·Invited Review·

The Puzzle of Rice Domestication

Tao Sang[1][*] and Song Ge[2]

([1]*Department of Plant Biology*, *Michigan State University*, East Lansing, MI 48824, USA;

[2]*State Key Laboratory of Systematic and Evolutionary Botany*, *the Institute of Botany*, *the Chinese Academy of Sciences*, Beijing 100093, China)

Abstract

The origin of cultivated rice has puzzled plant biologists for decades. This is due, at least in part, to the complex evolutionary dynamics in rice cultivars and wild progenitors, particularly rapid adaptive differentiation and continuous gene flow within and between cultivated and wild rice. The long-standing controversy over single versus multiple and annual versus perennial origins of cultivated rice has been brought into shaper focus with the rapid accumulation of genetic and phylogenetic data. Molecular phylogenetic analyses revealed ancient genomic differentiation between rice cultivars, suggesting that they were domesticated from divergent wild populations. However, the recently cloned domestication gene *sh4*, responsible for the reduction of grain shattering from wild to cultivated rice, seems to have originated only once. Herein, we propose two models to reconcile apparently conflicting evidence regarding rice domestication. The snow-balling model considers a single origin of cultivated rice. In this model, a core of critical domestication alleles was fixed in the founding cultivar and then acted to increase the genetic diversity of cultivars through hybridization with wild populations. The combination model considers multiple origins of cultivated rice. In this model, initial cultivars were domesticated from divergent wild populations and fixed different sets of domestication alleles. Subsequent crosses among these semi-domesticated cultivars resulted in the fixation of a similar set of critical domestication alleles in the contemporary cultivars. In both models, introgression has played an important role in rice domestication. Recent and future introgression of beneficial genes from the wild gene pool through conventional and molecular breeding programs can be viewed as the continuation of domestication.

Key words: *Oryza sativa*; *Oryza rufipogon*; *Oryza nivara*; phylogeny; population genetics; QTL; shattering; *sh4*.

Domestication that began approximately 10 000 years ago provided the foundation for civilization (Diamond 2002). Plant domestication primarily solved the problem of food shortage

and instability, and allowed steady human population growth and modernization of society. With only a small fraction of flowering plant species currently in cultivation, plant domestication will continue to meet the future needs of human society, such as the need for energy crops (Ragauskas et al. 2006). A better understanding of past domestication is instructive in ongoing efforts of crop domestication and improvement.

The biological, historical, and cultural dynamics of domestication have intrigued scientists for more than a century (Zeder et al. 2006). Our understanding of the genetic basis of domestication has recently leapt forward owing to rapid advances in genomics. In particular, the cloning of genes controlling key domestication traits in major crops has opened up a new avenue through which to investigate the molecular genetic mechanisms and population processes of domestication (Doebley et al. 2006).

Rice is the most important crop in Asia and provides staple

food for half of the world's population. Completion of the sequencing of the rice genome offered an unprecedented opportunity for studying rice genetics (Goff et al. 2002; Yu et al. 2002; International Rice Genome Sequencing Project 2005). Consequently, there have been many recent publications concerning rice domestication (e.g. Konishi et al. 2006; Li et al. 2006a, 2006b; Londo et al. 2006; Lu et al. 2006; Olsen et al. 2006; Sweeney et al. 2006; Tang et al. 2006; Zhu et al. 2007). In the present paper, we review the historical controversy and recent evidence regarding rice domestication, discuss factors confounding the study of rice domestication, propose models for rice domestication, and offer prospects for future investigations.

Controversy Regarding Rice Domestication

Going through the literature regarding rice domestication, one can find agreement that rice (*Oryza sativa*) was domesticated from wild Asian species belonging to the A-genome group of the genus *Oryza* (Chang 1976; Second 1982; Oka 1988; Wang et al. 1992; Khush 1997; Ge et al. 1999). However, controversy has persisted on two major issues. One is which wild species, *O. nivara* or *O. rufipogon*, served as the direct wild progenitor of cultivated rice and the other is whether rice was domesticated once or multiple times from divergent wild populations.

The controversy is due, at least in part, to the taxonomical inconsistency of the wild progenitors. *O. rufipogon* had been the species name widely used to accommodate the wild Asian A-genome taxa until *O. nivara* was recognized (Sharma and Shastry 1965). *O. nivara* was established for populations that were annual, photoperiod insensitive, predominantly self-fertilized, and adapted to seasonally dry habitats. *O. rufipogon* was retained for populations that were perennial, photoperiod sensitive, largely cross-fertilized, and adapted to persistently wet habitats (Sharma et al. 2000). Although both species were accepted in the recent classification of *Oryza* (Vaughan 1994; Lu et al. 2001), the argument that they should be treated as ecotypes or subspecies of *O. rufipogon* has continued (e.g. Morishima 2001; Cheng et al. 2003; Vaughan and Morishima 2003; Zhu et al. 2007). In the present review, the two taxa will be referred to as different species, namely *O. nivara* and *O. rufipogon*.

Regardless of the taxonomical status, these two ecologically distinct taxa are readily distinguished in nature. Seeds of *O. nivara* germinate at the beginning of the rainy season and plants flower and produce seeds before they decrease in the dry season, which starts in autumn. A suite of morphological traits associated with self-fertilization was taxonomically diagnostic for *O. nivara* (Vaughan 1989), including partially inserted, compact flowering panicles, small anthers (approximately 2

mm long), and a high seed set (> 90% in our greenhouse). *O. nivara* produces large and heavy seeds. These traits ensure a high reproductive allocation and maximize the transfer of photosynthetic product to seed production, which is consistent with the annual habit of the species (MA Grillo et al., unpubl. data, 2006).

