Reloading DNA History in Rice Domestication

Takeshi Izawa*

Laboratory of Plant Breeding and Genetics, Department of Agricultural and Environmental Biology, The University of Tokyo, Yayoi, Bunkyo-Ku, Tokyo, 113-8657 Japan

*Corresponding author: E-mail, takeshizawa@g.ecc.u-tokyo.ac.jp

(Received 2 February 2022; Accepted 2 June 2022)

Although crop domestication is a prehistoric event, DNA (or genome) sequences of modern cultivars and the accession lines of wild relatives contain information regarding the history of crop domestication and the breeding process. Accordingly, with plentiful genomic data, many new findings have been obtained concerning the crop domestication process, for which various (some controversial) interpretations exist. Since approximately 20 years ago, dozens of quantitative trait genes (QTGs) related to the domestication process have been cloned from several crops including rice, a global staple food. However, the determination of how and when these QTGs were involved in rice domestication requires a precise understanding of the DNA code. In addition to the identification of domestication-related QTGs, large-scale rice genome analysis based on short-read Illumina data (but with shallow depth) including more than 1,000 rice cultivars and hundreds of wild rice (or *Oryza rufipogon*) lines, along with extensive genome analysis including more than 3,000 cultivars with sufficient Illumina data, has been reported. From these data, the genome-wide changes during rice domestication have been explained. However, these genome-wide changes were not interpreted based on QTG changes for domestication-related traits during rice domestication. In addition, a substantial gap remains between the archeological hypothesis based on ancient relics and findings from DNA variations among current cultivars. Thus, this review reconsiders the present status of rice domestication research from a biologist’s perspective.

**Keywords:** Domestication • Introgression • Local adaptation • Rice • Seed color • Seed shattering • Taste trait

**Introduction**

Darwin first began scientific explanation on animal and plant domestication (Darwin 1868). In his book entitled ‘The variation of animals and plants under domestication’, the changes of biological traits in some animals such as doves were extensively documented; prehistoric processes were described as a form of evolution. However, researchers in the fields of molecular phylogenetics and population genetics have hesitated to accept such domestication processes as evolutionary events because they are based on unnatural (or biased) selection by ancient humans; moreover, they are very rapid compared with typical evolutionary trait changes. Instead, archeologists have preferred to study domestication using various relics obtained from prehistoric ruins and precise dating technologies based on radioisotope measurement. Recently, developments in the cloning of quantitative trait genes (QTGs) based on quantitative trait loci analyses and genome-wide association studies and in population genomics analyses using abundant next-generation sequencing data have enabled us to address various questions related to rice domestication. QTGs conferring changes in domestication-related traits such as seed shattering, seed color, plant architecture and seed size have been cloned from rice since 2006 (Izawa 2007, Sweeney and McCouch 2007, Izawa et al. 2009, Civán and Brown 2017, Chen et al. 2021). Thus far, dozens of QTGs related to rice domestication have been cloned. In addition to these QTGs, several extensive population genomics analyses have been conducted. In 2012, Huang et al. reported a comprehensive map of rice genome variations based on short-read Illumina sequence data for 446 accessions that originated from geographically diverse areas of the wild rice species *Oryza rufipogon*, the ancestral progenitor of cultivated Asian rice, and from 1,083 rice cultivars including both indica and japonica; however, the corresponding depth of Illumina data was very shallow (Huang et al. 2012). Furthermore, Wang et al. (2018) reported genomic variations of 3,010 cultivars of Asian rice with various origins. Although many papers concerning rice domestication have been published based on analyses of these data, the findings are often controversial (Huang et al. 2012, Huang and Han 2015, Choi et al. 2017, Civán and Brown 2017, 2018, Choi and Purugganan 2018). One major reason for controversy is the lack of information regarding genome variations in wild rice because there are scarce genomic data for *O. rufipogon* (Huang et al. 2012). Thus, regardless of available cultivar information, the data remain insufficient to address a wide range of questions concerning rice domestication. Additionally, there are problems with the setting (or definition) of issues and topics related to rice domestication. For example, although the location where rice domestication occurred was a
重大课题在于Huang等(2012)的工作中，指明了每一株野生稻种质的自然生长区域可能与今天的野生水稻有所不同。这可以与过去的10,000年发生。一个相似的趋势是真实的，因为农作区域在不同时期由不同的栽培品种构成。因此，在对水稻基因组数据的分析中，很难地址到野生水稻的原始区域通过农作研究，尽管一些关联可能存在于当今的栽培品种和野稻种质之间。

另一个挑战在于水稻农作研究，即有无农作和栽培有关的基因组广泛数据可用于定量性状核苷酸(QTNs)。这可能是由古代人类选定的。此外，如此多的基因组信息应该与考古学证据相结合，因为对时间点和地点的了解是通过考古学获得的。这项回顾重新考虑了水稻农作研究的现状。

What Are High-Priority Challenges for Rice Domestication Research?

