New Phytologist

Research

Association between nonsynonymous mutations of *starch synthase IIa* and starch quality in rice (*Oryza sativa*)

Guoqin Yu[1,2], Kenneth M. Olsen[1] and Barbara A. Schaal[1]

[1]Department of Biology, Washington University of St. Louis, Missouri, 63130, USA; [2]Present address: Institute of Bioscience and Biotechnology Research, University of Maryland, College Park, Maryland, 20742, USA

Author for correspondence:
Guoqin Yu
Tel: +1 240 314 6234
Email: guoqinyu@umd.edu

Received: *3 August 2010*
Accepted: *26 August 2010*

New Phytologist (2011) **189**: 593–601
doi: 10.1111/j.1469-8137.2010.03495.x

Key words: amylopectin, crop domestication, genotype–phenotype association, nonsynonymous mutation, *Oryza sativa* (rice), starch quality, starch synthase IIa.

Summary

- Starch quality is one of the most important agronomic traits in Asian rice, *Oryza sativa*. *Starch synthase IIa* (*SsIIa*) is a major candidate gene for starch quality variation. Within *SsIIa*, three nonsynonymous mutations in exon 8 have been shown to affect enzyme activity when expressed in *Escherichia coli*.
- To search for the variation in *SsIIa* that is responsible for starch quality variation in rice, we sequenced the *SsIIa* exon 8 region and measured starch quality as starch disintegration in alkali for 289 accessions of cultivated rice and 57 accessions of its wild ancestor, *Oryza rufipogon*. A general linear model and nested clade analysis were used to identify the associations between the three nonsynonymous single nucleotide polymorphisms (SNPs) and starch quality.
- Among the three nonsynonymous SNPs, we found strong evidence of association at one nucleotide site ('SNP 3'), corresponding to a Leu/Phe replacement at codon 781. A second SNP, corresponding to a Val/Met replacement at codon 737, could potentially show an association with increased sample sizes.
- Variation in *SsIIa* enzyme activity is associated with the cohesiveness of rice grains when cooked, and our findings are consistent with selection for more cohesive grains during the domestication of *tropical japonica* rice.

Introduction

Asian rice, *Oryza sativa*, is one of the most important crop species in the world, feeding about one-half of the world's population. Rice is diverse in terms of both morphology and underlying genetics. On the basis of morphological, physiological and ecological differences, rice has traditionally been classified into two subspecies: *O. sativa* ssp. *indica* and *O. sativa* ssp. *japonica* (Khush, 1997). More recent work using genetic markers has identified five genetically distinct cultivated rice groups: *aus* and *indica* rice (within the *indica* subspecies); and *tropical japonica*, *temperate japonica* and *aromatic* rice (within the *japonica* subspecies) (Garris *et al.*, 2005). Asian rice is one of the oldest domesticated species, originating in Asia at least 11 000 yr ago (Khush, 1997). Previous studies have indicated that *O. sativa* is derived from its wild ancestor, *O. rufipogon* (Khush, 1997). Recent work has suggested that there were at least two domestication events: one in south China for *japonica* rice, and another in the south and southwest of the Himalayan mountain range for *indica* rice. *Aus* rice may represent a third, independent domestication event (Londo *et al.*, 2006).

Starch, the major component of cereal grains, is a determinant of both the yield and quality of cereal crops. Starch is composed of amylose (a linear molecule of (1 → 4)-linked α-D-glucopyranosyl units) and amylopectin (a highly branched molecule of α-D-glucopyranosyl units) (Buléon *et al.*, 1998). *Indica* rice varieties tend to have higher ratios of amylose to amylopectin than do *japonica* rice varieties, with an amylose content of up to 30%. *Japonica* varieties tend to have an amylose content in the range 10–20%. Depending on the relative proportions of these two starches and on the branching structure of amylopectin, rice starch can be diverse in its chemical and physical properties (Umemoto *et al.*, 2004). The starch disintegration level in alkali has been widely used to assess the chemical and physical properties of the rice grain (Little *et al.*, 1958;

Umemoto & Aoki, 2005). High starch disintegration in alkali suggests either a high proportion of amylose or amylopectin with long branches; a low starch disintegration level in alkali suggests the opposite properties (Umemoto & Aoki, 2005). Rice varieties with high levels of amylose (such as *indica* rice) or amylopectin with long branches tend to have discrete, noncohesive grains when cooked, and correspondingly low disintegration levels in alkali, whereas varieties with more short-branched amylopectin tend to have cohesive grains when cooked and a correspondingly high disintegration in alkali.

Several candidate genes for starch quality variation in rice have been identified (Myers *et al.*, 2000). Among them, *starch synthase IIa* (*SsIIa*) is one of the major genes controlling variation in alkali starch disintegration between *indica* and *japonica* rice (Kudo, 1968; Umemoto *et al.*, 2002, 2004; Nakamura *et al.*, 2005; Tian *et al.*, 2009). The *SsIIa* enzyme elongates short chains of amylopectin clusters with a degree of polymerization (DP) of ≤ 10. Extremely low *SsIIa* enzyme activity results in S-type amylopectin, which is enriched in short chains (DP, 6–10) with few long chains (DP, 12–22), whereas high *SsIIa* enzyme activity produces L-type amylopectin (Umemoto *et al.*, 2004).

Four nonsynonymous single nucleotide polymorphisms (SNPs) have been observed in the *SsIIa* gene within cultivated rice (Umemoto *et al.*, 2004). Three of these SNPs are located in exon 8 and the fourth is in exon 1 (Fig. 1). Previous *SsIIa* expression studies in *Escherichia coli* (Nakamura *et al.*, 2005; Umemoto & Aoki, 2005) have suggested that the three SNPs in exon 8 could have a marked effect on *SsIIa* enzyme activity, whereas the SNP at exon 1 has no effect on phenotype. These studies suggest a relationship between the variation in *SsIIa* enzyme activity and nonsynonymous SNPs at *SsIIa* exon 8 in *E. coli* (Nakamura *et al.*, 2005).

