Copyright © 2002 by the Genetics Society of America

Molecular Evidence on the Origin and Evolution of Glutinous Rice

Kenneth M. Olsen[1] and Michael D. Purugganan

Department of Genetics, North Carolina State University, Raleigh, North Carolina 27695

Manuscript received June 19, 2002
Accepted for publication July 26, 2002

ABSTRACT

Glutinous rice is a major type of cultivated rice with long-standing cultural importance in Asia. A mutation in an intron 1 splice donor site of the *Waxy* gene is responsible for the change in endosperm starch leading to the glutinous phenotype. Here we examine an allele genealogy of the *Waxy* locus to trace the evolutionary and geographical origins of this phenotype. On the basis of 105 glutinous and nonglutinous landraces from across Asia, we find evidence that the splice donor mutation has a single evolutionary origin and that it probably arose in Southeast Asia. Nucleotide diversity measures indicate that the origin of glutinous rice is associated with reduced genetic variation characteristic of selection at the *Waxy* locus; comparison with an unlinked locus, *RGRC2*, confirms that this pattern is specific to *Waxy*. In addition, we find that many nonglutinous varieties in Northeast Asia also carry the splice donor site mutation, suggesting that partial suppression of this mutation may have played an important role in the development of Northeast Asian nonglutinous rice. This study demonstrates the utility of phylogeographic approaches for understanding trait diversification in crops, and it contributes to growing evidence on the importance of modifier loci in the evolution of domestication traits.

THE study of crop origins can provide unique insights into the evolution of morphological and developmental adaptations favored by early farming cultures. Phylogeographic analysis of allele genealogies has proved useful for examining crop origins (Olsen and Schaal 1999; Sanjur *et al.* 2002). When applied to crop systems, this approach typically uses noncoding, selectively neutral genetic variation to draw inferences about the history of population divergence and crop-progenitor relationships. A potential extension of the phylogeographic method is to examine genes directly responsible for the phenotypic variation within a crop species. This candidate-locus approach may provide a means of inferring the origin and dispersal of specific traits that have been favored over the course of domestication.

Glutinous rice exemplifies the capacity for plants to evolve phenotypic modifications in response to local cultural preferences. An important culinary and cultural component throughout East Asia, glutinous rice is generally reserved for use in festival foods and desserts, although it also serves as the staple food in upland regions of Southeast Asia (Golomb 1976; Roder *et al.* 1996). The evolutionary and geographical origins of glutinous rice have remained obscure. Because the glutinous phenotype is not detectable in the archeological record, the region of its earliest cultivation has not been documented. Efforts to trace its origins are further complicated by its long-standing cultural importance throughout a very wide geographical area in East Asia, which includes portions of China, Japan, Korea, and the countries of Southeast Asia. Laotian Buddhist legend places the origin of glutinous rice to ~1100 years ago (Terwiel 1994), although Chinese folklore indicates that it was in existence around the time of the death of the poet Qu Yuan >2000 years ago (Xu 1992).

Glutinous rice lacks the starch amylose, which constitutes up to 30% of the total starch in nonglutinous rice endosperm (Oka 1988; Morishima *et al.* 1992). The glutinous phenotype arises through the disrupted expression of the amylose biosynthesis gene *Waxy* (*Wx*), which encodes a granule-bound starch synthase (Sano 1984). Glutinous rice contains a G to T mutation at the 5′ splice site of *Wx* intron 1 that leads to incomplete post-transcriptional processing of *Wx* pre-mRNA (Wang *et al.* 1995; Bligh *et al.* 1998; Cai *et al.* 1998; Hirano *et al.* 1998; Isshiki *et al.* 1998). Glutinous rices do not have detectable levels of spliced mRNA as a result of this mutation (Wang *et al.* 1995; Bligh *et al.* 1998). In contrast to glutinous rice, nonglutinous (common) rices show wide variation in amylose content (*e.g.*, Ayres *et al.* 1997); this quantitative variation is affected by multiple loci. Higher amylose levels (~20–30%) are associated with many South Asian varieties that form discrete, noncohesive grains when cooked and are also found in the crop's wild progenitor. Lower amylose levels (~10–20%) are more common in East Asia, where a more cohesive cooked grain is often preferred.

While apparently required for the glutinous phenotype, the presence of the *Wx* intron 1 splice donor site

Sequence data from this article have been deposited with the EMBL/GenBank data libraries under accession nos. AY136760–AY136784.

[1]*Corresponding author:* Department of Genetics, Box 7614, North Carolina State University, Raleigh, NC 27695-7614.
E-mail: kmolsen@unity.ncsu.edu

mutation does not ensure that this phenotype is expressed. Some degree of amylose synthesis is restored in varieties that carry the mutation but that display cryptic splice site activation, which results in alternative *Wx* premRNA splicing patterns (Bligh *et al.* 1998; Cai *et al.* 1998; Isshiki *et al.* 1998). Thus, the *Wx* intron 1 splice donor site mutation is associated with all glutinous varieties as well as with some low-amylose nonglutinous varieties (Wang *et al.* 1995; Bligh *et al.* 1998; Cai *et al.* 1998).

Two alternative scenarios may be envisioned for the origin of the glutinous phenotype, depending in part on the number of times the *Wx* splice donor site mutation has arisen during domestication. One possibility is that the splice donor site mutation occurred a single time, with all glutinous varieties derived from this single mutational event. Under this scenario, glutinous rice would have a single geographical origin, and the current distribution of glutinous varieties would reflect dispersal from this center of origin. Alternatively, the glutinous phenotype could have arisen more than once across East Asia, either through independent mutations at the *Wx* intron 1 splice donor site or by other mechanisms involving *Wx* and/or other genes. The evolutionary origin of a major trait polymorphism has not previously been examined in any crop or other plant species, and there is no clear *a priori* evidence favoring either of these alternative scenarios.

In this article we describe a population genetic and phylogeographic analysis using the *Wx* gene to investigate the origin of the glutinous rice phenotype. We have analyzed DNA sequence variation in an ~2.7-kb portion of the *Wx* locus surrounding the intron 1 splice donor site mutation. Three questions are addressed:

i. Do patterns of nucleotide diversity at the *Wx* locus show evidence of selection associated with the rise and spread of the glutinous phenotype?
ii. How many times over the course of rice domestication has the splice donor site mutation arisen and been selected for?
iii. What can be inferred about the geographical origin(s) of glutinous rice?