The perennial species *O. rufipogon* lives in stable habitats and reproduces primarily by vegetative propagation. It flowers only under short day conditions, which begin in autumn in natural habitats. The cross-fertilization of *O. rufipogon* is likely to be promoted by morphological traits, including open flowering panicles distantly exserted from the leaf sheath and large anthers (>3 mm long). The seed set is relatively low and seeds are slender.

Along with the taxonomical controversy, it has long been debated which species or ecotype is the direct progenitor of cultivated rice (Chang 1976, 2003; Oka 1988; Morishima 2001). The hypothesis of an origin from *O. nivara* was based on the phenotypic similarity between *O. nivara* and *O. sativa*, including annuality, self-fertilization, and high reproductive allocation (Chang 1976; Khush 1997; Sharma et al. 2000). The hypothesis of an origin from *O. rufipogon* emphasized the benefit of higher genetic diversity of the out-crossing progenitor (Oka 1988).

However, either hypothesis alone does not seem to explain the considerable diversity of rice cultivars. Although the *O. nivara* origin accounts for the morphological and physiological similarity between cultivated rice and the wild progenitor, cultivars that are grown in deep water conditions or have relatively strong photoperiod sensitivity resemble *O. rufipogon*. Conversely, with the realization of the large phenotypic gap between *O. sativa* and *O. rufipogon*, authors who advocated the *O. rufipogon* origin did not rule out the possibility of rice domestication from populations intermediate between *O. nivara* and *O. rufipogon* (Oka 1988). The intermediate populations were found in usually disturbed habitats in Thailand (Sano et al. 1980). However, the evolutionary origins of these populations and their relationship to cultivated rice have not been subject to molecular phylogenetic investigation.

In addition to the complex situation in the wild progenitor, a high genetic diversity within cultivated rice has also posed challenges to the understanding of rice domestication (Second 1982; Glaszmann 1987). Two major types of cultivars, recognized as *O. sativa* subspecies *indica* and subspecies *japonica*, are not only morphologically and physiologically diverged, but are partially isolated by a post-zygotic barrier (Li et al. 1997; Chang 2003). Recent characterization of the genetic structure of cultivated rice further identified subdivisions within each subspecies, including *aus* and *indica* in traditionally defined *indica* and *temperate japonica*, *tropical japonica*, and *aromatic* in traditionally defined *japonica* (Garris et al. 2005).

A highly diverse crop species like rice almost automatically

invites arguments over single versus multiple origins. Alternative hypotheses include the independent origins of *indica* and *japonica* rice versus the origin of *japonica* within *indica* (e.g. Chang 1976; Second 1982; Oka 1988).

Recent Phylogenetic Analyses of Rice Domestication

Phylogenetic analyses of rice domestication have focused on addressing questions of when, where, and how many times rice was domesticated. Recent phylogenetic and population studies using a variety of molecular markers seemed to have reached the following consensus. First, the sampled accessions of the wild progenitors *O. rufipogon* and *O. nivara* did not form monophyletic groups within each species (Lu et al. 2002; Park et al. 2003; Zhu and Ge 2005; Kwon et al. 2006). Second, cultivated rice did not show a clearly closer relationship with one wild species compared with the other (Ge et al. 1999; Lu et al. 2002; Sun et al. 2002; Zhu and Ge 2005). Third, the major types of cultivars, such as subspecies *indica* and *japonica*, tended to form monophyletic groups separated by a relatively large genetic distance (Wang et al. 1992; Park et al. 2003; Garris et al. 2005; Zhu and Ge 2005; Kwon et al. 2006; Zhu et al. 2007).

Using restriction fragment length polymorphism (RFLP) markers and 93 accessions of 21 *Oryza* species, Wang et al. (1992) found that the *indica* and *japonica* cultivars were more closely related to the wild species than to each other. Sun et al. (2002) studied 75 cultivated varieties and 118 *O. rufipogon* strains using markers from nuclear, mitochondrial, and chloroplast genomes, and demonstrated a relatively high level of genetic differentiation between *indica* and *japonica* cultivars in both nuclear and cytoplasmic genomes. Zhu and Ge (2005) used sequences of four nuclear loci to investigate the phylogenetic relationships between the A-genome species of *Oryza*. They found that the cultivars fell into two monophyletic groups, with one containing *indica* rice and some accessions of *O. nivara* and *O. rufipogon*, and the other containing *japonica* rice and other accessions of both wild species.

The earliest effort to trace the geographic origin of rice with molecular markers was undertaken by using 40 allozyme loci (Second 1982). That study suggested that *japonica* rice was domesticated in China and *indica* rice was domesticated in other places of tropical Asia. A recent phylogeographic study of 203 cultivars and 129 wild accessions suggested that *indica* was domesticated within a region south of the Himalayan mountain range, including eastern India, Myanmar, and Thailand, whereas *japonica* may have originated from southern China (Londo et al. 2006). The results are consistent, to some extent, with archaeological findings of ancient rice cultivation in the Yangtze basin of southern China and northeastern India (Zhao 1998; Fuller 2006).

Molecular dating of divergence time between *indica* and *japonica* rice has yielded intriguing results. Three reports using different types of molecular data estimated the time of genomic divergence between *indica* and *japonica* to be between 0.4 and 0.2 million years ago (Ma and Bennetzen 2004; Vitte et al. 2004; Zhu and Ge 2005), which considerably predates the time of rice domestication approximately 10 000 years ago (Khush 1997; Zhao 1998). These studies all seem to suggest, in one way or another, that divergent rice cultivars were domesticated independently from genetically differentiated wild populations at different geographic locations.

