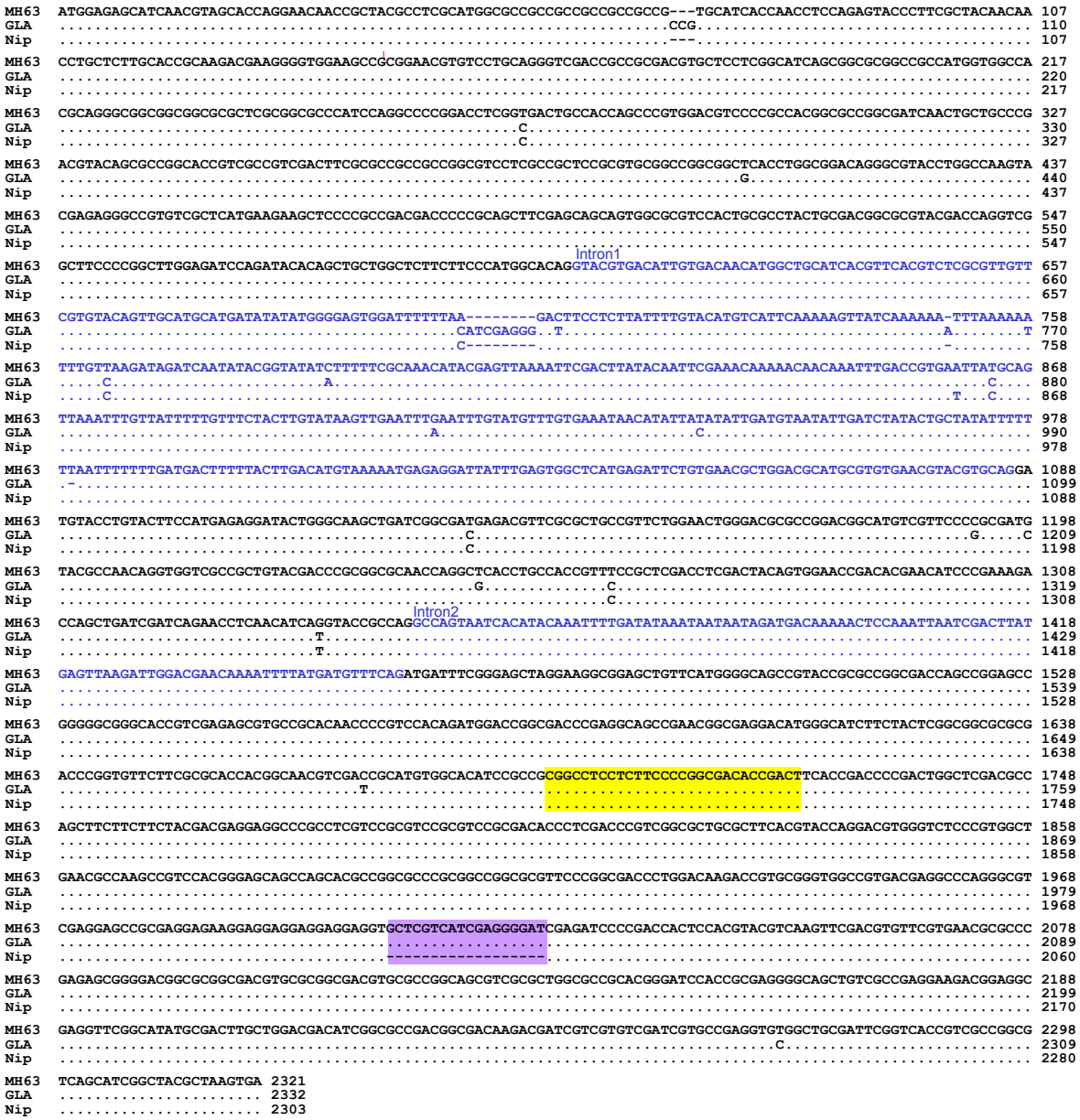
Supplemental Data. Yu et al. (2008). Independent Losses of Function in a Polyphenol Oxidase in Rice - Differentiation in Grain Discoloration between Subspecies and the Role of Positive Selection in Domestication.

