Evolution of crop species: genetics of domestication and diversification

Rachel S. Meyer[1] and Michael D. Purugganan[1,2]

Abstract | Domestication is a good model for the study of evolutionary processes because of the recent evolution of crop species (<12,000 years ago), the key role of selection in their origins, and good archaeological and historical data on their spread and diversification. Recent studies, such as quantitative trait locus mapping, genome-wide association studies and whole-genome resequencing studies, have identified genes that are associated with the initial domestication and subsequent diversification of crops. Together, these studies reveal the functions of genes that are involved in the evolution of crops that are under domestication, the types of mutations that occur during this process and the parallelism of mutations that occur in the same pathways and proteins, as well as the selective forces that are acting on these mutations and that are associated with geographical adaptation of crop species.

Domestication has always been considered a unique form of biological evolution — a co-evolutionary interaction that leads to the establishment of new domesticated species, the growth and reproduction of which are mostly controlled for the benefit of another species. Domestication has been documented to have evolved at least five times in evolutionary history, and classic examples include the cultivation of fungal species by attine ants, ambrosia beetles and termites[1]. However, the most prolific domesticators are humans, who have domesticated hundreds of plant species (BOX 1; see Supplementary information S1 (table)) and animal[2] species as sources of food and materials, and even for companionship and aesthetic value in the past 12,000 years. Crops, in particular, represent some of the most marked evolutionary transitions that are associated with domestication, which has prompted interest in their study since Darwin drew inspiration from domesticated species to illuminate genetic variation[3], evolution and the power of selection[4]. Research on such crop evolutionary processes is also driven by its cultural and economic importance for humans.

The genetic architecture of crop domestication and the nature of selection in domesticated species have been major foci of molecular genetic studies over the past two decades. A large number of domestication genes (or domestication-related genes) have been identified and isolated through candidate gene studies, quantitative trait locus (QTL) mapping and cloning, genome-wide association studies (GWASs) and, more recently, whole-genome resequencing studies. In these genes, widespread footprints of selection have been identified in the genomes of maize, rice, sunflower and several millet species, which allow us to better understand the forces of both conscious selection and unconscious selection. Recent population-level molecular analyses also enable us to clarify the demographic histories of the domestication process itself (for example, the processes of domesticating rice[5] and tomato[6]), which, together with expanded archaeological studies, can illuminate the origins and histories of crops[7,8]. Furthermore, the characterization of the mutations that lead to domestication gives an indication of the types of mutations and the functions of genes that are involved in the generation of domestication traits. Progress made in the past few decades now provides us with the foundation to examine patterns and processes that are associated with crop plant evolution, and to focus on the genetics of their domestication and diversification since the Neolithic period.

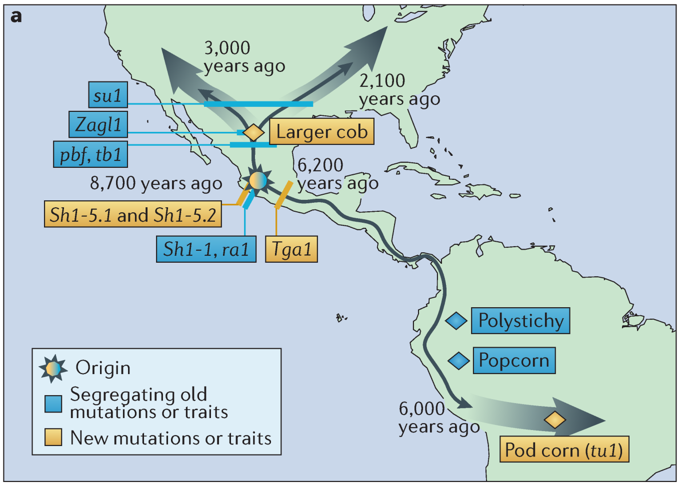
In this Review, we discuss the genetic architecture of crop plant domestication and investigate the evolutionary genomics of this important process. By compiling a list of known domestication and diversification genes, we dicuss patterns of selection over the course of the domestication process and also examine the origin and spread of domestication alleles. Finally, we show how these molecular genetic insights have led to a more robust characterization of the evolutionary development of crop species.

Quantitative trait locus
(QTL). A genomic region with a gene (or multiple linked genes) that contains mutations which result in phenotypic variation in populations.

Genome-wide association studies
(GWASs). Studies that use linkage disequilibrium between dense, usually single-nucleotide polymorphism, markers across the genome to identify significant associations between genes (or genomic regions) and trait phenotypes.

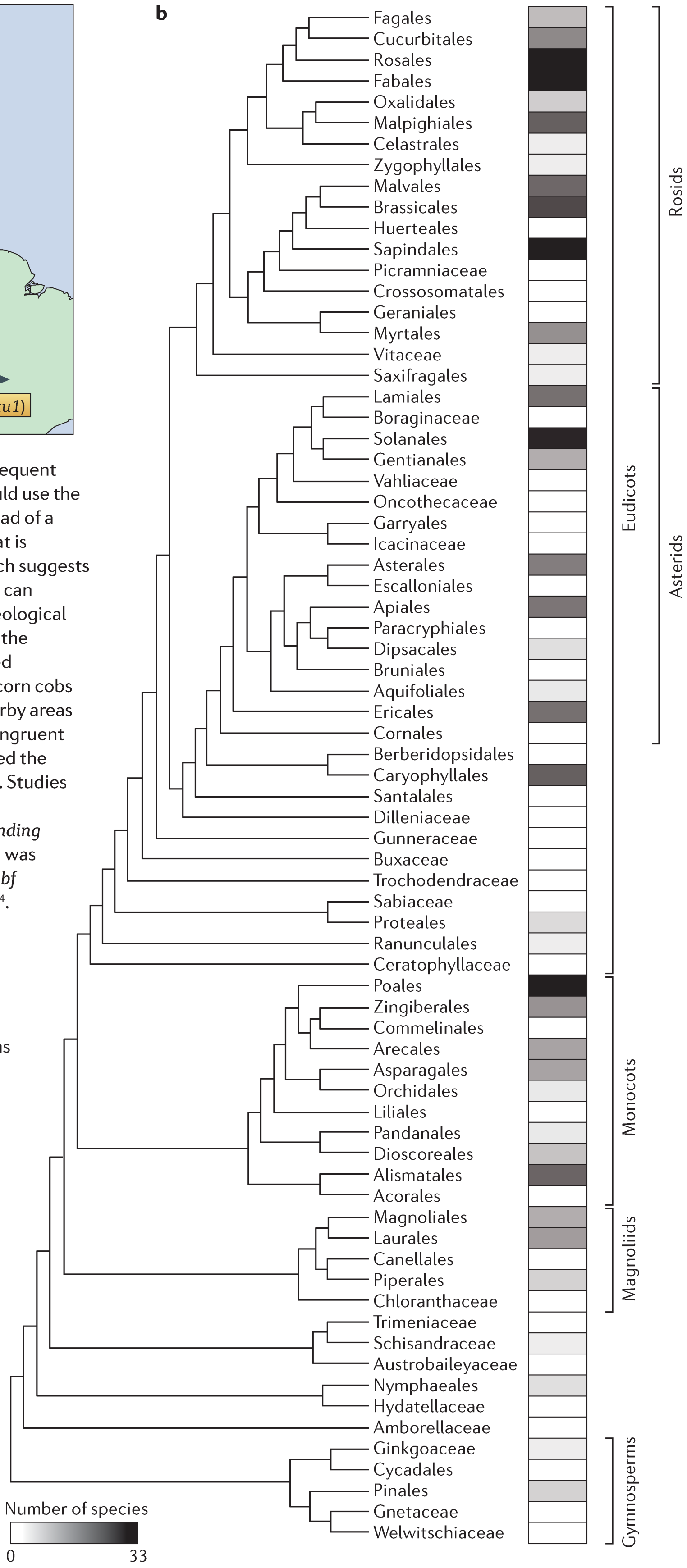
[1]*Center for Genomics and Systems Biology, Department of Biology, 12 Waverly Place, New York University, New York 10003, USA.*
[2]*Center for Genomics and Systems Biology, New York University Abu Dhabi Research Institute, Abu Dhabi, United Arab Emirates.*
Correspondence to M.D.P.
e-mail: mp132@nyu.edu
doi:10.1038/nrg3605

Box 1 | Geographical and phylogenetic distribution of domestication genes and species



Crop species are domesticated at particular locales, and their subsequent range expansion leads to wider geographical distribution. One could use the distribution of ancestral haplotypes to infer the origin and the spread of a particular domestication allele; for example, the rice haplotype that is ancestral to the glutinous allele is prevalent in Southeast Asia, which suggests that the phenotype originated in this region[85]. In special cases, one can directly examine the spread of alleles; for example, work on archaeological samples[93] and ancient DNA samples[94] of maize directly documents the accumulation, spread and fixation of mutations and their associated phenotypic evolution. Archaeological samples of ~8,700-year-old corn cobs have been recovered from southern Mexico, and samples from nearby areas dating 1,500 years later are the first to show a phenotype that is congruent with the *teosinte glume architecture1* (*Tga1*) mutation which exposed the grains and prevented the detachment of the grain from the cob[95,96]. Studies of ancient DNA samples reveal that another 2,500 years later, in northeastern Mexico, *teosinte branched1* (*tb1*) and *prolamin-box binding factor* (*pbf*) domestication alleles were fixed, whereas *sugary1* (*su1*) was fixed in North America much later (see the figure, part **a**). *tb1* and *pbf* altered meristem activity and kernel protein content, respectively[94]. Archaeological evidence shows that genes that are responsible for polystichy (that is, the whorled multiple-ranked maize ear that differs from the two-ranked inflorescence of teosinte) had already undergone selection as maize was brought into Peru ~6,200 years ago, and that these varieties retained the popcorn phenotype[21]. DNA sequence patterns in extant varieties indicate when mutations underwent selection: for mutations *Sh1-5.1*, *Sh1-5.2*, *Sh1-1* and ramosa1 (*ra1*) selection occurred at the onset of domestication, and for *Zagl1* and *tunicate* (*tu1*) it occurred during diversification.

The genetic changes that are seen in examples of food crop domestication have mostly been studied in the Poaceae family of grasses, but these represent only a small proportion of the total domesticated food crops. More than 160 plant families — mostly within the monocots and core eudicots, but also including many non-flowering plants and basal eudicots that are distant from model plants — have been found to contain domesticated species (see Supplementary information S1 (table)). Some families, such as Rosaceae, Malpighiaceae and Sapindaceae, contain the highest number of domesticated food crops, but few crops from these families have been studied to understand the genes and processes of domestication. Future foci in the topic of domestication will branch out to include non-model plant species from different environments that have different domestication traits. Part **b** of the figure shows a summary phylogeny[97,98] of land plants and the distribution of 353 domesticated food crops as shaded boxes (see Supplementary information S1 (table)).