Recent Genetic Analyses of Rice Domestication Traits

Quantitative trait locus (QTL) mapping has proven to be an effective initial step to reveal the genetic basis of phenotypic evolution under natural and human selection (Tanksley 1993; Mauricio 2001; Barton and Keightley 2002; Paterson 2002). To date, three QTL studies have been devoted to characterizing the genetic architecture of rice domestication. Two papers analyzed mapping populations derived from independent crosses between *indica* cultivars and the perennial wild species *O. rufipogon* (Xiong et al. 1999; Cai and Morishima 2002), and the third analyzed a cross between *indica* and the annual wild species *O. nivara* (Li et al. 2006a).

A suite of morphological and physiological transitions must have occurred during rice domestication, no matter which wild species was the direct wild progenitor. These include reduction in grain shattering and seed dormancy, synchronization of seed maturation, reduction in tiller number, increase in tiller erectness, increase in panicle length and branches, and reduction in coloration and awn length (Xiong et al. 1999; Cai and Morishima 2002; Thomson et al. 2003; Uga et al. 2003; Li et al. 2006a). Most of the changes are also found in other cereal crops, and were considered to be the domestication syndrome of cereals (Harlan 1975; Hancock 2004).

The reduction in shattering and seed dormancy and synchronization of seed maturation, which may have been selected unconsciously on most occasions (Harlan 1975), were necessary for improving the efficiency of harvest and planting to make the earliest efforts of cultivation worthwhile. The decrease in tiller numbers in cultivated rice facilitates a synchronous tiller production and maturation and, consequently, minimizes non-productive tillers that flower too late to contribute to harvest. Tiller erectness reduces the degree of tiller overlap and, thus, also increases the capacity of tiller cultivation per unit field. The increase in panicle length and branches resulted in the production of a larger number of grains per panicle. Therefore, selection for higher yield was most likely the driving

force for the morphological modifications.

Additional phenotypic changes had to occur if rice was domesticated from *O. rufipogon* rather than from *O. nivara*, including the switches from out-crossing to a selfing mating system, and from perennial to annual habit. An ongoing study has suggested that during the evolutionary origin of *O. nivara* from an *O. rufipogon*-like ancestor, the evolution of self-fertilization and annual life history has a complex genetic basis (MA Grillo et al., unpubl. data, 2006). Two previous QTL analyses of *indica* rice and *O. rufipogon* also identified loci underlying phenotypic difference associated with the mating system and life history. These included anther size, panicle shape and exsertion, and the presence of horizontal tillers (Xiong et al. 1999; Cai and Morishima 2002).

However, when *indica* was crossed with *O. nivara*, no segregation on mating system or life history related traits was observed in the F$_2$ population (Li et al. 2006a). There are two explanations of these results. First, the *indica* rice was domesticated from *O. nivara* because their phenotypic similarity was genetically based. Second, rice was domesticated from *O. rufipogon* and nature and humans have independently selected the same genes to give rise to a similar phenotype between *O. nivara* and *O. sativa.* Although we still do not have evidence to reject either hypothesis, it is more parsimonious to consider the origin of *indica* rice from the annual ancestor.

A QTL study of the domestication syndrome involving *O. nivara* indicated that at least some mutations of relatively large phenotypic effect were selected by humans to improve key domestication traits, such as shattering, seed dormancy, synchronization of seed maturation, plant architecture, and panicle morphology (Li et al. 2006a). Most strikingly, the QTL of the largest effect for almost all morphological traits were colocalized on the short arm of chromosome 7. These could represent a pleiotropic mutation or tightly linked mutations. In any event, selection for mutations of large effect, pleiotropic effect, or in tight linkage facilitates the development and maintenance of the domestication syndrome (Doganlar et al. 2002; Paterson 2002; Doebley 2004).

Cloning Shattering QTL and Implications for Rice Domestication

The cloning and subsequent molecular evolutionary analysis of a key domestication QTL, namely *tb1*, provided essential evidence for the single origin of maize (Doebley et al. 1997; Wang et al. 1999). The finding was soon substantiated by a detailed phylogenetic analysis of maize cultivars and wild relatives (Matsuoka et al. 2002). Similar analyses of other domestication genes in maize, barley, and tomato also yielded insights into the origins of the crops (Nesbitt and Tanksley 2002; Wang et al. 2005; Komatsuda et al. 2007). Thus, the recent cloning of the rice shattering QTL provides new opportunities for testing the hypotheses of rice domestication (Konishi et al. 2006; Li et al. 2006b).

With moderate morphological changes, rice domestication highlights the importance of physiological transitions, such as a reduction in shattering. The effort to genetically dissect the shattering trait recently led to the cloning of two shattering QTL. S*h4* was a shattering QTL identified from crosses between *indica* rice and wild species that explained 69% of phenotypic variance between the cultivar and *O. nivara* (Li et al. 2006a) and 15%–20% between the cultivar and *O. rufipogon* (Xiong et al. 1999; Cai and Morishima 2002). The other shattering QTL, *qSH1*, was identified from a cross between *indica* and *japonica* rice, with a phenotypic effect of 68.6% between these cultivars (Konishi et al. 2006).

S*h4*, mapped near the end of the long arm of chromosome 4, was positionally cloned and the functional mutation was confirmed to be a single nucleotide substitution in the first exon of a gene with a previously unknown function (Li et al. 2006b). This substitution led to an amino acid substitution from lycine to asparagine in the predicted MYB3 DNA-binding domain. The nuclear localization of sh4 protein, together with bioinformatic analysis, suggested that the gene was a transcription factor.