This study traces the evolutionary origin of a specific mutation that is directly responsible for a major trait polymorphism maintained within a crop species.

MATERIALS AND METHODS

Sampling and classification: Seed accessions from 105 varieties of *Oryza sativa* were sampled from the germplasm collection maintained by the International Rice Research Institute (IRRI, Los Baños, Philippines). Accessions were selected to maximize the geographical sampling of nonglutinous and glutinous traditional landraces across Asia and to represent both major "variety groups" traditionally recognized in Asian rice (indica and japonica; Kato *et al.* 1928; Table 1; appendix). In general, indica rices predominate in South Asia, and japonica varieties are more common in East Asia. Japonica varieties are sometimes further classified as temperate and tropical ("javanica") varieties (Khush 1997). Variety group designations in this study follow the IRRI germplasm database classification (see the appendix). In addition to *O. sativa* samples, one accession of a closely related species, *O. meridionalis*, was sampled for use in outgroup comparisons of DNA sequences.

Rice accessions were classified as either glutinous or nonglutinous on the basis of a colorimetric assay for amylose content. Five to 10 seeds were cut crosswise and stained in an ~1% iodine solution for 30 sec and then destained in dH_2O for 2–5 min and observed under a dissecting microscope. A bluish-black color indicates the presence of amylose. Accessions that differed from the IRRI database endosperm classification were retested to confirm the result (see the appendix).

PCR and DNA sequencing: Genomic DNA was extracted from a single 14-day seedling per accession using DNeasy plant mini kits (QIAGEN, Valencia, CA). An ~2.7-kb portion of the *Wx* gene was PCR amplified and sequenced (see also Hirano *et al.* 1998). This region of the gene surrounds the previously identified intron 1 splice donor site mutation and includes ~1.46 kb of promoter sequence, all of exon 1 and intron 1, and the 5′ end of exon 2; the region is entirely noncoding, as the start codon is within exon 2. The entire *Wx* transcriptional unit is ~5.4 kb.

PCR amplification was performed in two stages to amplify overlapping ~1.5-kb regions of the gene. The 5′ region was amplified with the primers WxU1F (5′-GCCGAGGGACCT AATCTGC-3′) and Wx1R (5′-TGGTGTGGGTGGCTATTTG TAG-3′); the 3′ region was amplified with primers Wx2Fa (5′-GCCCCGCATGTCATCGTC-3′) and Wx2R (5′-GTTGTCT AGCTGTTGCTGTGGA-3′). For each primer set, two 50-μl reactions were carried out per individual. Each reaction contained 1× PCR buffer (Promega, Madison, WI), 1.0 mm $MgCl_2$, 2 units *Taq* polymerase (Promega), 200 μm each dNTP, 0.4 μm each forward and reverse primer, and ~10–20 ng genomic DNA. Cycling conditions were 94° (2 min); then 30 cycles of 94° (30 sec), 55° (30 sec), 72° (2 min); and a final extension of 72° (5 min). The two PCR products were pooled and purified with QIAquick gel extraction kits (QIAGEN) and then directly sequenced using BigDye terminator reactions (Applied Biosystems, Foster City, CA) that were run on an ABI 3700 automated sequencer; reactions were run for each end primer plus one internal primer per amplicon: Wx1Fint (5′-TTGTCAGCACGTACAAGCA-3′) and Wx2Rint (5′-GCTA TATACATTTTCCTTTGACCAA-3′).

An ~1.0-kb portion of the *RGRC2* gene was also sequenced in all 105 rice accessions. This gene encodes cytosolic glutathione reductase, a key enzyme of all aerobic organisms that functions in oxidative stress response (Kaminaka *et al.* 1998). Whereas the *Wx* gene is located on chromosome 6, *RGRC2* is found on chromosome 5. The sequenced portion of *RGRC2* spans three introns and two small exons; most of this region is noncoding. Primers for PCR were GRC12Fb (5′-TGTTGC ACTGATGGAGGCTA-3′) and GRC15Rb (5′-TTTCATGACG GTCTTCTCCTG-3′). PCR and purification conditions were the same as for *Wx,* and direct sequencing was performed using end primers plus the internal primer GRC14R (5′-GTT CACTCAAGCCCACTACTGA-3′).

DNA sequences were visually aligned, and all polymorphisms were rechecked from chromatograms or by resequencing, with special attention to low-frequency polymorphisms. *O. sativa* is a predominantly self-fertilizing species, and no heterozygosity was observed in the genes sequenced. Sequences have been deposited in GenBank (accession nos. AY136760–AY136784).

Genetic diversity analysis: DNA sequences were analyzed in DnaSP 3.5 (Rozas and Rozas 1999). For both *Wx* and *RGRC2*, the average pairwise nucleotide diversity, π (Tajima 1983), and Watterson's (1975) estimator of θ (θ_W) were calculated

for glutinous accessions, nonglutinous accessions, and all accessions combined. Measures for *RGRC2* were based on silent polymorphisms only. Tajima's (1989) *D* statistic was used to compare these two diversity measures, which are expected to be equal under selective neutrality. Tajima's *D* statistic is sensitive to violations of the assumption that populations are at a mutation-drift equilibrium, including population bottlenecks that might arise during crop domestication; however, any such demographic effects should be equally observable at both loci, whereas the effects of selection should be locus specific. Statistical significance for Tajima's test of selection was determined by coalescent simulation (10,000 runs) under the assumption of no recombination and taking into account the number of segregating sites for each gene. The assumption of no recombination is appropriate for selfing species such as *O. sativa* and provides the most conservative criterion for testing the statistical significance of *D*.

A *Wx* haplotype tree (allele genealogy) was constructed from substitution polymorphisms, using a maximum-parsimony criterion (branch and bound search, stepwise addition) in PAUP* (Swofford 2000). The tree was rooted by the outgroup method, using the sequence from the closely related species *O. meridionalis*.