The definition of domestication

当我们报道了在2008年(Somura et al. 2008, Liu et al. 2017)，我们对六种农作相关QTG(qSW5, Rc, qSH1, sh4, waxy and Rd)的7个QTN进行了研究。我们发现，这些老品种的野生稻种质与原始QTN的选定有关，因此，人们认为，这些野生稻种质在农作过程中的选择事件以及早期培育的野生稻种质。然而，这可能与已知的野稻种质的驯化过程有关。

The date and place estimates of when the genome was obtained from genomic information. In addition, the resolution of dating differs between phylogenetic analysis using nucleotide sequences and archeological analysis using radioisotopes (Fuller et al. 2009, Choi et al. 2017). With these considerations, the data obtained from genome and QTN selection are expected to extensively overlap with archeological evidence. A similar tendency is true among molecular geneticists, who engage in endless controversy regarding the number of occurrences and locations of rice domestication. Therefore, we must clearly define domestication to address this question in a scientific manner.

Classification based on genome diversity and bottleneck selection

In a comprehensive map of genome variations among more than 3,000 rice cultivars (Wang et al. 2018), at least eight subgroups [XI-1, XI-2, XI-3, Aus, Bas, GJ-Subtrp, GJ-trp (or temperate japonica) and GJ-trp (or tropical japonica)] of rice cultivars are defined. These eight subgroups can be categorized into four major groups [XI (or indica), Aus, Bas and GJ (or japonica)]. This categorization suggests that at least four major bottleneck selection events, as well as a succession of minor bottleneck selection events, occurred to form the eight subgroups (or sub-species) during the domestication process or early breeding of rice. To precisely determine how the wild species of rice, O. rufipogon, contributed to these ancient selection events, more accurate genomic information is needed from diverse accessions of wild rice. Without such information, the domestication process would be simply an irritating logic problem. Nevertheless, many hypotheses have been proposed regarding the number of rice domestication events; one major subgroup, japonica (or GJ), diversified around 400,000 years ago from a common ancestor with another major subgroup of rice, known as indica (or XI). Apparently, this date was considerably earlier than the expected date of rice domestication. This finding was originally based on the evolutionary speed of long terminal repeat sequences of retrotransposons in the rice genome (Vitte et al. 2004); it has been supported by genome-wide assessments of japonica and indica (Choi et al. 2017). Because rice domestication presumably occurred fewer than 10,000 years ago (Fuller et al. 2009), some believe that sufficient genomic information is available at this moment from japonica (GJ) and indica (XI) cultivars to elucidate the rice domestication process. Thus, it remains controversial among rice molecular geneticists whether rice domestication occurred one, two or more times.

Basic selection processes during rice domestication

Let us imagine the beginning of human rice cultivation around 10,000 years ago. Until then, ancient humans gathered rice seeds from wild rice, which grew naturally in waterside areas, and then brought them home to utilize as food. Some of these gathered seeds grew into plants after they had been dropped at sites near the homes of those humans. A very strong genetic bottlenecks...
Fig. 1 Local adaptation of flowering-time genes in rice. Natural variations of at least eight heading-date (or flowering-time) genes are involved in local adaptation in Japan. The defective mutation in *Ghd7* is required for cultivation in a northern part of Japan (Hokkaido Island). Yumepirika is a cultivar adapted to the northern island of Japan, while Koshihikari and Nipponbare are cultivars adapted to the main island of Japan. Taichung 65 is a cultivar found in Taiwan. Genes with only lowercase letters have the selected QTNs. Cultivars having either functional (starting with uppercase letters) alleles or selected QTNs (starting with lowercase letters) are often found in the local area.

Critical introgressions between subgroups during rice domestication

In addition to these processes, substantial directional introgression, such as introgression events with *Rc* and *Kala4* seed color genes, should also be considered when establishing a comprehensive model of rice domestication (Civán and Brown 2018) (Fig. 2). First of all, consider *Kala4*, a black seed color gene (Oikawa et al. 2015), which is a distinct QTG to *Rc*, a well-known domestication gene that resulted in white color rice seed traits (Sweeney et al. 2006). Molecular analysis shows that the *Kala4* mutation occurred after rice domestication because wild rice accessions with black seed color are very rare (and to my knowledge none have been reliably published). Old studies state that the origin of black rice occurred at least 1,000 years ago because it was recorded as an imperial food in ancient China (Oikawa et al. 2015). When *Kala4* was cloned, black rice lines were found in various subgroups of rice, including both japonica and indica. Then, detailed sequencing data for *Kala4* genes from various rice
cultivars revealed that black rice may have originated from the ancient tropical japonica subgroup; however, black rice with an indica genetic background is most prevalent in China. This de novo insertion mutation in *Kala4*, conferring black seed color, was presumably introgressed into indica subgroups. The introgressed locus of *Kala4* might have been selected repeatedly over a few thousand years, resulting in the introgression of genome fragments of just a few hundred kb from tropical japonica to indica subgroups. The size of the introgressed fragments varies among indica black rice cultivars. Modern breeding using artificial crossing technologies in rice has enabled the replication of such selection. Thus, some black seed cultivars now exist in a temperate japonica background. In this context, the introgressed *Kala4* fragments have been further introgressed into temperate japonica modern cultivars. This process raises questions of how to explain *Kala4* in domestication research; namely, when and where was the *Kala4* gene involved in rice domestication or early rice breeding processes? Indeed, a similar process may have affected the *Rc* gene, a white seed color gene, because of the QTN selection of a 14-bp deletion; however, major cultivars of rice frequently possess this mutation (Sweeney et al. 2007), in contrast to the *Kala4* mutation. From such data, we can only estimate when such mutations first occurred; we cannot date the bottleneck selection event.