Sh4 is required for the development of the abscission zone between a grain and pedicle that controls programmed cell separation. The amino acid substitution weakened the function of *sh4* and caused the incomplete development of the abscission zone. However, the mutation did not eliminate abscission zone function and allowed manual separation of the grains from the pedicle, a part of the rice harvest process known as threshing. Thus, the early farmers selected a mutation that changed the way of seed dispersal from adaptive to natural habitats to cultivated conditions. It is conceivable that this type of mutation occurred at a low frequency and the chance for independent selections would have been very small.

A survey of 17 diverged rice cultivars that cover all five recognized subdivisions in *indica* and *japonica* rices revealed that the cultivars invariably carried the same functional mutation (Li et al. 2006b). In contrast, all wild species with conformed shattering phenotype had the wild-type *sh4.* Furthermore, the sequence variation of *sh4* was substantially reduced in cultivated rice compared with the wild species. The comparison of a 1.7-kb region of the gene indicated that there was only a 3-bp insertion/deletion in the 5' non-translated region among the 17 cultivars, whereas the nucleotide polymorphism among the wild A-genome species is comparable to other housekeeping genes.

We then compared the sequence of *sh4* and surrounding regions between the *indica* variety Guangluai 4 and the *japonica* variety Nipponbare, whose chromosome 4 had been sequenced (Feng et al. 2002). In the approximate 50-kb region containing *sh4*, there were 19 single nucleotide polymorphisms (SNP)

between the cultivars. The average SNP density in this region is calculated as 0.38 per kb. For the entire chromosome 4, the average SNP density was found to be 3.5 per kb (B Han, pers. comm., 2006). The nearly 10-fold reduction of sequence poly-morphism in the 50-kb region containing *sh4* further indicates a strong selection sweep at the major shattering locus. Thus, the evidence supports the single origin of the non-shattering *sh4* allele in cultivated rice.

Grains of *indica* are generally easier to thresh than those of *japonica* (Chang 2003). Thus, *indica* rice is more likely to be susceptible to yield loss due to shattering in the process of field harvest. This loss, however, could have been compensated for by easy threshing, especially for farmers who traditionally applied relatively weak force for threshing. *qSH1* was the major QTL accounting for the shattering difference between *indica* and *japonica* cultivars.

A functional SNP was identified approximately 12 kb upstream of the replumless (RPL) homolog of *Arabidopsis* (Konishi et al. 2006). In *Arabidopsis*, RPL was involved in the development of the valve margins of the siliques, which is essential for silique dehiscence. This putative regulatory mutation further weakened the development of the abscission zone between a rice grain and pedicel. Thus, the functional SNP was attributed to the selection for the non-shattering phenotype during the domestication of *japonica* rice (Konishi et al. 2006).

Rather than providing answers to the previous questions concerning rice domestication, the cloning of the two major shattering QTL seem to have brought the questions into sharper focus. Does the single origin of the non-shattering *sh4* allele and its fixation in all diverse cultivars surveyed so far mean a single origin of cultivated rice? If so, was the non-shattering *qSH1* allele derived when *japonica* rice was developed from existing cultivars that already carried the non-shattering *sh4* allele? If so, was *japonica* rice originated within *indica*? If *japonica* was domesticated directly from the wild species when *qSH1* was selected to reduce shattering, the non-shattering *sh4* allele must have been introgressed subsequently from other cultivars to *japonica* and became quickly fixed. Despite the rapid accumulation of molecular genetic and phylogenetic data over the past several years, it is remarkable how difficult it remains to reconcile the conflicting evidence. This indicates that rice domestication is, indeed, a very complex puzzle. Several factors that contribute to the difficulty in understanding rice domestication are worth discussing.

Difficulties in the Study of Rice Domestication

Continuous gene flow among cultivars and wild relatives has confounded the phylogenetic reconstruction of rice origin. Owing to the lack of a post-zygotic barrier among *O. sativa*, *O. nivara*, and *O. rufipogon*, gene flow between any two species could occur if they flower at the same time and within a close proximity. Although *O. nivara* and *O. rufipogon* are isolated prezygotically through ecogeographic and phonological barriers, there could be an overlap of flowering time in early autumn, especially for those *O. nivara* populations that flower late owing to delayed water availability (MA Grillo et al., unpubl. obs., 2006). Together with the recency of their divergence, introgression may have added another layer of difficulty to phylogenetically separating the two species.

The flowering time of *O. sativa* is largely determined by sowing time. The variation in sowing time provides ample opportunities for the overlap of flowering time between *O. sativa* and the wild species. Gene flow between cultivated and wild rice occurs in both directions. The introgression of valuable genes from wild to cultivar species has been a continuous breeding practice for cultivar improvement (Khush 1997, 2001). Meanwhile, constant gene flow from cultivars to the wild species, especially to out-crossing *O. rufipogon*, has produced weedy rice frequently seen near rice fields (Oka 1988; Morishima 2001).

Even though natural and artificial selection could guard species identity, they may operate effectively only at loci important for adaptation (Rieseberg et al. 2004; Wu and Ting 2004). The vast genomic regions that are adaptively neutral can potentially move freely among species given enough opportunities for gene flow and recombination. Thus, using neutral markers to trace the wild ancestor(s) of cultivated rice faces the considerable challenge of sorting out the correct phylogenetic history from that obscured by introgression.

