Supporting Information

Asano et al. 10.1073/pnas.1019490108

SI Methods

Plasmid Construction, Transformation, and Growth Conditions. To construct *N-SD1* and *K-SD1* in a pBI-Hm12 vector, the *SD1* genes of Nipponbare and Kasalath, which were cloned in BAC vectors, were digested with *Xba*I and inserted into pBluescript SK II(+). The resulting clone was redigested with *Xba*I and end-filled. The 10-kb fragment was then cloned into the *Sma*I site of the binary vector pBI-Hm12. The binary vectors were then introduced into *Agrobacterium tumefaciens* strain EHA 101 by electroporation. Rice transformation was performed as described previously (1) using Nipponbare. Transgenic plants were selected on a medium containing 50 mg/L of hygromycin. Hygromycin-resistant plants were transplanted to soil and grown at 30 °C under a 16-h light/8-h dark photoperiod.

Recombinant Protein Production and Enzyme Assay. The full coding region of both *SD1* alleles was obtained by PCR using cDNA of rice seedling as the template. The PCR products were ligated into the pCR4 Blunt-TOPO vector (Invitrogen), and their nucleotide sequences were confirmed. To construct SD1-GQ and SD1-ER, PCR was performed against the full-length N-SD1 (SD1-EQ) cDNA with set of mutagenized primers (Table S5) corresponding to each mutation. The parental methylated and hemimethylated DNA in the PCR reaction mixture was digested with *Dpn*I, and the mutated *SD1* cDNA that could not be digested with *Dpn*I was transformed into *Escherichia coli* strain XL10-Gold (Agilent Technology). After confirmation of nucleotide sequences, these products were inserted into the *Eco*RI site of pMAL-c2X (New England Biolabs) and transformed into the *E. coli* strain BL21 (DE3) pLyS Rosetta-gami2 (Novagen). A fresh overnight culture (10 mL) of *E. coli* was added to 500 mL of LB medium (5 g/L of NaCl, 5 g/L of yeast extract, and 10 g/L of tryptone) containing ampicillin (50 mg/L). The culture was incubated at 37 °C until midlog phase, after which isopropyl β-D-thiogalactoside (0.4 mM) was added, and the culture was incubated for another 20 h at 16 °C. The cells were collected by centrifugation and resuspended in 10 mL of 1× PBS (pH 7.4) and 10 mM β-mercaptoethanol. The suspension was kept at −80 °C overnight, thawed on ice, and sonicated. Soluble proteins were collected by centrifugation and used as crude extracts for the enzyme assay. The crude extracts were incubated at 30 °C with a GA substrate (GA_{53}, 875 nM) in 1 mL of 100 mM Tris (pH 7.9) containing 4 mM ascorbic acid, 4 mM 2-oxoglutaric acid, 0.5 mM $FeSO_4$, 4 mM DTT, 2 mg/mL BSA, and 1 mg/mL catalase. The reactions were stopped after overnight incubation by adding 125 μL of acetic acid. Next, 10 ng of deuterated [17,17-$^2H_2$] GA_{20} (internal standard) was added. All GAs used in this study were purchased from Dr. L.N. Mander (Australian National University, Australia). The solution was passed through a 1-mL C18-HD high-performance extraction cartridge (Empore). After the column was washed with 3 mL of water, substances retained on the column were eluted with 500 μL of methanol. The methanol eluate was evaporated with dry N_2 gas. After TMSi (trimethylsilyl) ester–TMSi ether derivatization with MSTFA (*N*-methyl-*N*-trimethylsilyl-trifluoroacetamide), the GA_{20} was analyzed by gas chromatography-selected ion monitoring–mass spectrometry.

GC-MS. GC-MS analysis was carried out with a JEOL JMS-K9 mass spectrometer and an Agilent Technology 6890N gas chromatograph. The trimethylsilylated derivatives (TMSi ester–TMSi ethers) were injected at 250 °C into an HP-5 ms column (0.32 mm i.d. × 30 m, 0.25-μm film thickness; Agilent Technology). The column temperature was kept at 100 °C for 2 min, and then increased to 260 °C at a rate of 30 °C/min and held for 1 min, and was followed by an increase at a rate of 30 °C/min to 300 °C. The flow rate of the carrier gas (He) was 1.5 mL/min. Spectra were acquired by scanning at *m*/*z* 50–750 at 70 eV.

Coalescent Simulation and Testing for Selection. To test for selection with our SNP data, we examined whether the observed level of genetic variation in *japonica* was significantly smaller than would be expected under neutrality. Two major factors cause a reduction in genetic diversity in a domesticated species: selection and bottleneck. To distinguish these two factors, we performed a statistical test by coalescent simulations with a commonly used two-population model of domestication (Fig. S3). The model assumes that there is a large stable population with constant size N_0, representing the wild progenitor species (*O. rufipogon* in this study). The founder population of the domesticated species (*O. sativa* ssp. *japonica* or *indica* in this study) with size N_1 was formed T_d generations ago. It is assumed that T_e generations ago, the size of this population changed to N_2. The time from T_d to T_e represents the duration of the bottleneck, and N_1 and this duration, $T_1 = T_d - T_e$, play major roles in determining the expected reduction of the nucleotide diversity in domesticated species. Previous theoretical studies demonstrated that the N_1-to-T_1 ratio accurately summarizes the expected reduction of genetic variation, so that $k = N_1/T_1$ represents the severity of bottleneck (2, 3). Here k is positively correlated with the expected relative genetic diversity in the domesticated species to that of the progenitor; that is, as k decreases, the genome-wide nucleotide diversity is further reduced by the bottleneck effect.

