Cell

The Molecular Genetics of Crop Domestication

John F. Doebley,[1,*] Brandon S. Gaut,[2] and Bruce D. Smith[3]
[1]Department of Genetics, University of Wisconsin, Madison, WI 53706, USA
[2]Department of Ecology and Evolutionary Biology, University of California, Irvine, CA 92697, USA
[3]Archaeobiology Program, National Museum of Natural History, Smithsonian Institution, NHB MRC 112, Washington, DC 20013, USA
*Contact: jdoebley@wisc.edu
DOI 10.1016.j.cell.2006.12.006

Ten thousand years ago human societies around the globe began to transition from hunting and gathering to agriculture. By 4000 years ago, ancient peoples had completed the domestication of all major crop species upon which human survival is dependent, including rice, wheat, and maize. Recent research has begun to reveal the genes responsible for this agricultural revolution. The list of genes to date tentatively suggests that diverse plant developmental pathways were the targets of Neolithic "genetic tinkering," and we are now closer to understanding how plant development was redirected to meet the needs of a hungry world.

Most members of our modern industrial societies have never seen and would not recognize the unpromising wild plants that are the progenitors of our remarkably productive crops. Very few members of these societies would survive if all they had were a field of wild grain and herbs and their own wits to sustain them. Yet, 10,000 years ago, people who could not read, write, or do calculus prospered on diets composed of wild plants and animals. Even more remarkably, these ancient peoples began a plant-breeding program that transformed hundreds of wild plant species into domesticated crops, including all of the most highly productive crops—rice, wheat, maize—on which human survival is dependent today.

In this review, we first summarize some basic observations about domestication and the origin of agriculture. How were crops modified during domestication? What was the breeding process by which wild species were converted to crops? And when and where did domestication take place? Next, we discuss the genes that have been identified to date as controlling key differences in plant structures and physiology that distinguish crops and their progenitors or different crop varieties from one another. We then discuss how population genetic analyses can be used to discover genes that were the targets of selection during plant domestication by humans. We end by summarizing what has been learned about how domestication modified plant development to produce today's crops and by giving some examples of how this knowledge is being exploited to drive crop improvements in ways not possible using traditional plant breeding methods.

The Domestication Syndrome

Plant domestication is the genetic modification of a wild species to create a new form of a plant altered to meet human needs. For many crops, domestication has rendered the plant completely dependent on humans such that it is no longer capable of propagating itself in nature. Maize and cauliflower are good examples of such highly modified forms. However, other crops, such as hemp, carrot, and lettuce, have been more modestly modified compared to their progenitors, and they can either revert to the wild or become self-propagating weeds.

There is a common suite of traits—known as the "domestication syndrome"—that distinguishes most seed and fruit crops from their progenitors (Hammer, 1984). Compared to their progenitors, food crops typically have larger fruits or grains, more robust plants overall, more determinate growth or increased apical dominance (the robust growth of the central stem in comparison to the side stems), and a loss of natural seed dispersal so that seeds remain attached to the plant for easy harvest by humans (Figure 1). Remarkably, crops often have fewer (although larger) fruits or grains per plant than their progenitors. A variety of physiological changes are also involved. These include a loss of seed dormancy, a decrease in bitter substances in edible structures, changes in photoperiod sensitivity, and synchronized flowering.

The Domestication Process

Most researchers believe that agriculture began as an attempt to modify the landscape and thereby encourage the growth of edible wild plants at the expense of less useful ones (Smith, 1998). Among hunter-gatherers such as the Australian Aborigines, burning of native vegetation was practiced because species of grasses favored as food thrived after the burning. From such a practice, it is a small step to burning an area devoid of useful plants and then sowing seed of favored species

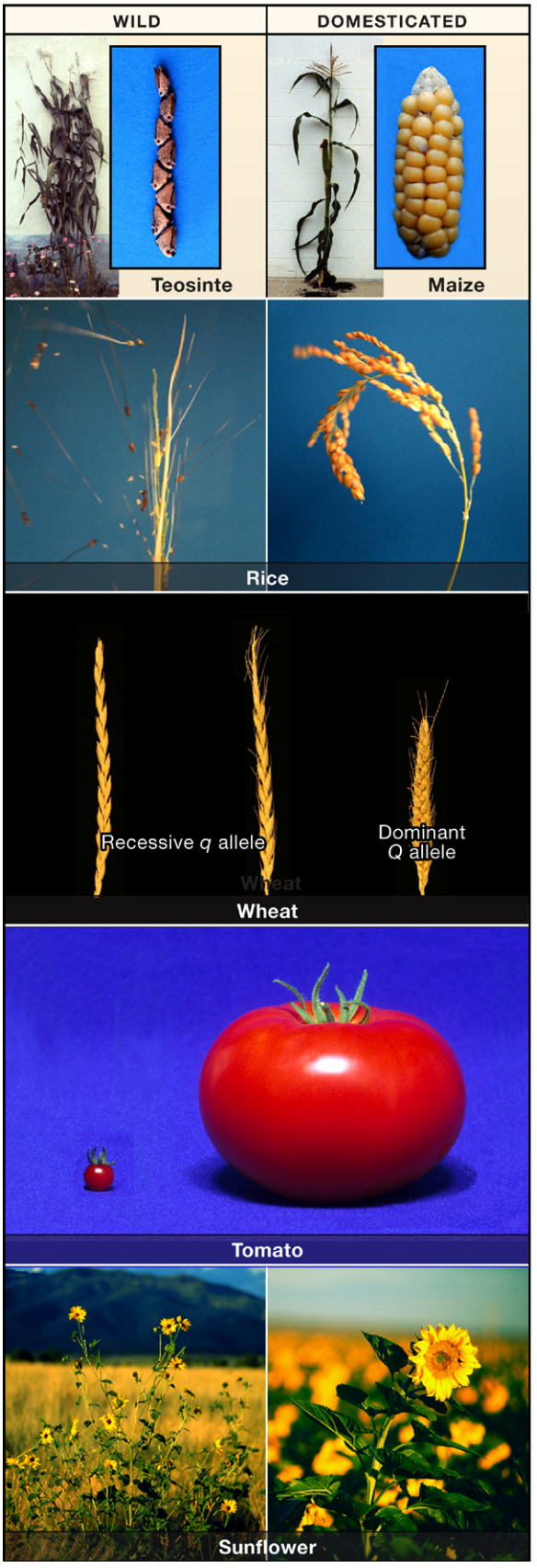

Figure 1. Phenotypes of Some Crops and Their Progenitors
(Top row) A plant of the maize progenitor, teosinte (left), with multiple stalks and long branches, is shown next to a plant of cultivated maize (right) with its single stalk. A maize ear (inset) bears its grain naked on the surface of the ear, whereas a teosinte ear (inset) has its grain (not visible) enclosed in the triangular casing that comprises the ear.
(Second row) Wild rice (left) has a panicle that shatters, whereas cultivated rice (rice) has a solid panicle of grain.
(Third row) Cultivated wheat with the dominate allele of the *Q* gene (right) has a condensed and tough spike. Cultivated wheat with the recessive allele *q* (center) and wild wheat (left) with the recessive allele have slender, fragile spikes.
(Fourth row) The massive fruit of cultivated tomato (right) next to the miniscule fruit of its progenitor (left).
(Fifth row) A wild sunflower plant (left) has many small heads borne on multiple slender stalks, whereas a cultivated sunflower plant (right) has a single large head borne on a thick stalk.

gathered from another location. Key to the domestication process would be a subsequent switch from allowing edible wild species to naturally resow themselves in burned fields, to sowing seed gathered the previous season. Once this practice was established, selection and crop improvement could begin.

Although cereals and other field crops were likely to have been domesticated in the context of large fields cleared by burning or by spring floods along rivers, other domesticates may have had their beginnings as weeds near seasonal campgrounds (Anderson, 1969). Hunter-gatherers often follow seasonal migratory schedules, visiting the same specific sites at given times every year. The disturbance of the natural vegetation and middens at these sites provided fertile ground for the types of colonizing species that were the progenitors of our crops. Seeds discarded with the "kitchen" trash one year would sprout into a new crop by the time the group returned the following year. If they preferentially collected seeds and fruit from plants with the most desirable traits, then over time the frequency of plants with these favored phenotypes would increase in their garden crop. Eventually, no new wild seeds and fruits would be collected and a switch to deliberate sowing of seeds would occur.