Another related difficulty involves taxonomical inconsistency and misidentification in germplasm collections (Virk et al. 1995; Ge et al. 2001). Recent molecular studies of rice domestication relied primarily on the extensive germplasm collection at the International Rice Research Institute. For various reasons, the identification of the collection is not free of taxonomical problems and misidentification has been reported (Wang et al. 1992; Martin et al. 1997; Aggarwal et al. 1999; Bao et al. 2005).

Despite the morphological, physiological, and ecological distinctions between *O. nivara* and *O. rufipogon*, correct identification based on herbarium specimens alone can be challenging. In addition, although typical individuals of the two species are clearly distinguishable in the field, intermediate forms have been found in Thailand and could be assigned to either species (Sano et al. 1980). Because of these difficulties and the hesitant acceptance of *O. nivara*, it is probably not uncommon that accessions typically belonging to *O. nivara* are still named *O. rufipogon*. All of these must have contributed to the obscured species boundary between *O. nivara* and *O. rufipogon* in germplasm collections and in molecular phylogenies.

This type of taxonomical problem is actually quite widespread in plants. Species delineation varies depending on morphological, biological, ecological, and phylogenetic species concepts.

Because different species definitions could affect the phylogenetic interpretation of rice origins, we suggest that any attempt to address questions concerning the phenotypic and ecological differences between these two wild species should grow the sampled accessions for correct identification. Otherwise, considering all accessions under *O. rufipogon* is an alternative and practical solution.

Another taxonomical problem that one may not so easily get away from is the classification of weedy rice. The introgression from cultivated to wild rice has created an array of phenotypic intermediates known as weedy rice (e.g. Bres-Patry et al. 2001). Although weedy rice has been classified as a subspecies of cultivated rice, such as *O. sativa* ssp. *spontanae* and ssp. *fatua* (Vaughan 1989), it must have also have been practically recognized as *O. rufipogon* or *O. nivara* in germplasm collections. For example, three accessions classified as *O. nivara* in the International Rice Research Institute (IRRI) collection carried the non-shattering *sh4* alleles. Greenhouse observation conformed that the three accessions had the non-shattering phenotype and showed a combination of traits of cultivated and wild species (Li et al. 2006b). These are very likely the cases where weedy rice was classified as *O. nivara*. No matter whether weedy rice is classified as cultivated or wild species, it is a source of misleading information in the molecular phylogenetic and population genetic studies of rice domestication.

Models of Rice Domestication

Despite the difficulties associated with the study of rice domestication, important new insights have been gained from recent studies. Multiple lines of phylogenetic evidence indicate that the genomes of major cultivar groups came from different wild populations that diverged considerably earlier than the time of rice domestication. However, the non-shattering allele of *sh4*, primarily responsible for the reduction of grain shattering from wild to cultivated rice, originated once during rice domestication.

Here we propose two models to reconcile the apparently conflicting evidence regarding rice domestication (Figure 1). Model I considers the single origin of cultivated rice. In this model, domestication started from a relatively small wild population and continued for a relatively long period of time. The earliest cultivar should have fixed alleles for the majority of critical domestication traits, including the non-shattering allele of *sh4*. This well-developed early cultivar was then introduced to other regions of Asia. During cultivation, introgression occurred frequently between the cultivar and local populations of *O. rufipogon* and *O. nivara* under primitive agricultural management. Modern cultivars, such as *indica* and *japonica*, were derived as hybrids between the early cultivar and

diverged wild populations. A different genomic background of wild populations was selected in different cultivars for optimized adaptation to different climatic conditions and agricultural practices. The major alleles underlying key domestication traits were maintained in all cultivars during the process.

Thus, the model can be best viewed as snow-balling. That is, the core of the domestication alleles, once fixed in the founding cultivar, was selected and maintained during the process of introgression and cultivar diversification. Meanwhile, introgression continued to bring beneficial genes from wild populations into cultivars. Domestication proceeded somewhat semi-automatically when the core of the domestication alleles rolled through the wild gene pool and enlarged the genetic diversity of cultivated rice.

Model II considers multiple origins of cultivated rice. In this model, rice domestication started from diverged wild populations. Rice was considered to be domesticated as long as it had a combination of alleles that made cultivation worthwhile for those early farmers. Alleles fixed for the same trait were different among cultivars. For example, a cultivar may be quite resistant to shattering because it fixed the allele at *sh4* but still had relatively strong seed dormancy because it only fixed alleles of small effect on seed dormancy. Another cultivar may be less resistant to shattering, but had substantially reduced seed dormancy and much improved yield owing to the fixation of major
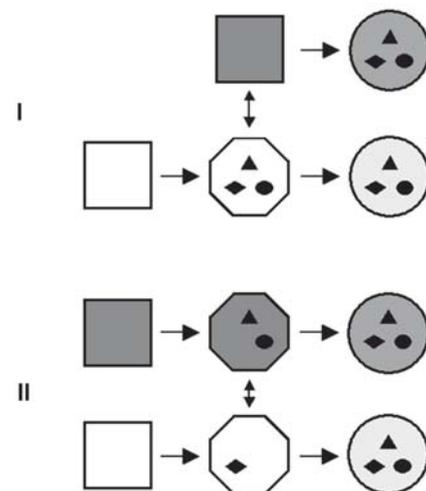

Figure 1. Schematic illustration of models of rice domestication.