Several questions remain unanswered with respect to the phenotypic effects of these *SsIIa* SNPs in rice. What is the contribution of these three nonsynonymous SNPs in *SsIIa* exon 8 to the variation in starch quality among rice varieties? What is the distribution of these nonsynoymous SNPs in wild and cultivated rice? Are there more functionally relevant nonsynonymous SNPs that would be revealed with more extensive sampling? What is the evolutionary relationship among these nonsynomous SNPs? Finally, how did starch quality evolve during rice domestication?

For this study, we sampled 289 *O. sativa* and 57 *O. rufipogon* accessions. Our objectives were as follows: to survey and identify the nonsynonymous SNPs at *SsIIa* exon 8 in wild and cultivated rice in the context of *SsIIa* evolution; to determine starch quality differences among genetically distinct rice varieties and *O. rufipogon*, using the starch disintegration level in alkali, to evaluate phenotypic differences in the context of rice domestication; and to determine the association between *SsIIa* haplotypes, as defined by exon 8 candidate SNPs, and starch quality variation in *O. rufipogon* and cultivated rice.

Materials and Methods

Plant materials

Sampled accessions of *O. sativa* and *O. rufipogon* are listed in Supporting Information Tables S1 and S2. *Oryza rufipogon* collections (57 accessions) cover the species' range, except for Australia, with extensive sampling from the centers of diversity: Thailand, India and China. Sampling of cultivated rice spans the five genetically distinct variety groups identified by Garris *et al.* (2005): *aus*, *indica*, *aromatic*, *tropical japonica* and *temperate japonica.* Not all accessions that were genetically characterized were also phenotyped because of limited seed availability. Two other *Oryza* species, *O. barthii* and *O. meridionalis*, were sampled to serve as outgroups. A summary of samples is given in Table 1.

DNA extraction, PCR and sequencing

DNA was extracted from dried leaves by a cetyl trimethylammonium bromide (CTAB) method with minor modifications (Doyle & Doyle, 1990). PCR primers were

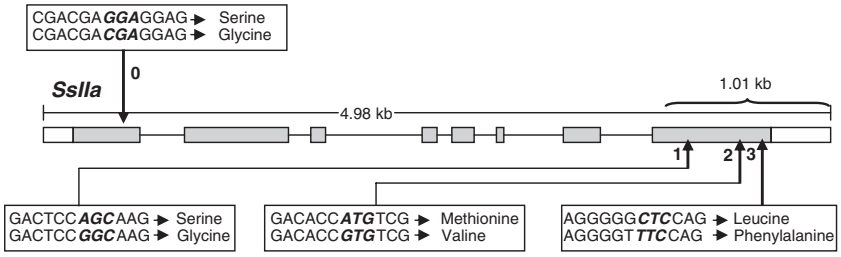

Fig. 1 Schematic representation of *starch synthase IIa* (*SsIIa*) showing the location of nonsynonymous polymorphisms observed in cultivated rice. The boxes represent exon regions: shaded regions are translated, and nonshaded regions are transcribed but not translated. Lines connecting boxes correspond to intron regions. The sizes of the gene and sequenced region are indicated above the boxes. The numbers 1, 2 and 3 indicate the locations of single nucleotide polymorphisms (SNPs) 1, 2 and 3 analyzed for phenotypic associations. Bold italic sequences indicate codons with nonsynonymous mutations.

Table 1 Summary of samples used in the analyses

Rice group	No. of accessions for phenotypic data	No. of accessions for genotypic data	No. of accessions for genotype–phenotype association
Oryza rufipogon	37	57	22
indica	93	50	46
tropical japonica	114	56	54
temperate japonica	51	29	26
aromatic	14	8	8
aus	16	8	8

designed using Primer3 (http://frodo.wi.mit.edu/primer3/) and the Nipponbare reference rice genomic sequence (available at http://www.gramene.org/). PCRs were conducted under the following conditions: 95°C for 5 min; 30 cycles of 95°C for 50 s, 53 or 58°C for 1 min (annealing temperature differs by primers) and 72°C for 2.5 min; finally, 10 min of extension at 72°C. Two pairs of primers (*SsIIa*_F1: GCACTCCTGCCTGTTTATCTG, *SsIIa*_R1: CGAGGCCACGGTGTAGTTG; *SsIIa*_F2: CGGGA GAACGACTGGAAGATGAAC, *SsIIa*_R2: CAGACACG AGAGCTAATGAAG) were used for PCRs. The annealing temperature was 53°C for the first primer pair and 58°C for the second pair. PCR products were cleaned using Exo1/SAP commercial kits, and then cycle sequenced using BigDye Terminator chemistry (Applied Biosystems, Carlsbad, CA, USA) and analyzed on an ABI 3130 capillary sequencer (Applied Biosystems) in the Biology Department sequencing facility of Washington University.

Genetic diversity analysis

Sequences were aligned and manually edited with the software Biolign version 4.0.6.2 (http://www.mbio.ncsu.edu/BioEdit/bioedit.html). The sequences of Nipponbare were downloaded from Genbank (http://www.ncbi.nlm.nih.gov/Genbank/) and included as a *temperate japonica* sample. Nucleotide diversity analyses (numbers of polymorphic synonymous, nonsynonymous and silent sites; pairwise nucleotide diversity θ_π; average number of segregating sites θ_W) (Watterson, 1975; Tajima, 1983) were performed in DnaSP version 5.0 (Librado & Rozas, 2009). Only silent sites were used for estimations of θ_π and θ_W. Neutrality tests, Tajima's *D*, and Fay and Wu's *H* were also calculated in DnaSP version 5.0 (Tajima, 1989; Fay & Wu, 2000; Rozas *et al.*, 2003). *Oryza meridionalis* served as the outgroup species for Fay and Wu's *H*. Coalescent simulations with 10 000 replications were conducted to determine the statistical significance.