The early agricultural practices just described have left their signatures on the patterns of genetic diversity in the genomes of crop plants. Because early farmers used only a limited number of individuals of the progenitor species, much of the genetic diversity in the progenitor was left behind. Moreover, with each generation during the domestication process, only seed from the best plants formed the next generation. This winnowing caused a genetic bottleneck, which reduced genetic diversity throughout the genome (Doebley, 1989). The extent of this loss of diversity depends on the population size during the domestication period and the duration of that period (Eyre-Walker et al., 1998). Notably, the loss in diversity was not experienced equally by all genes in the genome. For genes that do not influence favored phenotypes (which are called neutral genes), the loss in diversity is simply a function of the strength of the bottleneck

in terms of the population size and duration (Figure 2). However, genes that influence desirable phenotypes experienced a more drastic loss of diversity because plants carrying favored alleles contributed the most progeny to each subsequent generation and other alleles were eliminated from the population (Wright et al., 2005).

One unknown in the domestication process is the extent to which new mutations versus preexisting genetic variation in the wild species contributed to the evolution of crop phenotypes. In a few cases, crops possess alleles of major genes that disrupt seed shattering (Li et al., 2006) or the protective casing surrounding the seed (Wang et al., 2005) that are not found in the progenitor species. However, alleles of genes that contribute to increased fruit size in tomato (Nesbitt and Tanksley, 2002) or increased apical dominance in maize (Clark et al., 2004) are also found in their wild or feral relatives, although at lower frequencies. Given the large store of genetic variation in the progenitor species, it seems most reasonable that domestication largely involved filtering out the best alleles from standing allelic variation in crop ancestors, although new mutations in key developmental pathways may have been instrumental for some traits.

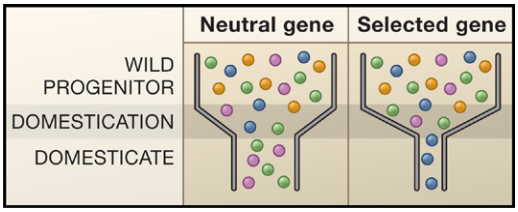

Figure 2. The Effects of the Domestication Bottleneck on Genetic Diversity

(Left) Population bottlenecks are a common important demographic event during domestication. Genetic diversity is represented by shaded balls; the bottleneck reduces diversity in neutral genes, as shown by the loss of the orange and blue variants.

(Right) Selection decreases diversity beyond that caused by the bottleneck, as shown by the loss of all but one genetic variant in the domesticated species. Note, however, that an exceptionally strong domestication bottleneck could leave little variation in neutral genes. In that case, it may be very difficult to distinguish selected from neutral loci.

Crop Origins and Diversification

More than a half dozen different independent centers of domestication have been identified to date (Figure 3). These centers comprise a promising comparative set of developmental trajectories in that they differ markedly in a number of important respects: their geographical size, the number and diversity of each region's locally domesticated species and their relative potential as food sources (both individually and as overall integrated food production systems), and how quickly each region's emerging domesticate-based economies initially developed and subsequently expanded into adjacent regions (Smith, 1998; Piperno and Pearsall, 1998). The Fertile Crescent region of the Near East, for example, witnessed the domestication of a remarkable set of plant and animal species that were formed relatively quickly into a powerful and expansive agricultural economy (e.g., goat, sheep, pig, cattle, einkorn and emmer wheat, barley, and lentils). In comparison, no animals were brought under domestication in eastern North America, and of the four plants that were domesticated, only summer squash and sunflower survived as domesticates into the 1800s.

Regional scale comparative analysis of agricultural origins is of course not restricted to the identified independent centers of domestication. Over the past 10,000 years most other areas of the world have also witnessed the transition to food production economies, as introduced domesticates were selectively recombined and integrated with local plants, resulting in a rich worldwide mosaic of agricultural systems. Aided by a range of new techniques and approaches, archaeologists and geneticists are documenting this long and complex process of agricultural expansion and the associated temporal and geographical patterns of crop diffusion in an increasing number of world regions (Harris, 1996).

Tracing the Origins of Crops

Over the past 20 years, there has been a concerted effort on the part of archaeologists and geneticists to answer a variety of questions regarding the histories of individual domesticated species, the basic building blocks of the agricultural transition (Smith, 2001; Zeder et al., 2006). What wild species and populations were ancestral to specific crops? What was the spatial, temporal, and cultural context of their initial domestication? What phenotypic changes occurred during domestication in the archaeological record and at what rate? The multidisciplinary, archaeological-genetic approach to these questions has proven remarkably informative, especially for crops like maize and wheat.

Maize (*Zea mays* ssp. *mays*) provides perhaps the best example of how parallel genetic and archaeological research can be combined to provide a reasonably detailed and comprehensive account of a species' initial domestication and subsequent dispersal. Genetic analysis has identified populations of the wild grass teosinte growing in the central Balsas river valley of southern Mexico as the closest modern relative of maize, indicating that this general region is a candidate for the location of the initial domestication of maize (Matsuoka et al., 2002). The oldest archaeological maize ears come from Guilá Naquitz Cave in the valley of Oaxaca, located only about 400 km northeast of the Balsas River, where two small cobs have been found dating to about 6300 BP (before present) (Piperno and Flannery, 2001).

Ongoing analyses continue to document changes in the morphology of maize and the development of regionally distinct land races. This work has shown that the full suite of morphological traits defining domesticated maize were already present in the 6300 year-old maize of Guilá Naquitz (Benz, 2001). At the same time, ancient

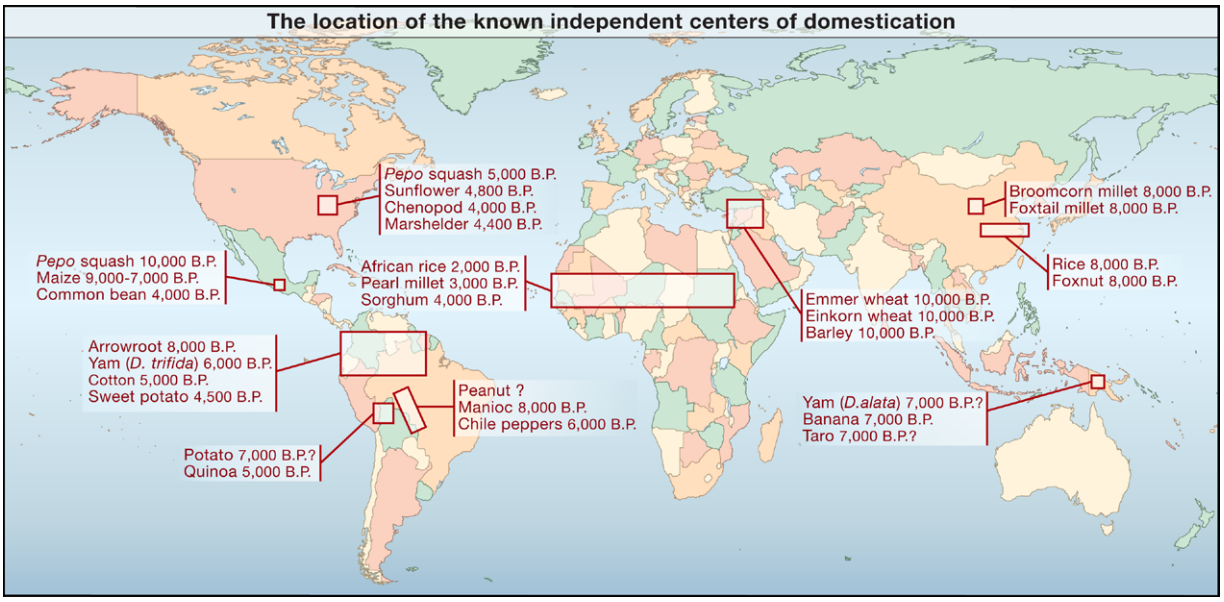

Figure 3. The Independent Centers of Domestication
For each region, principal crop plants and estimates of when they were brought under domestication based on currently available archaeological evidence are shown.

