New insights into the history of rice domestication

Michael J. Kovach, Megan T. Sweeney and Susan R. McCouch
Cornell University, Department of Plant Breeding and Genetics, Ithaca, NY 14853–1901, USA

The history of rice domestication has long been a subject of debate. Recently obtained genetic evidence provides new insights into this complex story. Genome-wide studies of variation demonstrate that the two varietal groups in Oryza sativa (indica and japonica) arose from genetically distinct gene pools within a common wild ancestor, Oryza rufipogon, suggesting multiple domestications of O. sativa. However, the evolutionary history of recently cloned domestication genes adds another layer of complexity to the domestication of rice. Although some alleles exist only within specific subpopulations, as would be expected if the domestications occurred independently, other major domestication alleles are common to all cultivated O. sativa varieties. Our current view of rice domestication supports multiple domestications coupled with limited introgression that transferred key domestication alleles between divergent rice gene pools.

From wild grass to staple crop
More than 10,000 years ago, ancient peoples began to gather and consume Oryza rufipogon, an unruly wild grass species that grew in the swamps and marshes throughout tropical and subtropical Asia (Box 1). Through a process of continuous selection for desirable features, these early farmers slowly transformed wild rice into Oryza sativa, which is now an essential staple crop for billions of people worldwide. Indeed, rice is now the primary source of food and livelihood for more than a third of the world’s population, and is produced on every continent with arable land [1].

Domestication involves a series of profound genetic changes resulting from selection that make a wild species more amenable for cultivation and consumption by humans. It is widely recognized that domestication is not a single ‘event’, but rather a dynamic evolutionary process that occurs over time and, in some species, continues to this day [2]. The traits that distinguish modern rice varieties from their wild ancestor can range from subtle to dramatic (Figure 1). In addition to traits that resulted in major alterations of plant structure and/or reproductive physiology, humans have selected for characteristics that made rice grains more appealing as a food source, including grain size, shape, color, fragrance and amylose content.

Another feature of domestication traits is that they are generally quantitative in nature, meaning they are under the control of numerous genetic components that often interact in complex biochemical and regulatory pathways. For example, in rice, there are at least five grain shattering loci (see Glossary) [3–5], a minimum of six dormancy

Glossary

Abscission layer: a zone of cells at the interface of plant organs that breaks down during senescence, promoting the organ to fall off.
Accession: a sample of seed collected to represent a species, population or variety.
Admixed: an organism whose genome contains DNA inherited from different subpopulations as the result of outcrossing and recombination.
Apiculus: the tip of the lemma or palea; a small point at the tip of the seed hull.
Dormancy: a physiological period of quiescence during which a mature seed will not germinate.
F$_{ST}$: a measure of population differentiation based on polymorphism data. It compares the genetic variability within and between populations. If two populations have completely divergent DNA sequences, their $F_{ST}$ value equals 1, indicating complete differentiation. If two populations have identical DNA sequences, their $F_{ST}$ value equals 0, indicating no differentiation. Therefore, when comparing two populations, the larger the $F_{ST}$ value (the closer to 1), the more differentiated those populations are from one another.
Glume: small bract found at the base of cereal grains.
Grain shattering: seed abscission; when ripe seed falls from the panicle before harvesting occurs.
Haplotype: a combination of alleles or sequence variants at multiple linked loci that are transmitted together.
Hybridization: the result of successful outcrossing and fertilization of two genetically distinct plants.
Introgression: the movement of a discrete portion of a genome from one genotype into another via meiotic recombination.
Isozymes: variants of an enzyme; proteins that differ in amino-acid sequence, but catalyze the same reaction.
Landrace: a primitive variety consisting of a heterogeneous mixture of genotypes selected and maintained by early farmers.
Life history habit: the reproductive cycle of a plant; annual habit refers to plants that live, reproduce and die in one year or season, whereas perennial habit refers to plants whose life cycle spans more than one year.
Molecular clock: using DNA polymorphisms between the sequences of two organisms to deduce the amount of time that has elapsed since they diverged from a common ancestor.
Panicle: a compound raceme; the inflorescence of a rice plant.
Pericarp: the outermost layer of cells in the seed coat.
Photoperiodic response: a plant’s ability to flower in response to increasing or decreasing day length.
Phylogenetic: referring to the evolutionary relationships between organisms.
Retrotransposon: a type of transposable element (mobile DNA) that moves through an RNA intermediate inserting new copies of itself throughout the genome of an organism.
RFLP marker: restriction fragment length polymorphism; a DNA marker that detects genetic differences between individuals based on the presence or absence of restriction enzyme sites.
SNP: single nucleotide polymorphism.
SSR marker: simple sequence repeat; a DNA marker that detects genetic differences between individuals based on the length of short repetitive sequences found throughout the genome.
Subpopulation: a population within a species that is genetically divergent from other populations within that species.
Tiller: a branch of the rice plant including roots, culms, leaves and panicle.
Thresh: the process of removing ripe grains from harvested panicles.
Transcription factor: a protein that affects the transcription and thus expression of specific target genes.
Box 1. The wild ancestor of *O. sativa*