Notably, trait selection due to standing variations would have begun upon the emergence of trait changes because the presence of both original allele and selected alleles in wild rice would arise due to the weak effects of the mutations. It is well known that all of the tested cultivars contain the weak *sh4* QTN and some wild rice also possesses it (Li et al. 2006, Zhang et al. 2009, Zhu et al. 2012). In addition, the effects of the *sh4* QTN were masked in some genetic backgrounds of wild rice (Ishikawa et al. 2017). Thus, ancient plants possessing the weak-*sh4* QTN for good agronomical traits would have become targets for ancient human selection during the early domestication process. Because there are large genetic distances among the rice subgroups, the weak *sh4* QTN may have been introgressed into other subgroups of rice from the original subgroup line. Notably, the dating of directional introgressions through the comparison of introgressed lines with recurrent parent lines is very difficult; this difficulty becomes apparent when trying to estimate the dates of introgressions using nearly isogenic lines. However, population genomics analysis may provide insights into such introgression events too (Choi et al. 2017). Since another seed-shattering QTN in *qSH1* was only found in some temperate japonica cultivars, the selection for the weak *sh4* QTN occurred prior to the *qSH1* selection (Konishi et al. 2006). Accordingly, the identification of QTNs that affect agronomic traits and the mapping of these QTNs to diverse genomes of cultivars and wild rice accessions provide clues to estimate the timing of allele propagation among distinct QTNs to elucidate rice domestication in terms of molecular genetics.
Fig. 3 Example of complex introgression in rice. This is a revised version of Fig. 4D in the work by Itoh et al. (2018). Complex introgression of the *Hd1* gene in rice cultivars is summarized.

### Potentials of QTN mapping in diverse rice lines

When gene sequences of certain QTGs in wild rice are compared with those in cultivars, various outcomes will be possible. In other words, obtaining an integrated history of each gene’s evolution during rice domestication is difficult without such sufficient gene sequence information from diverse wild rice accessions. In case that a gene group of cultivars differs from those of the wild rice accessions with certain genetic distances, such a gene may have a critical role in rice domestication, although some members of the group have resulted from absolute directional introgression. In other scenarios, in case that a group of genes map to two haplotype groups, one group overlaps only with japonica, while the other group overlaps with both indica and japonica, suggesting the two-origin of domestication in terms of gene evolution. This pattern would be representative of the rice domestication and subsequent breeding processes based on current knowledge.

DNA sequence variations of certain key genes in various rice cultivars have been analyzed (Sweeney et al. 2006, 2007, Fujino et al. 2010, Asano et al. 2011, Jiang et al. 2012, Shao et al. 2013, Hori et al. 2015, Civán and Brown 2017, Choi and Purugganan 2018, Itoh et al. 2018, Liu et al. 2018a, b, Wu et al. 2018, 2020, Huang et al. 2020, Li et al. 2021, Zhang et al. 2019, 2021, Wang et al. 2022). For example, variations in the *Heading date 1* (*Hd1*), a heading-date gene, or one of the rice orthologs of the *Arabidopsis thaliana* CONSTANS gene, were extensively analyzed; the findings demonstrated that a large part of the temperate japonica (GJ-tmp) group is close to a group of wild rice accessions designated as OrIIIa (Fig. 2). Using a large dataset of genome-wide information for wild rice accessions with shallow depth data (Huang et al. 2012), it has been demonstrated that the japonica (or G) subgroup genome is close to a group of wild rice accessions designated as OrIII (Fig. 2). Thus, a strong bottleneck selection must have occurred to establish the japonica (or G) group. Although...
Fig. 4 Some trials for the reconsideration of QTNs to explain the rice domestication process. (A) Frame-shift mutations in RFT1 and Hd1, obtained from public 3K genomic information (Wang et al. 2018). (B) Cultivar numbers with listed QTNs in japonica subgroups. (C) Cultivar numbers of japonica with listed QTNs in local areas of Asia. (D) Model of the establishment of japonica subgroups. Selection for the qSH1 QTN (Konishi et al. 2006) might have occurred in the Lower Yangtze valley around 7,000 years ago (Zheng et al. 2016). QTN information; Wx b (Isshiki et al. 1998), Wx in (Zhang et al. 2019), Hd1 del (Yano et al. 2000), Hd1 intro (Itoh et al. 2018) and qSH1 (Konishi et al. 2006).

additional big data analysis of diverse rice cultivars (Wang et al. 2018) revealed further categorization within the japonica group, including subtropical japonica (Gj-sbtrp), tropical japonica (Gj-trp) and temperate japonica (Gj-tmp), the contribution of wild rice to this categorization remains unknown (Fig. 2). In contrast, indica and aus were integrated into a large category containing wild accessions, thus indicating close relationships among the Or-I group, indica and aus (Huang et al. 2012). The ability to distinguish DNA sequences in indica and aus from DNA sequences of wild rice accessions depends on the contribution of each gene during the domestication and subsequent breeding of rice.