Haplotypes were inferred on the basis of three nonsynonymous SNPs within *SsIIa* exon 8 ('SNPs 1, 2 and 3'; see Results), and haplotype frequencies were calculated for each rice group. Differences in frequency distributions among rice groups were analyzed by nonparametric Friedman tests because of violation of the assumption of normality or equality of variance for the frequency distribution data. Haplotype networks were constructed by medium joining methods in software NETWORK version 4.5.1.0 (Bandelt *et al.*, 1999). The program was run under default parameters.

Estimation of disintegration of rice endosperm starch in alkali solution

In order to exclude the effect of environment on starch phenotype, seeds for all tested accessions were obtained from plants grown under uniform conditions in the glasshouse of Washington University in St. Louis, MO, USA. Seeds were harvested, dried (Equatherm incubator; Curtin Matheson Scientific Inc., Houston, Texas, USA), dehusked (Kett, model TR120, Villa Park, CA, USA) and polished (Kett pearlest grain polisher). Six polished seeds per accession were placed in 1.5% KOH solution for 23 h at room temperature. The degree of disintegration was quantified by a numerical scale of 1–7 (starch alkali spreading score, SASS), following the standard protocol of Little *et al.* (1958). Representative photographs of starch phenotypes in alkali and their corresponding SASS values are shown in Fig. S1. As a result of the limited availability of seeds, less than six seeds were used for some wild rice accessions. The average SASS value across seeds was used for each accession in subsequent analyses. Variation of SASS scores within rice accessions is very low; 273 of 325 rice accessions showed no variance in SASS values within accessions (Fig. S2).

Analyses of SASS variation

In order to determine differences in SASS values among rice groups, a nonparametric Kruskal–Wallis test and a parametric one-way analysis of variance (ANOVA) were performed. Variation in SASS values among inferred haplotype groups (with haplotypes determined by SNPs 1, 2 and 3) was also assessed by a Kruskal–Wallis test and an ANOVA. Pairwise SASS differences among rice groups and haplotype groups were analyzed by Student's *t*-tests. A significance level of 5% was used for all tests in this study.

Previous studies of the starch alkali phenotype have suggested that the seven SASS scores can be binned into classes (1–2, low; 3–5, medium; 6–7, high) which correspond to recognizably distinct rice cooking phenotypes (Bhattacharya, 1979; He *et al.*, 2006). Varieties with low SASS scores (1–2) tend to have discrete, noncohesive starch when cooked, whereas rice varieties with high SASS values (6–7) have noticeably cohesive starch. As a complementary analysis to tests treating SASS values as seven distinct categories, we also analyzed SASS as a qualitative trait (low, medium,

high). The distributions of these three SASS categories were assessed for each rice group. Frequency distributions were then compared among rice groups by Friedman tests.

Genotype–phenotype association analysis

In order to search for nonsynonymous polymorphic sites that are associated with starch disintegration variation, a general linear model (Bradbury *et al.*, 2007) and nested clade analysis (Templeton *et al.*, 1987) were performed. Both analyses were run separately for each of three major cultivated rice groups (*indica*, *tropical japonica*, *temperate japonica*) and *O. rufipogon* in order to exclude the confounding effect of population structure on association analysis (Marchini *et al.*, 2004). The *aus* and *aromatic* varieties were not included for analyses because of their small sample sizes. The general linear model analysis was conducted using the program TASSEL (Bradbury *et al.*, 2007). Statistical significance was determined at a 5% significance level with a Bonferroni correction for multiple comparisons.

Nested clade analysis (Templeton *et al.*, 1987) requires a haplotype network to define a hierarchy of evolutionary clades. It starts from the tips of the network by nesting 'zero-step clades' (the tip haplotypes) within 'one-step clades' (separated from tip haplotypes by one mutational change), and proceeds step by step until the final nesting includes the entire network. For the *SsIIa* gene, we only needed to nest 'zero-step clades' into 'one-step clades' to include all the clades of the entire network for nonsynonymous polymorphic sites. Nonsynonymous polymorphisms in *SsIIa* were used as they are the most likely candidates for association with starch disintegration diversity in rice. Moreover, the haplotype network for nonsynonymous polymorphic sites includes a sufficient number of samples in each clade for statistical analyses. Zero-step clades and their one-step clades were compared to determine whether the nonsynonymous mutations between them were statistically associated with a change in SASS phenotype. These comparisons were conducted by nonparametric Kruskal–Wallis tests with significance at the 5% level with a Bonferroni correction.

Results

Nucleotide variation at *SsIIa* exon 8 in rice

As population structure has been observed in *O. sativa*, nucleotide variation in *O. sativa* was examined separately in the five genetically distinct rice variety groups: *indica*, *tropical japonica*, *temperate japonica*, *aus* and *aromatic* rice. The observed nucleotide variation in *O. sativa* and its wild progenitor *O. rufipogon* is given in Table 2. The estimation of nucleotide diversity by θ_π and θ_W is highest in *O. rufipogon* (θ_π = 0.00508 and θ_W = 0.00486), followed in decreasing order by *indica*, *temperate japonica*, *tropical japonica*, *aromatic* and *aus* (where no polymorphisms were observed).