I, snow-balling model; II, combination model. Squares represent wild populations from which rice was domesticated; hexagons represent the earliest domesticated rice; circles represent contemporary rice cultivars. Differences in shadings indicate genomic divergence. Shapes inside (triangle, diamond, and ellipse) represent critical domestication alleles that are now fixed in the contemporary cultivars. Double-headed arrows indicate hybridization. Single arrowheads point to the progress of domestication.

alleles for plant architecture. Subsequent crosses between these semi-domesticated cultivars at an early stage of rice cultivation allowed farmers to select the best alleles for the critical domestication traits. The selection was so strong for those alleles that they became quickly and widely fixed in cultivated rice. This is a combination model.

Both models allow for the maintenance of a high genetic diversity and the fixation of the same set of major domestication alleles in cultivars. In either model, the crossability between early cultivars and between cultivars and wild species was essential. Hybridization may have been an important means of improving domestication traits through transgressive segregation, which has been shown to drive new adaptation in natural plant populations (Rieseberg et al. 2003). In fact, QTL alleles that could potentially contribute to higher yield have been identified in wild rice species (Xiao et al. 1996, 1998). The reproductive barrier between major cultivars, such as *indica* and *japonica*, could have developed much later when the maintenance of local adaptation of cultivars became an important issue.

Both models also allow an increase in the genetic diversity of cultivars following a usually strong genetic bottleneck during domestication, including rice (Zhu et al. 2007). Interestingly, the introgression has been facilitated continuously by breeding programs that more effectively moved beneficial genes from wild to cultivated plants. With technical advances in molecular breeding, wild gene pools will play an increasingly important role in the genetic improvement of cultivars (Tanksley and McCouch 1997; Khush 2001; Zeder et al. 2006). It is consistent with our models to view this as the continuation of domestication on a much finer genetic scale.

These models can be tested through cloning and phylogenetic analysis of major QTL alleles for key domestication traits that are now fixed in cultivated rice. If the origins of these alleles are traced back to the same or closely related wild populations, the snow-balling model is supported. If the origins of these alleles are traced back to divergent wild populations, the combination model is supported. Nevertheless, the two models are not mutually exclusive. We may end up with a case where most of critical domestication alleles had the same origin, whereas the rest were derived from diverged wild populations, which does not fit the conventional view of single or multiple origins. In any event, tracing domestication with domestication alleles may prove to be effective for unraveling the origin of a crop with a complex evolutionary history. In addition to the cloned domestication alleles (Konishi et al. 2006; Li et al. 2006b; Sweeney et al. 2006), other important domestication alleles are likely the remaining pieces of the rice domestication puzzle. We will have an increasingly clearer picture of the puzzle as these pieces are found and put together.

References

Aggarwal RK, Brar DS, Nandi S, Huang N, Khush GS (1999). Phylogenetic relationships among *Oryza* species revealed by AFLP markers. *Theor. Appl. Genet.* **98**, 1320–1328.

Bao Y, Lu BR, Ge S (2005). Identification of genomic constitutions of *Oryza* species with the B and C genomes species by the PCR-RFLP method. *Genet. Res. Crop Evol.* **52**, 69–76.

Barton NH, Keightley PD (2002). Understanding quantitative genetic variation. *Nat. Rev. Genet.* **3**, 11–21.

Bres-Patry C, Lorieux M, Clément G, Bangratz M, Ghesquière A (2001). Heredity and genetic mapping of domestication-related traits in a *temperate japonica* weedy rice. *Theor. Appl. Genet.* **102**, 118–126.

Cai HW, Morishima H (2002). QTL clusters reflect character associations in wild and cultivated rice. *Theor. Appl. Genet.* **104**, 1217–1228.

Chang TT (1976). The origin, evolution, cultivation, dissemination, and diversification of the Asian and African rices. *Euphytica* **25**, 425–441.

Chang TT (2003). Origin, domestication, and diversification. In: Smith CW, Dilday RH, eds. *Rice*: *Origin, History*, *Technology*, *and Production*. John Wiley & Sons, Hoboken, NJ. pp. 3–25.

Cheng C, Motohashi R, Tsuchimoto S, Fukuta Y, Ohtsubo H, Ohstubo E (2003). Polyphyletic origin of cultivated rice: Based on the interspersion pattern of SINEs. *Mol. Biol. Evol.* **20**, 67–75.

Diamond J (2002). Evolution, consequences and future of plant and animal domestication. *Nature* **418**, 700–707.

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

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

Doebley JF, Gaut BS, Smith BD (2006). The molecular genetics of crop domestication. *Cell* **127**, 1309–1321.

Doganlar S, Frary A, Daunay MC, Lester RN, Tanksley SD (2002). Conservation of gene function in the Solanaceae as revealed by comparative mapping of domestication traits in eggplant. *Genetics* **161**, 1713–1726.

Feng Q, Zhang Y, Hao P, Wang S, Fu G, Huang Y et al. (2002). Sequence and analysis of rice chromosome 4. *Nature* **420**, 316–320.

Fuller DQ (2006). Agricultural origins and frontiers in south Asia: A working synthesis. *J. World Prehist* **20**, 1–86.

Garris AJ, Tai TH, Coburn J, Kresovich S, McCouch S (2005). Genetic structure and diversity in *Oryza sativa* L. *Genetics* **169**, 1631–1638.

Ge S, Sang T, Lu BR, Hong DY (1999). Phylogeny of rice genomes with emphasis on origins of allotetraploid species. *Proc. Natl. Acad. Sci. USA* **96**, 14400–14405.

Ge S, Sang T, Lu BR, Hong DY (2001). Rapid and reliable identification of rice genomes by RFLP analysis of PCR-amplified *Adh* genes. *Genome* **44**, 1136–1142.

Glaszmann JC (1987). Isozymes and classification of Asian rice varieties. *Theor. Appl. Genet.* **74**, 21–30.