With these data, either single or multiple origins are possible for cultivated Asian rice, O. sativa, largely depending on the definition of rice domestication. Several conflicting models to explain the origin of this domesticated crop have been developed thus far (Huang et al. 2012, Huang and Han 2015, Civán et al. 2016, Choi et al. 2017, Civán and Brown 2017, 2018, Choi and Purugganan 2018). Some QTGs (or QTNs) involved in critical key domestication traits such as seed shattering (due to sh4 QTN) and pericarp color (due to Rc QTN) are shared among major subgroups, strongly supporting a single origin (Li et al. 2006, Sweeney et al. 2006, 2007, Sweeney and McCouch 2007). This hypothesis is supported by demographic modeling using DNA polymorphism information (Choi et al. 2017). In contrast, the two-origin model, in which major subgroups such as indica and japonica have independent origins from distinct O. rufipogon subpopulations, explains the genetic differences (or distances) identified in the genome-wide comparison of japonica and indica (Vitte et al. 2004, Civán et al. 2016). The earliest archeological evidence of domesticated rice is from approximately 8,000–9,000 years ago, while the japonica and indica genomes are separated by a mean genetic distance of around 400,000 years (Vitte et al. 2004, Fuller et al. 2009, Choi et al. 2017). The distinct genome clustering of japonica, indica and O. rufipogon suggests a complex origin story for cultivated Asian rice. Even based on the same genome-wide data from diverse accessions, conflicting models of both single and multiple origins have been proposed (Civán and Brown 2017; Huang and Han 2015). In addition to the differing definitions of rice domestication, this controversy may arise from the limited accuracy of available genomic information.
for both cultivars and wild rice accessions (Huang et al. 2012). Accurate information based on diverse accessions is restricted to cultivars; such information is lacking for wild rice accessions. Huang and Han (2015) proposed a reasonable model that explains both the key QTN selection events described above and the large genetic distances among subgroups of rice cultivars, which included a single origin and multiple introgressions. However, important questions remain unanswered, such as the number of key QTNs and the number of O. rufipogon accession groups involved in rice domestication. Thus, Purugganan and colleagues proposed a similar domestication model that uses de novo–assembled genomic information of five representative subgroup cultivars (one japonica, two indica, two aus and two wild rice accessions) in collaboration with genome biologists and archeologists, although genome information from the small amount of rice lines may bias the model and reduce accuracy (Choi et al. 2017). However, it is of note that the portions of genomes that are introgressed into the target genome have been estimated using both phylogenetic and coalescence-based modeling of demographic parameters in this work. Accordingly, around 17% and 15% of the genome were estimated to have introgressed from japonica into indica and aus, respectively. Furthermore, the times of bottleneck selection for indica and aus formation were estimated separately. From an archeobotanical perspective, no strong evidence has been found for proto-indica domestication traits until around 4,000 years ago, which is much later than the timing of japonica domestication. Reliable archeological evidence in the Ganges valley dates from approximately 9,000 years ago (Choi et al. 2017). This archeological evidence has been interpreted to indicate the early management of proto-indica rice, with this management (early cultivation) preceding the selection of domestication alleles of key QTGs. This explanation is often given by archeologists; it may arise from the differing definitions of domestication between archeology and molecular biology. For example, because there are subtle phenotypic differences in seed shattering deriving from the QTN in sh₄, one of the first domestication genes, their effects may be difficult to identify using archeological methods (Li et al. 2006, Zhu et al. 2012). Thus, archeologically defined pre-domestication management may have involved some QTN selection (or the bottleneck events described above). The application of genome data from diverse accessions allows estimation using traces of bottleneck selection (such as genome sweeps) in the genome with associated selection of standing variations and new useful de novo mutations, although the effects of directional introgressions after repeated natural crossings should be considered together. Here, traces of bottleneck selection can be readily identified through population genomics analysis (Huang et al. 2012, Wang et al. 2018). After numerous selection events, the local cultivation scale of rice increased alongside human community formation, which might explain the excavation of ruins indicating rice cultivation. Thus, an explanation gap must have existed between proto-domestication and archeological domestication. Some gaps between molecular and archeological lines of evidence are logically acceptable when a causal relationship can be established. Based on the hypothesis that various introgression events occurred during rice domestication and the early rice breeding process because of repeated natural crossings and subsequent selection processes, the definition of domestication origins cannot be determined in a simple manner.

Even ignoring clear introgression cases such as Kala₄, representative sequences of each gene in a subgroup can be categorized into a single or a few group(s) containing a distinct haplotype group, mainly because of haplotype formation following bottleneck selection events. Using this basic information, several key directional introgression events can be considered; the story of rice domestication from an integrated genomic perspective will become more complex.

In addition to clear introgression events like Kala₄ and Rc, typical selection of de novo mutations in a subgroup must have also occurred during the rice domestication process. In an example from recent temperate japonica breeding efforts, both the original and selected QTNs of three heading-date QTGs—Hd₆, Hd₁₆ and Hd₁₇ (Takahashi et al. 2001, Matsubara et al. 2012, Hori et al. 2013)—have been identified only in Japanese temperate japonica cultivars tested, indicating that these QTNs represent de novo mutations within a subgroup (Fig. 1). Each selected QTN affects the heading date of rice cultivated in Japan. The presence of these three QTNs in local varieties differs among cultivars; it is partially associated with the preferred cultivation area in Japan (Fig. 1). A similar pattern can be found in some key domestication QTGs. qSH1, a seed-shattering gene, is a good example for explaining this situation (Konishi et al. 2006). The selected qSH1 QTNs were found in some Chinese landraces of rice, but only in the temperate japonica subgroup. This qSH1 mutation in combination with the defective sh₄ mutation (Li et al. 2006) can lead to complete loss of the abscission layer at the base of each rice floret, which is required for seed shattering. In this case, the qSH1 mutation was also a de novo mutation selected in Chinese landraces; it was transmitted to primitive Japanese landraces in the temperate japonica subgroup around a few thousand years ago. Apparently, the qSH1 QTN was selected in China after sh₄ QTN selection but before the occurrence and selection of Hd₆, Hd₁₆ and Hd₁₇ during recent rice breeding efforts in Japan (Fig. 4D).