The wild ancestors of our cultivated crop species not only represent valuable sources of genetic diversity for crop improvement, but also provide a window through which we can catch a glimpse of the evolutionary history of our domesticated species. Several lines of evidence suggest that *O. sativa* was domesticated from *O. rufipogon*, also known as Asian common wild rice. Numerous genetic studies using molecular markers or DNA sequence information demonstrate that *O. sativa* cultivars are more closely related to *O. rufipogon* than to any other *Oryza* species, with *O. rufipogon* accessions often grouping together with *O. sativa* in phylogenetic (see Glossary) analyses [19,30]. The geographic ranges of *O. rufipogon* and *O. sativa* overlap throughout Asia, and there are no major reproductive barriers between them, leading to a continuous series of intermediate admixed (see Glossary) genotypes [37]. These findings clearly demonstrate the close relationship between cultivated Asian rice and *O. rufipogon*.

*O. rufipogon* is a complex species that includes both annual (*Oryza nivara*) and perennial (*O. rufipogon*) forms. Some researchers distinguish between *O. rufipogon* and *O. nivara* due to their variation in life history habit [63,64]. Others consider them to be ecotypes of a single species because they exhibit continuous variation in nature for their life history habit, coupled with the fact that there is no major reproductive barrier or significant genetic differentiation between them [19,31,37,65], as demonstrated by the abundance of shared sequence polymorphisms, lack of fixed differences, and very low *Fst* values [19,42,66]. As such, we will treat these two species as a single large gene pool referred to as *O. rufipogon*.

In addition to the suite of traits that were altered during domestication, the evolutionary history of rice is not complete without considering the deep population structure. Numerous genetic studies using molecular markers or DNA sequence information demonstrate that *O. sativa* cultivars are more closely related to *O. rufipogon* than to any other *Oryza* species, with *O. rufipogon* accessions often grouping together with *O. sativa* in phylogenetic (see Glossary) analyses [19,30]. The geographic ranges of *O. rufipogon* and *O. sativa* overlap throughout Asia, and there are no major reproductive barriers between them, leading to a continuous series of intermediate admixed (see Glossary) genotypes [37]. These findings clearly demonstrate the close relationship between cultivated Asian rice and *O. rufipogon*.

Table 1. Key domestication-related genes cloned in rice

<table>
<thead>
<tr>
<th>Gene</th>
<th>Chr*</th>
<th>Trait affected</th>
<th>Functional mutation</th>
<th>Molecular function</th>
<th>Refs</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sh4</td>
<td>4</td>
<td>Grain shattering</td>
<td>SNP causing an amino-acid substitution</td>
<td>Myb3 transcriptional regulator</td>
<td>[48,49]</td>
</tr>
<tr>
<td>qSH1</td>
<td>1</td>
<td>Grain shattering</td>
<td>SNP in regulatory region</td>
<td>BEL1-type homeobox transcriptional regulator</td>
<td>[50]</td>
</tr>
<tr>
<td>Rc</td>
<td>7</td>
<td>Grain pericarp color</td>
<td>Deletion causing protein truncation</td>
<td>Basic helix-loop-helix transcriptional regulator</td>
<td>[11]</td>
</tr>
<tr>
<td>Waxy</td>
<td>6</td>
<td>‘Sticky’ (glutinous) grains</td>
<td>SNP in intron affecting mRNA splicing</td>
<td>Granule-bound starch synthase</td>
<td>[52-54]</td>
</tr>
<tr>
<td>GS3</td>
<td>3</td>
<td>Grain size/shape</td>
<td>SNP causing protein truncation</td>
<td>Cellular signaling protein with a VWFC module</td>
<td>[55]</td>
</tr>
<tr>
<td>BAD2*</td>
<td>8</td>
<td>Grain fragrance/flavor</td>
<td>Deletion causing protein truncation</td>
<td>Betaine aldehyde dehydrogenase</td>
<td>[58]</td>
</tr>
<tr>
<td>Gn1a*</td>
<td>1</td>
<td>Grain number</td>
<td>Several possible mutations</td>
<td>Cytokinin oxidase/dehydrogenase</td>
<td>[57]</td>
</tr>
<tr>
<td>GW2*</td>
<td>2</td>
<td>Grain weight/width</td>
<td>Deletion causing protein truncation</td>
<td>RING-type protein with E3 ubiquitin ligase activity</td>
<td>[56]</td>
</tr>
</tbody>
</table>

*Chr, chromosome.