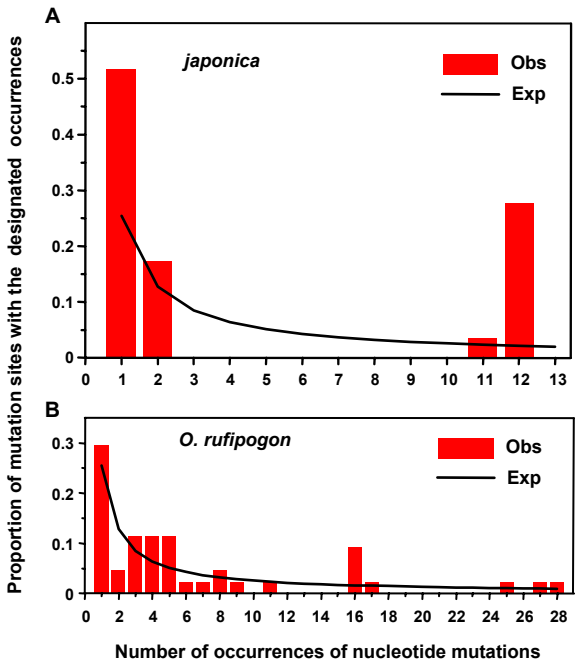


Supplemental Figure 1. Alignment of *Phr1* cDNA Sequences of *indica*-type Rice MH63 and GLA *with a japonica*-type Rice Nipponbare (Nip).

A dot stands for an identical nucleic acid base and a dash for a gap. The nucleic acid positions with 18-bp or 29-bp deletion in *japonica* cultivars are highlighted in pink and yellow, respectively. The red triangle indicates the position of 1-bp insertion in Tx36, a *japonica* cultivar. The blue letters indicate the intron sequence.

Supplemental Figure 2. Alignment of PHR1 with its homologous proteins identified in the rice genome.

Alignment of Phr1 with its homologous Phr1L1, Phr1L2 and Phr1L3 proteins. Blue boxes indicate the copper-binding domains A and B, with the red color showing the changed conserved amino acid residues in Phr1L1, Phr1L2 or Phr1L3.



Supplemental Figure 3. The Frequency Spectrum of *Phr1* Mutations**.**

(A) The frequency spectrum of *Phr1* mutations in the sample of 14 *japonica* samples. Each bar indicates the proportion of variant sites where the derived mutation is of the occurrence frequency indicated on the X-axis. *O. barthii* was used as the outgroup to infer ancestral character states of the *O. sativa*/*O. rufipogon* complex. The character states of segregating sites within *O. barthii* were resolved by checking the counterparts in more distant *O. alta* and *O. officianlis*. **(B)** The frequency spectrum of mutations in the sample of 27 *O. rufipogon* lines. Accessions of the ∆18/+ or ∆29/+ genotype (see Table 1) were counted as having two sampled genes, hence, bringing the sample size to 29.

Supplemental Table 1. PCR-based Molecular Markers Developed in This Study.

Name	BACs	Forward sequence	Reverse sequence
S2970	AL662970	GAGGTAAGGCATTGATGTGC	CCAAGAGCAGTTTGCATTGC
S6660	AL606660	CTTTGCTGTGTCCTCTGTGC	GCAATGCTGTCAACCACTCC
P80	AL606645	CGTCGACATTAACGACGATTCG	ATGCAGATGCAGGTTGCAGCTG
P168	AL606645	AAGGCTGAGATATGTGAACACC	TGCCGCTACGCGTGACGC
S6669	AL606669	CAAAGCATGAGCTACTCTGCG	ATTGCATTCACGAGTGCTCG
S2988	AL662988	AAGTACTCTCCCGTTTCAAA	CCTCCCATAAAAATCTTGTC
S100	AL606638	GGCACATCTCAATACTAACAGT	GAACAAAGGTTAGGTCCATGGT
S115	AL606445	AGCAAGCAGCCAAGCAGACC	TGCATATCACGCACATTGGC
F217	AL606645	GGGATGGAGGGAATATGTGTG	GACGTGAACGTGATGCAGCC
Cxp3	AL606645	ACCTGGCCAAGTACGAGAGG	CCTTCCTAGCTCCCGAAATCA
Cxp5	AL606645	CAACATCATGTACCGCCAGG	CGTCGAACTTGACGTACGTGG
c5311	AL606645	CTGCGCTTCACGTACCAGG	CCGTTATCCTGCGGTTGTG
pSTS18	AL606645	ACGGGAGCAGCCAGCAC	CGTCGAACTTGACGTACGTGG
pSTS29	AL606645	CACCACGGCAACGTCGAC	AAGAAGCTGGCGTCGAGC
P1	AL606645	gagctcCCGATATGATCAACGACGTTCG	aagcttCGTATGTGACATGCAGGAAGC
Actin		CCTCGTCTCGACCTTGCTGGG	GAGAACAAGCAGGAGGACGGC