Taken together, most of the genome of each rice subgroup was presumably fixed immediately after the first strong bottleneck selection event by early cultivators (or ancient humans) because of self-fertilization in the cultivated rice. However, many loci within each cultivar subgroup contain several alleles covering multiple haplotype groups, reflecting distinct evolutionary histories for each gene. This allelic diversity may be related to the size of the target population during the first bottleneck selection and subsequent local adaptation processes underling a subgroup (along with possible natural outcrossing effects), referred to as fixation and haplotype formation in Fig. 2. Thus, the presence of fixed alleles in a subgroup locus is an important factor for tracing the history of the subgroup, but
fixation of the locus is not necessarily evidence of selection for biological and agricultural traits.

**Perspective Shift from Genome-Wide Means to QTN Selection History in Rice Domestication Research**

To leverage the most genomic information for elucidating the rice domestication process, clarification of the QTN selection history in key QTGs would be a better approach than estimating the timing, location and number of events during crop domestication. One important reason for this proposal originates from several studies that describe new QTNs in previously identified genes. A defective allele with a 383-bp deletion in the Sa1 (**Semi dwarf** 1) gene, encoding a gibberellin biosynthesis enzyme, was selected to develop a series of semi-dwarf cultivars such as IR8 that are tolerant to lodging during the Green Revolution ([Asano et al. 2018]). This deletion originated from a Taiwanese cultivar named Dee-Geo-Woo-Gen. Recent work elucidating QTGs underlying deepwater rice traits revealed a new QTN in Sa1 that causes increased enzymatic activity with distinct substrate specificity; it is often found in wild rice accession lines ([Kuroha et al. 2018]). This finding suggests that selection for the haplotype including the QTN of Sa1 occurred during domestication because of a bottleneck effect or the selection of new mutations, causing compact stature prior to breeding selection during the Green Revolution. Thus, understanding the entire history of QTN selection for each key gene is a prerequisite; such genes may have undergone selection for a series of QTNs. Similar findings have been obtained for Wx (Waxy), a gene encoding a granule-bound starch synthase gene and conferring stickiness to cooked rice grains ([Zhang et al. 2019]) (Fig. 3). In this case, a few distinct alleles conferring moderate Wx enzymatic activity, which tends to be preferred by Asian people, were generated and selected repeatedly, resulting in at least six functionally distinct alleles of Wx in agricultural use. The identification of each QTN among various rice lines would be required to fully explain the rice domestication process (Fig. 4). Furthermore, Hd1 has also been extensively analyzed in terms of its polymorphisms among rice cultivars ([Fujino et al. 2010, Itoh et al. 2018, Wu et al. 2020]). In addition to various defective mutations, complex introgression patterns are characteristic of Hd1. Among the introgressions of Hd1, an allele from indica has introgressed into a substantial number of temperate japonica cultivars ([Itoh et al. 2018, Wang et al. 2021]) (Fig. 3). This is a very rare case of directional introgression from indica to temperate japonica, which may have occurred to support this local adaptation ([Choi et al. 2017]). In addition, as described above, population genomics data regarding a defective allele of Hd1 with a 2-bp deletion originally found in an aus cultivar ([Yano et al. 2000, Itoh et al. 2018]) across more than 3,000 cultivars of rice revealed that this mutation might have originated from tropical japonica ([Fig. 4D] [Wang et al. 2018]). In the assessments of defective mutations among heading-date genes, several such mutations of the RFT1 gene were found in cultivars of the aus subgroup. This observation can be explained genetically because the phenotype of the RFT1 gene is masked in hd1-defective mutants (Fig. 4A) ([Ogiso-Tanaka et al. 2013]).

Similar to previous trials ([Konishi et al. 2008]), QTN distributions for Wx, Hd1 and qSH1 were confirmed using new information concerning QTNs based on the 3K data (Fig. 4D). To define both fine subgroups of japonica and the local area where cultivars were bred, in this case, QTNs in the three genes were required. In this process, the presumed propagation of rice cultivation was largely northward based on the analysis of novel defective alleles (Fig. 1). This northward progression required substantial biological reduction of photoperiodic floral induction and floral repression in cold environments ([Izawa 2007, Cui et al. 2020, Fujino et al. 2022]). After these changes, cultivar performance often worsened when such cultivars were grown in southern areas, indicating that longitudinal change in the growth area required further biological selection even during early rice domestication. Therefore, when constructing a future model of rice domestication, such biological factors should be incorporated. Although popular methods such as phylogenetic tree construction from DNA changes in QTGs provide a good approach to highlight the selection of key QTNs, further developments are needed. A graphical method to integrate subgroup classification, gene evolution and QTN information could be possible for this purpose. Thus, the rice domestication process and early propagation of rice cultivation may be explained in the near future by obtaining abundant data for the QTN histories of key genes. Notably, the resulting models should be evaluated using unbiased DNA polymorphisms. Unbiased DNA polymorphisms have so far guided the construction of a model for rice domestication, but no good method is available for elucidating the hidden history of rice domestication because the biological aspect of target selection is often missing. The use of unbiased DNA polymorphism data alone to explain the domestication process is similar to the extensive usage of a cipher without decoding. Thus, a biological understanding of various QTN selection processes to explain rice domestication is urgently needed, although validations of models with unbiased information are ultimately required ([Zheng et al. 2022]).