DNA studies of archaeological maize from northeast Mexico and the southwest United States have shown that it is possible to track human selection for specific attributes that are not observable in the archaeological record (Jaenicke-Despres et al., 2003). This highlights the potential for combining genetic and archaeological research in order to reconstruct the evolution of crop plants spanning thousands of years. In similar fashion, the recent genetic data tracing the temporal and geographical radiation of maize from southern Mexico to the limits of cultivation in the Americas compare very closely to large scale efforts by archaeologists to track the gradual expansion of maize cultivation throughout the Americas (Matsuoka et al., 2002; Blake, 2006).

Einkorn wheat (*Triticum monococcum*) provides a similar, if less robust, example of cross illumination from genetic and archaeological research on the initial domestication of major crop plants. A genetic comparison of modern domesticated and wild einkorn populations across the Fertile Crescent identified the Karacadağ mountain region of southeastern Turkey as its likely heartland of domestication (Heun et al., 1997). Situated only about 200 km to the south, along the Euphrates River, the archaeological site of Abu Hureyra has yielded the earliest evidence (9600 BP) not only for domesticated einkorn wheat but also for emmer wheat (*Triticum araraticum*) and barley (*Hordeum vulgare*) (Smith, 1998). The subsequent radiation of these crop plants across the Fertile Crescent and then north and west throughout Europe has been documented in some detail in the archaeological record of the two regions (Harris, 1996).

The archaeological record of the domestication and early history of other major crop plants, such as rice (*Oryza sativa*), remains incomplete. The earliest evidence for domesticated rice has been recovered from the settlements of already sophisticated rice-farming societies along the middle and lower Yangtze River corridor in southern China. At the village settlement of Ho-mu-tu along the lower Yangtze River, for example, 1 m thick deposits of domesticated rice husks dating to about 7000 BP were recovered (Smith, 1998). These settlements predate any sign of rice cultivation elsewhere in East Asia by several thousand years. As recently developed methods of plant recovery and analysis are more widely applied to earlier sites along the Yangtze and throughout East Asia, the initial domestication of rice will quite likely be pushed back in time another 1000–2000 years or more. Unfortunately, archaeological rice grains cannot yet be assigned with any degree of confidence to particular varieties (e.g., *indica* and *japonica*) on the basis of morphology. As a result, any better understanding of the initial domestication and early history of *indica* versus *japonica* will probably be based on genetic analysis of ancient and present-day populations of domesticated and wild rice. For example, a recent analysis of DNA-sequence data has confirmed that *indica* and *japonica* are the products of separate domestication events, the former south of the Himalaya and the latter in southern China (Londo et al., 2006).

Another crop that has been the recent subject of landmark genetic research is the tomato (*Lycopersicon esculentum*). Unfortunately, this crop is much less well documented in the archaeological record. Given the

midelevation lomas zone habitat of the likely wild ancestor of tomato along the west coast of South America, it is surprising that evidence for domesticated tomato first appears far to the north, in Mexico, in contexts that are less than 1000 years old. However, genetic data suggest that the ancestor of tomato may have spread from South America to Mexico as a weed and then been domesticated in Mexico (Rick and Fobes 1975).

Genes Controlling Domestication Traits

Over the past decade, researchers have begun to identify the specific genes that control some of the most important morphological changes associated with domestication. Because the traits involved are mostly quantitative in nature, the path to identifying these genes started with the mapping of quantitative trait loci (QTL) in progenitor-crop hybrid populations, followed either by positional cloning or cloning using a combination of positional information and candidate gene analysis. Although the list of well-documented domestication genes is short, some generalities are beginning to appear. Below, we summarize what has been learned to date about genes that are known to have contributed to phenotypic differences in traits under selection during domestication.

Teosinte branched1 (*tb1*) of maize was identified as a major QTL controlling the difference in apical dominance between maize and its progenitor, teosinte (Doebley et al., 1997; Doebley, 2004). Maize plants typically have a single stalk with short branches tipped by ears, whereas teosinte plants are more highly branched (Figure 1). *tb1*, which controls these differences, is a founding member of the TCP family of transcriptional regulators, a class of genes involved in the transcriptional regulation of cell-cycle genes including PCNA and cyclins (Cubas et al., 1999; Kosugi and Ohashi, 2002). The current model is that *tb1* represses the outgrowth of the axillary meristems and branch elongation via its repressive effect on the cell cycle. This repression may result from competitive binding of TB1 (a repressor) to TCP-specific binding sites in the promoters of the cell-cycle genes, thus blocking other TCP genes from activating these cell-cycle genes (Li et al., 2005). Differences in *tb1* expression patterns between maize and teosinte indicate that human selection was targeted at regulatory differences that produced a higher level of *tb1* message in maize (Doebley et al., 1997; Wang et al., 1999). The lack of any fixed amino acid differences between maize and teosinte in the TB1 protein supports this hypothesis.

Fruitweight2.2 (*fw2.2*) was identified as a large effect QTL controlling 30% of the difference in fruit mass between wild and cultivated tomato (Figure 1) (Frary et al., 2000). The exact molecular function of *fw2.2* is not known, although it shares homology with the human *RAS* oncogenes in protein structure and contains two predicted transmembrane-spanning domains. *fw2.2* acts as a negative regulator of cell division in the fruit, perhaps via some role in cell-to-cell communication. Human selection appears to have favored a heterochronic shift in *fw2.2* expression such that the large-fruited allele is expressed at lower levels later in fruit development, thus allowing continued fruit growth through proliferative cell division (Cong et al., 2002). The large- and small-fruited alleles have no differences in protein sequence, supporting the model that changes in gene regulation underlie the evolution of tomato fruit size as controlled by *fw2.2* (Nesbitt and Tanksley, 2002).

Teosinte glume architecture1 (*tga1*) was identified as a QTL controlling the formation of the casing that surrounds the kernels of the maize ancestor, teosinte (Wang et al., 2005). *tga1* is a member of the *s*quamosa-promoter *b*inding *p*rotein (SBP) family of transcriptional regulators. Some members of this family directly regulate MADS-box transcriptional regulators (Cardon et al., 1999), suggesting that *tga1* may sit at the top of a cascade of transcriptional regulators. Consistent with this hypothesis, *tga1* has phenotypic effects on diverse traits including cell lignification, silica deposition in cells, three-dimensional organ growth, and organ size (Dorweiler and Doebley, 1997). The difference in function between the maize and teosinte alleles of *tga1* appears to be the result of a single amino acid change. The fact that there are no discernable differences in gene expression supports this interpretation.

Q is a major gene involved in wheat domestication that affects a suite of traits, including the tendency of the spike (ear) to shatter, the tenacity of the chaff surrounding the grain, and whether the spike is elongated as in wild wheat or compact like the cultivated forms (Figure 1) (Simons et al., 2006). Recently, *Q* has been identified as a member of the AP2 family of plant-specific transcriptional regulators. This gene family regulates a diverse set of developmental traits in plants, but especially traits related to inflorescence structure and flowering. The cultivated (*Q*) allele is expressed at a higher level than the wild (*q*) allele, and gene dosage analysis indicates that differences in expression could be sufficient to explain the difference in phenotype. However, these alleles also differ by a single amino acid change that affects protein dimerization, suggesting both regulator and protein function changes could be involved.

shattering4 (*sh4*) is a major QTL controlling whether the seed fall off the plant (shatter) as in wild rice or adhere to the plant as in cultivated rice. Li et al. (2006) have shown that *sh4* encodes a gene with homology to Myb3 transcription factors. Using transformation, Li et al. (2006) were able to confirm that a single amino change in the predicted DNA binding domain converts plants from shattering to nonshattering. A decrease in expression of the cultivated allele as compared to the wild allele may also be important. Notably, the cultivated *sh4* allele weakens, but does not fully eliminate, the shattering phenotype, which might be critical, because farmers need seed that stays on the plant long enough to be harvested but which can subsequently be freed from the plant by threshing. Although the downstream targets of *sh4* are unknown, they may be involved in programmed

Table 1. Genes of Interest in Crop Domestication and Improvement[1]