The pair of the pSTS18 forward primer and the primer (TCACTTAGCGTAGCCGATG) was used for RT-PCR analysis. P1 was applied to construct the complementation plasmid and the entire *Phr1* was sequenced with primers F217, cxp3, cxp5 and c5311. Primer P1 forward and primer p13PR (aagcttCCAGCAGCTGTGTATCTGGATC) were used to construct plasmid *p13p*.

Supplemental Table 2: Summary of Samples for Phr1 Study

Species	Taxon	TL	TS	Sequenced cultivars/lines	Genotype	PHR*
japonica	AA	35	14	Nipponbare[1], R4[1], Tx3[1], Jl506, Qfn, Saon, CJ06, Tzn, Zh11, Zhen5125, w11	Δ18	Negative
				Sun, Jnxn	Δ29	Negative
				Tx36[2]	1-bp Ins	Negative
			21 [TG]	Cj, Lg5, Bll, Ql5, Lzh, Xb, Tlb, Tgs, Ddt, Bws, Zgt, Sjq, Xby, Jzg, Dzg, Dn, Dfh1, Wnz, Yg3, Nhh, Zg7308	Δ18	Negative
indica	AA	20	7	93-11, GLA, MH63, Tx7, Tx9, dl[1]	+	Positive
				k2406	+/Δ18	Positive
			13[TG]	Nt, IR36, Mzt, Dzx, Hgx, Nbx, Yzg, Prq, Dax, Zyd, Ntz, Ntn, Xwx3	+	Positive
***O. glaberrim**a*	AA	1	1	G7[3]	+	Positive
O. nivara	AA	10	2	G12[4],A8[2]	+	Positive
O. rufipogon	AA	523	27	A3, w1, G25, G30, G40[4], G7113, G7134, G7232, G7251, G9014, G01037, G01049, G01054, G01060, G01067, G01084, G02016, G02068, G02104, G02115, G02177	+	Positive
				w0509[2], w1125[4]	Δ18	na
				w1862[6], w1727[6]	Δ29	na
				w0154[2]	+/Δ18	na
				w0634[7]	+/Δ29	na
O. barthii	AA	67	2	w6[2], w0009[2]	+	Positive[#]
O. longistaminata	AA	5			+	na
O. glumaepatula	AA	81			+	na
O. meridionalis	AA	16			+	na
O. panctata	BB	2			+	na
O. officinalis	CC	2	1	yaoyong[2]	+	Positive
O. alta	CCDD	2	1	G52030[2]	+	Positive
O. latifolia	CCDD	1			+	na
O. granulata	GG	1			+	na

The original localities of those lines collected outside of China are: [1], Japan; [2], not clear; [3], Guinea; [4], India; [5], Cambodia; [6], Thailand; [7], Burma. The "+" represents the functional *Phr1* without 18-bp (Δ18) or 29-bp (Δ29) deletion. na, not available; PHR, phenol reaction; TL, total number of lines screened for the 18-bp (Δ18) or 29-bp (Δ29) deletion; TS, total number of lines whose *Phr1* were sequenced. [TG], total number of lines whose *Phr1* were only genotyped with Δ18 and Δ29 specific primers. "*" PHR phenotype of the sequenced lines. "#" One line tested is PHR-positive and another is not determined due to unavailability of its seeds.

Supplemental Results

Additional considerations on the origin of the H29 haplotypes

The suggestion of introgression from *japonica* to *O. rufipogon* was not based solely on the low frequency in the latter. There are two other equally important considerations. First, while it is true that H29 comprises 3 different haplotypes as shown in Figure 5, the genealogical depth is very small. These 3 haplotypes differ from each other by 1 or 2 bp on a sequence of 2345 bp long. In other words, since the emergence of Δ29, some of the haplotypes acquired one single new mutation. The H29s cluster is indeed not old at all. Second, had Δ29 originated in *O. rufipogon* before domestication, Δ29 should have been recombined into many different haplotypes, given the outcrossing ability of this species (This is true even if there is selection against Δ29, keeping it low in *O. rufipogon*.). The very small genealogy of H29 haplotypes in *O. rufipogon* suggests that Δ29 could not have existed in that species for very long.