Five nonsynonymous SNPs were observed in *O. rufipogon* samples, three were observed in *indica*, *tropical japonica* and *temperate japonica* varieties, and one was observed in *aus* and *aromatic* varieties. We have designated the three nonsynonymous SNPs observed in *tropical japonica* and *temperate japonica* rice as SNPs 1, 2 and 3 (Fig. 1). The nonsynonymous SNP observed in *aus* and *aromatic* varieties is SNP 1. The three nonsynonymous SNPs observed in *indica* include SNP 1, 3 and a singleton which only exists in one *indica* individual in our samples. The five nonsynonymous SNPs observed in *O. rufipogon* include SNPs 1, 3 and three nonsynonymous SNPs detected only in *O. rufipogon* and at very low frequency. Among these three nonsynonymous SNPs, two are singletons and were detected as a homozygote in one individual of *O. rufipogon*. The other was found as a heterozygote in two *O. rufipogon* individuals.

Among all the rice varieties and *O. rufipogon*, Tajima's *D* values range from −1.055 to 0.615, and Fay and Wu's *H* values range from −9.428 to 1.404 (Table 2). No Tajima's *D* value deviates statistically significantly from neutral expectations. Fay and Wu's *H* value deviates significantly from neutral expectations only in *tropical japonica* varieties.

Haplotype networks

A haplotype network based on all polymorphic sites was constructed to infer the genealogical relationships among

Table 2 Genetic diversity of *starch synthase IIa* (*SsIIa*) exon 8 in *Oryza rufipogon* and rice variety groups

Polymorphism	*O. rufipogon*	*aus*	*indica*	*tropical japonica*	*temperate japonica*	*aromatic*
N	57	8	50	56	29	8
Nonsynonymous	5	1	3	3	3	1
Synonymous	5	0	3	2	2	1
Silent	11	0	8	5	5	3
θ_π	0.00508	0	0.00421	0.00173	0.00239	0.00262
θ_W	0.00486	0	0.00362	0.00222	0.00261	0.00236
Tajima's *D*	−0.64032	−1.05482	0.61507	−0.39692	−0.20385	0.48523
Fay and Wu's *H*	1.40351	0.21429	−1.11373	−9.42828**	−4.41133	1.14286

**, $P < 0.01$

haplotypes detected in rice varieties and *O. rufipogon* (Fig. S3). The number of haplotypes observed in *O. sativa* is less than that in *O. rufipogon*. However, some 'unique' haplotypes (haplotypes D to M; Fig. S3) were found in *O. sativa*. Moreover, haplotypes B and C occur at a high frequency in *O. sativa*, but are present at a low frequency in *O. rufipogon*. A second haplotype network, based only on the three major nonsynonymous SNPs (SNPs 1, 2 and 3), was constructed for the nested clade analysis (Fig. 2a,b). Five haplotypes in total were observed in this derived haplotype network. Haplotype frequencies for each rice group are given in Table 3. Haplotypes H1, H2 and H3 are present in high frequency, haplotype H4 was only detected in three *temperate japonica* individuals and one *tropical japonica*, and haplotype H5 was found in one *tropical japonica* and one *indica* individual. Rice groups varied in the frequencies of these haplotypes (Table 3). *Oryza rufipogon* as well as *indica*, *aromatic* and *aus* rice varieties have haplotype H1 in the highest frequency, whereas *tropical japonica* and *temperate japonica* rice varieties have haplotypes H2 and H3 in highest frequency, respectively. Friedman tests revealed no significant haplotype frequency differences among rice groups (Friedman chi square = 4.058, $P = 0.5410$, df = 5).

Starch quality variation

The SASS values for each rice group are shown in Fig. 3. The Kruskal–Wallis test did not show significantly different SASS values among rice varieties ($H = 9.377$, $P = 0.0949$, df = 5). However, ANOVA indicated statistically significant differences among these groups ($F = 2.617$, $P = 0.0245$, df = 5). In pairwise comparisons between rice groups, Student's *t*-tests indicated significant differences between *tropical japonica* and *aus* and between *tropical japonica* and *O. rufipogon* (Table 4), which is consistent with the significant result by the ANOVA test. None of the pairwise comparisons were statistically significant following Bonferroni correction.

The frequency distribution of SASS categories (based on binned SASS values) is shown for each rice group in

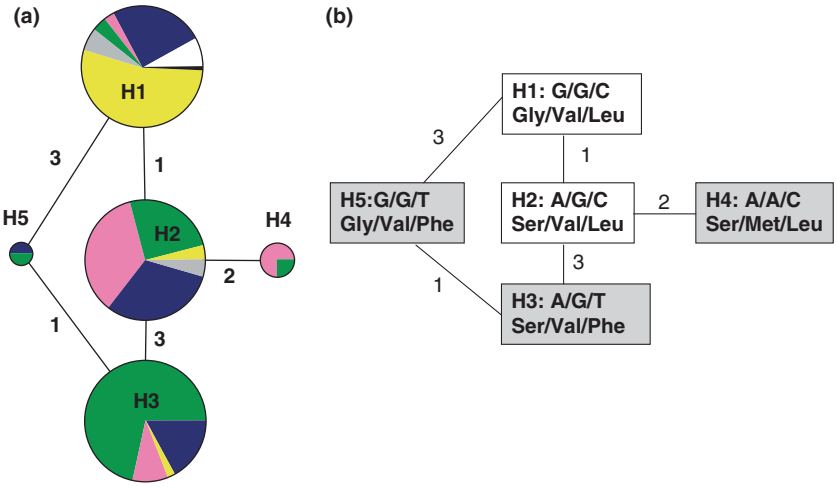

Fig. 2 Haplotype network for major nonsynonymous polymorphisms detected in the *starch synthase IIa* (*SsIIa*) exon 8 region. Numbers on the lines indicate nonsynonymous single nucleotide polymorphisms (SNPs) 1, 2 and 3 corresponding to Fig 1. (a) Labels H1–H5 indicate haplotype names. Colors indicate rice groups in which a haplotype was detected, as follows: *aus* (white), *indica* (blue), *tropical japonica* (green), *temperate japonica* (pink), *aromatic* (grey), *Oryza barthii* (black), *O. rufipogon* (yellow). The size of each haplotype circle is proportional to the number of individuals possessing that haplotype. (b) Relationship between major nonsynonymous SNPs (1, 2, 3), amino acid changes and starch phenotype for *SsIIa* haplotypes H1–H5. The nucleotides and amino acids distinguished by the three SNPs are shown in rectangles for each haplotype. Shaded and nonshaded rectangles correspond to haplotypes predominantly occurring in rice accessions with high and low starch alkali spreading score (SASS) values, respectively.