*Bracketed genes affect domestication-related traits, but there is no evidence at this time to determine whether these genes were part of the domestication process or whether they are novel mutations that have recently become the targets of selection.
Although *indica* and *japonica* represent the deepest genetic differentiation within *O. sativa*, five major subpopulations (see Glossary) are widely recognized (*indica*, *tropical japonica*, *temperate japonica*, *aus* and *aromatic*) [17,18] (Box 2). This population structure adds further complexity to our understanding of the domestication process that led to cultivated rice.

Modern genetics provides the necessary tools to understand the dynamics of the domestication process and to identify the key genes and alleles that were the targets of selection during rice domestication. Although only a limited number of genetic loci affecting key domestication traits in rice have been characterized at the gene level, an examination of the evolutionary history of these domestication genes provides novel insights into the movement of both cultivated rice and human populations across Asia. This review will critically examine the existing hypotheses used to describe the rice domestication process by integrating previous studies of genome-wide variation in rice with recent data on the evolutionary history of key domestication genes. The view that emerges is one that challenges us to rethink old paradigms as we continue to unravel the complex story of rice domestication.

**Prevailing hypotheses regarding rice domestication**

Several major hypotheses describe the process of rice domestication, each attempting to account for the evolution of the deeply differentiated subpopulations of *O. sativa*. First, several researchers have suggested that the *indica* varietal group was originally domesticated from *O. rufipogon* and the *japonica* group was derived later from *indica* [15,19,20]. The greater genetic diversity found within *indica* compared to *japonica* [17] and the close genetic relationship between *indica* and existing populations of both annual and perennial forms of *O. rufipogon* [21] provided evidence to support this hypothesis. The second theory is that the *indica–japonica* differentiation occurred as a result of adaptation to different ecological and geographical environments following a single domestication of *O. sativa* from *O. rufipogon* [22,23]. A third hypothesis, originally proposed by Chou [16], is that rice was domesticated independently at least twice from a pre-differentiated ancestral *O. rufipogon* gene pool [17,24–28].

An understanding of both the patterns of genome-wide variation in wild and cultivated rice and the number and origins of alleles associated with rice domestication genes

---

**Box 2. Subpopulation structure in *O. sativa***

There are two genetically distinct varietal groups within *O. sativa*, referred to as *indica* and *japonica*. Traditionally, the two groups have been distinguished based on morphological characters, including grain shape, apiculus (see Glossary) hair length, leaf color or through biochemical assays for reaction to phenol and sensitivity to potassium chloride [37,67]. There are also numerous reproductive barriers between the modern *indica* and *japonica* groups [60]. Yet the range of variation for any one of these phenotypic traits overlaps between the two groups, leading to confusion regarding the classification of particular genotypes [37].

Since publication of the first rice genetic map [68], DNA markers have been widely used to explore the genetic architecture of rice. RFLP markers (see Glossary) readily detect the differentiation between *indica* and *japonica* [69,70]. Isozymes, simple sequence repeat (SSR) markers (see Glossary), and single nucleotide polymorphisms (SNPs) provide a higher resolution of population structure. Using 169 SSR markers on a set of 234 diverse *O. sativa* genotypes, Garris et al. identified five subpopulations: *indica*, *aus*, *tropical japonica*, *temperate japonica* and *aromatic* [17] (Figure I). The same groups were identified using SNP markers derived from 111 randomly sequenced regions of the genome on a subset of 72 accessions [34]. These studies were both consistent with the original study by Glaszmann [18] using 15 isozyme markers on a larger set of 1700 diverse *O. sativa* genotypes. Of these five subpopulations, *indica* and *aus* belong to the *indica* varietal group whereas *tropical japonica*, *temperate japonica* and *aromatic* are closely related to the *japonica* varietal group. A high degree of differentiation exists between the groups, with pairwise estimates of $F_{st}$ ranging from 0.20 (tropical versus temperate *japonica*) to 0.45 (*indica* versus temperate *japonica*) [17].

