Supporting Information

Molina et al. 10.1073/pnas.1104686108

SI Text

DNA Extraction and Sequencing. DNA was extracted using a modified cetyl trimethylammonium bromide (CTAB) protocol (1), and the resequencing strategy consisted of PCR amplification and direct sequencing of ∼500-bp fragments from protein-coding genes spaced at ∼100-kb intervals on chromosomes 8, 10, and 12. Primers were anchored in exons and designed to span introns using Primer 3 (2) based on the Nipponbare reference genome (Rice Genome Annotation Project v. 6.0) (3). PCR amplification and sequencing were conducted by Cogenics, and subsequent editing was performed in a semiautomated process that incorporated Phred/Phrap (4, 5) for contig assembly and quality scoring and PolyPhred (6) for identification of heterozygotes. All PolyPhred quality scores <25 were assigned as ambiguous. Heterozygote base calls were checked, and International Union of Pure and Applied Chemistry (IUPAC) degenerate symbols were incorporated where appropriate. Multiple sequence alignments were generated with MUSCLE (7), and singletons and triallelic SNPs in the alignments were confirmed. Data is available at http://puruggananlab.bio.nyu.edu/Rice_data/.

Population Stratification and Summary Diversity Measures. The memberships of landrace accessions were evaluated by population structure analysis using Structure v.2.3.1 (8) on a subset of the SNP data using the procedure described in Caicedo et al. (9) but with a longer run length (300-k iterations; 100 k as burn-in) for each replicate. Summary population genetic statistics (10–12) were computed using the libsequence package (13).

Selection Mapping. We used two different methods to map selective sweeps related to rice domestication from *Oryza rufipogon* based on local reductions in diversity and the multipopulation allele frequency spectrum (AFS) used in the demographic inference.

Local reduction in diversity. To identify regions showing a local reduction in genetic diversity, we calculated the nucleotide diversity for each sequenced fragment. Nucleotide diversity was calculated as (Eq. S1)

$$\widehat{\theta}_\pi = \sum_{i=1}^{S}\left(1-\sum_{j=1}^{2}\frac{k_{ij}(k_{ij}-1)}{n_i(n_i-1)}\right), \qquad \textbf{[S1]}$$

where S is the number of segregating sites found in the fragment, k_{ij} is the count of allele j at site i, and n_i is the sample size for site i. Regions where there are k consecutive markers with $\theta_\pi = 0$ correspond to runs of monomorphic fragments with run length k ($k \geq 1$). From those, candidate regions were defined as regions where the total run length is in the top 5% of run lengths from the empirical distribution for the respective population (Fig. S2); this corresponded to a run length of four for *indica* and run length of eight for tropical *japonica*. We allowed for gaps of low variation in a candidate region by merging runs if they were separated by a fragment with only one segregating site.

Multipopulation AFS. We use the multipopulation AFS to derive a measure of how well the observed distribution of allele frequencies at a particular locus (the local AFS) fits the expected distribution derived from the AFS of the full dataset (the global AFS). The method is similar to the G2D test of Nielsen et al. (14), with two notable differences. First, we include all three study populations (i.e., our AFS is 3D), and second, we use a Poisson likelihood to calculate the composite likelihood ratio by modeling the expected entries of the AFS as independent Poisson variables (15).

For a given region of interest, we first obtain the local AFS $X = (X_1, X_2, \ldots, X_n)$, where X_I is the count of alleles for the ith category in the 3D AFS and n is the total number of categories. The categories correspond to the cells of the 3D matrix that makes up the AFS. For the following calculations, we consider all categories except sites that are fixed for the ancestral allele in all three populations [i.e., the entry $a(0,0,0)$ in the AFS]. Missing data are treated as in the demographic inference by removing sites with too few samples in each population and projecting down sites with larger sample sizes. Assuming that each entry in the AFS is an independent Poisson variable, a composite Poisson likelihood of the data can be calculated as (Eq. S2)

$$CL(X) = \prod_{i=1}^{n}\frac{e^{-\lambda_i}\lambda_i^{X_i}}{X_i!}. \qquad \textbf{[S2]}$$

We can then form a composite likelihood ratio (CLR) test comparing the observed local AFS to the expected local AFS as (Eq. S3)

$$CLR = \log\frac{\Pr(Data\,|\,H_A)}{\Pr(Data\,|\,H_0)} = \log\frac{CL(X\,|\,X_{\text{observed}})}{CL(X\,|\,X_{\text{expected}})}. \qquad \textbf{[S3]}$$

Substituting Eq. S2 into Eq. S3 results in (Eq. S4)

$$CLR = \sum_{i=1}^{n}\left(\left(-\lambda_{i,\text{expected}} + \lambda_{i,\text{observed}}\right) + X_i\cdot\log\frac{\lambda_{i,\text{observed}}}{\lambda_{i,\text{expected}}}\right). \qquad \textbf{[S4]}$$

Because $E[X_i] = \lambda_i$, we set $\lambda_{i,\text{expected}} = E\left[X_{i,\text{expected}}\right]$ and $\lambda_{i,\text{observed}} = E[X_{i,\text{observed}}] = X_i$.