Table 3 Haplotype frequencies (%) in *Oryza rufipogon* and cultivated rice variety groups

Haplotype	*aus*	*indica*	*tropical japonica*	*temperate japonica*	*aromatic*	*O. rufipogon*
H1	100.0	50.0	7.1	10.7	75.0	94.7
H2	0.0	30.0	21.4	60.7	25.0	3.5
H3	0.0	18.0	67.9	17.9	0.0	1.8
H4	0.0	0.0	1.8	10.7	0.0	0.0
H5	0.0	2.0	1.8	0.0	0.0	0.0

H1, H2, H3, H4 and H5 correspond to the haplotypes shown in Fig. 2.

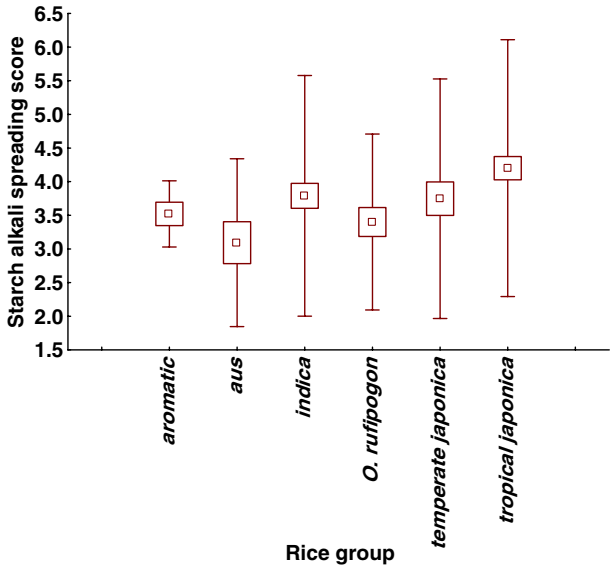

Fig. 3 Mean values of the starch alkali spreading score (SASS) in rice groups. SE is the standard error and SD is the standard deviation (□, mean; ☐, mean ± SE; ⌶, mean ± SD).

Table 5. All groups have the highest frequency at the medium SASS category (corresponding to SASS values of 3–5). The low SASS category (1–2) showed similar frequency (18.92–25%) in all rice varieties, except in *aromatic* varieties (7.14%). The medium SASS category was present at highest frequency in *aromatic* varieties (92.86%) and at lowest frequency in *tropical japonica* varieties (42.11%). The high SASS category (7–8) was absent in *aromatic* and *aus*, present at low frequency in *O. rufipogon* (5.41%), present at medium frequency in *indica* and *temperate japonica* (19.35%, 19.61%), and present at high frequency in *tropical japonica* (38.6%). Although there are frequency differences for low and high SASS categories among rice varieties, the difference is not statistically significant by the Friedman test (Friedman chi square = 0.7692, P = 0.9790, df = 5).

SASS values for each haplotype group are shown in Fig. 4. Haplotypes H3, H4 and H5 all have high SASS values, whereas haplotypes H1 and H2 have values in the low to medium range. The Kruskal–Wallis test (H = 105.2615, P = 0.0001, df = 4) and ANOVA (F = 91.971, P = 0.0001, df = 4) both indicate statistically significant differences in SASS values among haplotypes. Pairwise comparisons between haplotypes using Student's t-tests revealed that haplotypes H1 and H2 are both statistically significantly different from haplotypes H3, H4 and H5 ($P < 0.0021$), with statistical significance remaining after Bonferroni correction. However, not all the individuals with haplotypes H1 or H2 have low or medium SASS: one individual with haplotype H1 and one individual with haplotype H2 have high SASS. Similarly, not all individuals with haplotype H3 have high SASS, as one individual with H3 has medium SASS (Table S1).

Association between SASS phenotype and nonsynonymous SNPs at *SsIIa* exon 8

The results of association analyses between starch phenotype and nonsynonymous SNPs at *SsIIa* exon 8 by general linear models are shown in Table 6. Only SNP 3 is significantly associated with SASS variation in *indica*, *tropical japonica* and *temperate japonica* after Bonferroni correction. The results of association analyses between phenotype and haplotype by nested clade analysis are given in Table 7. Haplotype H5 is not included in the nested clade analysis as it is most likely the result of recombination between H1 and H3 (see Discussion). Only the phenotypic comparison between H2 and H3 is significant in *indica* and *tropical japonica* based on a 5% significance level with Bonferroni correction.

Table 4 *P* values of Student's *t*-tests comparing starch alkali spreading scores (SASS) between rice groups

	aus	*indica*	*tropical japonica*	*temperate japonica*	*aromatic*
indica	0.138529				
tropical japonica	0.021431*	0.077491			
temperate japonica	0.177222	0.891926	0.113499		
aromatic	0.550662	0.351509	0.081608	0.403848	
Oryza rufipogon	0.429825	0.232443	0.013466*	0.318538	0.858841

*, $P < 0.05$.