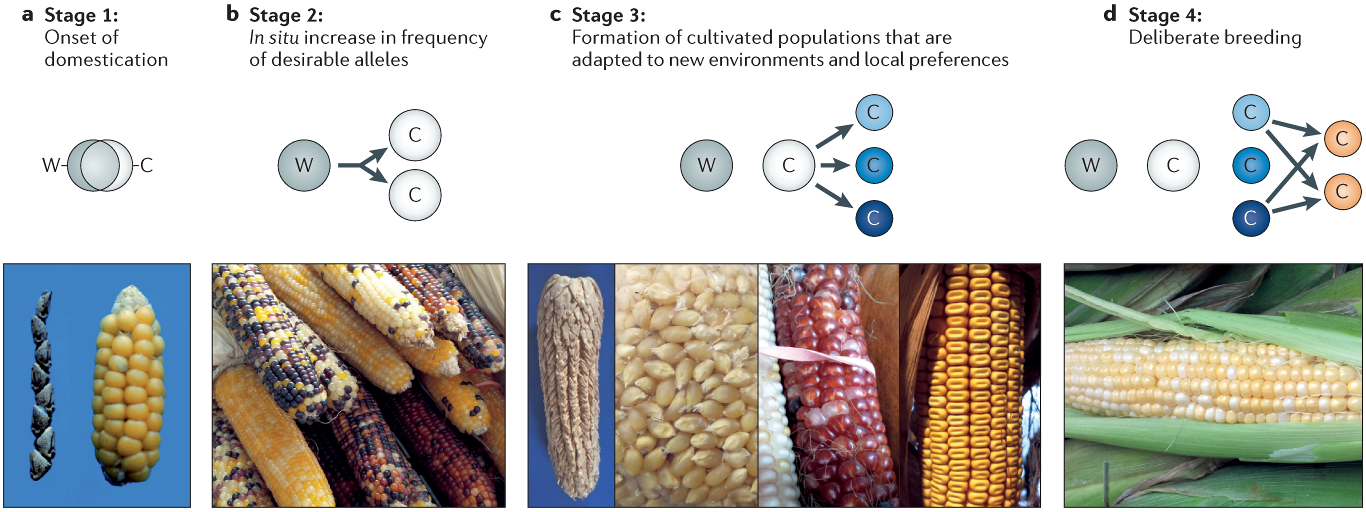


Figure 1 | **The evolutionary stages of domestication and diversification.** Plant exploitation involves harvesting and stewardship over wild stands with favourable traits (pre-Stage 1). The top panel schematically shows the evolutionary stages of crop plants, including the formation and the diversification of phenotypically distinct cultivated populations (C) from wild populations (W). Each circle represents a population and those of different phenotypic characteristics are shown in different shades of the same colour. Arrows represent the evolutionary establishment of derived populations from ancestral populations. This is illustrated by the example of *Zea mays* in the bottom panel. **a** | After extended tending of stands or the development and establishment of cultivation, selection occurs on the new crop in an agricultural ecosystem, which leads to the onset of domestication (Stage 1). Teosinte (left) and reconstructed primitive maize (right) are shown. Following Stage 1 is a crop diversification phase, which can encompass three non-exclusive stages. **b** | Stage 2 is the continuation of Stage 1 and involves the *in situ* amplification of populations with desirable alleles that lead to initial increases in yield, as well as the selection of favourable crop phenotypes. Trait variation also increases. Corn varieties that resemble those at its centre of origin (that is, the Mexican Highlands) are shown. **c** | As domesticated crops evolve and spread from their initial geographical range (Stage 3), crop populations are adapted to new diversified environments and local preferences. Pod corn (left) was selected for ceremonial use by Native Americans; popcorn (middle left) is preferred in Peru; Italian red sweet corn (middle right) has also been selected; and dent corn (right), which is used to make hominy and masa, is selected by Native Americans. **d** | Stage 4 is the deliberate breeding of crop varieties to maximize yield, ease of farming, uniformity and quality. Uniform improved corn varieties are produced through modern deliberate breeding efforts. Image in part **a** courtesy of J. Doebley, the University of Wisconsin–Madison, USA. Pod corn image in part **c** is reproduced, with permission, from REF. 120 © (2012) US National Academy of Sciences.

Domestication and diversification

Plant domestication by humans encompasses a broad spectrum of evolutionary changes that may decrease the fitness of a plant in the wild but increase it under human exploitation, and complete dependence on humans for survival is considered the fullest extent of domestication. The domestication process under this broad definition can span a wide range of features in crop species evolution; for the purpose of this Review, we use domestication to refer to the onset or the initiation of the process of evolutionary divergence from the wild ancestral species. We use diversification to refer to the subsequent evolution of new varieties, including greater improvement in yield, adaptation or quality in crop species.

Stages of domestication and diversification. Human-associated plant domestication began ~12,000 years ago in the Middle East and the Fertile Crescent, and subsequently in different parts of the world — China, Mesoamerica, the Andes, Near Oceania (all ~10,000 years ago), sub-Saharan Africa (~8,000 years ago) and eastern North America (~6,000 years ago[9]). The evolution of crop plants began as human behavioural ecology changed from food gathering to cultivation as the primary mode of supplying plant food resources[10]. Domesticated plant species are found in 160 taxonomic families (BOX 1; see Supplementary information S1 (table)), with estimates that 2,500 species have undergone domestication[11], and 250 species are considered as fully domesticated[2,12]. The evolutionary trajectory from wild species to crop species is a complex multi-staged process. Archaeological records suggest that there was a period of pre-domestication cultivation while humans first began to deliberate planting or caring for wild stands that have favourable traits (pre-Stage 1)[13]; as human-associated cultivation reshaped the evolutionary trajectories of these species, they were transformed into domesticated species (Stage 1) (FIG. 1). Little is known about the pre-domestication stage; although the domestication process itself was previously thought to be rapid[14], increasing numbers of studies suggest a protracted period for Stage 1 that could last as long as 2,000 years[15].

The diversification phase that follows initial domestication — sometimes referred to as the improvement phase[16] — involves the spread and adaptation of the

Conscious selection
The intentional choice, made by humans, of preferred phenotypes in cultivated plants for use and propagation.

Unconscious selection
Natural selection in crop species as a result of human cultivation practices and of growth in agro-ecological environments.

domesticated species to different agro-ecological and cultural environments. This phase leads to phenotypic and genetic divergence among domesticated populations, and it can be thought of as having multiple stages that are associated with varying selective pressures[17]. Some key post-domestication stages may include *in situ* amplification of populations that have desirable alleles (Stage 2); adaptation of a domesticated species to different environments and human cultural practices that accompany geographical radiation (Stage 3); and deliberate breeding to maximize yield, ease of farming and quality (Stage 4) (FIG. 1). Stages 1–3 have previously been described on the basis of the domestication history of seed crops[17]; although these stages are often sequentially presented, they may occur simultaneously. Conscious and deliberate breeding of plants in Stage 4 has been practised as far back as 11,400 years ago (for example, the hybrid breeding of figs[18]), but many traits in crop species during this stage are associated with modern breeding methods (for example, the Green Revolution).

Green Revolution
A series of research, breeding and technology transfer programmes in the mid-twentieth century that resulted in marked increases in agricultural productivity in developing countries.

Complementation
Introduction of a wild-type allele into a mutant individual, through either genetic crosses or transgenic methods, to confirm that a particular gene causes a specific phenotype.

Causative mutations
Mutations that lead to altered gene functions, which result in specific phenotypes.

Fixation
Increase in the frequency of an allelic variant until it is found in all individuals in a population.

Selective sweeps
Rapid increases in population frequencies of positively selected mutations and linked neutral mutations, which result in significant reductions in nucleotide diversity in localized regions of the genome that flank the selected mutations.

Introgression
Recurrent crossing that leads to the sharing of alleles between gene pools (which can be unidirectional), such as between domesticated and wild populations.

Genetic bottlenecks
Marked decreases in genetic diversity that are caused by reductions in effective population sizes.

Domestication traits. Which traits were selected during domestication or post-domestication diversification stages can vary depending on the species, as well as on the nature and the number of domestication events (FIG. 2). Domestication phenotypes are, by definition, traits that are selected during the initial transformation and establishment of the new domesticated species from its wild ancestor (or ancestors); these phenotypes often include the loss of dormancy, increases in seed size and changes in reproductive shoot architecture (TABLE 1). These traits can arise through human preferences for ease of harvest, growth advantages under human propagation and/or survival in deforested or disturbed habitats[17]. Both conscious selection by early farmers and unconscious selection as a result of agricultural practices or environments[19] are involved in the domestication process.

Diversification traits. Diversification traits among crop plant species can be even more varied (TABLE 1). They can be seen as variation in domesticated populations, as they result from crops that are adapting to fit specific uses, preferences and ecological growing conditions. For example, photoperiod sensitivity in wheat and barley arose as a phenotype when cultivars spread out of the Fertile Crescent[20]. Other traits, such as sticky or aromatic grains in rice and popcorn in maize[21], were selected and maintained by specific cultures. For many, if not most, of these diversification traits, it is likely that they evolve under conscious selection. Adding to this complexity in inferring whether a trait has been selected during crop evolution is the fact that the functional use and the specific organs that are targeted for selective change can differ over time[9]; for example, the initial domestication of lettuce in Egypt involved selection for oilseed production, whereas current selective breeding efforts focus on leaf characteristics[22].

Characterizing genetic architecture

Identifying domestication and diversification genes. Domestication or diversification genes have mostly been isolated through QTL fine-mapping studies and, more recently, by linkage disequilibrium mapping using GWASs, and transgenic or genetic complementation analyses are used to conclusively identify the relevant genes. This has primarily been undertaken in maize or rice, in which high-density genetic maps and molecular markers, as well as considerable genetic resources, allow thorough molecular characterization and high-resolution mapping.

Identifying causative mutations that lead to domestication or diversification phenotypes in these loci can be difficult. Few studies have used site-directed mutagenesis or transgenic complementation to directly test for the functional effects of specific mutations. However, in several studies, clear functional consequences of identified mutations in crop evolution genes — for example, premature stop codons, and insertion and deletions (indels) — have led to the inference that they are the causative mutations (see Supplementary information S2,S3 (tables)).

As a result of uncertainties in the phenotypes that are associated with specific stages in the evolution of domesticated species, it may also be problematic to distinguish genes that underlie domestication from those that give rise to subsequent diversification traits. Many genes that underlie phenotypes which distinguish a domesticated species from its wild ancestor have been labelled as domestication genes, although, in many cases, there is no evidence that these phenotypes arose as a result of selection during the domestication process[23].

We propose that a domestication gene should meet the following criteria. First, its function has been characterized and is known to underlie a trait — for example, a loss of seed dispersal and an increase in seed size — that is clearly associated with Stage 1 (that is, domestication) in the species of interest. Second, there is evidence of positive selection at that locus. Third, there should be complete or near-complete fixation of at least one causative mutation that is associated with the gene in all lineages from a single domestication event. Applying these criteria can prove difficult, as there may be multiple selective pressures that affect the same trait, domestication traits may be poorly characterized, and selection signatures can be difficult to detect. Moreover, soft selective sweeps on standing genetic variation rather than on new genetic variation, introgression or severe genetic bottlenecks can obscure the evolutionary and selective history of a locus. Thus, under our conservative criteria, we may not identify the full range of domestication genes in a crop species; nevertheless, these criteria can provide an initial appraisal of relevant genes that are associated with the origin of a crop species.