The expected entries for each category $E\left[X_{i,\text{expected}}\right]$ can be estimated by using (Eq. S5)

$$E\left[X_{i,\text{expected}}\right] = \theta_{region}\cdot\Pr(X_i) \qquad \textbf{[S5]}$$

[i.e., by distributing the total number of expected mutations at a particular locus according to their probability of observing them in a particular category $\Pr(X_i)$]. We obtain this probability from the AFS of the full dataset by (Eq. S6)

$$\Pr(X_i) = \frac{X_{i,\text{full}}}{\sum X_{i,\text{full}}}. \qquad \textbf{[S6]}$$

For the expected number of mutations in the region θ_{region}, we estimated the per base pair estimate of the population mutation rate θ_W from the sites observed in *O. rufipogon* in that region using Watterson's estimator (see below). For a particular locus, we then obtain (Eq. S7)

$$\theta_{region} = \theta_W\cdot N_{sites}, \qquad \textbf{[S7]}$$

where N_{sites} is the total number of sites with nonmissing data in the region. This leads to our final test statistic (Eq. S8)

$$CLR = \sum_{i=1}^{n}\left(\theta_{region}\cdot\Pr(X_i) + X_{i,\text{ observed}}\cdot\left(\log\frac{X_{i,\text{ observed}}}{\theta_{region}\cdot\Pr(X_i)} - 1\right)\right). \qquad \textbf{[S8]}$$

This statistic measures how well the data from a locus fits what would be expected for that locus, taking into account the observed genome-wide distribution of allele frequencies (i.e., demography) as well as the background mutation rate estimated from *O. rufipogon.*

As in Nielsen et al. (14), the method is sensitive to any deviation from the neutral expectation.

The CLR was calculated for each fragment, and empirical P values were assigned from the distribution of all CLR scores. To limit our results to regions with evidence for a selective sweep, we defined candidate regions as regions with a high CLR score combined with low diversity in at least one of the domesticated species. The cutoffs we used were an empirical CLR P value ≤ 0.1 as well as the number of segregating sites $S \leq 1$ in either *indica* or tropical *japonica.*

Demographic Modeling. ***SNP pruning for δaδi.*** Because our accessions were inbred lines, we randomly sampled one of two alleles for each individual, resulting in a dataset containing a total of 7,059 SNPs that were segregating in at least one population among *O. rufipogon*, *indica*, and tropical *japonica*. SNPs were then polarized using information from two outgroup species *O. meridionalis* and *O. barthii*, and for each SNP, we assigned the ancestral allele if at least one of the outgroup species matched one of the observed alleles in the data. To include only putatively neutral SNPs in the demographic inference, we only used noncoding and silent coding SNPs for the analysis, additionally removing all SNPs that were located in splice sites as well as UTR regions and those located in putative transposon regions. Furthermore, we included only SNPs that had nonmissing data in at least 13 *O. rufipogon*, 19 *indica* individuals, and 14 tropical *japonica* individuals, resulting in a total of 2,057 segregating sites in the AFS. For those sites where the samples sizes were larger, we performed a hypergeometric projection down to 13, 19, and 14 individuals, respectively, to generate a single AFS for numerical inference. After the projection, the final site frequency spectra used for the analysis contained 1,848 segregating sites, of which 1,574, 613, and 374 were segregating in *O. rufipogon*, *indica*, and tropical *japonica* populations, respectively.

Modeling. In each model, we modeled a Wright–Fisher ancient population of *O. rufipogon* that split, leading to a two-population epoch for $\tau_{2B} + \tau_2$-scaled generations, where τ_{2B} is the duration of the bottleneck and τ_2 is the length of time after the bottleneck and before the next population split. During these periods, the population size of the second species is $\eta_{1B} \times N_{\text{rufi}}$ and $\eta_1 \times N_{\text{rufi}}$ during and after the bottleneck, respectively. After the second population split, a three-population epoch extends for $\tau_B + \tau$-scaled generations, where the population size of the second species remains at $\eta_1 \times N_{\text{rufi}}$ and the population size of the third is $\eta_{2B} \times N_{\text{rufi}}$ for τ_B generations and $\eta_2 \times N_{\text{rufi}}$ for τ. Because bottleneck population size and bottleneck time are confounding variables, we fixed η_{1B} and η_{B2} to be 0.01 and inferred the severity of the bottleneck from the maximum likelihood estimates of τ_B and τ_{2B}.

In addition to estimating population size in contemporary *indica* and tropical *japonica*, we also allowed for migration during the final T-scaled generations of the model. Because Caicedo et al. (9) found no evidence for migration from *O. rufipogon* into *indica* and tropical *japonica*, we tested this scenario (no migration from *rufipogon*) along with three additional migration scenarios—no migration, symmetric migration, and asymmetric migration. Finally, to ensure that selective sweeps were not influencing our results from the demographic inference, we removed regions of the genome that were potentially under positive selection from the results of the selective sweep mapping and reanalyzed the demographic models with this AFS. We also inferred putative selective sweeps by applying the multipopulation CLR (XP-CLR) method, which is based on the expected patterns of population differentiation resulting from a selective sweep (16). However, it was unclear to what extent both the lack of a fine-scale recombination map as well as the type of the data (short resequenced unlinked fragments of highly inbred samples) affect this test, and therefore, they were excluded from the selective sweep results. In reanalyzing the demographic models without the sweeps, we did include those regions, which is a more conservative approach because a larger proportion of the data is excluded.