Table 5 Frequency distribution (%) of binned starch alkali spreading scores (SASS) in cultivated rice groups and *Oryza rufipogon*. Categories correspond to SASS values as follows: low (1,2); medium (3–5); high (7–8)

SASS	*aus*	*indica*	*tropical japonica*	*temperate japonica*	*aromatic*	*O. rufipogon*
Low	25.00	20.43	19.30	23.53	7.14	18.92
Medium	75.00	60.22	42.11	56.86	92.86	75.68
High	0.00	19.35	38.60	19.61	0.00	5.41

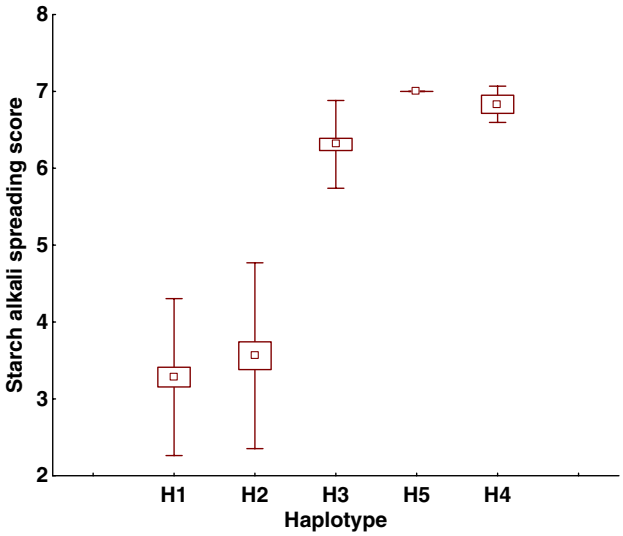

Fig. 4 Mean value of starch alkali spreading score (SASS) of each major *starch synthase IIa* (*SsIIa*) haplotype. SE is the standard error and SD is the standard deviation (□, mean; ☐, mean ± SE; ⌶, mean ± SD).

Table 6 *P* values for the association analyses between nonsynonymous single nucleotide polymorphisms (SNPs) and starch alkali spreading scores (SASS) by a general linear model

Site	*indica*	*tropical japonica*	*temperate japonica*	*Oryza rufipogon*
1	0.0348*	0.1291	0.0029*	0.3523
2	NA	0.2902	0.0097*	NA
3	1.10E-11**	4.96E-12**	7.12E-04**	NA

*, $P < 0.05$; **, Significant after Bonferroni correction.

Discussion

Patterns of nucleotide diversity at *SsIIa* exon 8

Previous studies have observed three nonsynonymous SNPs at *SsIIa* exon 8 in *O. sativa* (Umemoto & Aoki, 2005). Our study also detected these three SNPs in *O. sativa*. In addition, we also detected one unique nonsynonymous SNP in *indica* and three others in *O. rufipogon*. This finding suggests that more nonsynonymous SNPs at *SsIIa* exon 8 could be observed given more intense sampling of *O. sativa* and *O. rufipogon*. These newly reported nonsynonymous SNPs are present at very low frequency, potentially suggesting a relatively recent mutational origin.

As starch quality is one of the most important agronomic traits, and as the *SsIIa* exon 8 region has been implicated previously in contributing to starch quality differences between *indica* and *japonica* varieties (Umemoto *et al.*, 2002), selection at *SsIIa* exon 8 might be expected during domestication. In spite of this prediction, no strong statistical evidence of selection was found in this region in any cultivated rice group (see also Yu, 2009). Although an analysis of Fay and Wu's *H* in *tropical japonica* suggests an excess of derived variants, this pattern may potentially be explained by gene flow between cultivated rice and its wild ancestor. Crop–wild gene flow can potentially skew the proportions of what appear to be derived alleles in cultivated rice; evidence of genetic introgression between wild and cultivated rice has been observed in previous studies of rice (Sweeney & McCouch, 2007).

Haplotype network of *SsIIa* exon 8

We observed haplotypes which were detected in *O. sativa* but not in *O. rufipogon*. Haplotypes which have a high frequency in *O. sativa* but low frequency in *O. rufipogon* were also observed. Although these patterns would be unexpected for neutral variation in a crop and its wild progenitor, they are consistent with previous studies of haplotypes associated with important domesticated traits. The haplotypes associated with white pericarp color and nonshattering grains have also been observed in high frequency in cultivated rice, but not detected in the wild progenitor (Li & Sang, 2006; Sang & Ge, 2007; Sweeney *et al.*, 2007). There are two possible explanations for this observation with our *SsIIa* haplotypes. One possibility is that these haplotypes occur at low frequency in *O. rufipogon* but were missed in our sampling of the wild species. Another possibility is that these haplotypes originated through mutations in *O. sativa* after domestication. The second seems a less likely explanation. Asian rice was domesticated *c.* 11 000 yr ago (Khush, 1997), and it is unlikely that several mutations would have arisen in this gene during such a short evolutionary period of time.

The nonsynonymous SNP haplotype network contains one loop, which connects haplotype H5 to haplotypes H1 and H3. Loops in haplotype networks indicate ambiguities in genealogical relationships. These ambiguities can potentially reflect recurrent mutations, which create homoplasies in the network; alternatively, they may arise through intra-

Table 7 *P* values for association analyses between nonsynonymous single nucleotide polymorphisms (SNPs) and starch alkali spreading scores (SASS) by nested clade analyses

Site	Comparison	*indica*	*tropical japonica*	*temperate japonica*	*Oryza rufipogon*
1	H1 vs H2	1	1	0.430336	0.153407
2	H2 vs H4	NA	0.051782	0.102452	NA
3	H2 vs H3	0.001027**	0.000011**	0.020414*	NA

*, $P < 0.05$; **, Significant after Bonferroni correction. NA, not available.

genic recombination, in which case one haplotype is a recombinant of two others. Haplotype H5 was observed only in one *indica* and one *tropical japonica* individual, whereas the other haplotypes in the loop exist in *O. rufipogon* and occur at higher frequency than haplotype H5. This difference in haplotype frequencies is consistent with haplotype H5 having arisen as a result of recombination within exon 8 between SNPs 1 and 3.