Impact of Breeding Structure under Domestication Test

In general, the population genetic theories, such as the coalescence process and various test statistics (Tajima's D, Fay and Wu's H), can be applied to all breeding systems. We may see an inbred system as the equivalent of a (largely asexual) haploid population like *E. coli*. The theories are applicable to both, as well as the fully outbred diploid species. Where inbreeding may affect the application and interpretation of the theories are mainly the two areas. First, if there is heterosis at a locus where the heterozygote advantage is strong, then breeding structure has to be explicitly modeled. Fortunately, this is not a concern here as Phr1 locus does not show overdominance.

Second, inbreeding affects the population recombination rate, 4Nc (where N is effective population size and c is recombination rate). For example, the strength of hitchhiking is diminished as 4Nc increases (thus, hitchhiking is complete if c = 0). In a strongly but incompletely selfing species, c may be much smaller than that determined in genetic experiments. This is because a crossover event does not shuffle the genomes if the chromosomes are often identical by descent.

Importantly, selfing does not bias the results of the statistical tests. In other words, the perturbation of frequency spectrum 10 kb away from a sweeping variant in a selfing species may be the same as the perturbation only 1 kb away in an outcrossing species. Positive results from, say, Fay and Wu's H test would be equally biased or unbiased for selfing and outcrossing species. In our study, the contribution of inbreeding to hitchhiking strength is nevertheless modest as can be seen in the small amount of loss in genetic diversity near the *Phr1* gene. There is detectable reduction but, had hitchhikng been strong, the level of genetic diversity should be near zero. For elaboration, we discussed the population genetic implication of small 4Nc in selfing species in Lu et al (2006).

In the context of our study, the actual selfing rate of *O. rufipogon* is not crucial. We were trying to explain why *O. rufipogon* sequences show weaker recombination than cultivar sequences. Unless H18s in *O. rufipogon* are much younger than in the cultivars, the pattern makes no sense as explained below. The relevant parameter is 4Nc where N is the effective population size and c is the (effective) recombination rate. Since N is larger in *O. rufipogon* by about 50% and c should be at least 5 fold larger than in the cultivars, its 4Nc should be larger by at least 10 fold. (For calculating c, we assume 20-45% outcrossing in *O. rufipogon* as Reviewer pointed out; in our literature serach, the reported rate is 20-60%.) Hence, the observed low rate of recombination in H18s in *O. rufipogons* (relative to cultivars) suggests that the existence of H18s in the wild may have been substantially more recent than in the cultivars.

P-value Calculation and *DH* Test

Since the samples chosen for sequencing were not random, p-value calculation has to be done by resampling (as presented in Materials and Methods). In 100 rounds of resampling, the number of rounds in which the results were significant at 5% is, respectively, 67 for *japonica* (5 were expected), 2 for *indica* and 0 for *O. rufipogon*.

The reason *DH* is robust against demographical influences is explained in Zeng et al. (2006; 2007). Briefly, *D* and *H* are complementary tests for demographical factors. *D* is very sensitive to population growth but *H* is completely insensitive to it. *H* is mildly sensitive to population reduction and strong population subdivision whereas *D* is not sensitive to them at all. The nature of the compound test is that it is sensitive only to factors to which both component tests are sensitive. Hence, *DH* is rather insensitive to most demographical factors.

References

Fay, J.C., and Wu, C.I. (2000). Hitchhiking under positive Darwinian selection. Genetics. 155: 1405-1413.

Lu, J., Tang, T., Tang, H., Huang, J., Shi, S., and Wu, C.I. (2006). The accumulation of deleterious mutations in rice genomes: a hypothesis on the cost of domestication. Trends Genet. 22: 126-131.

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

Zeng, K., Shi, S., and Wu, C.I. (2007). Compound tests for the detection of hitchhiking under positive selection. Mol Biol Evol. 24: 1898-1908.

Zeng, K., Fu, Y.X., Shi, S., and Wu, C.I. (2006). Statistical tests for detecting positive selection by utilizing high-frequency variants. Genetics. 174: 1431-1439.