Goff SA, Ricke D, Lan TH, Presting G, Wang R, Dunn M et al. (2002). A draft sequence of the rice genome (*Oryza sativa* L. ssp. *japonica*). *Science* **296**, 92–100.

Hancock JF (2004). *Plant Evolution and The Origin of Crop Species*, 2nd edn. CABI Publishing, Cambridge, MA.

Harlan JR (1975). *Crops & Man*. American Society of Agronomy, Madison, WI.

International Rice Genome Sequencing Project (2005). The map-based sequence of the rice genome. *Nature* **436**, 793–800.

Khush GS (1997). Origin, dispersal, cultivation and variation of rice. *Plant Mol. Biol.* **35**, 25–34.

Khush GS (2001). Green revolution: The way forward. *Nat. Rev. Genet.* **2**, 815–822.

Komatsuda T, Pourkheirandish M, He C, Azhaguvel P, Kanamori H, Perovic D et al. (2007). Six-rowed barley originated from a mutation in a homeodomain-leucine zipper I-class homeobox gene. *Proc. Natl. Acad. Sci. USA* **104**, 1424–1429.

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

Kwon SJ, Lee JK, Hong SW, Park YJ, McNally KL, Kim NS (2006). Genetic diversity and phylogenetic relationship in AA *Oryza* species as revealed by Rim2/Hipa CACTA transposon display. *Gen. Genet. Syst.* **81**, 93–101.

Li C, Zhou A, Sang T (2006a). Genetic analysis of rice domestication syndrome with the wild annual species, *Oryza nivara*. *New Phytol.* **170**, 185–194.

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

Li Z, Pinson SRM, Paterson AH, Park WD, Stansel JW (1997). Genetics of hybrid sterility and hybrid breakdown in an intersubspecific rice (*Oryza sativa* L.) populations. *Genetics* **145**, 1139–1148.

Londo JP, Chiang YC, Hung KH, Chiang TY, Schaal BA (2006). Phylogeography of Asian wild rice, *Oryza rufipogon*, reveals multiple independent domestications of cultivated rice, *Oryza sativa*. *Proc. Natl. Acad. Sci. USA* **103**, 9578–9583.

Lu BR, Ge S, Sang T, Chen JK, Hong DY (2001). The current taxonomy and perplexity of the genus *Oryza* (Poaceae). *Acta Phytotax. Sin.* **39**, 373–388.

Lu BR, Zheng KL, Qian HR, Zhuang JY (2002). Genetic differentiation of wild relatives of rice as assessed by RFLP analysis. *Theor. Appl. Genet.* **106**, 101–106.

Lu J, Tang T, Tang H, Huang J, Shi S, Wu CI (2006). The accumulation of deleterious mutations in rice genomes: D hypothesis on the cost of domestication. *Trends Genet.* **22**, 126–131.

Ma J, Bennetzen JL (2004). Rapid recent growth and divergence of rice nuclear genomes. *Proc. Natl. Acad. Sci. USA* **101**, 12404–12410.

Martin C, Juliano A, Newbury HJ, Lu BR, Jackson MT (1997). The use of RAPD markers to facilitate the identification of *Oryza* species within a germplasm collection. *Genet. Res. Crop Evol.* **44**, 175–183.

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

Mauricio R (2001). Mapping quantitative trait loci in plants: Uses and caveats for evolutionary biology. *Nat. Rev. Genet.* **2**, 370–381.

Morishima H (2001). Evolution and domestication of rice. In: Khush GS, Brar DS, Hardy B, eds. *Rice Genetics IV. Proceedings of the Fourth International Rice Genetics Symposium.* International Rice Research Institute, Los Banos. pp. 63–77.

Nesbitt TC, Tanksley SD (2002) Comparative sequencing in the genus *Lycopersicon*: Implications for the evolution of fruit size in the domestication of cultivated tomatoes. *Genetics* **162**, 365–379.

Oka HI (1988). *Origin of Cultivated Rice.* Japan Scientific Society Press, Tokyo.

Olsen KM, Caicedo AL, Polato N, McClung A, McCouch S, Purugganan MD (2006). Selection under domestication: Evidence for a sweep in the rice *Waxy* genomic region. *Genetics* **173**, 975–983.

Park KC, Kim NH, Cho YS, Kang KH, Lee JK, Kim NS (2003). Genetic ariations of AA genome *Oryza* species measured by MITE-AFLP. *Theor. Appl. Genet.* **107**, 203–209.

Paterson AH (2002). What has QTL mapping taught us about plant domestication? *New Phytol.* **154**, 591–608.

Ragauskas AJ, Williams CK, Davison BH, Britovsek G, Cairney J, Eckert CA et al. (2006). The path forward for biofuels and biomaterials. *Science* **311**, 484–489.

Rieseberg LH, Raymond O, Rosenthal DM, Lai Z, Livingstone K, Nakazato T et al. (2003). Major ecological transitions in wild sunflowers facilitated by hybridization. *Science* **301**, 1211–1216.

Rieseberg LH, Church SA, Morjan CL (2004). Integration of populations and differentiation of species. *New Phytol.* **161**, 59–69.

Sano Y, Morishima H, Oka HI (1980). Intermediate perennial–annual populations of *Oryza perennis* found in Thailand and their evolutionary significance. *Bot. Mag. Tokyo* **93**, 291–305.

Second G (1982). Origin of the genic diversity of cultivated rice (*Oryza* spp.): Study of the polymorphism scored at 40 isozyme loci. *JPN J. Genet.* **57**, 25–57.