RESULTS AND DISCUSSION

Nucleotide variation and selection at the *Wx* gene in rice: We examined genetic variation in an ~2.7-kb portion of *Wx* in 37 glutinous and 68 nonglutinous rice accessions that are representative of landrace diversity across Asia (Table 1; appendix). Twenty-eight nucleotide substitution polymorphisms and two insertion/deletion polymorphisms were observed in the promoter region. Twenty-seven substitution polymorphisms and five insertion/deletion polymorphisms were observed in intron 1. A variable simple sequence repeat (SSR) region composed of 8–20 CT dinucleotide repeats was the only molecular variation observed in the 5′ untranslated exon 1 (Figure 1). The 55 substitution polymorphisms define a total of 18 different haplotypes (Figure 1).

Glutinous and nonglutinous accessions show a marked difference in levels of genetic variation at the *Wx* locus (Table 2). Average pairwise nucleotide diversity, π, is more than an order of magnitude greater in nonglutinous than in glutinous rice accessions. Similarly, θ_W is more than three times greater for the nonglutinous accessions. In contrast, no reduction in genetic diversity from nonglutinous to glutinous accessions is observed at the unlinked gene *RGRC2* (Table 2). These patterns suggest a reduction in genetic diversity specific to the *Wx* locus that is associated with the rise of glutinous rice from nonglutinous progenitors.

Further analysis of nucleotide diversity indicates that cultural selection for the glutinous phenotype is associated with selection at the *Wx* locus (Table 2). Tajima's (1989) test for selection indicates that glutinous rice accessions show a significant negative deviation from neutral expectations ($D = -2.336$, $P < 0.0001$), whereas nonglutinous varieties do not ($D = +0.4410$, $P > 0.10$). The significant negative value of Tajima's *D* in glutinous varieties arises from an excess of low-frequency nucleotide polymorphisms at this locus. This pattern is consistent with a recent selective sweep at the *Wx* gene associated with the origin and spread of the glutinous phenotype. In contrast, the unlinked gene *RGRC2* conforms to neutral expectations for both starch classes (Table 2). This difference between the two genes indicates that the observed deviation from neutrality at *Wx* is not due to changes in population size associated with the origin of glutinous varieties; demographic effects would be expected to be detectable at both loci. Similar evidence for domestication-related selection has also been reported for maize (*e.g.*, Wang *et al.* 1999) and *Brassica oleracea* (Purugganan *et al.* 2000).

TABLE 1

Summary of sampled rice accessions

	Nonglutinous	Glutinous
Northeast Asia		
China	10 (2)	3
Japan	8 (7)	2
Korea	6 (6)	4
Taiwan	0	3
Southeast Asia		
Brunei	0	2
Cambodia	1	2
Indonesia	4	2
Laos	0	4
Malaysia	6 (1)	2
Myanmar	0	2
Philippines	5	3
Thailand	3	3
Vietnam	4	3
South Asia		
Bangladesh	3	1
Bhutan	2	0
India	6	0
Nepal	3	1
Pakistan	2 (1)	0
Sri Lanka	5	0
Total	68	37

All glutinous accessions carry the *Wx* splice donor site mutation. Counts of nonglutinous accessions that also carry the mutation are shown in parentheses.

Origin of the *Wx* splice donor site mutation: A rooted haplotype genealogy was constructed from the substitution polymorphisms observed at *Wx* (Figure 2). Because the nucleotide sequence of one haplotype indicates that it is likely an intragenic recombinant (haplotype R; see below), this haplotype was not mapped onto the final tree. Two equally parsimonious arrangements were found for the remaining 17 haplotypes (consistency index = 0.97); the trees differ in the placement of a single homoplasious polymorphism (nucleotide position 1846; Figure 1), which minimally affects the tree topology (Figure 2).

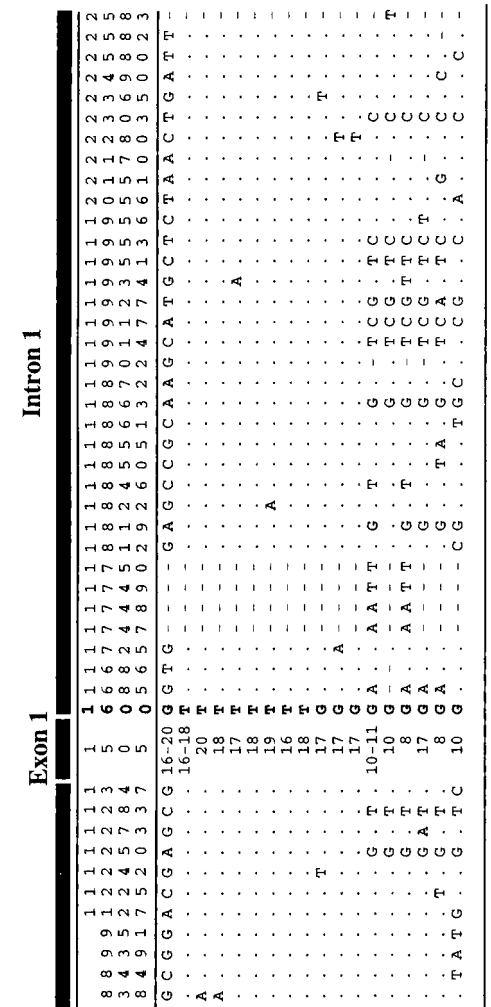

FIGURE 1.—Nucleotide polymorphisms at the *Wx* locus. Numbers in parentheses indicate the numbers of accessions observed per haplotype. The intron 1 splice donor site mutation is shown in boldface type. Numbers at nucleotide position 1505 indicate CT repeats within exon 1.

The *Wx* haplotype tree can reveal the number of times the splice donor site mutation has occurred, and it can therefore be used to distinguish between two alternative scenarios for the origin(s) of the glutinous phenotype: (i) a single-origin hypothesis with subsequent spread of glutinous varieties across Asia or (ii) a multiple-origins hypothesis across East Asia in response to local cultural selection. The splice donor site mutation maps to a single point on the *Wx* haplotype tree (Figure 2). Glutinous accessions are restricted to a cluster of five closely related haplotypes, with the splice donor site mutation directly leading to these haplotypes (Figure 2). Thus, the mutation responsible for the disruption of amylose synthesis appears to have a single origin in the glutinous landraces examined. Since all glutinous accessions contain the splice donor site mutation, this finding strongly supports the single-origin hypothesis for the glutinous phenotype.