Diversification or improvement genes are selected for after the domestication process in Stage 1 and are associated with Stages 2 to 4 (REF. 24). Defining that a gene is involved in diversification and not in domestication is aided by knowledge of the population structure of the domesticated species and by information on early cultivated forms from the archaeological record to delimit early evolving traits versus late evolving traits. Several loci, such as *FW2.2* (also known as *LOC101245309*) in the Solanaceae[25] and *suppressor of*

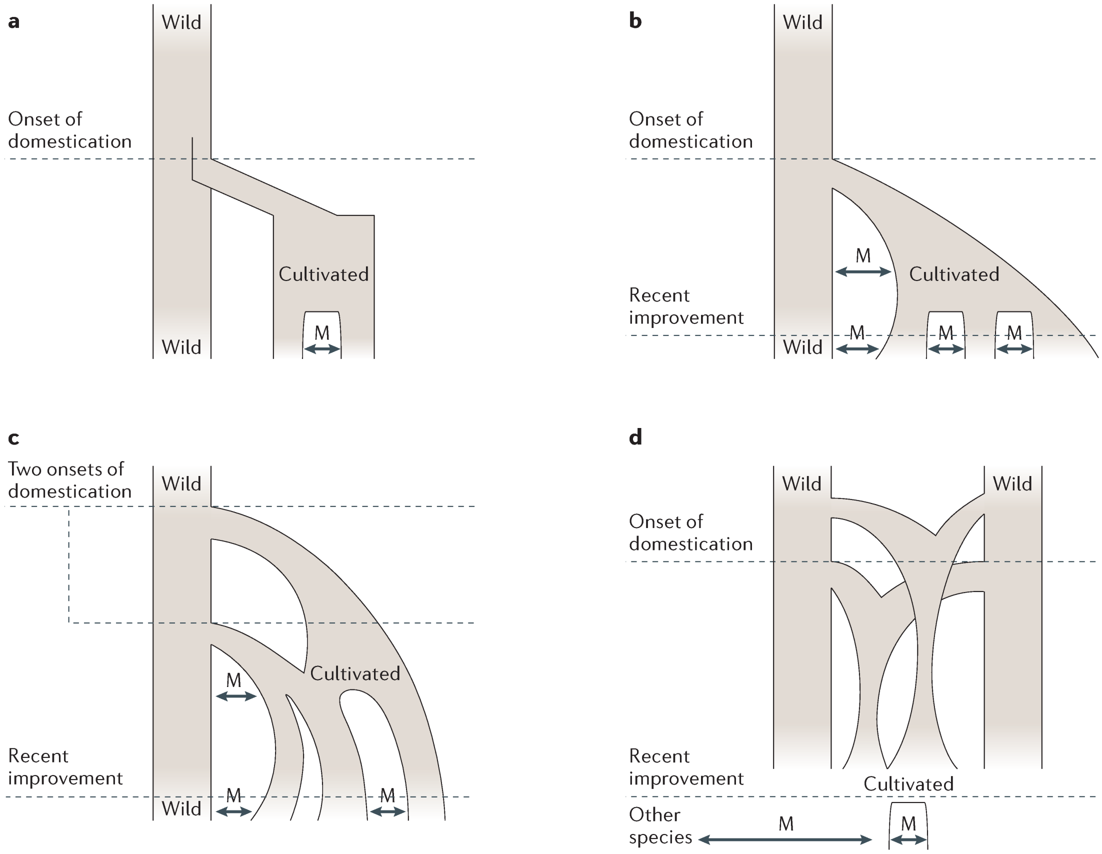


Figure 2 | **Demographic models of crop domestication.** The characterization of domestication in crop species is dependent on understanding the initiation and the course of the domestication process. The width of the channels represents population size and geographical range; $M = N_e m$, which is the product of effective population size (N_e) and the migration rate (m). **a** | Earlier models of domestication posited a single domestication event and suggested that domestication occurred through strong selection and severe genetic bottlenecks in a small population of the wild progenitor, which resulted in greater reproductive isolation between the wild species and the domesticated species[111]. As more archaeological and molecular data are now available and the evolutionary histories of more crops are better known, new general models for domestication have been proposed. New alternative demographic models of domestication suggest that the extent of the genetic bottleneck in the early evolution of crop species is variable — severe during the domestication of corn[43], but minimal for that of apple[112] and carrot[113]. Even after a domestication bottleneck, diversity can recover during the improvement or diversification phase through processes such as introgression from wild relatives[43]. Furthermore, strong reproductive isolation is not a necessary feature of domestication[114], and repeated introgression between crops and their wild progenitors or other related species have been suggested in more recent models. **b** | The importance of introgression between cultivated and wild relatives is indicated in alternative single domestication models. Many grain crops such as amaranths, common millet, foxtail millet, maize, pearl millet, rice and wheat, as well as many fruit crops (for example, apple and tomato) and root crops (for example, carrot), are thought to have undergone a single domestication event[9]. **c** | Alternatively, studies have also shown that multiple domestication events characterize the history of a quarter of the world's food crops[9], in which one wild species undergoes domestication in different regions or at different time points. This multiple domestication model is exemplified by barley, bottle gourd, coconut, common bean, aubergine and sorghum. **d** | A third alternative single domestication model has been proposed, in which crops are domesticated from interspecific hybridization followed by clonal propagation. This is especially common in tree crops[115], such as citrus and banana, but is also found in many short-lived species, such as peanut and strawberry.

sessile spikelets1 (*Sos1*) in maize[26], have been erroneously inferred to be domestication loci and are instead important in more recent diversification of cultivated species.

Genetic architecture of domestication and diversification. Despite the caveats described above, QTL mapping and genetic complementation analyses led to the isolation of the first domestication gene that has been characterized at the molecular level — the *teosinte branched1* (*tb1*) locus[27] — which controls differences in shoot architecture between maize and its wild teosinte progenitor. The identification of maize *tb1* as a domestication gene has been followed over the past two decades by the identification of numerous other domestication and diversification loci, most of which are in cereal crop species but with a few in non-grass species, such as beans, cole crops, grape, sunflower and tomato (see Supplementary information S2 (table)).

Table 1 | **Commonly observed traits in crops* accompanying domestication (Stage 1) and diversification (Stages 2–4)**

	Stage 1	Stage 2	Stage 3	Stage 4
Seed crop	• Larger seeds • Resource allocation • Thinner seed coat, and increased seed softening and ornamentation • Inflorescence architecture (including shape, number and determinacy) • Increased yield potential and productivity • Loss of dormancy • Determinate growth	• More seeds • Increased seed size variation • Pigment change • Flavour change • Change in starch content • Non-shattering seeds‡ • Reduced germination inhibition	• Reduced vernalization • Reduced photoperiod sensitivity • Modified hormone sensitivity • Synchronized flowering time • Shortened or extended life cycle • Dwarfism	• Increased yield • Increased abiotic stress tolerance • Increased biotic stress tolerance • Improved eating quality
Root and Tuber	• Flavour change • Resource allocation • Change in starch content • Ability to thrive in modified landscape • Reduced branching	• Reduced toxicity • Vegetative propagation and reduced sexual propagation • Abiotic stress tolerance • Biotic stress tolerance • Extended harvest season	• Hybridization using effect of heterosis • Promotion of allogamy • Increased yield	• Improved nutritional quality • Improve multiplication ability and rate
Fruit	• Flavour change • Resource allocation • Larger seed size • Larger fruit size • Shortened life cycle • Softer fruit	• Increased fruit size variation • Selfing breeding system	• Improved pollination success • Reduced fruit shedding • Continuous fruiting	• Delayed ripening • Increased post-harvest quality and delayed senescence • Increased yield • Increased abiotic stress tolerance • Increased resistance • Attractiveness and even ripening

*Examples in annual or short-lived perennial fruits, roots and seeds are shown. Fewer general traits could be identified for less well-characterized crops, such as leaf crops and long-lived perennial species, and these were therefore excluded. ‡A Stage 1 trait in some crop species.

QTL mapping studies (FIG. 3) were among the first attempts to dissect the genetic architecture of plant domestication and diversification, and such studies provided the initial steps to identifying specific genes that are involved. These early studies, which were mostly carried out in maize, rice and beans, indicated that only one or a few genes of large effect controlled many domestication traits[28,29], although this pattern was not universal; for example, in foxtail millet, both tillering and axillary branching are controlled by many loci of small effect[30]. Many QTL studies have also demonstrated that multiple key domestication traits are controlled by the same regions of the genome[31,32], which indicates that either pleiotropy or tight linkage among several loci may be an important attribute of the evolution of domesticated species.

The number of genes or QTLs that are thought to underlie traits of the domestication syndrome[33] is difficult to estimate. In maize, QTLs for 9 domestication traits ranged from 6 to 26 (REF. 34), and in rice, 13 domestication traits were associated with 76 QTLs[35]. Loci that are thought to underlie the diversification traits of the photoperiod response and of flowering time vary in number among members of the Poaceae family: 25 QTLs and 4 hotspot genomic regions were observed in maize[36], 16 in foxtail millet[37] and 14 in rice[38,39]. QTL analyses have also identified clusters of mapped loci for the same trait[32].

More recent GWASs have confirmed similar numbers and patterns of detectable associations. A GWAS in rice identified 80 loci for 14 agronomic traits[40], and in sorghum, 14 loci have been identified for the inflorescence branch length trait[41]. Similar numbers were also recently reported in a GWAS of foxtail millet varieties, in which 512 loci were found to be associated with 47 agronomic traits[42]. Despite the large number of domestication and/or diversification loci that have been identified by QTL mapping and GWASs, these may all be underestimates; for example, one study in maize[43] suggests that nearly 500 genomic regions, which are estimated to span up to 2,000 genes, show evidence of directional selection that is consistent with possible roles in domestication.

Domestication syndrome
The selection of traits that distinguish domesticated species from their wild progenitors; similar traits were often observed to occur in different crops, which led people to view them as a 'syndrome'.

Functions and mutations

Biological functions of domestication and diversification genes. We compiled data on 60 genes that have been reported as domestication and/or diversification loci (see Supplementary information S2 (table)). We chose these 60 genes because they have been included in population genetic studies and/or have been functionally validated. We also included various genes that have been investigated using a wide range of approaches to support their roles in crop evolution. Although this list is by no means comprehensive, it illustrates the state of the field.

As the roles of these genes have not necessarily been delimited by previous investigators, we re-evaluated the role of these 60 genes and categorized them as domestication or diversification loci. Using our criteria to examine these 60 genes, 23 genes were determined as probable domestication genes that are associated with evolution in Stage 1 (FIG. 1; see Supplementary information S2,S3 (tables)), and 32 genes were more plausible as diversification genes or early crop improvement genes (Stages 2 or 3). Five genes seem to have undergone selection in both domestication and diversification.