Estimating Effective Population Size. We estimated the effective population size for *O. rufipogon* by the population mutation rate $\theta = 4N_e\mu$. We used *O. meridionalis* as an outgroup, assuming a divergence time of $T_{\text{Div}} = 2$ Myr (17). We also estimated the mutation rate by the number of observed substitutions between *O. rufipogon* and *O. meridionalis* as $\mu = 3.34 \times 10^{-4}$ mutations per y over all sites or 3.83×10^{-9} per bp per generation, assuming a generation time of 1 y. From the total number of segregating sites, we can then estimate θ, and substituting into $\theta = 4N_e\mu$, we obtain the current effective population size N_e as 456,339 for *O. rufipogon*, 39,701 for *indica*, and 26,924 for tropical *japonica*. Alternatively, we can also fix the neutral substitution rate to 6.5×10^{-9} as previously reported (18). This leads to a slightly lower estimate for *O. rufipogon* of 268,715, with *indica* at 23,378 and tropical *japonica* at 15,854.

Phylogenetic Analyses. After discarding burn-in in the *BEAST analyses, we combined results from two independent runs of 50 million Markov Chain Monte Carlo (MCMC) chains each, sampling every 1,000th chain, and assessed convergence (i.e., both runs reaching the desired posterior distribution) with the program Tracer (19). The tree with the highest clade posterior probabilities (pp) was visualized using TreeAnnotator, which also outputs the clade pp and divergence times (i.e., mean ages taken from the entire sample of trees for that clade). To determine how *O. rufipogon* population structure would affect the multispecies coalescent (MSC) analysis, we also classified the wild rice accessions in the Yu et al. (20) dataset into India/Indochina and China groups based on the geographic regions described by Londo et al. (21).

1. Doyle JJ, Doyle JL (1989) Isolation of plant DNA from fresh tissue. *Focus* 12:13–15.
2. Rozen S, Skaletsky H (2000) Primer3 on the WWW for general users and for biologist programmers. *Methods Mol Biol* 132:365–386.
3. Ouyang S, et al. (2007) The TIGR Rice Genome Annotation Resource: Improvements and new features. *Nucleic Acids Res* 35:D883–D887.
4. Ewing B, Hillier L, Wendl MC, Green P (1998) Base-calling of automated sequencer traces using phred. I. Accuracy assessment. *Genome Res* 8:175–185.
5. Green P (1999) *Documentation for Phrap and Cross_match* (version 0.990319). Available at http://www.phrap.org/phredphrap/phrap.html. Accessed December 26, 2010.
6. Nickerson DA, Tobe VO, Taylor SL (1997) PolyPhred: Automating the detection and genotyping of single nucleotide substitutions using fluorescence-based resequencing. *Nucleic Acids Res* 25:2745–2751.
7. Edgar RC (2004) MUSCLE: Multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res* 32:1792–1797.
8. Falush D, Stephens M, Pritchard JK (2003) Inference of population structure using multilocus genotype data: Linked loci and correlated allele frequencies. *Genetics* 164: 1567–1587.
9. Caicedo AL, et al. (2007) Genome-wide patterns of nucleotide polymorphism in domesticated rice. *PLoS Genet* 3:1745–1756.
10. Nei M (1987) *Molecular Evolutionary Genetics* (Columbia University Press, New York).
11. Watterson GA (1975) On the number of segregating sites in genetical models without recombination. *Theor Popul Biol* 7:256–276.
12. Tajima F (1989) Statistical method for testing the neutral mutation hypothesis by DNA polymorphism. *Genetics* 123:585–595.
13. Thornton K (2003) Libsequence: A C++ class library for evolutionary genetic analysis. *Bioinformatics* 19:2325–2327.
14. Nielsen R, et al. (2009) Darwinian and demographic forces affecting human protein coding genes. *Genome Res* 19:838–849.
15. Sawyer SA, Hartl DL (1992) Population genetics of polymorphism and divergence. *Genetics* 132:1161–1176.
16. Chen H, Patterson N, Reich DE (2010) Population differentiation as a test for selective sweeps. *Genome Res* 20:393–402.
17. Zhu Q, Ge S (2005) Phylogenetic relationships among A-genome species of the genus *Oryza* revealed by intron sequences of four nuclear genes. *New Phytol* 167:249–265.
18. Gaut BS, Morton BR, McCaig BC, Clegg MT (1996) Substitution rate comparisons between grasses and palms: Synonymous rate differences at the nuclear gene *Adh* parallel rate differences at the plastid gene *rbcL*. *Proc Natl Acad Sci USA* 93: 10274–10279.
19. Drummond AJ, Rambaut A (2007) BEAST: Bayesian evolutionary analysis by sampling trees. *BMC Evol Biol* 7:214.