Gene(s)	Crop	Molecular and Phenotypic Function	Controls Phenotype[2]	Selection Evidence[3]	Causative Change
Genes Identified as Controlling Domestication Traits					
tb1	Maize	Transcriptional regulator (TCP); plant and inflorescence structure	Yes	Yes	regulatory change
tga1	Maize	Transcriptional regulator (SBP); seed casing	Yes	Yes	amino acid change
qSH1	Rice	Transcriptional regulator (homeodomain); abscission layer formation, shattering	Yes	N.T.	regulatory change
Rc	Rice	Transcriptional regulator (bHLH); seed color	Yes	N.T.	disrupted coding sequence
sh4	Rice	Transcriptional regulator (Myb3); abscission layer formation, shattering	Yes	N.T.	regulatory/amino acid change
fw2.2	Tomato	Cell signaling; fruit weight	Yes	N.T.	regulatory change
Q	Wheat	Transcriptional regulator (AP2); inflorescence structure	Yes	N.T.	regulatory/amino acid change
Genes Identified as Controlling Varietal Differences					
c1	Maize	Transcriptional regulator (MYB); kernel color	Yes	Yes	regulatory change
r1	Maize	Transcriptional regulator (bHLH); kernel color	Yes	N.T.	regulatory change
sh2	Maize	pyrophosphorylase; supersweet sweet corn	Yes	N.T.	transposon insertion
su1	Maize	isoamylase; sweet corn gene	Yes	Yes	amino acid change
y1	Maize	Phytoene synthase; carotenoid content	Yes	Yes	regulatory change
brix9-2-5	Tomato	Invertase; fruit soluble solid content	Yes	N.T.	amino acid change
ovate	Tomato	Unknown; fruit shape	Yes	N.T.	early stop codon
rin	Tomato	Transcriptional regulator (MADS); fruit ripening	Yes	N.T.	regulatory change
sp	Tomato	Cell signaling; determinant plant growth	Yes	N.T.	amino acid change
R	Pea	Starch branching enzyme; seed sugar content	Yes	N.T.	transposon insertion
ehd1	Rice	B-type response regulator; flowering time	Yes	N.T.	amino acid change
gn1	Rice	Cytokinin oxidase/dehydrogenase; grain number	Yes	N.T.	regulatory/early stop codon
hd1	Rice	Transcriptional regulator (zinc finger); flowering time	Yes	N.T.	disrupted coding sequence
hd6	Rice	Protein kinase; flowering time	Yes	N.T.	early stop codon
sd1	Rice	GA20 oxidase; plant height	Yes	Yes	disrupted coding sequence
waxy	Rice	Starch synthase; sticky grains	Yes	Yes	intron splicing defect
rht	Wheat	Transcriptional regulator (SH2); plant height	Yes	N.T.	early stop codon
vrn1	Wheat	Transcriptional regulator (MADS); vernalization	Yes	N.T.	regulatory change
vrn2	Wheat	Transcriptional regulator (ZCCT); vernalization	Yes	N.T.	amino acid change
Genes Identified by Selection Screens Targeted at Individual Candidate Genes					
boCal	Cauliflower	Transcriptional regulator (MADS); inflorescence structure	Candidate	Yes	early stop codon?
ba1	Maize	Transcriptional regulator (bHLH); plant and inflorescence structure	Candidate	Yes	—
ra1	Maize	Transcriptional regulator (MYB); inflorescence structure	Candidate	Yes	—
su1, bt2, ae1	Maize	Starch biosynthetic enzymes	Candidate	Yes	—

Table 1. Genes of Interest in Crop Domestication and Improvement[1] (continued)

Gene(s)	Crop	Molecular and Phenotypic Function	Controls Phenotype[2]	Selection Evidence[3]	Causative Change
Genes Identified through Untargeted Selection Screens					
zagl1	Maize	Transcriptional regulator (MADS)	Unknown	Yes	—
17 genes	Maize	Varied functions, including auxin response, growth factor, kinase, methyl binding protein, transcription factors, amino acid biosynthesis and a circadian gene	Unknown	Yes	—
~30 genes	Maize	Varied functions, including auxin response, cell elongation protein, F-box protein, growth factor, heat shock proteins, hexokinase, kinase, steroid biosynthesis, transcription factors, amino acid biosynthesis and a circadian rhythm gene	Unknown	Yes	—

[1]For a version of the table with references for each gene, see Supplemental Data.
[2]For genes listed as "candidate," it is known that major mutations at these genes affect phenotype, but it has not been shown that natural allelic variation controls agronomically important differences between crops and progenitors or between crop varieties. For genes listed as "unknown," there is no experimental evidence demonstrating an effect of these genes on agronomic phenotypes in the crop listed.
[3]N.T. signifies not tested.

cell death or the release of hydrolytic enzymes that dissolve the bonds between cells in the abscission layer that separates the grain from the plant.

qSH1 is another major QTL controlling shattering in rice that has recently been cloned and shown to encode a homeobox containing transcription factor (Konishi et al., 2006). By a combination of genetic approaches, Konishi and colleagues were able to demonstrate that a single nucleotide change in a *cis*-regulatory element of *qSH1* obliterated expression of the cells that form the shattering zone, thus preventing shattering. *qSH1* was first identified in a segregating population from *japonica* x *indica* cross, indicating that it differentiates these two subspecies and suggesting that the loss of shattering involved independent genetic changes during their domestications.

Although we have only a small sample of domestication genes for major domestication traits, it is notable that none of the six domestication genes discussed above resulted from a null or loss-of-function mutation. In three cases, regulatory changes are inferred, in one an amino acid substitution is found, and in two cases there are combined regulatory and protein changes. Also, it is notable that five of the six are transcription factors and the fifth a likely cellular signaling (regulatory) gene. A longer list is needed before firm conclusions can be drawn, though we expect transcriptional regulators will continue to be over-represented among the major genes controlling morphological differences between crops and their progenitors.

Genes Controlling Varietal Differences

In addition to genes controlling classic domestication traits, many genes controlling differences between varieties of a single crop or important agronomic traits have been clearly identified (Table 1). Some of these genes have been discovered as QTL, whereas others segregate as Mendelian loci. For morphological and structural traits, there are several excellent examples. Grain number differences between rice varieties are controlled by *grain number1* (*gn1*), which encodes an oxidase/dehydrogenase that degrades the plant hormone cytokinin (Ashikari et al., 2005). Regulatory changes in some alleles and a premature stop codon in another allele both contribute to functional variation at *gn1*. In tomato, the difference between varieties with pear-shaped versus round fruits is controlled by *ovate*, a novel regulatory gene with a putative nuclear localization signal and homology to human Von Willebrand factor genes (Liu et al., 2002). The functional polymorphism appears to be an early stop codon that conditions the pear shape. In cole crops (*Brassica oleracea*), the *BoCAL* gene (a member of the MADS box family of transcriptional regulators) appears to be involved in the unusual inflorescence morphologies of broccoli and cauliflower, possibly due to an early stop codon (Smith and King, 2000; Purugganan et al., 2000).

The list of known genes contributing to physiological or biochemical differences between crop varieties is much longer (Table 1). Here are a few well-characterized examples. Mendel's wrinkled seed gene (*r*), which converts the field pea into the garden pea, is the result of an Ac/Ds-like transposon insertion that disrupts the coding sequence of a starch-branching enzyme (Bhattacharyya et al., 1990). In maize, *yellow1* (*y1*) encodes a kernel specific phytoene synthase that produces yellow kernels with high levels of carotenoids, a precursor for vitamin A synthesis (Palaisa et al., 2003). The functional difference appears to involve a change in promoter sequences such that *y1*, which is normally expressed in leaves, is expressed in developing kernels. When it was discovered that yellow corn provides improved nutrition for

farm animals, virtually all US corn was rapidly converted from white to yellow. In rice, glutinous ("sticky") varieties lack amylose as a consequence of an altered intron splice donor site in the amylose synthesis gene, *waxy* (Wang et al., 1995; Olsen et al., 2006). The soluble solids content of tomatoes, a key determinant for the quality of tomato paste, is influenced by a QTL named *brix9-2-5*, which encodes an invertase—an enzyme that cleaves sucrose into simple sugars (Fridman et al., 2004). The functional difference between the alleles with high and low activity is an amino acid change. Finally, the colorful red and blue hues of maize kernels, which were selected for aesthetic or cultural reasons by ancient peoples, are the result of variants in two transcriptional regulators (*c1* and *r1*) (Hanson et al., 1996). For both genes, alterations in the 5′ regulatory sequences, probably mediated by transposable elements, are responsible for the activation of these genes and the anthocyanin pathway that they control in maize kernels.