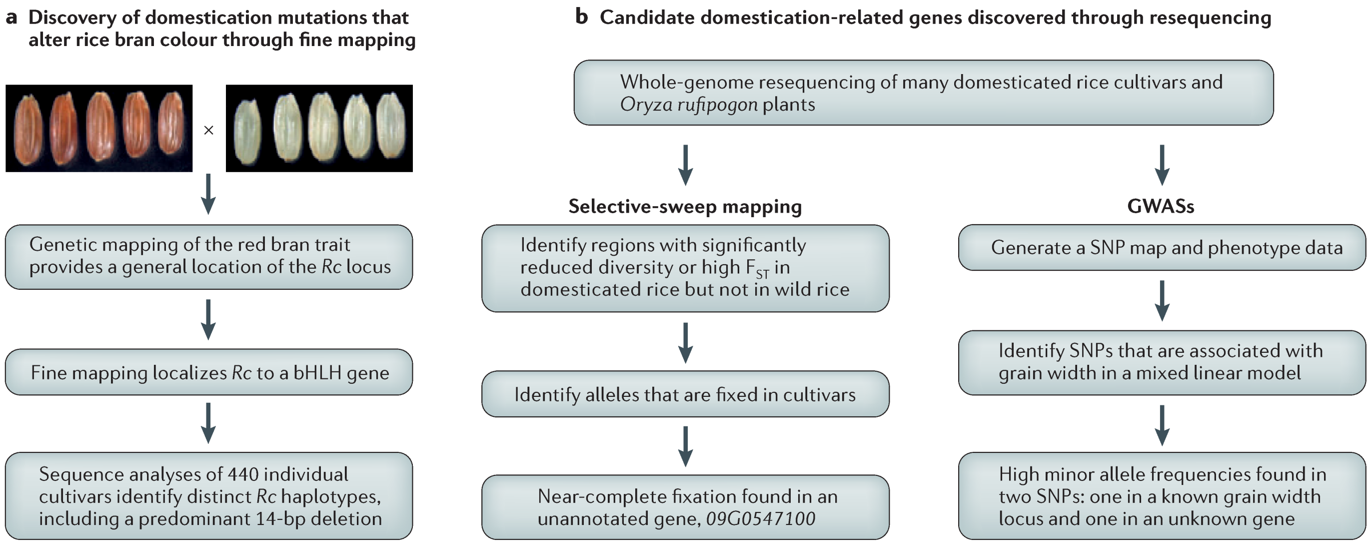


Figure 3 | **From discovery to characterization of domestication genes.** **a** | The process of discovering domestication genes and their underlying mutations are shown, and bran colour in *Oryza sativa* (rice) is used as an example. Disruption of the pigmentation pathway that leads to red bran colour occurred early in the domestication process for rice (Stage 1), possibly because contaminants in stored grain could be more easily identified against a white background. Red bran rice (that is, the *Rc* haplotype) and white bran rice (that is, the *rc* haplotype)[79] are shown. The white bran cultivars Jefferson and IR64 are crossed with the wild progenitor *Oryza rufipogon* (which has a red bran colour) in a fine-mapping study of the *Rc* quantitative trait locus[78], which localizes the *Rc* locus in mapping populations to a MYC-like basic helix–loop–helix (bHLH) gene. *Rc* haplotypes in 440 individual rice cultivars[82] have a 14-base pair deletion in the protein-coding region of the bHLH gene, which is the predominant mutation that disrupts gene function. Although this deletion is the predominant genotype under fixation, some varieties underwent parallel selection for different mutations in the same gene. **b** | Candidate domestication-related genes are identified by selective-sweep mapping and by genome-wide association studies (GWASs). Candidate genes that are under artificial selection are identified by whole-genome resequencing of 50 rice cultivars and their wild relatives at 15 times coverage[116]. Several regions (or genes) show reductions in diversity or high Wright fixation index (F_{ST}) on chromosome 1 in *O. sativa* ssp. *japonica* (that is, a domesticated variety group) relative to *O. rufipogon* (that is, the wild progenitor). The unannotated *09G0547100* gene is a candidate domestication-related gene because it shows a strong selective sweep in *japonica*[116]. *09G0547100* encodes a putative auxin-induced protein. Alternatively, GWAS can be carried out after resequencing. A GWAS in rice[39] led to the discovery of two single-nucleotide polymorphisms (SNPs) that are associated with grain width — a diversification trait — on chromosome 5 in a compressed mixed linear model. One SNP is within *qSW5*, which is a locus that is known to have a role in grain width, whereas the other SNP has not been previously studied with regard to grain width. The C→G SNP that is associated with the unknown gene has a minor allele frequency of 0.21 (REF. 39). Image in part **a** is reproduced, with permission, from REF. 78 © (2007) John Wiley & Sons, Inc.

Genes that are thought to be involved in domestication (Stage 1) contribute to various traits (see Supplementary information S2 (table)). They regulate inflorescence development (*Brassica oleracea CAL*; common bean *TFL1*; and maize *barren stalk1* (*ba1*), *ramosa1* (*ra1*), *tb1* and *Zagl1*), vegetative growth habit and height (maize *tb1*; and *Oryza sativa PROG1* and *LG1*), seed pigment, seed size, casing, ornamentation (rice *BH4*; barley *NUD*; and maize *teosinte glume architecture1* (*Tga1*) and *prolamin-box binding factor* (*pbf*)), seed retention (rice *SH4-1*; *Sorghum bicolor SH1*; and the wheat aspartic proteinase gene *WAP2*), nitrogen access and efficiency (*O. sativa AMT1;1*), and fruit flavour (strawberry *NES1* and *PINS*) (see Supplementary information S2 (table)).

Diversification genes also contribute to a range of phenotypes, and evolutionary changes include fruit shape and size (tomato *FW2.2*, *OVATE* (also known as *LOC543847*) and calmodulin-binding protein SUN-like (*SUN*)), inflorescence architecture (barley *VRS1*; soybean *TFL1B*; and maize *Sos1*) (see Supplementary information S2 (table)), colour (the grape myb-related transcription factor genes *MYBA1* and *MYBA2*; and *Brassica rapa TT8*) and starch composition traits (maize *sugary1* (*su1*); and *WAXY* in multiple species). Traits for specific cultural practices and preferences, such as dwarfism (*O. sativa SD1*), fragrance (rice *BADH2*) and pod corn (maize *MADS19* (*m19*)), were also selected. Moreover, genes that control flowering time diversity have been described (*O. sativa HD1*; barley *ELF3*; maize *CCT* (also known as *LOC100281853*); pearl millet *MADS11*; and strawberry *KSN*), and these genes are possibly associated with adaptation of crops to new environments (Stage 3).

There are numerous genes that are associated with recent breeding (Stage 4), which we have not enumerated. Nevertheless, some of these Stage 4 loci seem to have their origins in earlier stages in the crop evolutionary process; for example, in maize, the *yellow endosperm1* (*y1*) gene that colours endosperm yellow was strongly selected for in the 1920s in the United States, but this mutation can be traced to localized selection by Native

Americans in early diversification[44,45]. Another example is dwarfism in rice, particularly the reduction in culm length that is mostly attributable to the *semidwarf-1* gene (*SD1*). Plants with mutations in this gene were bred during the Green Revolution in the twentieth century, but evidence suggests that it was originally selected for by early Japanese farmers[46,47].

Molecular functions of domestication and diversification genes. The isolation of genes that underlie domestication and diversification traits provides an opportunity to examine some of the characteristics of the loci that are associated with the evolution of crop species. These loci show a wide range of functions — from transcription factors to metabolic enzymes — although many encode similar enzymes or are involved in the same pathways across species.

Mutations in regulatory genes, such as transcription factors, are thought to underlie phenotypic changes that are associated with domestication (reviewed in REFS 48,49). Of the 60 genes that we examined and that were reported to be involved in domestication or diversification, 37 genes (~62%) encode transcription factors, whereas 3 other genes encode transcription co-regulators. Enzyme-encoding genes make up the second largest class of loci (14 genes), whereas the remaining 6 genes encode transporter proteins and ubiquitin ligase.

Causative mutations in crop evolution loci have a range of functional effects (see Supplementary information S2 (table)). Many of these genes contain multiple mutations that have functional consequences, which indicates that, during crop evolution, multiple mutations that could be subject to selective pressures arise. Such mutations may be factors in the spread and modification of selected domestication and/or diversification phenotypes. On the basis of the genes that we have reviewed, nonsense mutations, premature truncations or other mutations that lead to null function (for example, frameshifts and splicing defects) are the predominant type of causative change (38 of 60 genes). The next major functional class of mutations are *cis*-regulatory mutations (26 of 60 genes) and, finally, missense mutations or other types of structural changes that alter protein function (10 of 60 genes).

These results suggest that both loss-of-function alleles and the alleles that alter gene expression are by far the most common types of functional changes that are observed during crop evolution. These types of alleles are likely to have large phenotypic effects, which is consistent with the marked phenotypic divergence that is observed during domestication and diversification[3,4]. A recent study in maize suggests that single-nucleotide polymorphisms (SNPs) that are associated with overall quantitative trait variation (~79%) are linked to gene regulatory regions within 5 kb upstream of protein-coding regions[50]. Thus, the pattern of mutations that we observe, particularly the preponderance of loss-of-function alleles in domestication and/or diversification loci, may be specific to crop evolutionary traits and may not be representative of overall causative variation in domesticated plant genomes.

Mutational lesions in domestication and diversification genes can range from SNPs, indels, transposon insertions and gene duplications to large-scale chromosomal rearrangements (FIG. 4). Of the 60 genes we examined (see Supplementary information S2 (table)), 35 genes had at least one causative SNP, 23 genes had indels and 9 genes had a transposable element among the causative mutations. For 4 of these 60 genes, a causative mutation has not been reported.

Overall, most causal SNPs in domestication or diversification genes were found to be nonsense mutations or were found to occur in regulatory regions such as the promoter, which causes putative *cis*-regulatory changes that are usually shown by altered expression and that are detected by PCR (FIG. 4; see Supplemenary information S2 (table)). Also common were genes with SNPs that produce altered, but presumably functional, proteins. Similarly, most indels formed null mutations either by inducing a translational frameshift or by inducing premature truncations of the translated protein, whereas only rarely were *cis*-regulatory changes induced by an indel. Interestingly, 15% of the genes had transposable element insertions that had functional effects, which suggests that transposable elements have an important mutational role in domesticated plant genomes.

Compared with SNPs and small indels, genomic changes that involve larger sequence alterations are less commonly observed. Copy-number variants have been observed only in the maize *m19* gene or the tomato *SUN* gene (see Supplementary information S2 (table)). An even rarer type of observed genetic change is large chromosomal rearrangements, as seen in *RRS2* in barley, in which the mutation is a genomic translocation that spans the domestication locus[51].