Possible Causal Link between Archeological Evidence from Ruins and DNA Traces in a Domestication-Related QTG

Ideally, archeological data can be logically explained based on changes in QTNs, which would require comprehensive understanding of the phenotypic changes related to causal QTN selection. One such example can be found in the seed-shattering trait of rice and its changes during the domestication process in rice. Two QTGs associated with seed-shattering traits have
been identified during rice domestication and the early spread of rice cultivation (Konishi et al. 2006, Li et al. 2006). The QTN selection of sh4 results in panicles with seeds upon maturation, whereas wild rice accessions have panicles devoid of mature seeds because of seed shattering. With respect to plant growth, partial abscission layers may form in cultivars with the selected sh4 QTN, while complete abscission layers form in wild rice accession lines. In contrast, qSH1 selection results in complete loss of abscission layers. Thus, the detached surface at the base of the rice seed grain became rough after qSH1 selection. Because the abscission layer forms during floret formation, the abscission process begins immediately before floret flowering. Thus, phenotypic changes related to seed shattering were very clear for qSH1 selection but not for sh4. In a paper published by Zheng et al. (2016), reduced shattering of archeological rice from the Lower Yangtze valley was extensively examined through morphological and histological comparisons of spikelet bases from four sites with distinct dates using scanning electron microscope imaging; the results revealed that selection for non-shattering seeds proceeded continuously among these sites, confirming a prolonged selection period for the non-shattering trait around 8,000–5,000 years ago in the Lower Yangtze valley (Fig. 4D). The authors described a possible QTN selection process for qSH1. Thus, earlier selection for sh4, which may have caused the first bottleneck in rice domestication, occurred more than 8,000 years ago. Such links between archeological data and phenotypic changes driven by the selection of QTNs are not easily observed; extensive biological analysis may enable more such links to be identified in the future. Notably, some archeological papers missed these botanical changes in seed-shattering traits in rice due to a lack of biological consideration for such genetic effects (Fuller et al. 2009). Yet, while ancient DNA may be isolated from rice remains in the future, to date such samples have not provided any reliable conclusions related to plant domestication. Thus, the link connecting qSH1 QTN selection with non-shattering seed selection in temperate japonica in the Lower Yangtze valley is a critical step in achieving an integrated view of rice domestication (Fig. 4D).

**How Domestication Genes and Improvement (or Diversification) Genes Can Be Distinguished in Rice Domestication**

As mentioned, both sh4 and qSH1 are related to seed shattering in rice, which has been considered as a typical domestication trait in many crops. However, while sh4 can be termed a domestication gene, qSH1 may be considered an improvement (or diversification) gene, i.e., a gene associated with post-domestication improvement or diversification. Thus, it may be not a good idea to consider specific traits, such as seed shattering, as criteria by which to distinguish them. In addition, like Sd1, Wx and Hd1, several key QTNs in the one gene are capable of being selected for several times during the domestication and improvement processes in rice; such a scenario complicates the distinction between domestication and improvement genes. Only after the entire mapping of QTNs among various rice accessions is achieved, will such distinctions become clearer. It is also worth noting that the selection for qSH1 might have occurred around 7,000 years ago. This was a rather old selection event when considering the entire history of rice domestication and breeding.

**A Message from the Past for the Future of Agriculture**

Using genome editing technology, desired mutations can be efficiently incorporated into crops (Rodríguez-Leal et al. 2017, Kwon et al. 2020), including possibly for orphan crops (Lemmon et al. 2018). In this context, information regarding crop domestication will be essential to support future breeding. An approach using such information was recently reported to describe the current status of the development of allotetraploid rice cultivars (Yu et al. 2021). During crop domestication, changes in cis-regulatory elements could lead to major QTNs (Lemmon et al. 2014). It is well known that the QTN for qSH1 selection occurred at a nucleotide in a cis-regulatory element around 12 kb upstream of the region encoding the qSH1 gene product. With the abundance of genomic information available from diverse cultivars, a novel method for identifying the cis-regulatory elements that underlie phenotypic changes from diverse DNA polymorphisms in cis-regulatory regions of QTGs is now needed.

**Data Availability**

No new datasets were generated in this study. Source data for figures are provided in the paper, while the data summary was easily obtained using publicly available data.

**Funding**

JSPS KAKENHI (grant numbers JP18H03948 and JP17H06246); the Human Frontier Science Program Organization (grant number RGP0011/2019); Cabinet Office, Government of Japan, Cross-ministerial Moonshot Agriculture, Forestry and Fisheries Research and Development Program, ‘Technologies for Smart Bio-industry and Agriculture’ (funding agency: Bio-oriented Technology Research Advancement Institution).