Association between *SsIIa* nonsynonymous SNPs and SASS values in rice

Previous studies have indicated that starch quality, as assessed by SASS, is distinguishable between the *indica* subspecies (comprising *aus* and *indica* varieties) and *japonica* subspecies (comprising *tropical japonica*, *temperate japonica* and *aromatic* varieties) (Warth & Darabsett, 1914; Umemoto *et al.*, 2004; Nakamura *et al.*, 2005). Our study also detected some degree of SASS variation among rice groups, including differences between *tropical japonica* and *indica* varieties, as well as between *tropical japonica* and *aus* varieties (e.g. Tables 4, 5). However, these differences are of marginal statistical significance, and there is substantial overlap in SASS values among rice groups (Fig. 3). In contrast, when we focus specifically on the major nonsynonymous SNPs in the *SsIIa* exon 8 region, rather than on rice varieties, a very strong association becomes apparent (Table 6, Fig. 4). Both general linear models and nested clade analyses in our study indicate that SNP 3 is associated statistically with variation between low and high SASS values, and that SNP 1 shows no such correlation. These results are consistent with *SsIIa* enzyme expression experiments conducted in *E. coli* (Nakamura *et al.*, 2005). The study by Nakamura and colleagues also suggested that SNP 2 may be associated with SASS variation; our results do not support such an association. The fact that we did not observe a correlation with SNP 2 may reflect the small sample sizes for accessions carrying the H4 haplotype, which is characterized by this SNP. Haplotype H4 is present at highest frequency in *temperate japonica* varieties, and our dataset undersampled *temperate japonica* rice from the regions of China and Japan. Thus, expanded sampling of *temperate japonica* samples from these regions may further our understanding of the effect of SNP 2 on starch quality in rice.

Why does nonsynonymous SNP variation in the *SsIIa* exon 8 region affect SASS values in rice? This gene region encodes the C-terminus of the *SsIIa* enzyme and has been identified as critical for substrate binding and catalysis in maize (Nichols *et al.*, 2000; Gao *et al.*, 2004). It has been suggested that SNPs 2 and 3, which result in amino acid changes at codons 737 and 781 of the *SsIIa* enzyme, respectively, are likely to alter the enzyme in terms of both activity and starch granule association (Umemoto & Aoki, 2005). Such changes would be expected to affect levels of S-type amylopectin synthesis and associated SASS values in rice (Nakamura *et al.*, 2005).

In addition to *SsIIa* exon 8 SNPs, there clearly are other factors that affect the SASS phenotype, as there is no absolute association between *SsIIa* haplotypes and SASS values. SNPs in other regions of *SsIIa* and other genes involved in the starch synthesis pathway might also play a role in SASS variation. Currently, over 20 genes involved in the starch synthesis pathway have been identified (Myers *et al.*, 2000). For example, *Waxy* is another major gene in the rice endosperm starch synthesis pathway. A mutation in the intron 1 region of *Waxy* causes alternative splicing of *Waxy* and results in an undetectable level of granule-bound starch synthase (Wang *et al.*, 1995; Bligh *et al.*, 1998). To date, no other mutations which could affect the amylose level or amylopectin structure in rice have been identified definitively. Further studies with more rice accessions and more starch candidate genes will be necessary to fully understand starch quality variation in rice.

Evolution of starch quality during rice domestication

Starch quality is one of the most important agronomic traits in cereals, and selection on starch quality during domestication has been detected in *Zea mays* (Whitt *et al.*, 2002). Likewise, we expect starch quality differences between rice's wild ancestor *O. rufipogon* and derived cultivated rice. Domesticated Asian rice is believed to have arisen through at least two domestication events, with varieties in the *indica* and *japonica* subspecies corresponding to two independent crop origins (Londo *et al.*, 2006). In the context of this domestication history, our findings for the *SsIIa* exon 8 region suggest differences in selective pressures on starch quality during these two domestications. We found significant differences in SASS values between *tropical japonica* varieties and *O. rufipogon*, consistent with selection for more cohesive cooked rice grains in the domestication of *japonica* rice. In contrast, no starch quality differences were detected between *O. rufipogon* and *indica* rice. These results suggest either that little selection for changes in starch quality occurred during the *indica* domestication process, or that other components of starch quality, not measured by SASS, were favored.

Acknowledgements

This work was supported by a grant from the National Science Foundation (DBI 0701382) to B.A.S. The authors thank three anonymous reviewers for helpful comments on the manuscript.

References

Bandelt HJ, Forster P, Rohl A. 1999. Median-joining networks for inferring intraspecific phylogenies. *Molecular Biology and Evolution* **16**: 37–48.

Bhattacharya KR. 1979. *Gelatinization temperature of rice starch and its determination*. Laguna, Los Banos, Philippines: International Rice Research Institute.

Bligh HFJ, Larkin PD, Roach PS, Jones CA, Fu HY, Park WD. 1998. Use of alternate splice sites in granule-bound starch synthase mRNA from low-amylose rice varieties. *Plant Molecular Biology* **38**: 407–415.

Bradbury PJ, Zhang Z, Kroon DE, Casstevens TM, Ramdoss Y, Buckler ES. 2007. TASSEL: software for association mapping of complex traits in diverse samples. *Bioinformatics* **23**: 2633–2635.

Buléon A, Colonna P, Planchot V, Ball S. 1998. Starch granules: structure and biosynthesis. *International Journal of Biological Macromolecules* **23**: 85–112.

Doyle JJ, Doyle JL. 1990. Isolation of plant DNA from fresh tissue. *Focus* **12**: 13–15.

Fay JC, Wu CI. 2000. Hitchhiking under positive Darwinian selection. *Genetics* **155**: 1405–1413.

Gao Z, Keeling P, Shibles R, Guan HP. 2004. Involvement of lysine-193 of the conserved 'K-T-G-G' motif in the catalysis of maize starch synthase IIa. *Archives of Biochemistry and Biophysics* **427**: 1–7.