In addition to the mapped haplotypes, the putative recombinant haplotype (R) also carries the splice donor site mutation. This haplotype is identical to the most common glutinous haplotype (G), except in the 5′ end of the promoter, where it is identical to distantly related haplotypes (haplotypes A and B; see Figures 1 and 2). This pattern suggests that haplotype R has arisen by intragenic recombination between two of the mapped haplotypes. A maximum-likelihood analysis supports the occurrence and boundary of this apparent recombination event (HOLMES *et al.* 1999). Thus, haplotype R most likely arose following the origin of the splice donor site mutation and is therefore very unlikely to represent an independent origin of the mutation.

Phylogeography of the *Waxy* gene: Because the *Wx* haplotype tree is rooted, the progenitor haplotype of the intron 1 splice donor site mutation can be unambiguously identified (haplotype F; Figure 2). Most of the glutinous accessions carry a haplotype that differs from this progenitor solely by the presence of the splice donor site mutation (haplotype G, $n = 31$). Five glutinous accessions carry additional mutations that arose subsequent to the splice donor site mutation (haplotypes I, K, L, and Q; Figure 2).

Since glutinous rice haplotypes are apparently all derived from the progenitor haplotype (F), the geographical distribution of this nonglutinous progenitor should reveal the geographical origin of glutinous rice. Nine of the 12 accessions carrying haplotype F come from Southeast Asia, while the remaining 3 are from South Asia. This distribution is significantly different from that expected by chance (Table 3), and it is the preponderance of Southeast Asian haplotypes that accounts for the significant result (Fisher's exact test, $P < 0.05$ for Southeast Asia *vs.* other regions; not significant for the other two pairwise comparisons). Within Southeast Asia, fully 39% of accessions lacking the splice donor site mutation possess the progenitor haplotype F. In contrast, only 15% of the South Asian and 0% of the North-

TABLE 2

Genetic diversity in glutinous and nonglutinous rice accessions

		π		θ_W		Tajima's *D*	
	n	*Waxy*	*RGRC2*	*Waxy*	*RGRC2*	*Waxy*	*RGRC2*
Nonglutinous	68	0.0045 (0.0002)	0.0024 (0.0002)	0.0040 (0.0011)	0.0021 (0.0008)	0.4410 (NS)	0.4301 (NS)
Glutinous	37	0.0003 (0.0002)	0.0025 (0.0002)	0.0012 (0.0005)	0.0019 (0.0007)	−2.336*	0.9916 (NS)
All	105	0.0039 (0.0003)	0.0025 (0.0001)	0.0039 (0.0010)	0.0019 (0.0007)	0.0275 (NS)	0.6946 (NS)

Standard deviations are shown in parentheses; values for θ were calculated using the conservative assumption of no recombination.

*$P < 0.0001$; NS, not significant; $P > 0.10$.

east Asian accessions carry this haplotype (Table 3). Thus, the progenitor haplotype is found in highest frequency in Southeast Asia. This finding strongly suggests that the splice donor site mutation initially arose in Southeast Asia and was subsequently spread northward across East Asia.

Our finding of a Southeast Asian origin for glutinous rice is consistent with Asian cultural practices. While consumed throughout eastern Asia, only in Southeast Asia does glutinous rice attain the importance of a staple crop (Golomb 1976; Roder *et al.* 1996; Rao *et al.* 2002). Thus, for this trait the area of major cultivation coincides with the geographical origin of domestication. Ethnographic studies suggest that glutinous rice cultivation is associated with upland agriculture in mainland Southeast Asia, particularly among the Tai-speaking peoples who migrated to the region between 1100 and 1500 years ago (Golomb 1976; Roder *et al.* 1996). Indeed, ~200,000 square miles of mainland Southeast Asia (encompassing portions of Laos, Myanmar, Thailand, Cambodia, and Vietnam) are referred to as the "glutinous rice zone," reflecting the importance of this type of rice to the economy and culture of the area.

The dispersal and proliferation of glutinous landraces from Southeast Asia has left a clear molecular signature at the *Wx* locus, including reductions in nucleotide diversity as well as an excess of low-frequency polymorphisms (Table 2). Both of these effects are specific to the *Wx* locus and are consistent with a selective sweep associated with the origin and spread of the glutinous phenotype. Because the *Wx* sequences do not conform to molecular clock assumptions (likelihood-ratio test = 56.34, d.f. = 16, $P < 0.001$; Posada and Crandall 1998), they are not suitable for calculating a specific date for the origin of glutinous rice. However, the fact that several mutations have occurred following the splice donor site mutation (Figure 2) suggests that the glutinous phenotype probably did not arise in the very recent past.

Variety groups: The accessions included in this study represent both of the major variety groups (indica and japonica) traditionally recognized within Asian rice (ap-

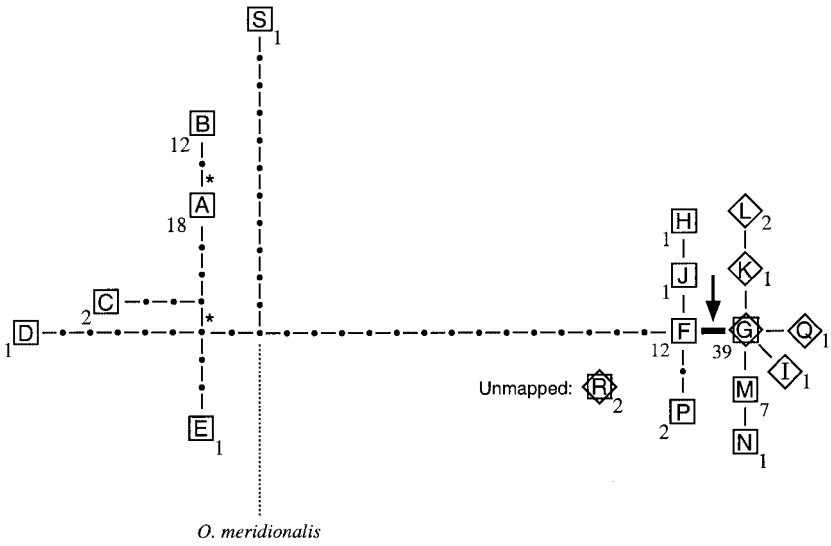

Figure 2.—The *Wx* haplotype tree. Letters correspond to haplotype designations in Figure 1. Numbers indicate accession counts per haplotype. Short lines represent mutational changes corresponding to substitution polymorphisms, and solid circles indicate inferred intermediate haplotypes. The intron 1 splice donor site mutation is indicated in boldface type with an arrow. Squares indicate haplotypes found in nonglutinous accessions, and diamonds indicate haplotypes found in glutinous accessions. The tree is one of two equally parsimonious arrangements that differ solely in the placement of the homoplasious polymorphism indicated with an asterisk.