**Disclosures**

The authors have no conflicts of interest to declare.

**References**

Alam, O., Gutaker, R.M., Wu, C.C., Hicks, K.A., Bocinsky, K., Castillo, C.C., et al. (2021) Genome analysis traces regional dispersal of rice in Taiwan and Southeast Asia. *Mol. Biol. Evol.* 38: 4832–4846.

Asano, K., Yamasaki, M., Takuno, S., Miura, K., Katagiri, S., Ito, T., et al. (2011) Artificial selection for a green revolution gene during japonica rice domestication. *Proc. Natl. Acad. Sci. U.S.A.* 108: 11034–11039.
Ishikawa, R., Nishimura, A., Htun, T.M., Nishioka, R., Oka, Y., Tsujimura, Y., Huang, X., Kurata, N., Wei, X., Wang, Z.X., Wang, A., Zhao, Q., et al. 2015) Rice domestication occurred through single Hori, K., Nonoue, Y., Ono, N., Shibaya, T., Ebana, K., Matsubara, K., et al. 2018) Japonica rice carried to, not from, Southeast Fuller, D.Q., Qin, L., Zheng, Y., Zhao, Z., Chen, X., Hosoya, L.A., et al. (2009) The domestication process and domestication rate in rice: spikelet bases from the Lower Yangtze. Science 323: 1607–1610. Fuller, D.Q. and Sato, Y. (2008) Rapid improvement of domestication traits in an orphan crop by genome editing. Nat. Plants 4: 766–770. Li, C., Zhou, A. and Sang, T. (2006) Rice domestication by reducing shattering. Science 311: 1936–1939. Li, J., Zeng, Y., Pan, Y., Zhou, L., Zhang, Z., Guo, H., et al. (2021) Stepwise selection of natural variations at CTB2 and CTB4a improves cold adaptation during domestication of japonica rice. New Phytol. 231: 1056–1072. Liu, C., Ou, S., Mao, B., Tang, J., Wang, W., Wang, H., et al. (2018a) Early selection of bZIP73 facilitated adaptation of japonica rice to cold climates. Nat. Commun. 9: 3302. Liu, H., Li, Q. and Xing, Y. (2018b) Genes contributing to domestication of rice seed traits and its global expansion. Genes (Basel) 9: 489. Liu, J., Chen, J., Zheng, X., Wu, F., Lin, Q., Heng, Y., et al. (2017) GWS5 acts in the brassinosteroid signalling pathway to regulate grain width and weight in rice. Nat. Plants 3: 17043. Lu, L., Yan, W., Xue, W., Shao, D. and Xing, Y. (2012) Evolution and association analysis of Ghd7 in rice. PLoS One 7: e40201. Matsubara, K., Ogiso-Tanaka, E., Horii, K., Ebana, K., Ando, T. and Yano, M. (2012) Natural variation in Hdi7, a homolog of Arabidopsis ELF3 that is involved in rice photoperiodic flowering. Plant Cell Physiol. 53: 709–716. Ogiso-Tanaka, E., Matsubara, K., Yamamoto, S., Nonoue, Y., Wu, J., Fujisawa, H., et al. (2013) Natural variation of the RICE FLOWERING LOCUS T 1 contributes to flowering time divergence in rice. PLoS One 8: e75959. Oikawa, T., Maeda, H., Oguchi, T., Yamaguchi, T., Tanabe, N., Ebana, K., et al. (2015) Melanins are involved in the origin of cultivated rice. Nature 490: 497–501. Ishikawa, R., Nishimura, A., Htun, TM., Nishioka, R., Oka, Y., Tsujimura, Y., et al. (2016) The characterization of Hdi17, a homolog of Arabidopsis ELF3 that is involved in rice photoperiodic flowering. Plant Cell Physiol. 53: 709–716. Ogiso-Tanaka, E., Matsubara, K., Yamamoto, S., Nonoue, Y., Wu, J., Fujisawa, H., et al. (2013) Natural variation of the RICE FLOWERING LOCUS T 1 contributes to flowering time divergence in rice. PLoS One 8: e75959. Oikawa, T., Maeda, H., Oguchi, T., Yamaguchi, T., Tanabe, N., Ebana, K., et al. (2015) The birth of a black rice gene and its local spread by introgression. Plant Cell 27: 2401–2414. Qin, P., Lu, H., Du, H., Wang, H., Chen, W., Chen, Z., et al. (2021) Population genomics analysis of 33 genetically diverse rice accessions reveals hidden genomic variations. Cell 178: 3542–3558.e16. Rodríguez-Leal, D., Lemmon, Z.H., Man, J., Bartlett, M.E. and Lippman, Z.B. (2017) Engineering quantitative trait variation for crop improvement by genome editing. Cell 171: 470–480.e8. Shao, G., Tang, S., Chen, M., Wei, X., He, J., Luo, J., et al. (2013) Haplotype variation at Badh2, the gene determining fragrance in rice. Genomics 101: 157–162. Shomura, A., Izawa, T., Ebana, K., Ebitani, T., Kanegae, H., Konishi, S., et al. (2008) Deletion in a gene associated with grain size increased yields during rice domestication. Nat. Genet. 40: 1023–1028.