A notable feature of this partial list of genes controlling varietal differences is the high frequency of loss-of-function alleles (Table 1). However, there are several examples of regulatory change as well. An important caution concerning this list is that most of the genes characterized to date represent the "low-hanging fruit" that were easily cloned and characterized because they are major mutants in intensively studied biochemical pathways. At this point, the list is too short to draw firm conclusions beyond the observation that a diversity of functional classes is represented and null alleles are common.

Tests for Selection on Domesticated Species

QTL cloning of domestication genes is slow and labor intensive, and as a consequence relatively few domestication genes have been discovered by these means. An alternative, less costly approach is to ask whether a gene has been the target of human selection during domestication using population genetic analyses. The logic of this approach is straightforward: if a gene was the target of selection because it favorably influences a domestication or crop improvement trait, then it may show a decrease in nucleotide diversity, increased linkage disequilibrium (LD), and/or altered population frequencies of polymorphic nucleotides in the gene and linked regions (Smith and Haigh, 1974). If nucleotide polymorphism reveals such evidence, one can infer that the gene in question, or a closely linked gene, has been the target of human selection.

Crops are good systems to detect selection for two reasons (Wright and Gaut, 2005). First, selection has been intense and recent, and thus the signature of selection should be strong. Second, diversity can also be measured in the crop's extant wild relative. In many cases, the wild relative is a reasonable representative of the ancestral, predomestication population of the crop. By using extant wild populations as a proxy, genetic diversity can be contrasted before and after the domestication bottleneck. Recent studies have constructed statistical tests for selection that either take advantage of this contrast (Vigouroux et al., 2002; Tenaillon et al., 2004; Wright et al., 2005) or that control for demographic history (Nielsen et al., 2005). These approaches are relatively new, and several challenges remain (Hamblin et al., 2006). Nonetheless, they permit improved identification of genes with historical importance (Nielsen, 2005) and can be applied to any organism that experienced a recent bottleneck, including domesticated crops and animals (Pollinger et al., 2005) as well as species like humans (Akey et al., 2004) and *Drosophila* (Thornton and Andolfatto, 2006), both of which experienced a bottleneck during migration out of Africa.

Selection on Candidate Genes and Linked Regions

Tests for selection have been applied most commonly to data from genes for which there has been a priori evidence of a role in domestication or crop improvement. One example is *tb1* in maize, where the pattern of nucleotide polymorphism was particularly striking (Wang et al., 1999). As expected after a domestication bottleneck, the coding region of *tb1* contains less genetic diversity in maize than teosinte; the maize coding region retains ~40% of the genetic diversity in teosinte. The more surprising observation was that the reduction in diversity was far more severe in the 5′ untranslated region (UTR), where maize retains only 2% of teosinte diversity. Further, the pattern changed abruptly over a narrow ~100 base pair region. Based on these observations, Wang et al. (1999) made two conclusions. First, they concluded that selection targeted the *tb1* 5′ UTR during domestication, consistent with previous observations that *tb1* expression differs between maize and teosinte (Doebley et al., 1997). Second, based on the abrupt shift in the pattern of diversity, they concluded that recombination had been sufficient to uncouple the history of the 5′ UTR from the coding region.

Population genetic analyses have confirmed a history of selection in several more genes that were first identified functionally (Table 1). The work thus far is heavily biased toward maize and represents genes that contribute to plant and inflorescence architecture (*tb1*, *tga1*, *ba1*, *ra1*), to plant and kernel color (*c1*, *y1*), and to kernel composition (*bt2*, *ae1*, *su1*). Fewer examples of selected loci exist in other domesticated plants, but important nonmaize examples include the *waxy* gene in rice and the *BoCal* gene in cauliflower (Table 1). It is worth noting that most of these analyses have relied on standard tests of selection, which do not correct for demographic history and therefore can be misleading. Nonetheless, for most of these genes, the combination of functional and evolutionary analysis makes a credible case for a role in domestication or crop improvement.

Thus far, it is not clear whether recombination typically limits the effect of selection to a very small genomic region or whether large genomic regions are "dragged along" with selected genes via linkage (hitchhiking). To date, this question has only been addressed in two cases

in maize and one in rice. In one, Clark et al. (2004) measured nucleotide polymorphism upstream of *tb1* of maize. They found that the exceptionally low diversity typical of the 5′ UTR extended 60–90 kb upstream from *tb1*, and then diversity returned to normal levels. The hitchhiked region consisted only of intergenic DNA, and thus apparently strong selection on *tb1* did not affect the history of other genes (Clark et al., 2004). Similarly, Palaisa et al. (2004) characterized selection around the maize *y1* gene, which was a target of selection in the 1930s when the US farmers switched to yellow corn because of its nutritional value (Palaisa et al., 2003). Diversity levels suggest that hitchhiking has affected nucleotide polymorphism ~600 kb downstream and ~200 kb upstream from *y1* (Palaisa et al., 2004). This 800 kb region contains a handful of genes, demonstrating that selection on one gene (*y1*) can alter polymorphism at other genes in the case of extremely strong and rapid selection. Similarly, in rice, the *waxy* gene, which was the target of recent and strong selection, is flanked by a 250 kb region, including six other genes with reduced diversity as the result of hitchhiking (Olsen et al., 2006). Nonetheless, the general picture suggests that hitchhiking affects relatively few genes beyond those targeted by selection.

When Population Genetics Fails

Population genetic analyses have failed to verify selection on several candidate domestication genes such as *opaque2*, which influences kernel lysine content (Henry et al., 2005), and *zfl2*, which affects inflorescence structure (Bomblies and Doebley, 2005). For each of these genes, functional or QTL evidence suggests they contribute to agronomically important traits. Why do nucleotide polymorphism data fail to uncover a history of selection? There are at least four potential reasons. First, the gene may not have been a target of selection. The gene may contribute to a trait in a particular QTL study but may not have been prominent historically or in the genetic backgrounds available during domestication. Second, the study may have assayed polymorphism in the wrong genic region. For genes like *tb1*, a survey of polymorphism in the coding region would miss strong evidence for selection in the nearby promoter. Third, the statistical power to detect selection depends critically on sampling design and underlying levels of diversity. If diversity levels are low, it can be difficult to distinguish neutral from selected genes (Figure 2).

Finally, the ability to detect selection also depends on the history of the favored allele. Selection can be difficult to detect if the beneficial variant pre-existed as a common neutral polymorphism prior to domestication (Innan and Kim, 2004; Przeworski et al., 2005). In this special case, the variant had the opportunity to recombine onto a number of haplotypes prior to the onset of selection. When selection commenced, it favored the variant and dragged along multiple linked haplotypes. These different haplotypes may encompass substantial genetic diversity. As a result, selection does not substantially reduce genetic diversity around the selected site, and nucleotide polymorphism data may not provide a clear signature of a selection event. However, it is not clear whether this model conforms to reality. Many mutations for domestication traits, such as shattering, would have been deleterious in the wild population; thus, it is unlikely that such variants pre-existed as common, neutral alleles in wild populations.

Large-Scale Screens to Identify Selected Genes

Candidate gene approaches have several shortcomings. For example, you need a candidate, there may be an unmanageably large number of candidates, and if you nominate the wrong candidates, you may be wasting your efforts while the true selected genes pass undetected. Thus, an approach that ignores a priori information about gene function and instead assays large numbers of genes in a less biased manner can enable one to learn more than a candidate approach. This unbiased method, sometimes called a "selection screen," is now practical, given technology that allows the collection of polymorphism data on large numbers of genes and individuals. For some organisms, such as humans, whole genome selection screens are already possible.