20. Yu G, Olsen KM, Schaal BA (2011) Molecular evolution of the endosperm starch synthesis pathway genes in rice (*Oryza sativa* L.) and its wild ancestor, *O. rufipogon* L. *Mol Biol Evol* 28:659–671.

21. 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*. *Proc Natl Acad Sci USA* 103:9578–9583.

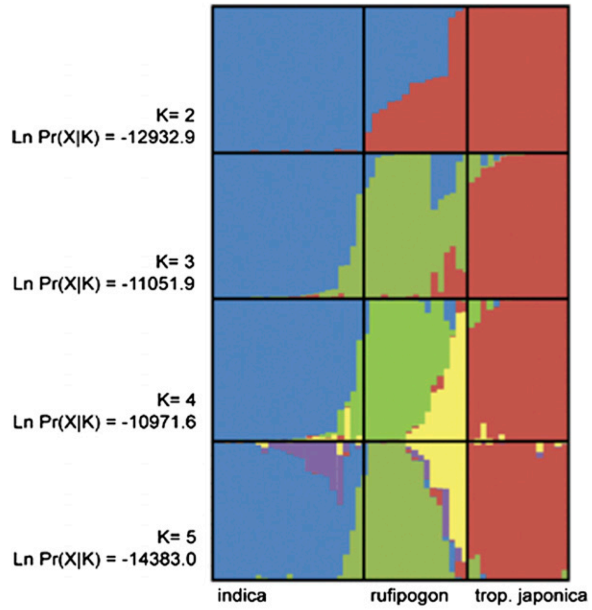

Fig. S1. Structure analysis. Model-based clustering analysis with STRUCTURE was used to identify population substructure and verify group membership of accessions. Using the method by Evanno et al. (1) to detect the correct number of clusters (K), δK increased at $K = 2$, wherein *O. rufipogon* accessions are shown as an admixture of alleles of the two domesticated subspecies. However, δK was highest at $K = 4$, which splits the accessions into *indica*, tropical *japonica*, and two *O. rufipogon* clusters. However, there was only a marginal difference between $K = 3$ and $K = 4$. Substructure within *O. rufipogon* cannot be explained by geographic partitions.

1. Evanno G, Regnaut S, Goudet J (2005) Detecting the number of clusters of individuals using the software STRUCTURE: a simulation study. *Mol Ecol* 14:2611–2620.