Using this model, we set the parameters to be suitable for the domestication process of *japonica*. Because the genome-wide average of synonymous nucleotide diversity is ∼0.005 in *O. rufipogon* (4), we assumed $4N_0\mu = 0.005$, where μ is the mutation rate per site per generation. If we assume that $\mu = 10^{-8}$ (5, 6), then N_0 is 125,000, which is similar to previous estimates (4, 7). Estimating the current effective population size of domesticated species is difficult; thus, we assumed that $N_2 = N_0$. However, additional simulations with a wide range of N_2 demonstrated that N_2 has very little effect on the simulation results, a finding in agreement with previous studies (2, 7). For the time of the domestication event, we used several values (T_d = {7,500, 9,000, 10,000, 12,000}) to cover the possible range reported in the archaeological literature (8–11). The severity of the bottleneck for the *japonica* domestication process was estimated as k = 0.445–0.9 (4, 7). To conservatively test for the effect of domestication selection, we used $k = 0.445$, which predicts that the level of nucleotide diversity in the domesticated species is roughly 15% that of the progenitor. Thus, an observed diversity in *japonica* significantly lower than predicted under this bottleneck model suggests the possibility of selection.

To test this prediction, we performed coalescent simulations using Hudson's ms software (12) with a few modifications. First, because both species are highly selfing, the algorithm of Nordborg and Donnelly (13) was incorporated. Second, Hudson's ms software was modified such that a specific selfing rate is given in each population, because the selfing rates differ between cultivated and wild species. The selfing rate of *japonica* and *indica* was assumed to be 95% (14). The selfing rate for annual plants in *O. rufipogon* may be as high as that of *japonica* (95%), but may be low in perennial plants (40%) (15, 16). Because our sample is a mixture of annual and perennial individuals, we used the weighted

average (60%) in our simulation. We further incorporated the algorithm of Nordborg (17) to adjust the recombination rate with the selfing rate in each population. The recombination rate was assumed to be 4 cM/Mb across the genome (18). Then, with the demographic setting described above, we simulated a large number of SNP patterns in a 400-kb region, which corresponds to the region from genes 3–15 in Fig. 3. For each replication, we computed the nucleotide diversities for the cultivated and progenitor species, denoted by π_J' and π_R', respectively. We compared with the observed synonymous nucleotide diversities for *japonica* and *O. rufipogon* ($\pi_{J\_obs} = 0.0000536$; $\pi_{R\_obs} = 0.00524$). It is important to note that all sequenced cultivars and accessions were treated as haplotypes, because they underwent several generations of inbreeding and had almost no heterozygous sites. As a summary statistic for the test for selection, we used the nucleotide diversity in the cultivated species conditional on that of the progenitor species being close to π_{R_obs}. In practice, we collected 10,000 simulation replications that satisfied $|\pi_R' - \pi_{R\_obs}| < 0.025\ \pi_{R\_obs}$, from which we obtained the null distribution of π_J'. It turned out that the proportion of the simulation replications with $\pi_J' < \pi_{J\_obs}$ was $P < 0.01$ for all values of T_d. When we used $k = 0.9$, we obtained a smaller P value ($P < 0.001$). We also tested neutrality in *indica*. The observed nucleotide diversity in *indica*, π_{I_obs}, was 0.00243. The k value in *indica* was estimated to be 0.471–1.5 (4, 7), and for this range of k, neutrality was not rejected.