TABLE 3

Regional distributions of nonglutinous haplotypes lacking the intron 1 splice donor site mutation

Region	Progenitor haplotype (F)	Nonprogenitor haplotypes
Northeast Asia	0	9
Southeast Asia	9	14
South Asia	3	17

Chi square = 6.774, $P < 0.05$.

pendix). The intron 1 splice donor site mutation has previously been proposed to have accompanied the origin and spread of the japonica variety group (Hirano *et al.* 1998). Our data tentatively support this hypothesis. Within our samples, 24 of 25 temperate japonica varieties carry the mutation. Eleven of 15 tropical japonicas (javanicas) also carry the mutation, and the 4 others all carry the progenitor haplotype F (appendix). This pattern suggests that the mutation could have arisen within tropical japonicas and then subsequently spread northward as the temperate japonicas came to be distributed across Northeast Asia. Temperate japonica varieties are generally believed to have originated in Southeast Asia or South China with subsequent dispersal northward (Khush 1997), which is consistent with this scenario.

A number of indica varieties, however, also possess the splice donor site mutation. In our sample, these represent 17 of the 62 varieties classified as indica (27%). These may represent indicas into which the mutation has been introgressed through selective breeding for the glutinous phenotype. The glutinous phenotype is far more common in japonica rices than in indicas (*e.g.*, IRRI germplasm collection), and the high proportion of glutinous indicas seen here reflects our deliberate sampling for them in this study (to test for separate origins of the splice donor site mutation in japonica and indica varieties). Thus, these accessions may be relatively rare cases that are not typical of indica rices overall. Additional phylogenetic analysis based on selectively neutral genes will be helpful for resolving the connection between amylose content and the indica/japonica division.

On the *Wx* haplotype tree, haplotypes are differentiated into three distinct groups, which are separated from each other by 13 or more mutational steps: haplotypes A–E, haplotype S, and haplotypes F–N/P–Q (Figure 2). Whereas japonica accessions are restricted to the third group, indica accessions are represented in all three divergent groups. Thus, the indica accessions examined here are genetically much more diverse than the japonica varieties. Measures of nucleotide diversity confirm this observation, not only for the nonneutral *Wx* gene (indica, $\pi = 0.0044 \pm 0.002$; japonica, $\pi = 0.0003 \pm 0.0001$), but also for the neutrally evolving *RGRC2* gene (indica, $\pi = 0.0023 \pm 0.0002$; japonica, $\pi = 0.0015 \pm 0.0004$). These findings are consistent with previous studies, which also report greater genetic diversity in indica than in japonica varieties (*e.g.*, Glaszmann 1987). Overall, there is little obvious geographical structuring of accessions corresponding to the three haplotype groups (see the appendix); the one notable exception involves Northeast Asian nonglutinous varieties, which are predominantly represented in the third haplotype group (discussed below).

Rice domestication and the *Waxy* gene: Rice is the single most important food crop in the world, serving as the staple food for over one-third of the world's population (Khush 1997). Archaeological evidence of rice cultivation dates to well over 10,000 years ago, suggesting that it may also be the oldest crop in continuous use (Huke and Huke 1990). The development of rice-based cultures across Asia has been accompanied by tremendous phenotypic diversification within the crop, including both agroecological and grain characteristics. Glutinous rice represents one such phenotypic variant, which has come to play an important role in the cultural traditions of much of eastern Asia (*e.g.*, Xu 1992; Wada *et al.* 1999). In this study we have traced the evolutionary and geographical origins of this major domestication trait.

Nonglutinous Wx splice donor mutants: Besides glutinous accessions, 17 nonglutinous landraces also carry the splice donor site mutation (Table 1; Figure 2). Eight of these accessions share the haplotype found in most glutinous varieties (haplotype G); eight possess haplotypes derived from this haplotype (haplotypes M and N); and one carries the recombinant haplotype (R). Interestingly, these nonglutinous accessions carrying the mutation constitute over one-half of the nonglutinous accessions sampled from Northeast Asia and 93% of samples from this region that were classified as japonica rice. Moreover, these nonglutinous *Wx* mutants are almost entirely restricted to this geographical region (Table 1; Fisher's exact test, $P < 0.05$ for Northeast Asia *vs.* other regions; not significant for the other two pairwise comparisons).

Previous molecular genetic studies have established that nonglutinous varieties carrying the splice donor site mutation show reduced levels of amylose synthesis (Wang *et al.* 1995; Bligh *et al.* 1998; Cai *et al.* 1998; Hirano *et al.* 1998; Isshiki *et al.* 1998); in these varieties, reduced levels of spliced *Wx* mRNA and low levels of waxy enzymatic activity are observed. This partial suppression of the *Wx* mutant phenotype arises through cryptic splice site activation (Bligh *et al.* 1998; Cai *et al.* 1998; Isshiki *et al.* 1998) and is controlled in part by modifier genes, including mutations at the *Dull* locus (Dung *et al.* 2000; Isshiki *et al.* 2000; see also Mikami *et al.* 2000).

In our sample, the striking localization of nonglutinous *Wx* mutant accessions to Northeast Asia (Table 1; appendix) suggests that partial suppression of the *Wx*

splice donor site mutation may have been an important mechanism by which the low-amylose phenotype was selected for in this region (see also Hirano *et al.* 1998). Within Asia, nonglutinous rices with low-to-intermediate amylose content (producing cohesive cooked grains) are generally more preferred in East Asia, while high-amylose rices are more typical of South Asia. While multiple loci are known to affect quantitative amylose variation, within Northeast Asia it appears that the *Wx* locus may have played a key role in the development of nonglutinous varieties. Reverse transcriptase-PCR analysis confirms that the nonglutinous *Wx* mutants in our sample do produce spliced *Wx* mRNA despite the presence of the mutation (M. Purugganan and K. Olsen, unpublished observations).