To date, there have been few large-scale selection screens in crop plants. One study examined polymorphism at microsatellites found within ESTs (expressed sequence tags) (Vigouroux et al., 2002). The study examined 501 microsatellites, which prior work indicated were invariant in US maize. The authors argued that this lack of diversity in maize suggested that these 501 ESTs may be enriched for selected genes. Microsatellite diversity was then examined in a larger and geographically more diverse sample of maize and teosinte. The authors employed a battery of statistical tests, including standard neutrality tests and coalescent simulations that mimicked the domestication bottleneck, to detect selection from the polymorphism data. Microsatellites in 15 genes exhibited evidence of selection. Importantly, the authors verified the microsatellite-based inference by studying sequence polymorphism in one gene, a MADS-box transcription factor (Table 1). Similar screens of microsatellite diversity has been successfully used to identify genomic regions that show the signature of selection in both sorghum (Casa et al., 2005) and sunflower (Burke et al., 2005). In both studies, there was tendency for "selected" microsatellite to lie near known QTL for domestication traits; however, because of extended LD in these crops, the identification of the candidate selected genes will require further work.

A selection screen has also been applied to maize using nucleotide polymorphism data. Wright et al. (2005) compared sequence diversity between maize inbreds and teosintes in 774 genes. By implementing a novel approach, these authors circumvented two statistical pitfalls: (1) the problem of circularity in using the same genes for demographic estimation and for selection

tests and (2) the problem of maintaining statistical power when performing multiple tests. The approach also estimated the proportion of selected genes and ranked each gene by the posterior probability (PP) that it belonged in the selected group. Two to four percent of loci in the data set harbored the signature of selection. If this proportion is representative of the entire genome of ~59,000 genes (Messing et al., 2004), then ~1,200 maize genes bear a signature of selection, due either to domestication or recent crop improvement. The top 4% of candidate genes (about 30 genes) encompassed a range of predicted functions but appeared to be enriched for functions related to transcriptional regulation, plant growth, and amino acid biosynthesis (Table 1). Notably, these selected genes clustered near known maize domestication QTL in a statistically significant way.

In the third large-scale selection screen, Yamasaki et al. (2005) obtained nucleotide polymorphism data for 1200 maize genes. They focused on 35 genes with no nucleotide diversity in a set of 14 diverse maize inbreds, reasoning that this group with low diversity likely included selected genes. After assaying diversity in teosinte, 17 of 35 genes exhibited some evidence of selection, and eight of these showed selection under very stringent statistical criteria. The eight genes included an auxin-response factor, a transcription factor, genes related to amino acid biosynthesis, and a circadian rhythm gene (Table 1). In addition to identifying these genes, the study used a new strategy to distinguish between types of selection. Specifically, using a stratified sample of primitive and elite maize, this study was able to determine whether selection occurred early in the domestication process or later as a consequence of crop improvement.

Sorghum provides the only nonmaize examples of large-scale selection screens for DNA sequence diversity (Hamblin et al., 2004, 2006). Among the 371 loci that have been examined, none show compelling evidence for positive selection, although these genes appear not to be evolving in a strictly neutral manner. Hamblin et al. suggest that the non-neutral pattern of diversity is the result of demographic factors, such as population structure and migration, rather than selection during domestication. Thus, the success of selection screens in maize may not be realized in all other crops.

Altogether, selection screens in maize have identified >50 genes that have a history of selection consistent with contribution to agronomic traits. Their putative functions vary widely (Table 1), but some of the gene families and biological processes essential to maize domestication and improvement are becoming apparent. It is also statistically unavoidable that some selection candidates will be false positives and other candidates may show evidence of selection solely because they are linked to another gene that was the actual gene under selection. For these reasons, functional characterization of candidate genes is necessary, and additional characterization promises to yield important insights into the architecture of agronomic traits.

Perspective

In this article, we have reviewed two approaches to understanding the genetic changes that underlie crop domestication and improvement: (1) a classical genetic approach of starting with phenotype and working back to the gene and (2) a population genetic approach of starting with genes and asking whether these genes were targets of selection. From the list of genes identified using these approaches, we draw several conclusions.

First, the fact that five of the six major genes controlling morphological and structural changes during domestication are transcriptional regulators suggests to us that this class of genes played a central role in domestication. Although based on limited evidence, this conclusion is not surprising in that transcriptional regulators are also the dominant class of genes that regulate morphological development in plants (Doebley and Lukens, 1998). Given that domestication typically involved increases in organ size (fruit, seed, etc.), it appears likely that genes controlling cell division should also be over-represented among major domestication genes. *fw2.2* of tomato falls in this class. It has also been suggested that genes controlling meristem size and patterning, which are a function of cell division, will be important genes for domestication traits (Bomblies et al., 2003; Bommert et al., 2005), although, to date, there are no conclusive examples supporting this hypothesis.

Second, a diverse set of genes have contributed to varietal differences. For morphological traits and flowering time, transcriptional regulators are again over-represented, although a kinase and a cytokinin oxidase/reductase have also been identified. For the nutrient composition of seeds and fruits, basic enzymatic genes in the biosynthetic pathways predominate. It is striking that many of the genes involved in varietal differences harbor early stop codons or other lesions that disrupt the coding sequence, suggesting that these are loss-of-function mutations. However, regulatory changes are also frequent, indicating that changes in the levels and pattern of gene expression have been important.

Third, there is also a clear difference in the control of complex phenotypes (morphology, flowering time) versus simple phenotypes (accumulation of a specific metabolite). For the former class, transcriptional regulators predominate—10 of 17 genes are transcriptional regulators, and the remaining seven are a mixed group of regulatory genes (kinase, response regulator, hormone biosynthetic enzymes). For simple phenotypes, biosynthetic enzymes predominate—six of nine are biosynthetic enzymes, and three are transcriptional regulators of the pathway in question. This set of 26 genes is beginning to approach a large enough sample to expect that these proportions may remain roughly the same as the list of genes of large effect expands.

Fourth, selection screens appear to be revealing aspects of crop domestication and improvement that could have been missed by a classic phenotype-to-gene approach. We note that the range of molecular functions

among the selected genes is quite broad (Table 1). It is likely in our view that some of these genes were under selection for aspects of morphology or physiology that have never been considered targets of selection during crop domestication and improvement. Some examples of exceptional selected genes include an F-Box protein involved in the ubiquitin pathway, a methyl binding protein involved in epigenetic gene regulation, and heat-shock proteins. Clearly, much remains to be learned about how domestication sculpted crop genomes.

Can a knowledge of the genes contributing to past crop domestication and improvement guide future breeding efforts? We believe the answer is yes, and, in fact, plant breeding companies are already putting this knowledge into action. Some current efforts among leading plant breeding companies include up- and downregulating all the transcription factors in the genome, based on an understanding of the importance of this class of genes in regulating plant phenotypes. Other companies are applying selection screens to their own breeding lines to identify the genes that have contributed to the success of their best varieties. Similarly, QTL cloning for key agronomic genes is a focus. For known agronomic genes, companies are practicing "allele mining" or screening unimproved varieties and wild relatives to recover superior alleles that failed to pass through the domestication and improvement bottlenecks (Tanksley and McCouch 1997).

Can an understanding of the genetics of domestication and plant development in general catalyze the development of new crops or novel varieties of existing crops? Yes is the clear answer. Knowing that a naturally occurring allele of a phytoene synthase gene in maize produces high levels of provitamin A in kernels encouraged the development of crops like "golden rice" to help fight vitamin A deficiency in developing countries (Al-Babili and Beyer, 2005). The discovery that a MADS-box transcription factor (*FRUITFALL*) controls pod shattering in *Arabidopsis* allowed the development of canola with reduced pod-shattering via overexpression of *FRUITFALL* (Østergaard et al., 2006). Some authors have gone so far as to propose that the maize gene *tga1*, which regulates cob development, could be used to produce wheat, rye, and barley "on the cob" (Lev-Yadun et al., 2002). Although perhaps a fanciful idea, the extent to which plant development can be modified to meet human needs has certainly not yet reached it limits.