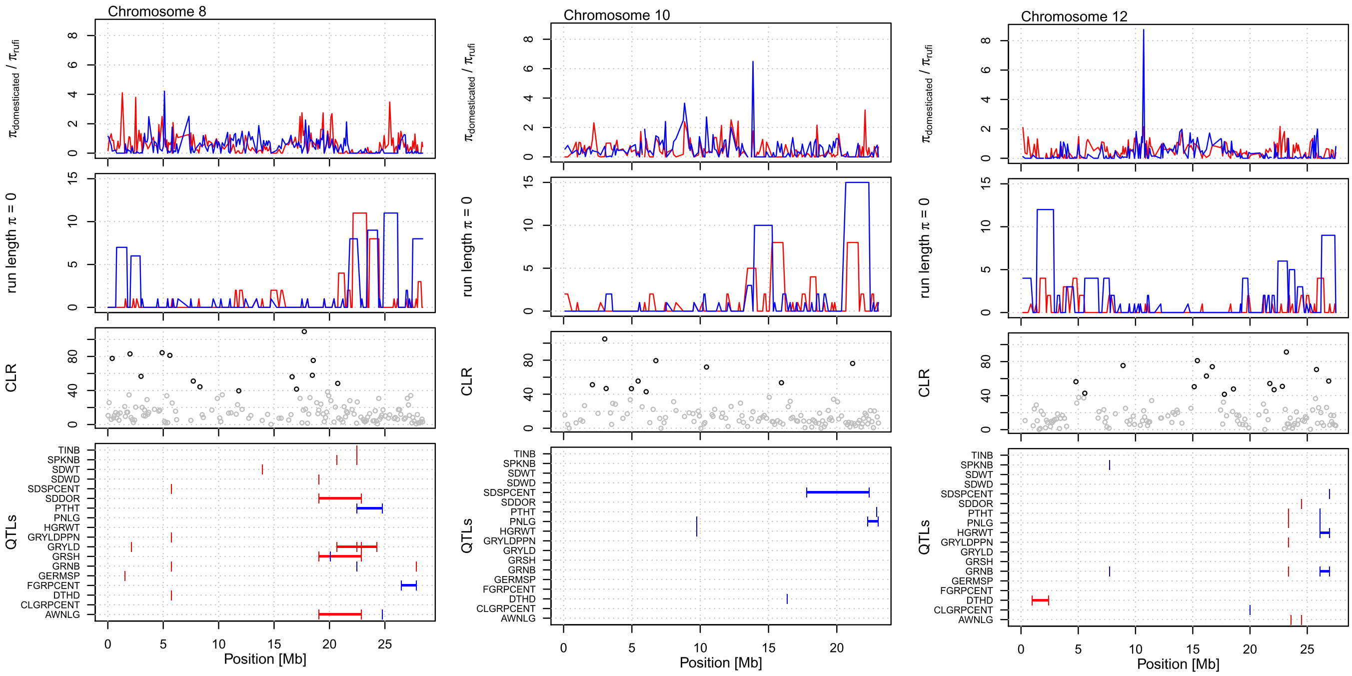

Fig. S2. Summary of selective sweep mapping results for chromosomes 8, 10, and 12. The *x* axis is the position along the chromosome. Results for different subspecies are indicated in red for *indica* and blue for tropical *japonica*. The ratio of nucleotide diversity in domesticates vs. wild rice is shown in the first row. Run length of consecutive monomorphic fragments is shown in the second row. CLR score for selective sweep mapping is shown in the third row. Plot symbols with darker shading indicate windows with empirical $P \leq 0.1$. Quantitative trait locus (QTL) regions associated with putative selective sweeps are shown in the fourth row. Plotted are the extents of QTL regions with size < 5 Mb involving either crosses of *O. rufipogon* and *O. sativa* ssp. *indica* (red) or tropical *japonica* (blue). QTL data is taken from www.gramene.org. AWNLG, awn length; CLGRPCENT, colored grain percentage; DTHD, days to heading; FGRPCENT, filled grain percentage; GERMSP, germination speed; GRNB, grain number; GRSH, grain shattering; GRYLD, grain yield; GRYLDPPN, grain yield per panicle; HGRWT, 100-grain weight; PNLG, panicle length; PTHT, plant height; SDDOR, seed dormancy; SDSPCENT, seed set percent; SDWD, seed width; SPKNB, spikelet number; TINB, tiller number.