The results of these studies highlight the importance of utilizing molecular markers to assign genotypes to the appropriate subpopulation for genetic analyses. For example, the Basmati rice (known for their aroma and specialized grain quality) have traditionally been classified as members of the *indica* varietal group because of their long, slender grain shape. However, the Basmati rice genotypes form a genetically distinct subpopulation (*aromatic*) that is more closely related to *japonica* than *indica* ($F_{st}$ for *tropical japonica–aromatic* is 0.23; $F_{st}$ for *indica–aromatic* is 0.39) [17]. Similarly, the *aus* varieties have traditionally been referred to as *indica* for lack of distinguishing morphological differences, but they are genetically divergent from *indica* ($F_{st}$ for *indica–aus* is 0.26) [17].

**Figure I.** Subpopulation structure in *O. sativa*. *O. sativa* is characterized by the presence of deep genetic differentiation. This unrooted phylogenetic tree was constructed from data using 169 nuclear SSR and two chloroplast markers on 234 landraces of *O. sativa* [17]. The branch structure reflects the phylogenetic relationships based on the nuclear SSR markers. The branch color corresponds to the chloroplast haplotype of each accession. This tree clearly illustrates the major division between the two varietal groups (*indica* and *japonica*), which are further subdivided into the five rice subpopulations: *indica*, *aus*, *tropical japonica*, *temperate japonica* and *aromatic*. Reproduced and modified, with permission, from Ref. [17].
will enable us to consider the validity of the proposed domestication hypotheses. According to the first hypothesis, if *japonica* genotypes were derived from *indica*, we would expect *O. rufipogon* genotypes to be more closely related to *indica* than to the *O. rufipogon* ancestor. Additionally, under this hypothesis we would expect to see a common set of domestication alleles in both groups with evidence that the alleles had originated in *indica* and were later introgressed into *japonica*. In the second scenario, if a single domestication from *O. rufipogon* were followed by post-domestication divergence between *indica* and *japonica*, we would expect sequence differences between *indica* and *japonica* to postdate the time of domestication. Again under this theory, a common set of domestication alleles would be shared between the two groups. According to the third hypothesis, if *indica* and *japonica* were domesticated independently from a pre-differentiated *O. rufipogon* ancestor, we would expect sequence differences between the two groups to predate the time of domestication. We would also expect there to be unique sets of *indica*- and *japonica*-specific domestication alleles originating from genetically distinct *O. rufipogon* populations. These group-specific alleles would each contribute to the same vital domestication phenotypes in their respective varietal groups. It is highly unlikely that an identical mutation would have arisen and been selected for in these two gene pools, had they been independently domesticated.

**Evolutionary history of genome-wide variation in *O. rufipogon* and *O. sativa***

To address these hypotheses, various studies have examined *O. rufipogon* for evidence of differentiation into *indica* and *japonica* types. Genetic analysis of *O. rufipogon* chloroplasts revealed two distinct haplotypes (see Glossary), which have clear frequency differences between the *indica* and *japonica* groups of *O. sativa* [26,29]. Support for a differentiated *O. rufipogon* gene pool also comes from multiple studies using isozymes (see Glossary), restriction fragment length polymorphisms (RFLPs), single nucleotide polymorphisms (SNPs; see Glossary), transposable elements and the published genomic sequence, all demonstrating that *indica* and *japonica* accessions (see Glossary) are more closely related to certain *O. rufipogon* accessions than to each other [19,24,26,30–34].

The diverse geography and ecology of the broad region that is home to *O. rufipogon* (East, South and Southeast Asia) suggests that this ancestral species was broadly adapted to diverse climates and would have provided ample genetic diversity from which to have domesticated *O. sativa* on numerous occasions. Studies exploring the diversity within *O. rufipogon* have recently reported genetically identifiable population sub-structure with at least some hint of geographic associations [25,26,34,35]. The degree of genetic differentiation between populations of *O. rufipogon* increases with geographical distance, suggesting that geographical isolation played a major role in establishing the predifferentiated gene pools within *O. rufipogon* [35]. Also, the inbreeding habit that accompanied rice domestication and transformed the highly outcrossing *O. rufipogon* (outcrossing rate of 30–50%) [36,37] to the almost exclusively inbreeding *O. sativa* (outcrossing rate ~2%) [37,38] would have further promoted isolation between rice subpopulations.

A further source of evidence supporting multiple domestications comes from recent work estimating the divergence time between *indica* and *japonica*. The availability of complete genomic sequence for both *japonica* and *indica* [39–41] has allowed the estimation of the divergence time between these two groups. Two independent studies used a molecular clock approach (see Glossary) to calculate the *indica*- *japonica* divergence time to be around 0.4 million years ago (mya) [42,43]. Another study evaluating the patterns of retrotransposon insertion (see Glossary) in the two varietal groups estimated the divergence time to be at least 0.2 mya [44]. This discrepancy may be due to the fact that the level of divergence varies across the genome [45]. In any case, these molecular clock estimates are several orders of magnitude earlier than the first archaeological evidence for rice consumption by humans (Box 3), providing strong evidence that the divergence of the two cultivated rice gene pools predated rice domestication. Taken together, these studies contradict the first two hypotheses and support multiple domestications of *O. sativa* from a predifferentiated *O. rufipogon* ancestor.