Processes of evolution

Selection at the molecular level. Selection is a hallmark of domestication and should leave molecular footprints in the genomes of crop species. The first domestication gene that was isolated — the maize *tb1* locus — has a 60–90-kb selective sweep that occurred upstream of the 5′ end of the protein-coding region[52]. This sweep, which is defined as an extended region of low nucleotide diversity, spans the *Hopscotch* transposable element insertion (FIG. 4b) in the *cis*-regulatory region that regulates *tb1* expression[53]. Early genome-scale surveys in maize suggested that as many as 2–4% of genes in this cereal crop species were under positive selection[54], but recent work indicates that a much larger percentage (~7.6%) of the maize genome has been affected by domestication and diversification[43].

Recent studies also reveal that selective sweeps are prevalent in the genomes of other crops, such as mungbean[55], rice[5,56] and tomato[57,58]. The largest crop genome resequencing study so far, in which the genomes of 1,529 wild and cultivated rice accessions were analysed, identified 55 selective sweeps, including those that are associated with the domestication genes *BH4* (which causes a loss of hull colour) and *SH4-1* (which causes a loss of seed shattering)[59] that show fixation of causal mutational variants in cultivated samples[5].

Nonsense mutations
Point mutations that transform amino acid-encoding codons into premature stop codons, which result in the generation of truncated proteins.

cis-regulatory mutations
Mutations in linked, usually non-coding portions, of genes that alter levels and/or patterns of transcription of the linked gene.

Missense mutations
Point mutations that change the identities of encoded amino acids, which result in changes in protein sequences.

Nucleotide diversity
The number of single-nucleotide polymorphism in a genomic region, usually estimated as the mean level of pairwise nucleotide divergence in a sample or a population.

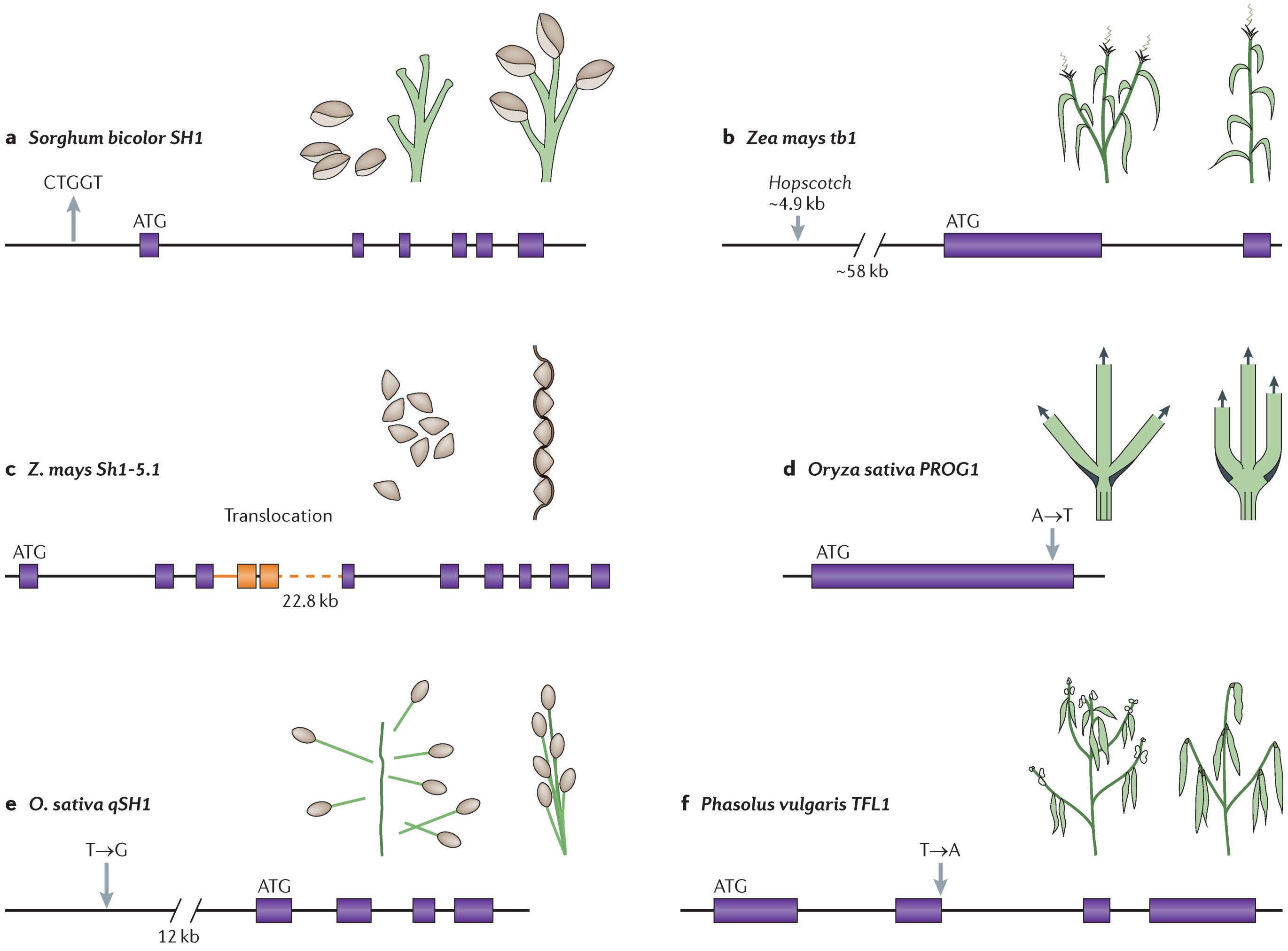


Figure 4 | **Types of mutations in crop domestication and diversification genes.** For each crop, phenotypic changes that correspond to the mutations of wild varieties (left) versus hypothetical early domesticated varieties (right) are shown. **a** | A deletion in the *cis*-regulatory region (that is, the promoter) of the *Sorghum bicolor SH1* gene[75] results in the non-shattering phenotype. **b** | Insertion of the *Hopscotch* transposable element results in a *cis*-regulatory mutation in the *Zea mays teosinte branched1* (*tb1*) gene, which leads to altered shoot architecture. **c** | Translocation leads to the fusion of two exons from an unknown gene after exon 3 of the *Z. mays Sh1-5.1* gene[75], which results in the loss of the YABBY domain and a reduction in shattering. **d** | A missense single-nucleotide polymorphism (SNP) in the *Oryza sativa PROG1* gene[76] results in erect growth in domesticated Asian rice varieties. **e** | A SNP in the *cis*-regulatory region (that is, the promoter) of the *O. sativa qSH1* gene[117,118] results in the non-shattering phenotype. **f** | A SNP in a splice site of the *Phasolus vulgaris TFL1* gene[119] results in determinate inflorescences.

Several studies reveal that many genes that seem to underlie domestication phenotypes — such as the barley *RRS2* locus that confers disease resistance and the pearl millet *TB1* locus that confers reproductive meristem identity — show evidence of partial sweeps in which the causal alleles are not fixed within the species but are found at moderate frequencies[30,60]. Several factors can preclude allele fixation and maintain allelic diversity in domesticated populations; for example, the tomato *LC* and *OVATE* loci, which are both thought to confer the initial increase in fruit size during domestication, may cause seed sterility if both alleles are found together[61]. Another possibility is that multiple independent genes underlie the domestication trait, and that different genes lead to the selected phenotype in different crop populations.

Partial selective sweeps are also observed in diversification loci, in which culture-specific selection of desirable traits leads to fixation of alleles in varieties but not across the entire domesticated species. A classic example is the rice *WAXY* gene. Mutations that confer the sticky rice phenotype are prized by some East and Southeast Asian cultures. The mutation at the splice donor site of intron 1 of this gene is associated with a ~240-kb selective sweep, but this is mostly found in the temperate *japonica* variety group of rice that is popular in Japan and Korea (BOX 2).

Old versus new mutations. A major issue is whether mutations that lead to domestication or diversification phenotypes are new mutations that arise near-contemporaneously with the onset of positive selection,

Box 2 | **Glutinous grains — parallel evolution across species**

A waxy endosperm results when starch in cereal crop grains has low or no amylose and contains greater amounts of amylopectin, which produces a sticky glutinous grain on boiling. Waxy grains are found among many domesticated cereals and pseudocereals[99]. The *WAXY* gene encodes the granule-bound starch synthase (GBSS) enzyme[84,100]. In rice, a G→A single-nucleotide polymorphism at the splice donor site of intron 1 is responsible for the reduction in GBSS activity, which leads to glutinous rice in some varieties. This mutation arose only once in glutinous rice varieties, possibly in mainland Southeast Asia, and spread to temperate *japonica* varieties that have reduced amylose levels in grains[85,101]. Results from studies on the *WAXY* gene in various species suggest that mutant phenotypes are rare in the wild and that many cultivar alleles probably arose through novel mutations[102].

Several cultures are partial to sticky grains, and this phenotype has repeatedly evolved in different cereal crop species. In sub-Saharan Africa, sorghum (*Sorghum bicolor*) *waxy* mutants underwent selection during diversification[99,103]. Northeast Asian cultural preferences for sticky grains[104] also seem to have driven parallel selection on the *waxy* mutants in numerous species. Subsequent to glutinous rice being incorporated into Japanese culture, the grain crop Job's tears that has a waxy phenotype was domesticated[105]. In northern China, three mutations in the two copies of *WAXY* are found in tetraploid broomcorn millet, and these mutations probably underwent selection as this crop spread into East Asia[86,106], where sticky rice already existed. In East Asia, mutations in *WAXY* also arose in foxtail millet[107] and in barley[108], and they were preferentially selected for in Japanese culture.

In the New World, sticky grain amaranths were used to make cakes as part of Aztec human sacrifice rituals in Mexico, where the domestication of both *Amaranthus cruentus* and *Amaranthus hypochondriacus* was thought to occur[109]. *waxy* mutants have also been selected in at least three *Amaranthus* spp. pseudocereals in Central and South American — *Amaranthus caudatus* in Peru and *A. hypochondriacus* in Mexico during domestication, and *A. cruentus* in Mexico during diversification[110]. *A. hypochondriacus* was domesticated after *A. cruentus*, and the *waxy* allele is nearly completely fixed in Mexican *A. hypochondriacus* cultivars, which suggests that it was a domestication gene in this species. There are many cases other than the example of *WAXY*, in which processing technology or cultural practices were adopted around a particular diversification mutation in one crop, and these innovations may have influenced selection for similar mutations in other new crops.

Parallel evolution
Independent evolution of the same trait in different species.

or old mutations that have a long history of segregation in populations before the advent of selection. Whether selection has affected old or new mutations has implications for both the nature of the selective sweeps and the dynamics of the evolution of crop species; for example, selective sweeps on standing variation (rather than on new mutations) are expected to leave a weaker signature of selection in the genome, which highlights the necessity to investigate gene polymorphisms in both wild and domesticated populations[62–64].

