

# Molecular evidence for a single evolutionary origin of domesticated rice

Jeanmaire Molina<sup>a,1</sup>, Martin Sikora<sup>b,1</sup>, Nandita Garud<sup>b</sup>, Jonathan M. Flowers<sup>a</sup>, Samara Rubinstein<sup>a</sup>, Andy Reynolds<sup>b</sup>, Pu Huang<sup>c</sup>, Scott Jackson<sup>d</sup>, Barbara A. Schaal<sup>c,2</sup>, Carlos D. Bustamante<sup>b</sup>, Adam R. Boyko<sup>b</sup>, and Michael D. Purugganan<sup>a,2</sup>

<sup>a</sup>Department of Biology and Center for Genomics and Systems Biology, New York University, New York, NY 10003; <sup>b</sup>Department of Genetics, Stanford University, Stanford, CA 94305; <sup>c</sup>Department of Biology, Washington University, St. Louis, MO 63130; and <sup>d</sup>Department of Agronomy, Purdue University, West Lafayette, IN 47907

Contributed by Barbara A. Schaal, March 25, 2011 (sent for review January 28, 2011)

Asian rice, *Oryza sativa*, is one of world's oldest and most important crop species. Rice is believed to have been domesticated ~9,000 y ago, although debate on its origin remains contentious. A single-origin model suggests that two main subspecies of Asian rice, *indica* and *japonica*, were domesticated from the wild rice *O. rufipogon*. In contrast, the multiple independent domestication model proposes that these two major rice types were domesticated separately and in different parts of the species range of wild rice. This latter view has gained much support from the observation of strong genetic differentiation between *indica* and *japonica* as well as several phylogenetic studies of rice domestication. We reexamine the evolutionary history of domesticated rice by resequencing 630 gene fragments on chromosomes 8, 10, and 12 from a diverse set of wild and domesticated rice accessions. Using patterns of SNPs, we identify 20 putative selective sweeps on these chromosomes in cultivated rice. Demographic modeling based on these SNP data and a diffusion-based approach provide the strongest support for a single domestication origin of rice. Bayesian phylogenetic analyses implementing the multispecies coalescent and using previously published phylogenetic sequence datasets also point to a single origin of Asian domesticated rice. Finally, we date the origin of domestication at ~8,200–13,500 y ago, depending on the molecular clock estimate that is used, which is consistent with known archaeological data that suggests rice was first cultivated at around this time in the Yangtze Valley of China.

Domestication is a complex evolutionary process in which human use of plant and animal species leads to genetically based morphological and/or physiological diversification of domesticated taxa from their wild ancestors (1). The process of domestication provides insights into the nature of selection and the rise of species differences (1). Understanding the origins of domesticated species impacts our understanding of the evolutionary mechanisms surrounding domestication (e.g., founder events, selection, and parallel evolution) and the cultural context by which human societies cultivate and become dependent on specific species for food, fiber, and other uses.

Asian rice, *Oryza sativa* L., is one of world's oldest and most important crop species, having been domesticated beginning some ~8,000–9,000 y ago (2–4). Asian rice feeds more than one-half of the global population and has become a key model system for plant biology (5). Several genetic studies have shown that *O. rufipogon*, which remains extant in South and Southeast Asia, is the wild progenitor of domesticated rice (6). Although others have suggested that *O. nivara* may be the progenitor of rice (7), there is evidence to indicate that this species is an annual ecotype of *O. rufipogon* (6, 8, 9).

Genetic analysis has established that rice consists of several genetically differentiated variety groups, with the two main groups being *indica* and *japonica* (10). Sometimes described as subspecies, *indica* and *japonica* have been recognized since ancient China (11) and are the most widely grown rice varieties. Several studies have shown strong genetic differentiation between *indica* and *japonica* (8, 12–20), and molecular studies suggest divergence

time estimates of 86–440 ky between these two variety groups (14, 15, 17, 21), far older than the ~9,000-y archaeological estimate for rice cultivation (2, 4).

Despite recent advances in genetics and archaeology, there continues to be debate on the origin(s) of domesticated rice (22, 23). Several models to explain the origin of rice have been advanced over the last half century; these models can be broadly classified as advocating either a single origin or multiple origins for this important crop species. Single-origin models posit that domesticated rice originated from wild rice (Fig. 1), with differentiation of *indica* and *japonica* occurring after domestication of the cultivated species (24). Molecular evidence for this model is largely based on recent studies that show that the key domestication gene *sh4* that confers nonshattering (25) and the *progl1* locus responsible for the erect habit (26) have nearly identical sequences shared by both subspecies of rice. There is also recent molecular evidence for the single-origin model from a Bayesian demographic analysis of multilocus microsatellite data (27).