Spengler, R.N., 3rd, Stark, S., Zhou, X., Fuks, D., Tang, L., Mir-Makhamad, B., et al. (2021) A journey to the west: the ancient dispersal of rice out of East Asia. *Rice (N Y)* 14: 83.
Sweeney, M. and McCouch, S. (2007) The complex history of the domestication of rice. *Ann. Bot.* 100: 951–957.
Sweeney, M.T., Thomson, M.J., Cho, Y.G., Park, YJ, Williamson, S.H., Bustamante, C.D., et al. (2007) Global dissemination of a single mutation conferring white pericarp in rice. *PloS Genet.* 3: e133.
Sweeney, M.T., Thomson, M.I., Pfeil, B.E. and McCouch, S. (2006) Caught red-handed: *Rc* encodes a basic helix-loop-helix protein conditioning red pericarp in rice. *Plant Cell* 18: 283–294.
Takahashi, Y., Shomura, A., Sasaki, T. and Yano, M. (2001) *Hd6*, a rice quantitative trait locus involved in photoperiod sensitivity, encodes the alpha subunit of protein kinase CK2. *Proc. Natl. Acad. Sci. U.S.A.* 98: 7922–7927.
Vitte, C., Ishii, T., Lamy, F., Brar, D. and Panaud, O. (2004) Genomic paleontology provides evidence for two distinct origins of Asian rice (*Oryza sativa* L.). *Mol. Genet. Genomics* 272: 504–511.
Wang, M., Chen, J., Zhou, F., Yuan, J., Chen, L., Wu, R., et al. (2021) The ties of brotherhood between japonica and indica rice for regional adaptation. *Sci. China Life Sci.* Online ahead of print.
Wang, W., Gao, H., Liang, Y., Li, J. and Wang, Y. (2022) Molecular basis underlying rice tiller angle: current progress and future perspectives. *Mol. Plant* 15: 125–137.
Wang, W., Mauleon, R., Hu, Z., Chebotarav, D., Tai, S., Wu, Z., et al. (2018) Genomic variation in 3,010 diverse accessions of Asian cultivated rice. *Nature* 557: 43–49.
Wei, F.J., Tsai, Y.C., Wu, H.P., Huang, L.T., Chen, Y.C., Chen, Y.F., et al. (2016) Both *Hd1* and *Ehd1* are important for artificial selection of flowering time in cultivated rice. *Plant Sci.* 242: 187–194.
Wei, X., Qiu, J., Yong, K., Fan, J., Zhang, Q., Hua, H., et al. (2021) A quantitative genomics map of rice provides genetic insights and guides breeding. *Nat. Genet.* 53: 243–253.
Wu, C.C., Wei, F.J., Chiou, W.Y., Tsai, Y.C., Wu, H.P., Gotarkar, D., et al. (2020) Studies of rice *Hd1* haplotypes worldwide reveal adaptation of flowering time to different environments. *PloS One* 15: e0239028.
Wu, Y., Zhao, S., Li, X., Zhang, B., Jiang, L., Tang, Y., et al. (2018) Deletions linked to PROG1 gene participate in plant architecture domestication in Asian and African rice. *Nat. Commun.* 9: 4157.
Xue, W., Xing, Y., Weng, X., Zhao, Y., Tang, W., Wang, L., et al. (2008) Natural variation in *Ghd7* is an important regulator of heading date and yield potential in rice. *Nat. Genet.* 40: 761–767.
Yano, M., Katayose, Y., Ashikari, M., Yamanouchi, U., Monna, L., Fuse, T., et al. (2000) *Hd1*, a major photoperiod sensitivity quantitative trait locus in rice, is closely related to the *Arabidopsis* flowering time gene *CONSTANS*. *Plant Cell* 12: 2473–2484.
Yu, H., Lin, T., Meng, X., Du, H., Zhang, J., Liu, G., et al. (2021) A route to de novo domestication of wild allotetraploid rice. *Cell* 184: 1156–1170.e14.
Zhang, C., Zhu, J., Chen, S., Fan, X., Li, Q., Lu, Y., et al. (2019) *Wx*(lv), the ancestral allele of rice waxy gene. *Mol. Plant* 12: 1157–1166.
Zhang, L.B., Zhu, Q., Wu, Z.Q., Ross-Ibarra, J., Brandon, S., Gaut, B.S., et al. (2009) Selection on grain size genes by RapMap reveals directional selection during rice domestication. *Nat. Commun.* 12: 5673.
Zheng, X., Pang, H., Wang, J., Yao, X., Song, Y., Li, F., et al. (2022) Genomic signatures of domestication and adaptation during geographical expansions of rice cultivation. *Plant Biotechnol. J.* 20: 16–18.
Zheng, Y., Crawford, G.W., Jiang, L. and Chen, X. (2016) Rice domestication revealed by reduced shattering of archaeological rice from the Lower Yangtze valley. *Sci. Rep.* 6: 28136.
Zhu, Y., Ellstrand, N.C. and Lu, B.R. (2012) Sequence polymorphisms in wild, weedy, and cultivated rice suggest seed-shattering locus sh4 played a minor role in Asian rice domestication. *Ecol. Evol.* 2: 2106–2113.