Genetics, phylogeography, and the evolution of crop species: The occurrence of the *Wx* intron 1 splice donor mutation in all glutinous rice varieties that have been examined (Table 1; see also Wang *et al.* 1995; Cai *et al.* 1998) suggests that this mutation is required for the glutinous phenotype. At the same time, the prevalence of the mutation in Northeast Asian nonglutinous accessions (Table 1) highlights the fact that this mutation alone does not guarantee expression of the glutinous phenotype. Thus, the *Wx* intron 1 splice donor site mutation is apparently required for, but does not ensure, the glutinous phenotype.

In this respect, the evolution of *Wx* and the glutinous rice phenotype resembles other domestication traits that have been investigated at the molecular evolutionary level. In the domesticated cauliflower (*B. oleracea* ssp. *botrytis*), for example, the inflorescence head characterizing the crop is associated with a nonsense allele of the MADS-box regulatory gene *BoCAL*; however, the presence of the nonsense allele alone does not ensure the cauliflower phenotype (Purugganan *et al.* 2000). Similarly, the origin of the shoot architecture characterizing cultivated maize (*Zea mays* ssp. *mays*) has been accompanied by selection on the floral developmental gene *tb1*, yet there are no fixed genetic differences between maize *tb1* alleles and those found in the crop's progenitor (Doebley *et al.* 1997). Together, these studies suggest a trend in the genetic architecture of domestication traits. It appears that while domestication leads to selection on genes of major effect, the actions of interacting loci also strongly influence the domestication phenotype. This idea is further supported by QTL studies of genetic differences between maize and its progenitor (*e.g.*, Doebley *et al.* 1990), which reveal the central importance not only of genes of major effect, but also of secondary modifier loci in the evolution of domestication phenotypes.

This study demonstrates the utility of phylogeographic analysis of genes associated with domestication in studying crop origins. Previous studies have elucidated the molecular evolutionary basis of traits associated with domestication, including anthocyanin synthesis in maize (Hanson *et al.* 1996), inflorescence and vegetative architecture in maize (*e.g.*, Dorweiler *et al.* 1993; Doebley *et al.* 1997), and inflorescence development in *B. oleracea* (Purugganan *et al.* 2000). Other studies have used neutral genetic variation to trace the population structure and phylogeographic histories of crops (Eyre-Walker *et al.* 1998; Hilton and Gaut 1998; Olsen and Schaal 1999; Sanjur *et al.* 2002). This study represents a synthesis of these two approaches. By examining the phylogeography of a gene directly responsible for the presence or absence of amylose, we have been able to infer a single origin of the glutinous phenotype and the geographical region where this phenotype most likely arose. This approach offers a promising means of investigating the origin and diversification of traits accompanying both crop domestication and evolution in wild species.

The authors thank the International Rice Research Institute for providing rice accessions, Zhaohui Xue of the Duke University East Asia Library Collection for references on glutinous rice origins, and members of the Purugganan laboratory and Carolyn J. Ferguson for critical reading of the manuscript. Funding for this work was provided by a grant from the Alfred P. Sloan Foundation to M.D.P.

LITERATURE CITED

Ayres, N. M., A. M. McClung, P. D. Larkin, H. F. J. Bligh, C. A. Jones *et al.*, 1997 Microsatellites and a single-nucleotide polymorphism differentiate apparent amylose classes in an extended pedigree of US rice germ plasm. Theor. Appl. Genet. **94:** 773–781.

Bligh, H. F. J., P. D. Larkin, P. S. Roach, C. A Jones, H. Fu *et al.*, 1998 Use of alternate splice sites in granule-bound starch synthase mRNA from low-amylose rice varieties. Plant Mol. Biol. **38:** 407–415.

Cai, X.-L., Z.-Y. Wang, Y.-Y. Xing, J.-L. Zhang and M.-M. Hong, 1998 Aberrant splicing of intron 1 leads to the heterogeneous 5′ UTR and decreased expression of waxy gene in rice cultivars of intermediate amylose content. Plant J. **14:** 459–465.

Doebley, J., A. Stec, J. Wendel and M. Edwards, 1990 Genetic and morphological analysis of a maize-teosinte F2 population: implications for the origin of maize. Proc. Natl. Acad. Sci. USA **87:** 9888–9892.

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

Dorweiler, J., A. Stec, J. Kermicle and J. Doebley, 1993 Teosinte glume architecture 1: a genetic locus controlling a key step in maize evolution. Science **262:** 233–235.

Dung, L.-V., I. Mikami, E. Amano and Y. Sano, 2000 Study on the response of *dull endosperm 2–2*, *du2–2*, to two *wx* alleles in rice. Breed. Sci. **50:** 215–219.

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

Glaszmann, J. C., 1987 Isozymes and the classification of Asian rice varieties. Theor. Appl. Genet. **74:** 21–30.

Golomb, L., 1976 The origin, spread and persistence of glutinous rice as a staple crop in mainland Southeast Asia. J. Southeast Asian Stud. **7:** 1–15.

Hanson, M. A., B. S. Gaut, A. O. Stec, S. I. Fuerstenberg, M. M. Goodman *et al.*, 1996 Evolution of anthocyanin biosynthesis in maize kernels: the role of regulatory and enzymatic loci. Genetics **143:** 1395–1407.

Hilton, H., and B. S. Gaut, 1998 Speciation and domestication in maize and its wild relatives: evidence from the *globulin-1* gene. Genetics **150:** 863–872.

Hirano, H.-Y., M. Eiguchi and Y. Sano, 1998 A single base change altered the regulation of the *waxy* gene at the posttranscriptional

level during the domestication of rice. Mol. Biol. Evol. **15:** 978–987.

Holmes, E. C., M. Worobey and A. Rambaut, 1999 Phylogenetic evidence for recombination in dengue virus. Mol. Biol. Evol. **16:** 405–409.

Huke, R. E., and E. H. Huke, 1990 *Rice: Then and Now*. International Rice Research Institute, Manila, The Philippines.

Isshiki, M., K. Morino, M. Nakajima, R. J. Okagaki, S. R. Wessler *et al.*, 1998 A naturally occurring functional allele of the rice *waxy* locus has a GT to TT mutation at the 5′ splice site of the first intron. Plant J. **15:** 133–138.