**Evolutionary history of cloned rice domestication genes**

To determine whether the history of rice domestication alleles is consistent with multiple domestications, we will now examine several examples of cloned rice domestication genes whose evolutionary histories have recently been explored. If *indica* and *japonica* were independently domesticated as the patterns of genome-wide variation suggest, we would expect to see different domestication alleles present in the different varietal groups.

**Grain shattering: Sh4**

One of the most universal changes during the domestication of the cereal crops was a reduction in grain shattering. This change was essential for humans to efficiently harvest the crop, but rendered the plant almost entirely dependent on humans for dispersal [46]. In a cross between an *indica* cultivar and an annual form of *O. rufipogon*, Li et al. identified a major QTL, **shattering4** (*sh4*), which affects grain shattering [47]. The *Sh4* gene was subsequently cloned [48,49] and, although its exact function remains unknown, SH4 is thought to be a transcription factor (see Glossary) involved in cell wall degradation and/or establishment of the abscission layer (see Glossary) that releases the grain from the panicle (see Glossary) [48,49]. The functional nucleotide polymorphism in the gene was identified as an SNP causing a single amino acid change within the predicted DNA-binding domain of SH4.

The *sh4* mutation is found in all five *O. sativa* subpopulations, whereas the dominant *Sh4* allele exists in all wild, shattering accessions of *O. rufipogon* along with six other closely related species of *Oryza* [47]. Lin et al. recently examined *sh4* (which they refer to as *sha1*) in 96 *indica* and 112 *japonica* cultivars and found that they all contained the same functional SNP associated with the domesticated allele [49]. Additionally, *sh4* cDNA sequencing
Box 3. Effects of geography and climate on rice diversity

As early as the late 19th Century, De Candolle suggested that rice cultivation began in China, where the oldest historical records could be found [71]. By contrast, Vavilov designated India as the location of rice domestication [72]. Since then, countless studies have generally agreed that the birthplace of rice domestication lies within a broad arc extending from eastern China, through the foothills of the Himalayas in Vietnam, Thailand and Myanmar, to eastern India [20,37]. More recently, archaeological studies have strongly supported the Middle and lower Yangtze River Valley in China as the site where rice was first domesticated (Figure II). Some hypothesize that the rice remains discovered at these sites in China represent foraged wild forms instead of truly domesticated forms of rice [46]. Yet, there is mounting evidence to the contrary. A new study investigating an archaeological site in Lower Yangtze region found evidence of paddy rice cultivation, suggesting that the inhabitants of this region had already formed agricultural societies around rice by 7,700 years ago [73]. Also, analysis of phytolith samples (silicon fossils of rice plant cells) from the Diaotonghuan site in Northern Jiangxi province (China) dated the earliest known samples of domesticated rice, based on a morphological indicator (double-peaked glume (see Glossary) cell phytoliths), between 9,000-10,000 years ago [74].

Despite this extensive archaeological evidence supporting the Yangtze River Valley in China as a region where rice was first domesticated, the wild O. rufipogon populations in this region today exhibit very low genetic diversity. An examination of twelve wild-rice populations in China demonstrated that the lowest genetic diversity existed in the northern-most populations near the Yangtze River [38]. It is also well known that the japonica varietal group, believed to have undergone domestication in this region, has dramatically less genetic diversity than the indica group [17]. The reasons for the lower genetic diversity of japonica relative to indica can be explained by reviewing the history of the region. A major short-term climate change during a period known as the Younger Dryas resulted in a return to glacial-like conditions across Northern Asia from 11,500 to 13,000 years ago [75,76]. The colder climate would have eliminated a large portion of the japonica-like wild ancestors in the Yangtze River Valley. As a result, humans would have been forced to rely on a narrowing gene pool, motivating a more rapid movement toward domestication of rice in this region. By contrast, the warmer tropics of South and Southeast Asia would have maintained larger, more diverse populations of indica-like O. rufipogon that would have been foraged by humans for a longer time, resulting in a more gradual domestication process. Under this scenario, the domestication bottleneck would have been more abrupt and severe in the japonica gene pool than in the indica gene pool, consistent with the available genetic data [17].

revealed an identical allele in four indica and four japonica accessions. The authors suggest that these results indicate the non-shattering sh4 allele arose and was selected before the differentiation of indica and japonica from a common O. rufipogon ancestor. Although this is a logical conclusion, an alternative hypothesis is that the domesticated sh4 allele arose and was selected during the domestication of one of the O. sativa subpopulations and was subsequently introgressed into all rice genotypes. To clarify the origin and routes of global dispersal of the non-shattering sh4 allele, it will be necessary to examine the haplotype structure around the Sh4 gene in diverse populations of O. rufipogon and O. sativa.