Some domestication or diversification genes, such as the rice *LG1* gene that is associated with a closed panicle trait[65,66] or the *SUN* gene duplication in tomato that regulates organ shape[61], seem to be novel alleles in domesticated cultivars that are absent in wild accessions. However, many domestication alleles occur in low to moderate frequencies in wild progenitor species. Although the presence of domestication alleles in wild populations could have resulted from crop-to-wild gene flow, several studies have indicated that some of these are indeed ancestral alleles found in the wild species that underwent positive selection in the derived crop. For example, the *B. oleracea CAL* gene encodes a MADS box transcription factor that regulates floral meristem development, and a nonsense mutation leads to the proliferation of floral meristems in domesticated cauliflower (*B. oleracea* ssp. *botrytis*) and broccoli (*B. oleracea* ssp. *italica*) (see Supplementary information S3 (table)). This mutation is either fixed or at high frequency in these domesticated subspecies, but it is also present at low frequency in wild *B. oleracea*. Other examples of possibly old mutations that are important in crop evolution include those in the *tb1* and *Zagl1* genes in maize[67], the *INTERMEDIUM-C* (*INT-C*) gene[68] and the *PPD-H1A* haplotype in barley[69], and the *LC* gene in tomato[61]. This suggests that many domesticated traits arise not from new mutations but rather from mutations that are segregating in ancestral wild populations of crop species[70].

Multiple mutations and parallelism at the molecular level. It is not uncommon to observe morphological homoplasy in nature[71–73], which naturally leads to the question: does selection for particular phenotypes affect the same genes or distinct genes in different species? Domesticated species provide excellent models to study this question. Selective pressures across multiple independently evolved domesticated populations or species can act on the same traits, such as the loss of seed dispersal or increased seed size, and the ancestral states for these traits are well characterized for these domesticated taxa. Darwin used these 'analogous variations' to describe changes in parallel evolution[4], and Vavilov developed the Law of Homologous Series[74] through the study of domesticated plant species.

Parallelisms at the molecular level provide a basis for Darwin's observations and for Vavilov's Law. In a single species, there are cases of multiple mutations that cause the same domestication phenotype in cultivated species; these represent independent origins of the domestication trait. In *S. bicolor*, unique haplotypes of *SH1* characterize each of the three separate origins of the loss of seed shattering in this species[75]. In this context, the discovery of independent mutations in domestication loci adds support to the hypothesis that multiple domestications of *S. bicolor* occurred.

Other domestication genes have also been shown to have multiple causal mutations, but in these cases it is generally believed that only one mutation is fixed and is associated with domestication, whereas other mutations are in low to moderate frequencies across the species. For example, the *O. sativa PROG1* gene may have 10 non-synonymous SNPs and 6 indels in the protein-coding region, as well as 27 SNPs and 2 indels in the 5′ flanking region. However, a single A→T mutation that causes a threonine-to-serine change in the carboxyl terminus of the protein was shown to be sufficient to cause an erect plant habit by altering the binding properties of this transcription factor[76]. This is consistent with phylogenetic analyses of the *PROG1* gene that supports the monophyly of cultivars that have *PROG1* alleles arising from a single population of the wild progenitor species *Oryza rufipogon*, which indicates that selection on this gene during domestication occurred once[5]. In addition, at least four other mutations in the promoter region have been proposed to regulate gene expression levels that result in intraspecific phenotypic variation[77], and these may represent parallel modifier mutations that are fixed in smaller populations.

Another example of multiple domestication mutations is in the domestication *Rc* gene (FIG. 3), which has three causal variants that contribute to regulatory changes in the production of anthocyanin in the rice grain. These three mutations are associated with the elimination of the dominant red pigment seed colour that is found in wild *O. rufipogon*. Only one mutation, a 14-base pair deletion in exon 7 that leads to a translational frameshift, is consistently found in all white seeded domesticate species and is absent in all wild accessions[78]. This suggests that this deletion is the only causal variant that is associated with domestication, whereas the two other variants seem to be diversification mutations. One of these variants is fixed only in *japonica* cultivars[5], and the other is not fixed but actually leads to a light red (as opposed to white) grain colour that is prized in certain varieties[79]. Other mutations have also been found in the *Rc* locus[80,81], albeit at very low frequencies, one of which restores the function of the RC protein to produce fully red seeds. The history of *Rc* variants[82] suggests that, as rice cultivation spread, parallel selection towards an increase in colour diversity was applied to new mutations, as well as introgressed from other progenitor populations.

The same gene can also undergo parallel selection in multiple crop species and may be a recurring target of selection; for example, comparative genomics studies in the Poaceae family have shown the correspondence of QTLs for several independently selected domestication or diversification traits among genera[83]. Mutations at the *WAXY* locus, which encodes the granule-bound starch synthase enzyme for amylose synthesis, is altered in rice[84,85], broomcorn millet[86], foxtail millet[87] and three *Amaranthus* spp. pseudocereals[88] to produce sticky grains (BOX 2). Other examples of parallel selection during diversification include the fruit-weight locus *FW2.2* (REF. 25) in tomato, chilli pepper and aubergine; the orthologues of both the shattering gene *SH3* and the *Rc* gene in Asian rice (*O. sativa*) and African rice (*Oryza glaberrima*)[89]; and *tb1* orthologues in maize (*tb1*), pearl millet (*TB1*) and barley (*INT-C*)[68]. There are also examples of parallel selection for genes within the same gene families (see Supplementary information S2 (table)), such as the APETALA2 transcription factors *SH1* in rice and the paralogous *WAP2* gene in wheat, both of which reduce shattering by the same mechanism[90,91].

Gene flow in domestication and diversification. In recent years, there has been a greater appreciation of the role of hybridization between domesticated species and their wild ancestors, or even between distinct populations, in the spread of domestication or diversification phenotypes (FIG. 2). The role of gene flow in the dynamics of domestication has been underscored by the idea that domestication, coupled with long-range movement of plants through human migrations and trade, is a prolonged process with cultivars and wild relatives occasionally occurring in sympatry; for example, a recent molecular study in rice suggests that it was domesticated once in China, which gave rise to the *japonica* variety group. *Indica* rice — a genetically distinct variety group — arose through subsequent hybridization of *japonica* with a putative proto-*indica* or *O. rufipogon* in South Asia[56], which resulted in the introgression of domestication genes into *indica*. The *rc* allele that confers white pigmentation is an example of a domestication gene that spread into *indica* by hybridization from *japonica*[5].

Diversification genes also spread to various varieties through hybridization as alleles move to new places and cultures. The *BADH2* locus is responsible for aromatic rice; although there were multiple causative mutations that arose in *japonica,* a single mutation in the *badh2.1* allele recombined into *indica*. This recombination resulted in fragrant *indica* cultivars that then continued to spread across several geographical regions[92]. The *waxy* splice site mutation originated in glutinous rice in tropical Southeast Asia, but subsequently moved into the low-amylose temperate *japonica* variety of Northeast Asia[85] (FIG. 2).

Perspectives

With the continued interest in domesticated taxa that arise as a result of their agricultural value, there are now detailed analyses of the genetics of numerous crop species, which provide opportunities to examine general patterns and to infer the dynamics of the evolutionary processes that are associated with crop origins and diversification. We can begin to discern some general outlines regarding the genetics of the evolution of domesticated plant species. We do find that, as previously suggested[49,70], many genes that underlie crop evolutionary traits are regulatory in nature, with either transcription factors or cofactors being the targets of selection and *cis*-regulatory mutations having a key role in evolutionary divergence. Most genes also have mutational lesions that lead to loss of function, including nonsense mutations or frameshift indels, which is consistent with the large phenotypic effects that are observed during crop evolution. Transposable element insertions, which have been thought to have a key role in plant evolution, also account for causative mutations in 15% of the domestication and diversification genes reviewed in this paper. Finally, many loci have more than one functional mutation that segregates in populations of crop species, which indicates that genes associated with crop domestication and diversification are subject to recurrent mutations that are possibly selective targets during evolution.

Although we can now begin to discern some general patterns of the molecular evolution of species, the challenge remains to obtain greater interspecific and intraspecific molecular genetic data, to use the information to develop and test more realistic models of origin and diversification, and to expand the research beyond the well-studied cereal crop domesticates. Researchers are now investigating the genetics of domestication in non-model crops and perennial crops, which increases our understanding of the domestication process and will probably lead to the discovery of novel domestication genes and evolutionary trajectories. Finally, we are making great advances in the understanding of how cultivation by ancestral farmers in the Neolithic period led to the origination and adaptation of new species with yields that are capable of sustaining human population growth.

Domestication provides a fascinating model for the study of evolution, and genetic and archaeological advances in the last decade have replaced simplistic ideas with more robust and complex models on the origin of crop species. We can now begin to see what lessons can be learnt in the quest to feed the world in the face of growing population pressures and changing climates.