1. Hiei Y, Ohta S, Komari T, Kumashiro T (1994) Efficient transformation of rice (*Oryza sativa* L.) mediated by *Agrobacterium* and sequence analysis of the boundaries of the T-DNA. *Plant J* 6:271–282.
2. Eyre-Walker A, Gaut RL, Hilton H, Feldman DL, Gaut BS (1998) Investigation of the bottleneck leading to the domestication of maize. *Proc Natl Acad Sci USA* 95:4441–4446.
3. Wright SI, et al. (2005) The effects of artificial selection on the maize genome. *Science* 308:1310–1314.
4. Caicedo AL, et al. (2007) Genome-wide patterns of nucleotide polymorphism in domesticated rice. *PLoS Genet* 3:1745–1756.
5. Wolfe KH, Li WH, Sharp PM (1987) Rates of nucleotide substitution vary greatly among plant mitochondrial, chloroplast, and nuclear DNAs. *Proc Natl Acad Sci USA* 84:9054–9058.
6. Gaut BS, Clegg MT (1991) Molecular evolution of alcohol dehydrogenase 1 in members of the grass family. *Proc Natl Acad Sci USA* 88:2060–2064.
7. Gao LZ, Innan H (2008) Nonindependent domestication of the two rice subspecies, *Oryza sativa* ssp. *indica* and ssp. *japonica*, demonstrated by multilocus microsatellites. *Genetics* 179:965–976.
8. Sweeney M, McCouch S (2007) The complex history of the domestication of rice. *Ann Bot (Lond)* 100:951–957.
9. Kovach MJ, Sweeney MT, McCouch SR (2007) New insights into the history of rice domestication. *Trends Genet* 23:578–587.
10. Sang T, Ge S (2007) Genetics and phylogenetics of rice domestication. *Curr Opin Genet Dev* 17:533–538.
11. Fuller DQ, et al. (2009) The domestication process and domestication rate in rice: Spikelet bases from the Lower Yangtze. *Science* 323:1607–1610.
12. Hudson RR (2002) Generating samples under a Wright–Fisher neutral model of genetic variation. *Bioinformatics* 18:337–338.
13. Nordborg M, Donnelly P (1997) The coalescent process with selfing. *Genetics* 146: 1185–1195.
14. Oka HI (1988) *Origin of Cultivated Rice* (Elsevier, Tokyo, Japan).
15. Oka HI, Morishima H (1967) Variation in the breeding systems of a wild rice, *Oryza perennis*. *Evolution* 21:249–258.
16. Morishima H, Barbier P (1990) Mating system and genetic structure of natural populations in wild rice *Oryza rufipogon*. *Plant Species Biol* 5:31–39.
17. Nordborg M (2000) Linkage disequilibrium, gene trees and selfing: An ancestral recombination graph with partial self-fertilization. *Genetics* 154:923–929.
18. Matsumoto T, et al.; International Rice Genome Sequencing Project (2005) The map-based sequence of the rice genome. *Nature* 436:793–800.
19. Tajima F (1983) Evolutionary relationship of DNA sequences in finite populations. *Genetics* 105:437–460.
20. Watterson GA (1975) On the number of segregating sites in genetical models without recombination. *Theor Popul Biol* 7:256–276.
21. Tajima F (1989) Statistical method for testing the neutral mutation hypothesis by DNA polymorphism. *Genetics* 123:585–595.

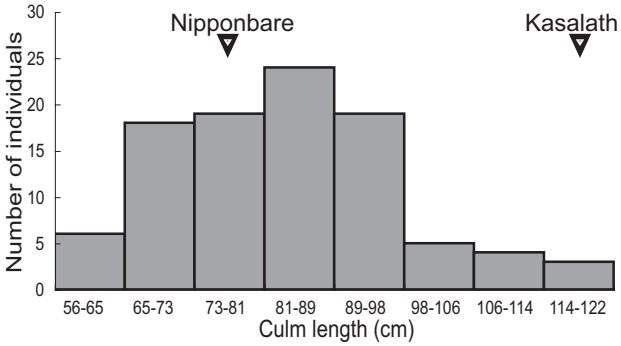

Fig. S1. Distribution frequency of CL in 98 BILs with respect to the parental lines Nipponbare and Kasalath.

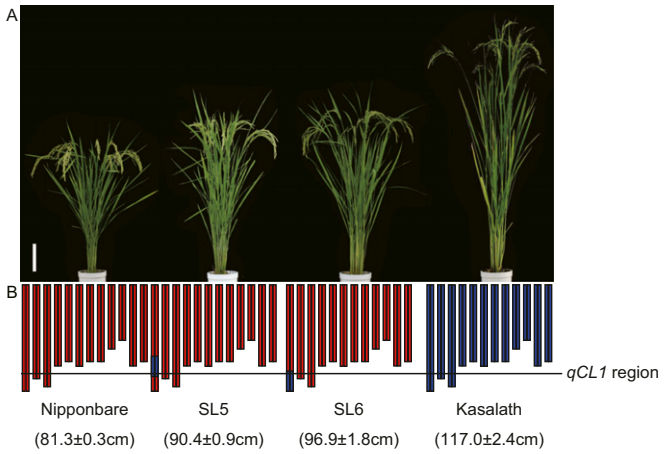

Fig. S2. Confirmation of QTL analysis by comparing CL of near-isogenic lines. Gross morphology (*A*) and graphical genotypes (*B*) are shown for each line, from left to right, Nipponbare, SL5, SL6, and Kasalath. The red and blue bars indicate Nipponbare and Kasalath chromosomal segments, respectively. (Scale bar in *A*: 20 cm.)

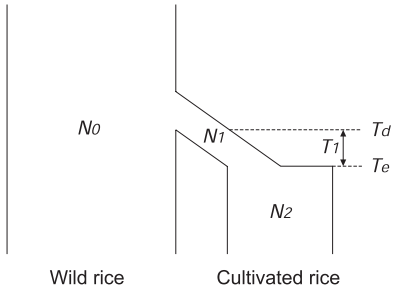

Fig. S3. Schematic representation of the coalescent models used in the simulation analyses.

Other Supporting Information Files

Table S1 (DOCX)
Table S2 (DOCX)
Table S3 (DOCX)
Table S4 (DOC)
Table S5 (DOCX)