Grain shattering: qSH1

In addition to the Sh4 gene, a second major shattering QTL, qSH1, was recently identified from an intra-specific cross between japonica (cultivar Nipponbare) and aus (cultivar Kasalath) [50]. The functional mutation was a regulatory SNP 12 kilobases upstream of a BEL1-type homeobox transcription factor, which decreased expression of the transcription factor only at the provisional abscission layer, resulting in reduced shattering.

By examining the haplotypes around the qSH1 gene in 118 rice lines including five O. rufipogon accessions, Konishi et al. deduced the subpopulation origin of this allele [50]. Two clear haplotype groups were found, corresponding to the indica and japonica varietal groups. The japonica haplotypes carrying the non-shattering allele were most closely related to shattering O. rufipogon accession W1943, which originated in China and grouped closely with japonica cultivars from a previous study [32]. This led to the conclusion that the causative mutation in the domesticated qsh1 allele occurred in early domesticates of the japonica varietal group [50]. The non-shattering allele was prevalent in the temperate japonica subpopulation, but was absent in accessions from tropical japonica and indica, suggesting that this allele was not widely disseminated following its fixation within a portion of the temperate japonica gene pool. This distribution pattern is also
consistent with the finding that *temperate japonica* varieties tend to be more difficult to thresh (see Glossary) than varieties from other subpopulations, reflecting the presence of a non-shattering allele in addition to *sh4*.

**Pericarp color: Rc**

All wild *Oryza* species and some of the early *O. sativa* landraces (see Glossary) exhibit red pigmentation of the seed coat. However, virtually all modern rice cultivars lack red pigmentation and appear white or beige. Classical genetic analysis identified a single locus in rice, *Rc*, that conditioned the change from red to white pericarp (see Glossary) [13]. The *Rc* gene was recently cloned and shown to encode a basic helix–loop–helix (bHLH) transcription factor [11]. The *rc* allele differs from the wild-type allele by a 14-base pair (bp) deletion that results in truncation of the protein before the bHLH domain and causes the change from red to white pericarp [11,12].

To trace the origin and distribution of the *rc* allele, Sweeney et al. [51] evaluated 440 geographically and genetically diverse rice cultivars, representing landraces and modern varieties from all five of the subpopulations defined by Garris et al. [17] (Box 2). They determined that 98% (330/337) of rice accessions with white pericarp contained the 14-bp deletion, whereas the mutation was not found in any of the landrace varieties with red pericarp (n = 103). To determine the subpopulation origin of the 14-bp mutation, Sweeney et al. [51] used diverse red-grained varieties to identify polymorphisms in a 6.5-kilobase (kb) region around the *Rc* gene whose alleles were specific to either red *indica* or red *japonica*. They observed clear haplotype differentiation between *indica* and *japonica* across the region, making it possible to trace the origin of the *rc* allele conferring white pericarp. The white *rc* haplotype differed from the ancestral *japonica* haplotype by only the 14-bp deletion (Figure 2). This analysis suggests that the mutation associated with the change from red to white pericarp in 98% of the *O. sativa* varieties surveyed originated in a *japonica* ancestor [51]. This study presents clear genetic evidence for a domestication allele arising in the *japonica* varietal group and subsequently spreading across genetic, geographic and ethnic boundaries to become fixed in all cultivated rice.

**Pericarp color: Rc-s**

Whereas 98% of surveyed rice accessions with white pericarp had the same 14-bp deletion, two *aromatic* and five *aus* varieties had white pericarp but did not contain the *rc* allele. These seven varieties were sequenced to look for alternative mutations in the *Rc* gene that could explain the lack of pigments [51]. A comparison of the sequences from these seven varieties to nine control varieties with red pericarp (from all five subpopulations) identified a single SNP in close proximity to the where the 14-bp deletion was in all the other white accessions. This second mutation (*Rc*-s) introduces a premature stop codon that truncates the protein before the bHLH domain, similar to the *rc* mutation [51]. Interestingly, none of the plants contained a complete knockout of this gene, suggesting that the protein is critical to plant function independently of its ability to promote red grain color. All seven individuals that contained the *Rc*-s allele had an identical haplotype, which differed from the ancestral (red pericarp) *aus* haplotype only at the functional SNP. Thus, it was concluded that the *Rc*-s allele, which was only found in *aus* and a small percentage of *aromatic* accessions, had its origins in the *aus* subpopulation [51]. This clearly demonstrates that different mutations in the same gene can arise in different subpopulations over the course of domestication. Both *rc* and *Rc*-s apparently became the targets of selection and are virtually indistinguishable at the phenotypic level. It is interesting to note that, unlike the *rc* and *sh4* alleles, but
similar to the non-shattering qSH1 allele, the Rc-s allele was not widely disseminated during the domestication of O. sativa.