1. Mueller, U. G., Gerardo, N. M., Aanen, D. K., Six, D. L. & Schultz, T. R. The evolution of agriculture in insects. *Annu. Rev. Ecol. Evol. Syst.* **36**, 563–595 (2005).
2. Duarte, C. M., Marba, N. & Holmer, M. Rapid domestication of marine species. *Science* **316**, 382–383 (2007).
3. Darwin, C. *The Variation of Animals and Plants Under Domestication* (John Murray, 1868).
4. Darwin, C. *On the Origin of Species by Means of Natural Selection, or the Preservation of Favoured Races in the Struggle for Life* (John Murray, 1859).
5. Huang, X. *et al.* A map of rice genome variation reveals the origin of cultivated rice. *Nature* **490**, 497–501 (2012).
This paper uses more than 1,500 accessions in whole-genome resequencing and high-resolution mapping to reveal the origin of rice, and the selective sweeps and fixed domestication mutations.
6. Blanca, J. *et al.* Variation revealed by SNP genotyping and morphology provides insight into the origin of the tomato. *PLoS ONE* **7**, e48198 (2012).
7. Fuller, D. Q. *Biodiversity in Agriculture: Domestication, Evolution, and Sustainability.* 110–135 (Cambridge Univ. Press, 2012).
8. Fuller, D. Q., Willcox, G. & Allaby, R. G. Early agricultural pathways: moving outside the 'core area' hypothesis in Southwest Asia. *J. Exp. Bot.* **63**, 617–633 (2012).
9. Meyer, R. S., DuVal, A. E. & Jensen, H. R. Patterns and processes in crop domestication: an historical review and quantitative analysis of 203 global food crops. *New Phytol.* **196**, 29–48 (2012).
This study provides an overview of the features of domesticated plant species.
10. Devore, I. & Lee, R. B. *Man the Hunter* (Aldine De Gruyter, 1999).
11. Dirzo, R. & Raven, P. H. Global state of biodiversity and loss. *Annu. Rev. Environ. Resour.* **28**, 137–167 (2003).
12. Gepts, P. *et al.* (eds.) *Biodiversity in Agriculture: Domestication, Evolution, and Sustainability* (Cambridge Univ. Press, 2012).
13. Willcox, G. in *Biodiversity in Agriculture: Domestication, Evolution, and Sustainability* (eds Gepts, P. *et al.*) 92–109 (Cambridge Univ. Press, 2012).
14. Hillman, G. C. & Davies, M. S. Measured domestication rates in wild wheats and barley under primitive cultivation, and their archaeological implications. *J. World Prehist.* **4**, 157–222 (1990).
15. Purugganan, M. D. & Fuller, D. Q. Archaeological data reveal slow rates of evolution during plant domestication. *Evolution.* **65**, 171–183 (2011).
16. Simmonds, N. W. *Evolution of Crop Plants* (Longman, 1976).
17. Fuller, D. Q. & Allaby, R. G. Seed dispersal and crop domestication: shattering, germination, and seasonality in evolution under cultivation. *Annu. Plant Rev.* **38**, 238–295 (2009).
18. Kislev, M. E., Hartmann, A. & Bar-Yosef, O. Early domesticated fig in the Jordan Valley. *Science.* **312**, 1372–1374 (2006).
19. Zohary, D. Unconscious selection and the evolution of domesticated plants. *Econ. Bot.* **58**, 5–10 (2004).
20. Knüpffer, H., Terentyeva, I., Hammer, K., Kovaleva, O. & Sato, K. Ecogeographical diversity — a Vavilovian appraoch. *Dev. Plant Genet. Breed.* **7**, 53–76 (2003).
21. Grobman, A. *et al.* Preceramic maize from Paredones and Huaca Prieta, Peru. *Proc. Natl Acad. Sci. USA* **109**, 1755–1759 (2012).
22. De Vries, I. M. Origin and domestication of *Lactuca sativa* L. *Genet. Res. Crop Evol.* **44**, 165–174 (1997).
23. Fuller, D. Q. Contrasting patterns in crop domestication and domestication rates: recent archaeobotanical insights from the Old World. *Ann. Bot.* **100**, 903–924 (2007).
24. Takeda, S. & Matsuoka, M. Genetic approaches to crop improvement: responding to environmental and population changes. *Nature Rev. Genet.* **9**, 444–457 (2008).
25. Doganlar, S., Frary, A., Daunay, M. C., Lester, R. N. & Tanksley, S. D. Conservation of gene function in the Solanaceae as revealed by comparative mapping of domestication traits in eggplant. *Genetics* **161**, 1713–1726 (2002).
26. Doebley, J., Stec, A. & Kent, B. *Suppressor of sessile spikelets1 (Sos1)*: a dominant mutant affecting inflorescence development in maize. *Am. J. Bot.* **82**, 571–577 (1995).
27. Doebley, J., Stec, A. & Hubbard, L. The evolution of apical dominance in maize. *Nature* **386**, 485–481 (1997).
This paper reports the isolation of one of the first domestication genes.
28. Koinange, E. M. K., Singh, S. P. & Gepts, P. Genetic control of the domestication syndrome in common bean. *Crop Sci.* **36**, 1037–1045 (1996).
29. Gepts, P. Crop domestication as a long-term selection experiment. *Plant Breed. Rev.* **24**, 1–44 (2004).
30. Remigereau, M. S. *et al.* Cereal domestication and evolution of branching: evidence for soft selection in the *Tb1* orthologue of pearl millet (*Pennisetum glaucum* [L.] R. Br.). *PLoS ONE* **6**, e22404 (2011).
31. Poncet, V. *et al.* Comparative analysis of QTLs affecting domestication traits between two domesticated x wild pearl millet (*Pennisetum glaucum* L., *Poaceae*) crosses. *Theor. Appl. Genet.* **104**, 965–975 (2002).
32. Cai, H. & Morishima, H. QTL clusters reflect character associations in wild and cultivated rice. *Theor. Appl. Genet.* **104**, 1217–1228 (2002).
33. Hammer, K. Das domestikationssyndrom. *Die Kulturpflanze* **32**, 11–34 (in German) (1984).
34. Doebley, J. & Stec, A. Inheritance of the morphological differences between maize and teosinte: comparison of results for two F_2 populations. *Genetics* **134**, 559–570 (1993).
This is one of the classic papers on QTL mapping of domestication traits.
35. Thomson, M. J. *et al.* Mapping quantitative trait loci for yield, yield components and morphological traits in an advanced backcross population between *Oryza rufipogon* and the *Oryza sativa* cultivar Jefferson. *Theor. Appl. Genet.* **107**, 479–493 (2003).
36. Xu, J. *et al.* The genetic architecture of flowering time and photoperiod sensitivity in maize as revealed by QTL review and meta analysis. *J. Integr. Plant. Biol.* **54**, 358–373 (2012).
37. Mauro-Herrera, M. *et al.* Genetic control and comparative genomic analysis of flowering time in *Setaria* (Poaceae). *G3* **3**, 283–295 (2013).
38. Yano, M. Genetic and molecular dissection of naturally occurring variation. *Curr. Opin. Plant Biol.* **4**, 130–135 (2001).
39. Zhao, K. *et al.* Genome-wide association mapping reveals a rich genetic architecture of complex traits in *Oryza sativa*. *Nature Commun.* **2**, 467 (2011).
40. Huang, X. *et al.* Genome-wide association studies of 14 agronomic traits in rice landraces. *Nature Genet.* **42**, 961–967 (2010).
41. Morris, G. P. *et al.* Population genomic and genome-wide association studies of agroclimatic traits in sorghum. *Proc. Natl Acad. Sci. USA* **110**, 453–458 (2013).
This is a large-scale study that identifies regions with selective sweeps that indicate potential diversification genomic regions.
42. Jia, G. *et al.* A haplotype map of genomic variations and genome-wide association studies of agronomic traits in foxtail millet (*Setaria italica*). *Nature Genet.* **45**, 957–961 (2013).
43. Hufford, M. B. *et al.* Comparative population genomics of maize domestication and improvement. *Nature Genet.* **44**, 808–811 (2012).
44. Palaisa, K., Morgante, M., Tingey, S. & Rafalski, A. 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 (2004).
45. Benz, B., Perales, H. & Brush, S. Tzeltal and Tzotzil farmer knowledge and maize diversity in Chiapas, Mexico. *Curr. Anthropol.* **48**, 289–300 (2007).
46. Asano, K. *et al.* Artificial selection for a Green Revolution gene during *japonica* rice domestication. *Proc. Natl Acad. Sci. USA* **108**, 11034–11039 (2011).
47. Asano, K. *et al.* Genetic and molecular analysis of utility of *sd1* alleles in rice breeding. *Breed. Sci.* **57**, 53–58 (2007).
48. Olsen, K. M. & Wendel, J. F. A bountiful harvest: genomic insights into crop domestication phenotypes. *Annu. Rev. Plant. Biol.* **64**, 47–70 (2013).
49. Purugganan, M. D. & Fuller, D. Q. The nature of selection during plant domestication. *Nature.* **457**, 843–848 (2009).
50. Li, X. *et al.* Genic and non-genic contributions to natural variation of quantitative traits in maize. *Genome Res.* **22**, 2436–2444 (2012).
51. Hanemann, A., Schweizer, G. F., Cossu, R., Wicker, T. & Roder, M. S. Fine mapping, physical mapping and development of diagnostic markers for the *Rrs2* scald resistance gene in barley. *Theor. Appl. Genet.* **119**, 1507–1522 (2009).
52. Clark, R. M., Linton, E., Messing, J., Doebley, J. F. Pattern of diversity in the genomic region near the maize domestication gene *tb1*. *Proc. Natl Acad. Sci. USA* **101**, 700–707 (2004).
53. Studer, A., Zhao, Q., Ross-Ibarra, J. & Doebley, J. Identification of a functional transposon insertion in the maize domestication gene *tb1*. *Nature Genet.* **43**, 1160–1163 (2011).
This is a functional analysis of a *cis*-regulatory polymorphism that results in a domesticated trait.
54. Wright, S. I. *et al.* The effects of artificial selection on the maize genome. *Science* **308**, 1310–1314 (2005).
This is one of the first systematic estimates of the number of genes in a domesticated plant genome that shows evidence of positive selection.
55. Isemura, T. *et al.* Construction of a genetic linkage map and genetic analysis of domestication related traits in mungbean (*Vigna radiata*). *PLoS ONE* **7**, e41304 (2012).
56. Molina, J. *et al.* Molecular evidence for a single evolutionary origin of domesticated rice. *Proc. Natl Acad. Sci. USA* **108**, 8351–8356 (2011).
57. Cong, B., Liu, J. & Tanksley, S. D. Natural alleles at a tomato fruit size quantitative trait locus differ by heterochronic regulatory mutations. *Proc. Natl Acad. Sci. USA* **99**, 13606–13611 (2002).
58. Paran, I. & Van Der Knaap, E. Genetic and molecular regulation of fruit and plant domestication traits in tomato and pepper. *J. Exp. Bot.* **58**, 3841–3852 (2007).
59. Li, C., Zhou, A. & Sang, T. Rice domestication by reducing shattering. *Science* **311**, 1936–1939 (2006).
60. Fu, Y. B. Population-based resequencing analysis of wild and cultivated barley revealed weak domestication signal of selection and bottleneck in the *Rrs2* scald resistance gene region. *Genome* **55**, 93–104 (2012).
61. Rodriguez, G. R. *et al.* Distribution of *SUN, OVATE, LC,* and *FAS* in the tomato germplasm and the relationship to fruit shape diversity. *Plant Physiol.* **156**, 275–285 (2011).
62. Teshima, K. M., Coop, G. & Przeworski, M. How reliable are empirical genomic scans for selective sweeps? *Genome Res.* **16**, 702–712 (2006).
63. Ross-Ibarra, J., Morrell, P. L. & Gaut, B. S. Plant domestication, a unique opportunity to identify the genetic basis of adaptation. *Proc. Natl Acad. Sci. USA* **104**, 8641–8648 (2007).
64. Innan, H. & Kim, Y. Detecting local adaptation using the joint sampling of polymorphism data in the parental and derived populations. *Genetics* **179**, 1713–1720 (2008).
65. Ishii, T. *et al.* *OsLG1* regulates a closed panicle trait in domesticated rice. *Nature Genet.* **45**, 462–465 (2013).
This is an excellent example of fine mapping, identification of a selective sweep, and functional characterization and identification of a causative mutation that validates *LG1* as a domestication gene.