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

He Y, Han Y, Jiang L, Xu C, Lu J, Xu M. 2006. Functional analysis of starch-synthesis genes in determining rice eating and cooking qualities. *Molecular Breeding* **18**: 277–290.

Khush GS. 1997. Origin, dispersal, cultivation and variation of rice. *Plant Molecular Biology* **35**: 25–34.

Kudo M. 1968. Genetical and thremmatological studies of characters, physiological or ecological, in the hybrids between ecological rice groups. *Bulletin of the National Institute of Agricultural Sciences Series D* **19**: 1–8.

Li C, Sang T. 2006. Rice domestication by reducing shattering. *Science* **311**: 1936–1939.

Librado P, Rozas J. 2009. DnaSP v5: a software for comprehensive analysis of DNA polymorphism data. *Bioinformatics* **25**: 1451–1452.

Little R, Hilder G, Dawson E. 1958. Differential effect of dilute alkali on 25 varieties of milled white rice. *Cereal Chemistry* **35**: 111–126.

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*. *Proceedings of the National Academy of Sciences, USA* **103**: 9578–9583.

Marchini J, Cardon LR, Phillips MS, Donnelly P. 2004. The effects of human population structure on large genetic association studies. *Nature Genetics* **36**: 512–517.

Myers AM, Morell MK, James MG, Ball SG. 2000. Recent progress toward understanding biosynthesis of the amylopectin crystal. *Plant Physiology* **122**: 989–997.

Nakamura Y, Francisco PB, Hosaka Y, Sato A, Sawada T, Kubo A, Fujita N. 2005. Essential amino acids of starch synthase IIa differentiate amylopectin structure and starch quality between *japonica* and *indica* rice varieties. *Plant Molecular Biology* **58**: 213–227.

Nichols DJ, Keeling PL, Spalding M, Guan HP. 2000. Involvement of conserved aspartate and glutamate residues in the catalysis and substrate binding of maize starch synthase. *Biochemistry* **39**: 7820–7825.

Rozas J, Sànchez-DelBarrio JC, Messequer X, Rozas R. 2003. DnaSP, DNA polymorphism analyses by the coalescent and other methods. *Bioinformatics* **19**: 2496–2497.

Sang T, Ge S. 2007. The puzzle of rice domestication. *Journal of Integrative Plant Biology* **49**: 760–768.

Sweeney M, McCouch S. 2007. The complex history of the domestication of rice. *Annals of Botany* **100**: 951–957.

Sweeney MT, Thomson MJ, Cho YG, Park YJ, Williamson SH, Bustamante CD, McCouch SR. 2007. Global dissemination of a single mutation conferring white pericarp in rice. *PLoS Genetics* **3**: e133.

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.

Templeton AR, Boerwinkle E, Sing CF. 1987. A cladistic-analysis of phenotypic associations with haplotypes inferred from restriction endonuclease mapping. 1. Basic theory and an analysis of alcohol-dehydrogenase activity in *Drosophila*. *Genetics* **117**: 343–351.

Tian ZX, Qian Q, Liu QQ, Yan MX, Liu XF, Yan CJ, Liu GF, Gao ZY, Tang SZ, Zeng DL *et al.* 2009. Allelic diversities in rice starch biosynthesis lead to a diverse array of rice eating and cooking qualities. *Proceedings of the National Academy of Sciences, USA* **106**: 21760–21765.

Umemoto T, Aoki N. 2005. Single-nucleotide polymorphisms in rice starch synthase IIa that alter starch gelatinisation and starch association of the enzyme. *Functional Plant Biology* **32**: 763–768.

Umemoto T, Aoki N, Lin HX, Nakamura Y, Inouchi N, Sato Y, Yano M, Hirabayashi H, Maruyama S. 2004. Natural variation in rice starch synthase IIa affects enzyme and starch properties. *Functional Plant Biology* **31**: 671–684.

Umemoto T, Yano M, Satoh H, Shomura A, Nakamura Y. 2002. Mapping of a gene responsible for the difference in amylopectin structure between *japonica*-type and *indica*-type rice varieties. *Theoretical and Applied Genetics* **104**: 1–8.

Wang ZY, Zheng FQ, Shen GZ, Gao JP, Snustad DP, Li MG, Zhang JL, Hong MM. 1995. The amylose content in rice endosperm is related to the posttranscriptional regulation of the waxy gene. *Plant Journal* **7**: 613–622.

Warth F, Darabsett D. 1914. Disintegration of rice grains by means of alkali. *Bulletin of Agricultural Research Institute* **38**: 1–9.

Watterson GA. 1975. On the number of segregating sites in genetical models without recombination. *Theoretical Population Biology* **7**: 256–276.

Whitt SR, Wilson LM, Tenaillon MI, Gaut BS, Buckler ES. 2002. Genetic diversity and selection in the maize starch pathway. *Proceedings of the National Academy of Sciences, USA* **99**: 12959–12962.

Yu G. 2009. Evolution of endosperm starch synthesis pathway genes in the context of rice (*Oryza sativa*) domestication. PhD thesis, Washington University in St. Louis, St. Louis, MO, USA.

Supporting Information

Additional supporting information may be found in the online version of this article.

Fig. S1 Representative phenotypes for starch alkali spreading score (SASS) analysis.

Fig. S2 Distribution of variance of starch alkali spreading score (SASS) values within rice samples.

Fig. S3 Haplotype network for polymorphisms detected in the *starch synthase IIa* (*SsIIa*) exon 8 region.

Table S1 Collections of *Oryza rufipogon* and *O. sativa* from the International Rice Research Institute (IRRI) used in this study

Table S2 Collections of *Oryza rufipogon* from the field used in this study

Please note: Wiley-Blackwell are not responsible for the content or functionality of any supporting information supplied by the authors. Any queries (other than missing material) should be directed to the *New Phytologist* Central Office.