Amylose content: Wx
Glutinous or ‘sticky’ rice is used as a staple food in Southeast Asia and has long-standing culinary and cultural importance in China, Japan and Korea. Japonica rice is typically valued for their low amylose content, leading to a soft, sticky grain texture ideal for eating with chopsticks, whereas indica rice has a higher amylose content leading to firm grains that flake apart when cooked. The glutinous phenotype is the result of a defect in the Waxy (Wx) gene, which encodes a granule-bound starch synthase responsible for amylose biosynthesis in the grain [52]. There are two functional forms of Wx (non-glutinous), with a clear divergence between the two varietal groups in rice: Wx\(^a\) is predominant in indica and Wx\(^b\) is predominant in japonica. The Wx\(^a\) allele was fixed in all O. rufipogon accessions examined [53], indicating that the Wx\(^b\) allele is the wild type and Wx\(^b\) is the derived form, having been selected during the domestication of japonica.

Unlike the Wx\(^a\) allele, the Wx\(^b\) allele contains an SNP at the 5′ splice site of the first intron, causing incomplete post-transcriptional processing and lower amylose production in these genotypes [52,54]. A second mutation in either Wx\(^a\) or Wx\(^b\) creates the wx allele, which is recessive and fully glutinous. Glutinous rice strains can have either Wx\(^a\)- or Wx\(^b\)-derived wx alleles, but the Wx\(^b\)-derived wx allele is predominant, as it was found in 97% of 353 glutinous samples from both varietal groups collected from diverse regions across Asia [53]. This finding demonstrates that the majority of glutinous rice cultivars from both indica and japonica carry the japonica-derived allele, mirroring the situation described above for the rc mutation. Also, similar to the Rc-s and qsh1 alleles, glutinous genotypes derived from the indica Wx\(^a\) allele were not widely disseminated.

Other domestication-related traits
The examples presented above include only the rice domestication genes that have been cloned and investigated in terms of their evolutionary origins. Several other genes that affect domestication-related traits have been cloned in rice, including GS3 and GW2, which condition grain size and weight [55,56], Gn1a, associated with grain number [57], and BAD2, which conditions fragrance [58] (Table 1). Further examination of these genes will determine whether they were involved in the rice domestication process or have been the targets of more recent selection.

What we can interpret from the examples presented above is that several key domestication alleles had a single origin and subsequently became fixed across all rice subpopulations. In the case of white pericarp, the rc allele arose in the japonica gene pool and was then introgressed into the other subpopulations [51]. This story is repeated for the Wx\(^b\)-derived glutinous rice mutation (wx) that originated in the japonica group and was also introgressed and dispersed [53]. In the future, we might find that the sh1 allele for non-shattering followed a similar evolutionary path. By contrast, several other domestication alleles were not widely disseminated across subpopulations. The Rc-s allele for white pericarp is largely restricted to the aus subpopulation, with some introgression (see Glossary) into the aromatic subpopulation [51]. The qsh1 allele for non-shattering and the Wx\(^a\)-non-glutinous allele both remain largely contained within the temperate japonica subpopulation [50,53].

Current observations regarding the history of rice domestication
The examples of rice domestication genes, coupled with the studies involving genome-wide variation in both O. rufipogon and O. sativa, cannot be reconciled with any of the three prevailing hypotheses about rice domestication. Available genetic data suggest that there were predifferentiated gene pools within the O. rufipogon ancestor and that these gene pools gave rise to the indica and japonica varietal groups. Yet, the presence of identical domestication alleles in both groups strongly suggests that these domestications were not entirely independent. Gene flow across the distinct gene pools of rice appears to have been crucial to the domestication process. Taken together, the current data supports the recently proposed ‘combination model’ for rice domestication [59]. In this model, the early japonica and indica cultivars were domesticated from divergent O. rufipogon populations and subsequent introgression of key domestication alleles between these early domesticates resulted in a common set of domestication alleles being fixed in all modern varieties. Although all of the evolutionary data from domestication genes thus far demonstrates gene flow from japonica to indica, only a limited number of rice domestication alleles have thus far been described. This situation is likely to change as the costs of high-resolution genotyping and resequencing continue to decline, and the development of new statistical approaches improves our ability to identify and characterize the targets of selection in diverse germplasm. Whether future genetic studies involving larger numbers of domestication genes will substantiate this observation of gene flow from japonica to indica or will identify ubiquitous domestication alleles arising in other subpopulations remains to be determined.

