement loci in maize by extended DNA sequencing. *Plant Genome 1*(1).

Zhang, X., A. Sallam, L. Gao, T. Kantarski, J. Poland, L.R. DeHaan, D.L. Wyse, and J.A. Anderson. 2016. Establishment and optimization of genomic selection to accelerate the domestication and improvement of intermediate wheatgrass. *The Plant Genome* 9:1–18.

Zohary D. 1999. Monophyletic vs. polyphyletic origin of the crops on which agriculture was founded in the Near East. *Genet. Resour. Crop. Evol.* 46:133–142.

Zsögön, A., T. Cermak, D. Voytas, and L.E.P. Peres. 2017. Genome editing as a tool to achieve the crop ideotype and de novo domestication of wild relatives: case study in tomato. *Plant Sci.* 256:120–130.

Zsögön, A., T. Čermák, E.R. Naves, M.M. Notini, K.H. Edel, S. Weinl, L. Freschi, D.F. Voytas, J. Kudla, and L.E.P. Peres. 2018. De novo domestication of wild tomato using genome editing. *Nat. Biotechnol.* 36:1211–1216.