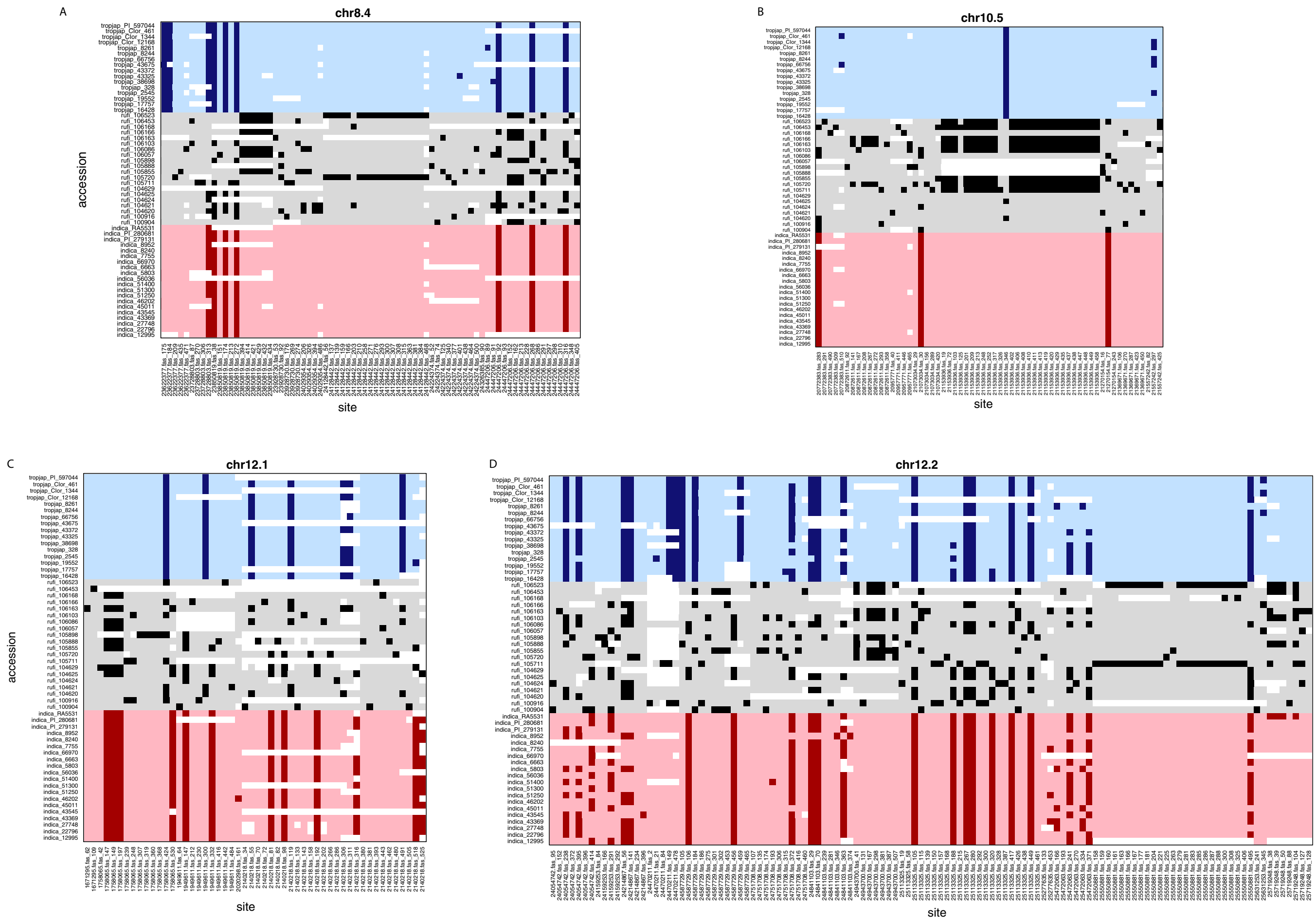

Fig. S3. Haplotype variation at four putative selective sweep regions. Each row in the plot corresponds to a haplotype for an individual with accession in the row legend, whereas the columns indicate the observed alleles for a particular polymorphic position along the chromosome. Haplotypes are colored according to the population of origin for the accessions (blue, tropical *japonica*; gray, *O. rufipogon*; red, *indica*). Alleles for a particular SNP are distinguished with light and dark shading of the respective color. Color shading was assigned according to the minor allele frequency in *O. rufipogon* (light shading, major allele in *O. rufipogon*; dark shading, minor allele in *O. rufipogon*) to assist visualization of positions showing high differentiation between *O. rufipogon* and *indica*/*japonica*. White cells indicate missing data.

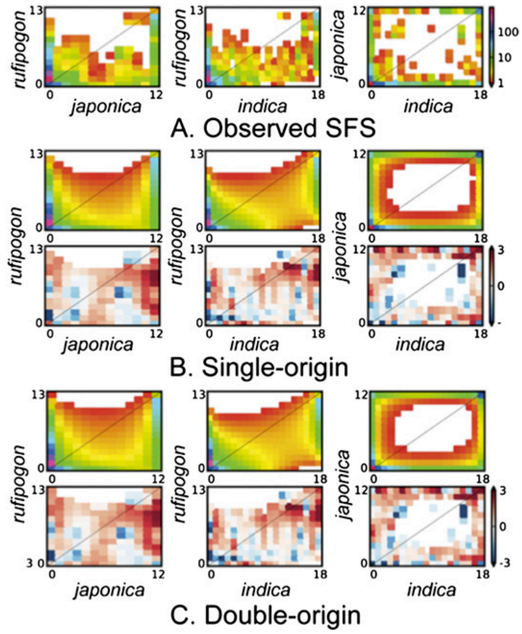

Fig. S4. Observed and modeled SNP 3D site frequency spectrum (SFS) of *O. rufipogon*, *O. sativa* ssp. *indica*, and *O. sativa* ssp. tropical *japonica*. (*A*) Observed SFS across 1,848 putatively neutral SNPs. (*B* and *C*) Expected SFS (*Upper*) and residuals between the observed SFS and expected SFS (*Lower*) under the best-fit (*B*) single-founder and (*C*) double-founder models. Although likelihoods were computed based on the 3D SFS, we show the 2D projection for each population pair for clarity.