If we accept the ‘combination model’ for rice domestication, how can we reconcile the maintenance of deep genetic differentiation between rice varietal groups with the movement of domestication genes? A recent study involving SNP data on 20 accessions of O. rufipogon and 72 landrace and modern O. sativa cultivars demonstrated that O. rufipogon has the greatest level of heterozygosity, followed by the landraces, with very low levels observed in modern cultivars. These data are consistent with a gradual change from an outcrossing to an inbreeding habit during the domestication process; a change that coincided with the rise of the deeply differentiated subpopulations of O. sativa [34]. Numerous reproductive barriers help maintain this genetic differentiation in modern O. sativa [60], but cross-hybridization was likely to have been easier in early landrace varieties. Evidence of indica x japonica hybridization (see Glossary) comes from the recovery of early domesticates containing indica-like organellar genomes in
The story of rice domestication is one of both genetic exchange and genetic containment. It involved episodes of hybridization and introgression between ancient gene pools against a backdrop of emerging inbreeding and sterility barriers that progressively restricted gene flow between subpopulations. These competing evolutionary forces, shaped by human selection, gave rise to an array of interconnected, well-differentiated subpopulations of *O. sativa*. Although these subpopulations are intimately associated with their specific geographies, ecologies and cultures, the history of rice domestication is a dynamic rather than static story.

Today, these same forces of genetic exchange and containment are at work, and the tools of modern genetics provide new opportunities for the identification and the mobilization of genetic variation. High-yielding hybrid rice varieties take advantage of the traditional subpopulation structure to systematically create new combinations of genes by bringing together inbred parents from different subpopulations. The development of inter- and intraspecific introgression lines mimics the ancient road to domestication by allowing only selected segments of the donor genome to be introgressed into adapted genetic backgrounds. Genomics techniques offer a rapid way of identifying useful variation, which can then be exploited with transgenic technology. Yet, with all of these possibilities for genetic manipulation available to the modern plant breeder, the most valuable source of genetic variation today is still the wild germplasm itself. Learning how humans once domesticated crops from wild ancestors might provide more than a retrospective. As we seek new ways of harnessing genetic variation in the context of the 21st century, understanding the dynamics of crop domestication might offer important lessons for the future.

Acknowledgements

We acknowledge financial support from the Plant Genome Program of the National Science Foundation (Award Numbers #0606461 and #0110004) to S.R.M. We are grateful to Lois Swales for formatting and administrative help, and Jennifer Kimball-Pevo and Lisa Polewczak for photographic contributions. Also, many thanks to Laurent Sagart for numerous conversations concerning rice domestication.

References

587

Five things you might not know about Elsevier

1. Elsevier is a founder member of the WHO’s HINARI and AGORA initiatives, which enable the world’s poorest countries to gain free access to scientific literature. More than 1000 journals, including the Trends and Current Opinion collections and Drug Discovery Today, are now available free of charge or at significantly reduced prices.

2. The online archive of Elsevier’s premier Cell Press journal collection became freely available in January 2005. Free access to the recent archive, including Cell, Neuron, Immunity and Current Biology, is available on ScienceDirect and the Cell Press journal sites 12 months after articles are first published.

3. Have you contributed to an Elsevier journal, book or series? Did you know that all our authors are entitled to a 30% discount on books and stand-alone CDs when ordered directly from us? For more information, call our sales offices:

   +1 800 782 4927 (USA) or +1 800 460 3110 (Canada, South and Central America)
   or +44 (0)1865 474 010 (all other countries)

4. Elsevier has a long tradition of liberal copyright policies and for many years has permitted both the posting of preprints on public servers and the posting of final articles on internal servers. Now, Elsevier has extended its author posting policy to allow authors to post the final text version of their articles free of charge on their personal websites and institutional repositories or websites.

5. The Elsevier Foundation is a knowledge-centered foundation that makes grants and contributions throughout the world. A reflection of our culturally rich global organization, the Foundation has, for example, funded the setting up of a video library to educate for children in Philadelphia, provided storybooks to children in Cape Town, sponsored the creation of the Stanley L. Robbins Visiting Professorship at Brigham and Women’s Hospital, and given funding to the 3rd International Conference on Children’s Health and the Environment.

www.sciencedirect.com