66. Zhu, Z. *et al.* Genetic control of inflorescence architecture during rice domestication. *Nature Commun.* **4**, 2200 (2013).
67. Weber, A. L. *et al.* The genetic architecture of complex traits in teosinte (*Zea mays* ssp. *parviglumis*): new evidence from association mapping. *Genetics* **180**, 1221–1232 (2008).
68. Ramsay, L. *et al. INTERMEDIUM-C*, a modifier of lateral spikelet fertility in barley, is an ortholog of the maize domestication gene *TEOSINTE BRANCHED1*. *Nature Genet.* **43**, 169–172 (2011).
69. Jones, H. *et al.* Population-based resequencing reveals that the flowering time adaptation of cultivated barley originated east of the Fertile Crescent. *Mol. Biol. Evol.* **25**, 2211–2219 (2008).
70. Doebley, J. F., Gaut, B. S. & Smith, B. D. The molecular genetics of crop domestication. *Cell* **127**, 1309–1321 (2006).
71. Hennig, W. *Phylogenetic Systematics* (Univ. of Illinois Press, 1966).
72. Wood, T. E., Burke, J. M. & Rieseberg, L. H. Parallel genotypic adaptation: when evolution repeats itself. *Genetica* **123**, 157–170 (2005).
73. Ralph, P. & Coop, G. Parallel adaptation: one or many waves of advance of an advantageous allele. *Genetics* **186**, 647–668 (2010).
74. Vavilov, N. I. The law of homologous series in variation. *J. Genet.* **12**, 47–89 (1922).
    This is a classic paper on the recurring traits that are seen among crops and their influence on the development of core evolutionary concepts.
75. Lin, Z. *et al.* Parallel domestication of the *Shattering1* genes in cereals. *Nature Genet.* **44**, 720–724 (2012).
76. Jin, J. *et al.* Genetic control of rice plant architecture under domestication. *Nature Genet.* **40**, 1365–1369 (2008).
77. Tan, L. *et al.* Control of a key transition from prostrate to erect growth in rice domestication. *Nature Genet.* **40**, 1360–1364 (2008).
78. Furukawa, T. *et al.* The *Rc* and *Rd* genes are involved in proanthocyanidin synthesis in rice pericarp. *Plant J.* **49**, 91–102 (2007).
79. Sweeney, M. T., Thomson, M. J., Pfeil, B. E. & McCouch, S. Caught red-handed: *Rc* encodes a basic helix-loop-helix protein conditioning red pericarp in rice. *Plant Cell.* **18**, 283–294 (2006).
80. Brooks, S. A., Yan, W., Jackson, A. K. & Deren, C. W. A natural mutation in *rc* reverts white-rice-pericarp to red and results in a new, dominant, wild-type allele: *Rc-g*. *Theor. Appl. Genet.* **117**, 575–580 (2008).
81. Gross, B. L. *et al.* Seeing red: the origin of grain pigmentation in US weedy rice. *Mol. Ecol.* **19**, 3380–3393 (2010).
82. Sweeney, M. T. *et al.* Global dissemination of a single mutation conferring white pericarp in rice. *PLoS Genetics* **3**, e133 (2007).
    This paper shows how introgression between populations leads to the spread of a domestication trait.
83. Paterson, A. H. *et al.* Convergent domestication of cereal crops by independent mutations at corresponding genetic loci. *Science* **269**, 1714–1718 (1995).
84. Wang, Z. Y. *et al.* The amylose content in rice endosperm is related to the post-transcriptional regulation of the *waxy* gene. *Plant J.* **7**, 613–622 (1995).
85. Olsen, K. M. & Purugganan, M. D. Molecular evidence on the origin and evolution of glutinous rice. *Genetics* **162**, 941–950 (2002).
86. Hunt, H. V., Denyer, K., Packman, L. C., Jones, M. K. & Howe, C. J. Molecular basis of the waxy endosperm starch phenotype in broomcorn millet (*Panicum miliaceum* L.). *Mol. Biol. Evol.* **27**, 1478–1494 (2010).
87. Fukunaga, K., Kawase, M. & Kato, K. Structural variation in the *Waxy* gene and differentiation in foxtail millet [*Setaria italica* (L.) P. Beauv.]: implications for multiple origins of the waxy phenotype. *Mol. Genet. Genom.* **268**, 214–222 (2002).
88. Park, Y. J., Nishikawa, T., Tomooka, N. & Nemoto, K. The molecular basis of mutations at the *Waxy* locus from *Amaranthus caudatus* L.: evolution of the waxy phenotype in three species of grain amaranth. *Mol. Breed.* **30**, 511–520 (2012).
89. Gross, B. L., Steffen, F. T. & Olsen, K. M. The molecular basis of white pericarps in African domesticated rice: novel mutations at the *Rc* gene. *J. Evol. Biol.* **23**, 2747–2753 (2010).
90. Hofmann, N. R. *SHAT1*, a new player in seed shattering of rice. *Plant Cell.* **24**, 839 (2012).
91. Zhou, Y. *et al.* Genetic control of seed shattering in rice by the APETALA2 transcription factor SHATTERING ABORTION1. *Plant Cell* **24**, 1034–1048 (2012).
92. Kovach, M. J., Calingacion, M. N., Fitzgerald, M. A. & McCouch, S. R. The origin and evolution of fragrance in rice (*Oryza sativa* L.). *Proc. Natl Acad. Sci. USA* **106**, 14444–14449 (2009).
93. Piperno, D. R., Ranere, A. J., Holst, I., Iriarte, J. & Dickau, R. Starch grain and phytolith evidence for early ninth millennium B.P. maize from the Central Balsas River Valley, Mexico. *Proc. Natl Acad. Sci. USA* **106**, 5019–5024 (2009).
94. Jaenicke-Despres, V. *et al.* Early allelic selection in maize as revealed by ancient DNA. *Science.* **302**, 1206–1208 (2003).
95. Dorweiler, J. & Doebley, J. Developmental analysis of *teosinte glume architecture1*: a key locus in the evolution of maize (*Poaceae*). *Am. J. Bot.* **84**, 1313 (1997).
96. Wang, H. *et al.* The origin of the naked grains of maize. *Nature.* **436**, 714–719 (2005).
97. Soltis, D. E. *et al.* Angiosperm phylogeny: 17 genes, 640 taxa. *Am. J. Bot.* **98**, 704–730 (2011).
98. Stevens, P. F. *Angiosperm Phylogeny Website, Version 12.* [online], http://www.mobot.org/MOBOT/research/APweb/ (2012).
99. Kempton, J. H. Waxy endosperm in coix and sorghum. *J. Hered.* **12**, 396–400 (1921).
100. Sano, Y. Differential regulation of *waxy* gene expression in rice endosperm. *Theor. Appl. Genet.* **68**, 467–473 (1984).
101. Olsen, K. M. *et al.* Selection under domestication: evidence for a sweep in the rice *waxy* genomic region. *Genetics* **173**, 975–983 (2006).
    This is an early molecular population genetic analysis of a crop diversification gene.
102. Kilian, B. *et al.* Haplotype structure at seven barley genes: relevance to gene pool bottlenecks, phylogeny of ear type and site of barley domestication. *Mol. Genet. Genom.* **276**, 230–241 (2006).
103. de Alencar Figueiredo, L. F. *et al.* Phylogeographic evidence of crop neodiversity in sorghum. *Genetics* **179**, 997–1008 (2008).
104. Sakamoto, S. in *Redefining Nature: Ecology, Culture and Domestication* (eds Ellen, R. & Fukui, K.) 215–231 (Berg, 1996).
105. Hachiken, T. *et al.* Deletion commonly found in *Waxy* gene of Japanese and Korean cultivars of Job's tears (*Coix lacryma-jobi* L.). *Mol. Breed.* **30**, 1747–1756 (2012).
106. Araki, M., Numaoka, A., Kawase, M. & Fukunaga, K. Origin of waxy common millet, *Panicum miliaceum* L. in Japan. *Genet. Res. Crop Evol.* **59**, 1303–1308 (2012).
107. Kawase, M., Fukunaga, K. & Kato, K. Diverse origins of waxy foxtail millet crops in East and Southeast Asia mediated by multiple transposable element insertions. *Mol. Genet. Genom.* **274**, 131–140 (2005).
108. Nakao, S. On waxy barleys in Japan. *Seiken Jiho.* **4**, 111–113 (in Japanese) (1950).
109. Sauer, J. D. The grain amaranths and their relatives: a revised taxonomic and geographic. *Ann. Missouri Bot. Gard.* **54**, 103–137 (1967).
110. Jimenez, F. R. *et al.* Assessment of genetic diversity in Peruvian amaranth (*Amaranthus caudatus* and *A. hybridus*) germplasm using single nucleotide polymorphism markers. *Crop Sci.* **53**, 532–541 (2013).
111. Haudry, A. *et al.* Grinding up wheat: a massive loss of nucleotide diversity since domestication. *Mol. Biol. Evol.* **24**, 1506–1517 (2007).
112. Cornille, A. *et al.* New insight into the history of domesticated apple: secondary contribution of the European wild apple to the genome of cultivated varieties. *PLoS Genet.* **8**, e1002703 (2012).
113. Iorizzo, M. *et al.* Genetic structure and domestication of carrot (*Daucus carota* subsp. *sativus*) (Apiaceae). *Am. J. Bot.* **100**, 930–938 (2013).
114. Dempewolf, H., Hodgins, K. A., Rummell, S. E., Ellstrand, N. C. & Rieseberg, L. H. Reproductive isolation during domestication. *Plant Cell.* **24**, 2710–2717 (2012).
115. Miller, A. J. & Gross, B. L. From forest to field: perennial fruit crop domestication. *Am. J. Bot.* **98**, 1389–1414 (2011).
    This paper is an overview of the state of understanding about perennial crop domestication traits and demographic histories.
116. Xu, X. *et al.* Resequencing 50 accessions of cultivated and wild rice yields markers for identifying agronomically important genes. *Nature Genet.* **30**, 105–111 (2011).
117. Konishi, S. *et al.* An SNP caused loss of seed shattering during rice domestication. *Science* **312**, 1392–1396 (2006).
    This is a classic paper on the isolation of a gene for non-shattering, which is a major domestication trait.
118. Konishi, S., Ebana, K. & Izawa, T. Inference of the *japonica* rice domestication process from the distribution of six functional nucleotide polymorphisms of domestication-related genes in various landraces and modern cultivars. *Plant Cell Physiol.* **49**, 1283–1293 (2008).
119. Repinski, S. L., Kwak, M. & Gepts, P. The common bean growth habit gene *PvTFL1y* is a functional homolog of *Arabidopsis TFL1*. *Theor. Appl. Genet.* **124**, 1539–1547 (2012).
120. Wingen, L. U. *et al.* Molecular genetic basis of pod corn (*Tunicate* maize). *Proc. Natl Acad. Sci. USA* **109**, 7115–7120 (2012).

Acknowledgements

The authors thank the members of the Purugganan laboratory for their feedback, particularly J. Flowers, A. Plessis and U. Rosas. R.S.M. is supported by a postdoctoral fellowship from the US National Science Foundation Plant Genome Research Program (NSF PGRP). Work on domestication in the Purugganan laboratory is also funded by grants from NSF PGRP and the New York University Abu Dhabi Research Institute, United Arab Emirates.

Competing interests statement

The authors declare no competing interests.

FURTHER INFORMATION

Plants for a future: http://pfaf.org

SUPPLEMENTARY INFORMATION

See online article: S1 (table) | S2 (table) | S3 (table)

ALL LINKS ARE ACTIVE IN THE ONLINE PDF