Isshiki, M., M. Nakajima, H. Satoh and K. Shimamoto, 2000 *dull*: rice mutants with tissue-specific effects on the splicing of the waxy pre-mRNA. Plant J. **23:** 451–460.

Kaminaka, H., S. Morita, M. Nakajima, T. Masumura and K. Tanaka, 1998 Gene cloning and expression of cytosolic glutathione reductase in rice (*Oryza sativa* L.). Plant Cell Physiol. **39:** 1269–1280.

Kato, S., H. Kosaka and S. Hara, 1928 On the affinity of rice varieties as shown by fertility of hybrid plants. Bull. Sci. Fac. Agric. Kyushu Univ. **3:** 132–147.

Khush, G. S., 1997 Origin, dispersal, cultivation and variation of rice. Plant Mol. Biol. **35:** 25–34.

Mikami, I., L.-V. Dung, H.-Y. Hirano and Y. Sano, 2000 Effects of the two most common *Wx* alleles on different genetic backgrounds in rice. Plant Breed. **119:** 505–508.

Morishima, H., Y. Sano and H.-I. Oka, 1992 Evolutionary studies in cultivated rice and its wild relatives, pp. 135–184 in *Oxford Surveys in Evolutionary Biology*, Vol. 8, edited by D. Futuyma and J. Antonovics. Oxford University Press, Oxford.

Oka, H.-I., 1988 *Origin of Cultivated Rice*. Elsevier, Amsterdam.

Olsen, K. M., and B. A. Schaal, 1999 Evidence on the origin of cassava: phylogeography of *Manihot esculenta*. Proc. Natl. Acad. Sci. USA **96:** 5586–5591.

Posada, D., and K. A. Crandall, 1998 MODELTEST: testing the model of DNA substitution. Bioinformatics **14:** 817–818.

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

Rao, S. A., C. Bounphanousay, J. M. Schiller and M. T. Jackson, 2002 Collection, classification, and conservation of cultivated and wild rices of the Lao PDR. Genet. Res. Crop Evol. **49:** 75–81.

Roder, W., B. Keoboulapha, K. Vannalath and B. Phouaravanh, 1996 Glutinous rice and its importance for hill farmers in Laos. Econ. Bot. **50:** 401–408.

Rozas, J., and R. Rozas, 1999 DnaSP: an integrated program for molecular population genetics and molecular evolution analysis, version 3. Bioinformatics **15:** 174–175.

Sanjur, O. I., D. R. Piperno, T. C. Andres and L. Wessel-Beaver, 2002 Phylogenetic relationships among domesticated and wild species of *Cucurbita* (Cucurbitaceae) inferred from a mitochondrial gene: implications for crop plant evolution and areas of origin. Proc. Natl. Acad. Sci. USA **99:** 535–540.

Sano, Y., 1984 Differential regulation of waxy gene expression in rice endosperm. Theor. Appl. Genet. **68:** 467–473.

Swofford, D. L., 2000 *PAUP*. Phylogenetic Analysis Using Parsimony (*and Other Methods)*, version 4.0b8a. Sinauer, Sunderland, MA.

Tajima, F., 1983 Evolutionary relationship of DNA sequences in finite populations. Genetics **105:** 437–460.

Tajima, F., 1989 Statistical method for testing the neutral mutation hypothesis by DNA polymorphism. Genetics **123:** 585–595.

Terwiel, B. J., 1994 Rice legends in mainland Southeast Asia: history and ethnography in the study of myths of origin. Contrib. Southeast Asian Ethnogr. **10:** 5–36.

Wada, J., B. Pogosian, Y. Sato and L. E. Grivetti, 1999 Japanese festival foods. Ecol. Food Nutr. **38:** 307–350.

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

Wang, Z.-Y., F.-Q. Zheng, G.-Z. Shen, J.-P. Gao, D. P. Snustad *et al.*, 1995 The amylose content in rice endosperm is related to the post-transcriptional regulation of the *waxy* gene. Plant J. **7:** 613–622.

Watterson, G. A., 1975 On the number of segregating sites in genetical models without recombination. Theor. Popul. Biol. **7:** 256–276.

Xu, F. (Editor), 1992 *Encyclopedia of Chinese Customs (Zhongguo feng su ci dian)*. Shanghai ci su chu ban she, Shanghai, China.