Sharma SD, Shastry SVS (1965). Taxonomic studies in genus *Oryza* L. III. *O. rufipogon* Griff. *sensu stricto* and *O. nivara* Sharma et Shastry *nom. nov. Indian J. Genet. Plant Breed.* **25**, 157–167.

Sharma SD, Tripathy S, Biswal J (2000). Origin of *O. sativa* and its ecotypes. In: Nanda JS, ed. *Rice Breeding and Genetics*: *Research Priorities and Challenges*. Science Publications, Enfield,

NH. pp. 349–369.

Sun C, Wang X, Yoshimura A, Doi K (2002). Genetic differentiation for nuclear, mitochondrial and chloroplast genomes in common wild rice (*Oryza rufipogon* Griff.) and cultivated rice (*Oryza sativa* L.). *Theor. Appl. Genet.* **104**, 1335–1345.

Sweeney MT, Thomson MJ, Pfeil BE, McCouch S (2006). Caught red-handed: *Rc* encodes a basic helix-loop-helix protein conditioning red pericarp in rice. *Plant Cell* **18**, 283–294.

Tang T, Lu J, Huang J, He J, McCouch SR, Shen Y et al. (2006). Genomic variation in rice: Genesis of highly polymorphic linkage blocks during domestication. *PLoS Genet.* **2**, 1824–1833.

Tanksley SD (1993). Mapping polygenes. *Annu. Rev. Genet.* **27**, 203–233.

Tanksley SD, McCouch SR (1997). Seed banks and molecular maps: Unlocking genetic potential from the wild. *Science* **277**, 1063–1066.

Thomson MJ, Tai TH, McClung AM, Lai X, Hinga ME, Lobos KB et al. (2003). Mapping quantitative trait loci for yield, yield components and morphological traits in an advanced backcross population between *Oryza rufipogon* and *Oryza sativa* cultivar jefferson. *Theor. Appl. Genet.* **107**, 479–493.

Uga Y, Fukuta Y, Cai HW, Iwata H, Ohsawa R, Morishima H et al. (2003). Mapping QTLs influencing rice floral morphology using recombinant inbred lines derived from a cross between *Oryza sativa* L. and *Oryza rufipogon* Griff. *Theor. Appl. Genet.* **107**, 218–226.

Vaughan DA (1989). *The Genus Oryza* L.: *Current Status of Taxonomy*. International Rice Research Institute, Manila, Philippines.

Vaughan DA (1994). *The Wild Relatives of Rice*: *A Genetic Resources Handbook*. International Rice Research Institute, Manila, Philippines.

Vaughan DA, Morishima H (2003). Biosystematics of the genus *Oryza*. In: Smith CW, ed. *Rice*: *Origin, History, Technology, and Production*. John Wiley & Sons, Hoboken, NJ. pp. 27–65.

Virk PS, Ford-Lloyd BV, Jackson MT, Newbury HJ (1995). Use of RAPD for the study of diversity within plant germplasm collections. *Heredity* **74**, 170–179.

Vitte C, Ishii T, Lamy F, Brar D, Panaud O (2004). Genomic paleontology provides evidence for two distinct origins of Asian rice (*Oryza sativa* L.). *Mol. Gen. Genet.* **272**, 504–511.

Wang H, Nussbaum-Wagler T, Li BL, Zhao Q, Vigouroux Y, Faller M et al. (2005). The origin of the naked grains of maize. *Nature* **436**, 714–719.

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

Wang ZY, Second G, Tanksley SD (1992). Polymorphism and phylogenetic relationships among species in the genus *Oryza* as determined by analysis of nuclear RFLPs. *Theor. Appl. Genet.* **83**, 565–581.

Wu CI, Ting CT (2004). Genes and speciation. *Nat. Rev. Genet.* **5**, 114–122.

Xiao J, Li J, Grandillo S, Ahn SN, McCouch SR, Tanksley SD (1996). Genes from wild rice improve yield. *Nature* **384**, 223–224.

Xiao J, Li J, Grandillo S, Ahn SN, Yuan L, Tanksley SD (1998). Identification of trait-improving quantitative trait loci alleles from a wild rice relative, *Oryza rufipogon*. *Genetics* **150**, 899–909.

Xiong L, Liu K, Dai K, Xu C, Zhang Q (1999). Identification of genetic factors controlling domestication-related traits for rice using an F_2 population of a cross between *Oryza sativa* and *O. rufipogon*. *Theor. Appl. Genet.* **98**, 243–251.

Yu J, Hu S, Wang J, Wong GKS, Li S, Liu B et al. (2002). A draft sequence of the rice genome (*Oryza sativa* L. ssp. *indica*). *Science* **296**, 79–92.

Zeder MA, Emshwiller E, Smith BD, Bradley DG (2006). Documenting domestication: The intersection of genetics and archaeology. *Trends Genet.* **22**, 139–155.

Zhao Z (1998). The middle Yangtze region in China is one place where rice was domesticated: Phytolith evidence from the Diaotonghuan Cave, northern Jaingxi. *Antiquity* **72**, 885–897.

Zhu Q, Ge S (2005). Phylogenetic relationships among A-genome species of the genus *Oryza* revealed by intron sequences of four nuclear genes. *New Phytol.* **167**, 249–265.

Zhu Q, Zheng X, Luo J, Gaut BS, Ge S (2007). Multilocus analysis of nucleotide variation of *Oryza sativa* and its wild relatives: Severe bottleneck during domestication of rice. *Mol. Biol. Evol.* **24**, 875–882.

(Handling editor: Xiao-Quan Wang)