Table S1. Accession information

Taxon	Country of origin	IRGC	GRIN	Other no.	Variety name
Oryza barthii	Chad	104119	—	RA2740	—
Indica	Thailand	5803	—	—	Pin kaeo
Indica	India	6663	—	—	Mudgo
Indica	Sri Lanka	7755	—	—	Kalukantha
Indica	Taiwan	8240	—	—	Hsia-chioh-keh-tu
Indica	Sri Lanka	8952	PI 584605	RA4911	Rathuwee
Indica	Loas	12995	—	—	I moume
Indica	Cambodia	22796	—	—	Damnoeub ansang
Indica	Thailand	27748	—	—	Khao dawk mali-105
Indica	Indonesia (S. Sumatra)	43369	—	—	Cere air
Indica	Indonesia (E. Kalimantan)	43545	—	—	Popot-165
Indica	India/Bangladesh	45011	—	RA4991	Badkalamakati
Indica	India	46202	—	—	Lalaman
Indica	China	51250	PI 584576	—	Ai-chiao-hong
Indica	China	51300	PI 584577	—	Guan-yin-tsan
Indica	China	51400	PI 584578	—	Pao-tou-hung
Indica	Vietnam	56036	—	—	Chau
Indica	Philippines	66970	—	—	Ir64
Indica	Taiwan	—	PI 279131	RA 5344	Deegeowoogen
Indica	Philippines	—	PI 280681	RA5374	Taducan
Indica	China	—	—	RA5531	93–11
Oryza meridionalis	Australia	101148	—	RA2661	—
Oryza nivara	Nepal	105706	—	—	—
Oryza rufipogon	Thailand	100904	—	—	—
Oryza rufipogon	China	100916	—	—	—
Oryza rufipogon	China	104620	—	W1952	—
Oryza rufipogon	China	104621	—	W1953(-3)	—
Oryza rufipogon	China	104624	—	W1956(-2)	—
Oryza rufipogon	China	104625	—	W1957	—
Oryza rufipogon	China	104629	—	RA4790	—
Oryza rufipogon	India	105711	—	—	Kozhinelli
Oryza rufipogon	Cambodia	105720	—	—	—
Oryza rufipogon	Thailand	105855	—	—	Khao nok
Oryza rufipogon	Bangladesh	105888	—	—	Uri
Oryza rufipogon	Bangladesh	105898	—	—	Uri dan
Oryza rufipogon	India	106057	—	—	Balunga
Oryza rufipogon	India	106086	—	—	Uri dan
Oryza rufipogon	India	106103	—	—	—
Oryza rufipogon	Laos	106163	—	—	—
Oryza rufipogon	Vietnam	106166	—	—	Khao noc pit
Oryza rufipogon	Vietnam	106168	—	—	—
Oryza rufipogon	Indonesia	106453	—	—	—
Oryza rufipogon	Papua New Guinea	106523	—	—	—
Tropical *japonica*	Philippines	328	—	RA5535	Azucena
Tropical *japonica*	Japan	2545	—	RA4882	Kotobuki mochi
Tropical *japonica*	Philippines	8244	—	RA4901	Davao
Tropical *japonica*	Indonesia	8261	PI 584546	RA4905	Padi kasalle
Tropical *japonica*	Indonesia	16428	—	RA5297	Gundil kuning
Tropical *japonica*	Indonesia	17757	—	RA5353	Jambu
Tropical *japonica*	Maylasia	19552	—	—	Tassangih
Tropical *japonica*	Pakistan	38698	PI 584568	RA4948	Npe-844
Tropical *japonica*	Indonesia (West Java)	43325	PI 584570	RA4951	Arias
Tropical *japonica*	Indonesia (Bali)	43372	—	RA4955	Cicih beton
Tropical *japonica*	Indonesia (East Java)	43675	—	RA4988	Trembese
Tropical *japonica*	Texas	66756	—	RA4998	Lemont
Tropical *japonica*	Philippines	—	Clr 12168	RA5396	Sinampaga selection
Tropical *japonica*	Louisiana	—	Clr 1344	RA5045	Fortuna
Tropical *japonica*	Philippines (introduced)	—	Clr 461	RA5333	Asse y pung
Tropical *japonica*	Thailand	—	PI 597044	RA5294	Ku115

Oryza accessions used in this study with their voucher/accession numbers and countries of origin. Clr, cereal investigation; GRIN, Germplasm Resources Information Network; IRGC, International Rice GenBank Collection; PI, plant introduction; RA: rice accession numbers from the McCouch Laboratory, Cornell University, Ithaca, NY.

Table S2. Overview of selective sweep regions predicted by two different methods

Region	Population	Chromosome	Start (kB)	End (kB)	Size (kB)	Method	QTLs
Chr8.1	*Japonica*	8	5,598	5,598	0.5	CLR	
Chr8.2	*Indica*	8	20,828	21,326	498	Diversity	AWNLG, GRSH, GRYLD, SDDOR
Chr8.3	*Indica, japonica*	8	21,841	23,333	1,492	Diversity	AWNLG, GRNB, GRSH, GRYLD, PTHT, SDDOR, SPKNB, TINB
Chr8.4	*Indica, japonica*	8	23,445	24,448	1,003	Diversity	GRYLD, PTHT
Chr8.5	*Japonica*	8	24,932	26,138	1,208	Diversity	
Chr8.6	*Japonica*	8	27,518	28,406	888	Diversity	FGRPCENT, GRNB
Chr10.1	*Japonica*	10	3,115	3,116	0.5	CLR	
Chr10.2	*Indica, japonica*	10	13,467	15,265	1,798	Diversity	
Chr10.3	*Indica*	10	15,264	16,053	789	Diversity, CLR	
Chr10.4	*Indica*	10	18,070	18,462	392	Diversity	SDSPCENT
Chr10.5	*Indica, japonica*	10	20,656	22,342	1,686	Diversity, CLR	PNLG, SDSPCENT
Chr12.1	*Indica, japonica*	12	1,419	2,835	1,417	Diversity	DTHD
Chr12.2	*Indica, japonica*	12	4,556	4,871	315	Diversity, CLR	
Chr12.3	*Indica, japonica*	12	5,535	5,735	200	CLR	
Chr12.4	*Japonica*	12	15,126	15,126	0.5	CLR	
Chr12.5	*Indica*	12	16,196	16,196	0.5	CLR	
Chr12.6	*Japonica*	12	22,102	22,102	0.5	CLR	
Chr12.7	*Japonica*	12	22,823	22,823	0.5	CLR	
Chr12.8	*Japonica*	12	23,176	23,176	0.5	CLR	
Chr12.9	*Indica, japonica*	12	25,891	27,505	1,614	Diversity, CLR	GRNB, HGRWT, PNLG, PTHT

Candidate regions with overlapping signals for either of the populations or different methods were collapsed into one region. Note that regions showing evidence for both *indica* and *japonica* as well as multiple methods do not automatically imply that all groups show evidence for all tests within the same region. For example, region *chr12.12* shows evidence for all methods; however, the CLR signal was only found for tropical *japonica*. Fig. S2 contains the QTL abbreviations.