Communicating editor: S. W. Schaeffer

APPENDIX

Rice accessions examined

IRRI-IRGC no.	Country	Name	Endosperm	Variety group	*Wx* haplotype
26631	Bangladesh	Gouji	Nonglutinous	Indica	A
64766	Bangladesh	Baulan	Nonglutinous	Indica	B
64774	Bangladesh	Dharia Boalia	Nonglutinous	Indica	B
26708	Bangladesh	Lal Binni	Glutinous	Indica	G
62183	Bhutan	Ray Konjtshey	Nonglutinous	Indica	H
62154	Bhutan	Asu	Nonglutinous	Indica	B
13476	Brunei	Badus	Glutinous	Javanica	G
13765	Brunei	Menjalin	Glutinous	Indica	G
5965	Cambodia	Kraya K 101	Nonglutinous	Indica	A
23004	Cambodia	Lovoan	Glutinous	Indica	G
23036	Cambodia	Neang Chaul	Glutinous	Javanica	G
77317	China	Ai Gu	Nonglutinous	Indica	A
77336	China	Gi Yue Zhan	Nonglutinous	Indica	A
77347	China	Hong Jiao Zhan	Nonglutinous	Indica	A
77366	China	Jing Bao Ying	Nonglutinous	Indica	A
77376	China	Mao Zhan	Nonglutinous	Indica	A
77385	China	Qing Gao Er	Nonglutinous	Indica	A
77435	China	He Jiang 5	Nonglutinous	Japonica	P
51391	China	Pai-Li-Huang	Nonglutinous	Indica	A
60072	China	Ye Miao Dao	Nonglutinous	Japonica	G
61776	China	Da Xiang Nuo	Nonglutinous[a]	Indica	G
63960	China	49-1	Glutinous	Indica	L
64003	China	727031	Glutinous	Intermed	L
10348	China	K. Assaw 4546	Glutinous	Japonica	G
52336	India	Kusale Bhat (White)	Nonglutinous	Indica	S
38323	India	Mutant 65	Nonglutinous	Indica	F
51705	India	Arc 15297	Nonglutinous	Indica	B
51746	India	Arc 18475	Nonglutinous	Indica	B
74751	India	Kasapur	Nonglutinous	Indica	C
74716	India	Sayari	Nonglutinous	Indica	C
10594	Indonesia	Sigardis	Nonglutinous	Indica	A
66535	Indonesia	Alur Kuning	Nonglutinous	Indica	B
66590	Indonesia	Lekat Rentik	Nonglutinous	Indica	A
66545	Indonesia	Emping Ara	Nonglutinous	Indica	B
13512	Indonesia	Ketan Gadjih	Glutinous	Javanica	G
48588	Indonesia	Pulorogo	Glutinous	Indica	I
10383	Japan	Take	Nonglutinous	Japonica	G
10795	Japan	Kokumasari	Nonglutinous	Japonica	M
10879	Japan	Isaribi	Nonglutinous	Japonica	G
10889	Japan	Yamabiko	Nonglutinous	Japonica	M
10896	Japan	Nishikaze	Nonglutinous	Japonica	M
11117	Japan	Chokoto	Nonglutinous	Indica	P
10389	Japan	A 83	Nonglutinous	Japonica	M
10390	Japan	A 108	Nonglutinous	Japonica	G
10397	Japan	A 58	Glutinous	Japonica	G
11118	Japan	Yakeiko	Glutinous	Japonica	G
77648	Korea	Damajo	Nonglutinous	Japonica	G
77654	Korea	Geum	Nonglutinous	Japonica	M
77656	Korea	Gokrangdo	Nonglutinous	Japonica	M
77658	Korea	Gumi	Nonglutinous	Japonica	M
77647	Korea	Daejichal 2	Glutinous[b]	Japonica	G
19692	Korea	An Nam Zo	Glutinous	Japonica	G
19704	Korea	Baek Gok Na	Glutinous	Japonica	G
77638	Korea	Aeguk	Nonglutinous	Japonica	G
77668	Korea	Monggeunchal	Nonglutinous	Japonica	G
77649	Korea	Doyajichal	Glutinous[b]	Japonica	G
11515	Laos	Do Makkham	Glutinous[b]	Javanica	G

(*continued*)

APPENDIX

(Continued)

IRRI-IRGC no.	Country	Name	Endosperm	Variety group	*Wx* haplotype
12936	Laos	I Houm B	Glutinous[b]	Javanica	G
11508	Laos	Dam-Do	Glutinous	Javanica	G
11620	Laos	Mack Kouk	Glutinous	Javanica	G
13442	Malaysia	Padi Pulot	Nonglutinous[a]	Indica	R
13976	Malaysia	Radin Benua 126	Nonglutinous[a]	Indica	A
71554	Malaysia	Lekatan	Glutinous	Indica	G
71587	Malaysia	Pulut Apila	Glutinous	Indica	G
71503	Malaysia	Baganan Asalao	Nonglutinous	Indica	F
71529	Malaysia	Indurok	Nonglutinous	Indica	F
77952	Malaysia	Bidor	Nonglutinous	Indica	F
77956	Malaysia	Berteh	Nonglutinous	Unclassified	F
33432	Myanmar	Nathasi	Glutinous	Indica	G
33436	Myanmar	Natpyihmwe	Glutinous	Indica	G
16201	Nepal	Rato Marsi	Nonglutinous	Indica	J
16184	Nepal	Masino	Nonglutinous	Indica	F
16192	Nepal	Marinaker Anase	Nonglutinous	Indica	B
61934	Nepal	Co Anadi	Glutinous	Javanica	G
70922	Pakistan	Kolai	Nonglutinous	Indica	F
70932	Pakistan	Saray Koli	Nonglutinous	Japonica	N
39375	Philippines	IR3941-77	Nonglutinous	Indica	A
11169	Philippines	Bakaw	Nonglutinous	Javanica	F
12050	Philippines	Balatinaw	Nonglutinous[a]	Javanica	F
23365	Philippines	T. Q. Qi. Qinqullatiw-B	Nonglutinous	Javanica	F
38831	Philippines	Ratossa	Glutinous	Indica	G
47125	Philippines	Azucena	Glutinous	Indica	G
38713	Philippines	Alaminos	Glutinous	Indica	R
72474	Philippines	P.T.Q. Qimbanig	Nonglutinous	Javanica	F
12091	Sri Lanka	Madael Galle	Nonglutinous[a]	Indica	B
11670	Sri Lanka	Murungan	Nonglutinous[a]	Indica	B
11692	Sri Lanka	Kalukarayal	Nonglutinous[a]	Indica	B
11713	Sri Lanka	Batapolawee	Nonglutinous[a]	Indica	B
11724	Sri Lanka	Oddavalan (Straw)	Nonglutinous[a]	Indica	A
10410	Taiwan	Kokko	Glutinous	Indica	G
10444	Taiwan	Shinchiku-Mochi-Iku 40	Glutinous	Japonica	G
10461	Taiwan	Taichu Iku 61	Glutinous	Japonica	K
78235	Thailand	Tne 39	Nonglutinous	Indica	E
78240	Thailand	Chaw Pin Gaew	Nonglutinous	Indica	A
10872	Thailand	Daw Leuang 133-3-88	Glutinous	Indica	G
11476	Thailand	Nahng Cha-Lawng	Glutinous	Javanica	G
15064	Thailand	Ku281	Glutinous	Javanica	Q
78254	Thailand	Khao Tah Ruay	Nonglutinous	Indica	F
11115	Vietnam	Te-Tep	Nonglutinous[a]	Indica	D
78320	Vietnam	Khau Can	Glutinous	Javanica	G
78325	Vietnam	Khau Lao	Glutinous	Intermed	G
10224	Vietnam	Muong Trai	Glutinous	Indica	G
78291	Vietnam	Ba Danh	Nonglutinous	Indica	A
78377	Vietnam	Quyet Tam 813	Nonglutinous	Indica	A
78380	Vietnam	Te Kha Trang Hoa Binh	Nonglutinous	Indica	A

Variety group designations follow the IRRI germplasm classification.

[a] Classified as glutinous in IRRI germplasm database.

[b] Classified as nonglutinous in IRRI germplasm database.