Over 10,000 years ago, human societies worldwide began to switch from economies based on gathering to ones based on agriculture. In doing so, they began a tradition of crop improvement that continues today. Ancient breeders accepted some trade-offs in this process, such as the 50% reduction in the protein content of domesticated cereal grains as compared to wild cereals in exchange for an increase in yield. However, it is they who built the foundation for the remarkably stable and plentiful food supply that we have today. There is every reason to expect that by using the full range of available tools modern breeders can further modify crops to improve their productivity and nutrition and to reduce the impact of humans on our environment.

Supplemental Data

Supplemental Data include one table and Supplemental References and can be found with this article online at http://www.cell.com/cgi/content/full/127/7/1309/DC1/.

ACKNOWLEDGMENTS

We thank E. Buckler, J. Faris, S. McCouch, D. Piperno, M. Purugganan, L. Rieseberg, T. Sang, and M. Zeder for helpful comments or other assistance. Work in our laboratories is supported by NSF grant DBI-0096033 and DBI-0321467 and NIH grant GM58816.

REFERENCES

Akey, J.M., Eberle, M.A., Rieder, M.J., Carlson, C.S., Shriver, M.D., Nickerson, D.A., and Kruglyak, L. (2004). Population history and natural selection shape patterns of genetic variation in 132 genes. PLoS Biol. *2*, e286.

Al-Babili, S., and Beyer, P. (2005). Golden Rice–five years on the road – five years to go? Trends Plant Sci. *10*, 565–573.

Anderson, E. (1969). Plants, man and life (Berkeley: University of California Press).

Ashikari, M., Sakakibara, H., Lin, S., Yamamoto, T., Takashi, T., Nishimura, A., Angeles, E.R., Qian, Q., Kitano, H., and Matsuoka, M. (2005). Cytokinin oxidase regulates rice grain production. Science *309*, 741–745.

Benz, B.F. (2001). Archaeological evidence of teosinte domestication from Guila Naquitz, Oaxaca. Proc. Natl. Acad. Sci. USA *98*, 2104–2106.

Bhattacharyya, M.K., Smith, A.M., Ellis, T.H., Hedley, C., and Martin, C. (1990). The wrinkled-seed character of pea described by Mendel is caused by a transposon-like insertion in a gene encoding starch-branching enzyme. Cell *60*, 115–122.

Blake, M. (2006). Dating the initial spread of *Zea mays*, In Histories of Maize, J. Staller, R. Tykot, and B. F. Benz, eds. (New York: Elsevier).

Bomblies, K., and Doebley, J.F. (2005). Molecular evolution of *FLORICAULA/LEAFY* orthologs in the Andropogoneae (Poaceae). Mol. Biol. Evol. *22*, 1082–1094.

Bomblies, K., Wang, R.L., Ambrose, B.A., Schmidt, R.J., Meeley, R.B., and Doebley, J. (2003). Duplicate FLORICAULA/LEAFY homologs *zfl1* and *zfl2* control inflorescence architecture and flower patterning in maize. Development *130*, 2385–2395.

Bommert, P., Lunde, C., Nardmann, J., Vollbrecht, E., Running, M., Jackson, D., Hake, S., and Werr, W. (2005). *thick tassel dwarf1* encodes a putative maize ortholog of the Arabidopsis *CLAVATA1* leucine-rich repeat receptor-like kinase. Development *132*, 1235–1245.

Burke, J.M., Knapp, S.J., and Rieseberg, L.H. (2005). Genetic consequences of selection during the evolution of cultivated sunflower. Genetics *171*, 1933–1940.

Cardon, G., Hohmann, S., Klein, J., Nettesheim, K., Saedler, H., and Huijser, P. (1999). Molecular characterisation of the Arabidopsis SBP-box genes. Gene *237*, 91–104.

Casa, A.M., Mitchell, S.E., Hamblin, M.T., Sun, H., Bowers, J.E., Paterson, A.H., Aquadro, C.F., and Kresovich, S. (2005). Diversity and selection in sorghum: simultaneous analyses using simple sequence repeats. Theor. Appl. Genet. *111*, 23–30.

Clark, R.M., Linton, E., Messing, J., and Doebley, J.F. (2004). Pattern of diversity in the genomic region near the maize domestication gene

tb1. Proc. Natl. Acad. Sci. USA *101*, 700–707.

Cong, B., Liu, J., and Tanksley, S.D. (2002). Natural alleles at a tomato fruit size quantitative trait locus differ by heterochronic regulatory mutations. Proc. Natl. Acad. Sci. USA *99*, 13606–13611.

Cubas, P., Lauter, N., Doebley, J., and Coen, E. (1999). The TCP domain: a motif found in proteins regulating plant growth and development. Plant J. *18*, 215–222.

Doebley, J. (1989). Isozymic evidence and the evolution of crop plants, In Isozymes in Plant Biology, D. Soltis, and P. Soltis, eds. (Portland, Oregon: Dioscorides Press), pp. 165-191.

Doebley, J. (2004). The genetics of maize evolution. Annu. Rev. Genet. *38*, 37–59.

Doebley, J., and Lukens, L. (1998). Transcriptional regulators and the evolution of plant form. Plant Cell *10*, 1075–1082.

Doebley, J., Stec, A., and Hubbard, L. (1997). The evolution of apical dominance in maize. Nature *386*, 485–488.

Dorweiler, J., and Doebley, J. (1997). Developmental analysis of *teosinte glume architecture1*: a key locus in the evolution of maize (Poaceae). Am. J. Bot. *87*, 1313–1322.

Eyre-Walker, A., Gaut, R.L., Hilton, H., Feldman, D.L., and Gaut, B.S. (1998). Investigation of the bottleneck leading to the domestication of maize. Proc. Natl. Acad. Sci. USA *95*, 4441–4446.

Frary, A., Nesbitt, T.C., Grandillo, S., Knaap, E., Cong, B., Liu, J., Meller, J., Elber, R., Alpert, K.B., and Tanksley, S.D. (2000). *fw2.2*: a quantitative trait locus key to the evolution of tomato fruit size. Science *289*, 85–88.

Fridman, E., Carrari, F., Liu, Y., Fernie, A., and Zamir, D. (2004). Zooming in on a quantitative trait for tomato yield using interspecific introgressions. Science *305*, 1786–1789.

Hamblin, M.T., Mitchell, S.E., White, G.M., Gallego, J., Kukatla, R., Wing, R.A., Paterson, A.H., and Kresovich, S. (2004). Comparative population genetics of the panicoid grasses: sequence polymorphism, linkage disequilibrium and selection in a diverse sample of sorghum bicolor. Genetics *167*, 471–483.

Hamblin, M.T., Casa, A.M., Sun, H., Murray, S.C., Paterson, A.H., Aquadro, C.F., and Kresovich, S. (2006). Challenges of detecting directional selection after a bottleneck: lessons from *Sorghum bicolor.* Genetics *173*, 953–964.

Hammer, K. (1984). Das Domestikationssyndrom. Kulturpflanze *32*, 11–34.

Hanson, M.A., Gaut, B.S., Stec, A.O., Fuerstenberg, S.I., Goodman, M.M., Coe, E.H., and Doebley, J.F. (1996). Evolution of anthocyanin biosynthesis in maize kernels: the role of regulatory and enzymatic loci. Genetics *143*, 1395–1407.

Harris, D. (1996). The origins and spread of agriculture and pastoralism in Eurasia. (Washington, D.C.: Smithsonian Institution Press).

Henry, A.M., Manicacci, D., Falque, M., and Damerval, C. (2005). Molecular evolution of the Opaque-2 gene in Zea mays L. J. Mol. Evol. *61*, 551–558.

Heun, M., Schäfer-Pregl, R., Klawan, D., Castagna, R., Accerbi, M., Borghi, B., and Salamini, F. (1997). Site of Einkorn Wheat Domestication Identified by DNA Fingerprinting. Science *278*, 1312–1314.

Innan, H., and Kim, Y. (2004). Pattern of polymorphism after strong artificial selection in a domestication event. Proc. Natl. Acad. Sci. USA *101*, 10667–10672.

Jaenicke-Despres, V., Buckler, E.S., Smith, B.D., Gilbert, M.T., Cooper, A., Doebley, J., and Paabo, S. (2003). Early allelic selection in maize as revealed by ancient DNA. Science *302*, 1206–1208.