Table S3. Maximum likelihood parameters for the four models (with and without SNPs in putative sweep regions) that we tested assuming symmetric migration

Model	η_i	η_j	τ	τ_2	τ_B	τ_{B2}	m_{ir}	m_{jr}	m_{ij}
With sweeps									
Japonica from *indica*	0.08	0.06	1.00	0.00	0.00	0.04	3.13	1.32	1.99
Indica from *japonica*	0.08	0.06	1.00	0.00	0.00	0.04	3.14	1.29	1.92
Indica first	0.09	0.07	1.00	0.01	0.02	0.03	2.71	1.17	1.80
Japonica first	0.07	0.07	1.00	0.01	0.02	0.03	3.69	1.24	2.15
Without sweeps									
Japonica from *indica*	0.09	0.07	1.00	0.00	0.00	0.04	3.84	1.57	2.16
Indica from *japonica*	0.07	0.07	1.00	0.00	0.00	0.07	5.07	2.01	2.35
Indica first	0.11	0.10	1.00	0.00	0.01	0.01	3.51	1.41	1.59
Japonica first	0.10	0.08	1.00	0.00	0.01	0.00	4.02	1.85	1.37

Maximum likelihood parameter estimates for the single- and double-founder domestication models tested in ∂a∂i. All parameters were inferred by ∂a∂i, although we fixed the *indica* and tropical *japonica* bottleneck population sizes to 1% of the *O. rufipogon* population size for all models. τ_{B2} and τ_2 represent the length of time of the bottleneck and time thereafter for the two-population epoch. Likewise, τ_B and τ represent the length of the bottleneck and time thereafter during the three-population epoch. All models had symmetric migration in the three-population epoch. Symmetric migration between *O. rufipogon* and *indica* is represented by m_{ir}, migration between tropical *japonica* and *O. rufipogon* is represented by m_{jr}, and migration between *indica* and tropical *japonica* is indicated by m_{ij}.

Table S4. Parameter estimates for the different migration models tested (SNPs in sweep regions included)

Migration	Model	Log likelihood	η_i	η_j	τ	τ_2	m_{ir}	m_{ri}	m_{jr}	m_{rj}	m_{ij}	m_{ji}
Symmetric migration	*Japonica* from *indica*	−1,214.7	0.05	0.05	1.00	0.11	5.79		1.25		3.30	
	Indica from *japonica*	−1,216.8	0.06	0.05	1.00	0.05	4.95		1.26		2.80	
	Indica first	−1,233.9	0.05	0.06	1.00	0.03	4.91		1.17		2.58	
	Japonica first	−1,236.0	0.08	0.06	1.00	0.05	3.21		1.16		2.17	
Asymmetricmigration	*Japonica* from *indica*	−1,190.7	0.05	0.08	1.00	0.16	4.07	4.38	1.38	2.59	0.44	1.53
	Indica from *japonica*	−1,178.6	0.13	0.04	0.98	0.15	0.46	4.32	4.43	1.19	1.44	1.59
	Indica first	−1,230.5	0.11	0.06	1.00	0.27	0.34	2.97	3.77	0.90	0.85	0.16
	Japonica first	−1,227.8	0.16	0.03	1.00	0.43	0.02	5.63	8.83	2.60	1.88	1.23
No migration from *O. rufipogon*	*Japonica* from *indica*	−1,318.1	0.05	0.05	0.05	0.01	7.22		1.12		0.64	
	Indica from *japonica*	−1,318.2	0.05	0.05	0.05	0.00	6.27		0.99		0.61	
	Indica first	−1,321.4	0.05	0.05	0.06	0.00	5.97		1.11		1.31	
	Japonica first	−1,319.8	0.05	0.05	0.05	0.02	6.10		0.40		1.58	
No migration	*Japonica* from *indica*	−1,335.7	0.08	0.04	0.04	0.00						
	Indica from *japonica*	−1,334.5	0.08	0.04	0.04	0.00						
	Indica first	−1,342.2	0.08	0.04	0.05	0.00						
	Japonica first	−1,339.5	0.07	0.06	0.04	0.02						

Table S5. Domestication models and their log likelihoods with weak positive selection

Type of model	Model	Log likelihood (with sweeps)	Log likelihood (no sweeps)
Single founder	*Japonica* from *indica*	−1,226.19	−1,004.65
Single founder	*Indica* from *japonica*	−1,226.18	−1,004.65
Double founder	*Indica* first	−1,266.83	−1,019.79
Double founder	*Japonica* first	−1,263.67	−1,049.88

All models had founding bottlenecks, with a bottleneck population size fixed to 1% of the *O. rufipogon* population size, and all models had symmetric migration in the three-population epoch. Weak positive selection was added to the model (selection parameter = +1). Note that log-likelihood values from models with and without sweeps cannot be directly compared.