Konishi, S., Izawa, T., Lin, S.Y., Ebana, K., Fukuta, Y., Sasaki, T., and Yano, M. (2006). An SNP caused loss of seed shattering during rice domestication. Science *312*, 1392–1396.

Kosugi, S., and Ohashi, Y. (2002). DNA binding and dimerization specificity and potential targets for the TCP protein family. Plant J. *30*, 337–348.

Lev-Yadun, S., Abbo, S., and Doebley, J. (2002). Wheat, rye, and barley on the cob? Nat. Biotechnol. *20*, 337–338.

Li, C., Potuschak, T., Colon-Carmona, A., Gutierrez, R.A., and Doerner, P. (2005). Arabidopsis TCP20 links regulation of growth and cell division control pathways. Proc. Natl. Acad. Sci. USA *102*, 12978–12983.

Li, C., Zhou, A., and Sang, T. (2006). Rice domestication by reducing shattering. Science *311*, 1936–1939.

Liu, J., Van Eck, J., Cong, B., and Tanksley, S.D. (2002). A new class of regulatory genes underlying the cause of pear-shaped tomato fruit. Proc. Natl. Acad. Sci. USA *99*, 13302–13306.

Londo, J.P., Chiang, Y.C., Chiang, T.Y., and Schall, B.A. (2006). Phylogeography of Asian wild rice, *Oryza rufipogon*, reveals multiple independent domestications of cultivated rice, *Oryza sativa.* Proc. Natl. Acad. Sci. USA *103*, 9578–9583.

Matsuoka, Y., Vigouroux, Y., Goodman, M.M., Sanchez, G.J., Buckler, E., and Doebley, J. (2002). A single domestication for maize shown by multilocus microsatellite genotyping. Proc. Natl. Acad. Sci. USA *99*, 6080–6084.

Smith, M.J. and Haigh, J. (1974). The hitch-hiking effect of a favourable gene. Genet. Res. *23*, 23–25.

Messing, J., Bharti, A.K., Karlowski, W.M., Gundlach, H., Kim, H.R., Yu, Y., Wei, F., Fuks, G., Soderlund, C.A., Mayer, K.F., and Wing, R.A. (2004). Sequence composition and genome organization of maize. Proc. Natl. Acad. Sci. USA *101*, 14349–14354.

Nesbitt, T.C., and Tanksley, S.D. (2002). Comparative sequencing in the genus Lycopersicon. Implications for the evolution of fruit size in the domestication of cultivated tomatoes. Genetics *162*, 365–379.

Nielsen, R. (2005). Molecular signatures of natural selection. Annu. Rev. Genet. *39*, 197–218.

Nielsen, R., Bustamante, C., Clark, A.G., Glanowski, S., Sackton, T.B., Hubisz, M.J., Fledel-Alon, A., Tanenbaum, D.M., Civello, D., White, T.J., et al. (2005). A scan for positively selected genes in the genomes of humans and chimpanzees. PLoS Biol. *3*, e170.

Olsen, K.M., Caicedo, A.L., Polato, N., McClung, A., McCouch, S., and Purugganan, M.D. (2006). Selection under domestication: evidence for a sweep in the rice waxy genomic region. Genetics *173*, 975–983.

Østergaard, L., Kempin, S., Bies, D., Klee, H., and Yanofsky, M. (2006). Pod shattering-resistant *Brassica* fruit produced by ectopic expression of the *FRUITFALL* gene. Plant Biotechnol J. *4*, 45–51.

Palaisa, K., Morgante, M., Tingey, S., and Rafalski, A. (2004). Long-range patterns of diversity and linkage disequilibrium surrounding the maize *Y1* gene are indicative of an asymmetric selective sweep. Proc. Natl. Acad. Sci. USA *101*, 9885–9890.

Palaisa, K.A., Morgante, M., Williams, M., and Rafalski, A. (2003). Contrasting effects of selection on sequence diversity and linkage disequilibrium at two phytoene synthase loci. Plant Cell *15*, 1795–1806.

Piperno, D.R., and Pearsall, D. (1998). The origins of agriculture in the lowland Neotropics (New York: Academic Press).

Piperno, D.R., and Flannery, K.V. (2001). The earliest archaeological maize (*Zea mays* L.) from highland Mexico: new accelerator mass spectrometry dates and their implications. Proc. Natl. Acad. Sci. USA *98*, 2101–2103.

Pollinger, J.P., Bustamante, C.D., Fledel-Alon, A., Schmutz, S., Gray,

M.M., and Wayne, R.K. (2005). Selective sweep mapping of genes with large phenotypic effects. Genome Res. *15*, 1809–1819.

Przeworski, M., Coop, G., and Wall, J.D. (2005). The signature of positive selection on standing genetic variation. Evolution Int. J. Org. Evolution *59*, 2312–2323.

Purugganan, M.D., Boyles, A.L., and Suddith, J.I. (2000). Variation and selection at the *CAULIFLOWER* floral homeotic gene accompanying the evolution of domesticated *Brassica oleracea.* Genetics *155*, 855–862.

Rick, C., and Fobes, J. (1975). Allozyme variation in cultivated tomato and closely related species. Bull. Torrey Bot. Club *102*, 376–384.

Simons, K.J., Fellers, J.P., Trick, H.N., Zhang, Z., Tai, Y.S., Gill, B.S., and Faris, J.D. (2006). Molecular characterization of the major wheat domestication gene *q.* Genetics *172*, 547–555.

Smith, B.D. (1998). The emergence of agriculture (New York: W. H. Freeman).

Smith, B.D. (2001). Documenting plant domestication: the consilience of biological and archaeological approaches. Proc. Natl. Acad. Sci. USA *98*, 1324–1326.

Smith, L., and King, G. (2000). The distribution of *BoCal-a* alleles in *Brassica oleracea* is consistent with a genetic model for curd development and domestication of the cauliflower. Mol. Breed. *6*, 603–613.

Tanksley, S.D., and McCouch, S.R. (1997). Seed banks and molecular maps: unlocking genetic potential from the wild. Science *277*, 1063–1066.

Tenaillon, M.I., Uren, J., Tenaillon, O., and Gaut, B.S. (2004). Selection versus demography: a multilocus investigation of the domestication process in maize. Mol. Biol. Evol. *21*, 1214–1225.

Thornton, K.R., and Andolfatto, P. (2006). Approximate Bayesian Inference Reveals Evidence for a Recent, Severe Bottleneck in a Netherlands Population of *Drosophila melanogaster.* Genetics *172*, 1607–1619.

Vigouroux, Y., McMullen, M., Hittinger, C.T., Houchins, K., Schulz, L., Kresovich, S., Matsuoka, Y., and Doebley, J. (2002). Identifying genes of agronomic importance in maize by screening microsatellites for evidence of selection during domestication. Proc. Natl. Acad. Sci. USA *99*, 9650–9655.

Wang, H., Nussbaum-Wagler, T., Li, B., Zhao, Q., Vigouroux, Y., Faller, M., Bomblies, K., Lukens, L., and Doebley, J.F. (2005). The origin of the naked grains of maize. Nature *436*, 714–719.

Wang, R.L., Stec, A., Hey, J., Lukens, L., and Doebley, J. (1999). The limits of selection during maize domestication. Nature *398*, 236–239.

Wang, Z.Y., Zheng, F.Q., Shen, G.Z., Gao, J.P., Snustad, D.P., Li, M.G., Zhang, J.L., and Hong, M.M. (1995). The amylose content in rice endosperm is related to the post-transcriptional regulation of the waxy gene. Plant J. *7*, 613–622.

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

Wright, S.I., and Gaut, B.S. (2005). Molecular population genetics and the search for adaptive evolution in plants. Mol. Biol. Evol. *22*, 506–519.

Yamasaki, M., Tenaillon, M.I., Bi, I.V., Schroeder, S.G., Sanchez-Villeda, H., Doebley, J.F., Gaut, B.S., and McMullen, M.D. (2005). A large-scale screen for artificial selection in maize identifies candidate agronomic loci for domestication and crop improvement. Plant Cell *17*, 2859–2872.

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.