Multiple-origin proponents, however, attribute this sharing of key domestication genes between *indica* and *japonica* as arising from hybridization between the two variety groups some time after their independent domestication (3, 22, 23). They suggest a model in which *indica* and *japonica* were domesticated separately from predifferentiated ancestral *O. rufipogon* populations (Fig. 1). This multiple-origin model can readily explain genetic differentiation observed in domesticated rice and thus has gained support from phylogenetic analyses that show distinct clades in *O. sativa* for *indica* and *japonica*, with different *O. rufipogon* accessions associated with each clade (9, 15–19). Moreover, archaeological studies show evidence for rice domestication in the Yangtze Valley beginning ~8,000–9,000 y ago (2–4, 28) as well as early (and putatively separate) cultivation of rice in the Ganges in India beginning ~4,000 y ago (3).

The recent origins of domesticated crop species (<10,000 y) increase the likelihood that ancestral polymorphisms persist in domesticated taxa through incomplete lineage sorting, which results in sequence similarities that do not necessarily reflect species and population relationships (29, 30). Simulation studies have shown that data concatenation can generate misleading phylogenetic relationships, because it ignores the different evo-

Author contributions: J.M., J.M.F., S.J., B.A.S., C.D.B., A.R.B., and M.D.P. designed research; J.M., J.M.F., S.R., and P.H. performed research; A.R., P.H., and B.A.S. contributed new reagents/analytic tools; J.M., M.S., N.G., J.M.F., C.D.B., and A.R.B. analyzed data; and J.M., M.S., N.G., B.A.S., A.R.B., and M.D.P. wrote the paper.

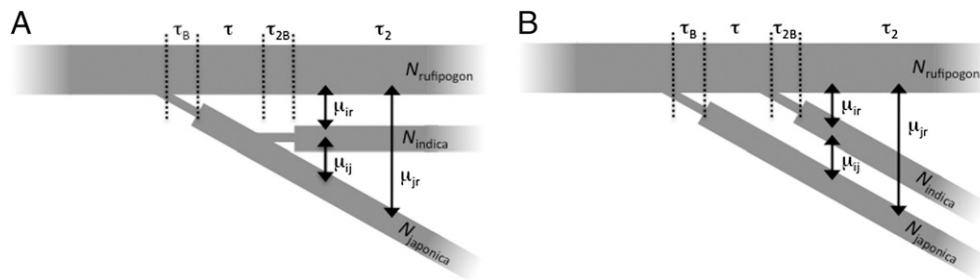
The authors declare no conflict of interest.

Data deposition: Because of the complexity of the data as multiple sequence alignments, there is no public database that can accommodate the format. We are, therefore, making the data available as a zipped file at [http://puruggananlab.bio.nyu.edu/Rice\\_data/](http://puruggananlab.bio.nyu.edu/Rice_data/) as indicated in Results.

<sup>1</sup>J.M. and M.S. contributed equally to this work.

<sup>2</sup>To whom correspondence may be addressed. E-mail: [shaal@wustl.edu](mailto:shaal@wustl.edu) or [mp132@nyu.edu](mailto:mp132@nyu.edu).

This article contains supporting information online at [www.pnas.org/lookup/suppl/doi:10.1073/pnas.1104686108/-DCSupplemental](http://www.pnas.org/lookup/suppl/doi:10.1073/pnas.1104686108/-DCSupplemental).



**Fig. 1.** Schematics of the single- (A) vs. double- (B) founder models. (A) In the single domestication event, both domesticated subspecies originated from the same *O. rufipogon* ancestral population. (B) In the double-founder model, *indica* and tropical *japonica* were domesticated independently from different *O. rufipogon* populations. *O. rufipogon*, *O. sativa* ssp. *indica*, and *O. sativa* ssp. tropical *japonica* are indicated by the subscripts *r*, *i*, and *j*, respectively. The times  $\tau_B$  and  $\tau$  represent the length of the bottleneck and time thereafter during the two-population epoch. Likewise,  $\tau_{2B}$  and  $\tau_2$  represent the length of time of the bottleneck and time thereafter for the three-population epoch. Symmetric migration ( $\mu$ ) between the populations is represented by arrows, and  $N$  is the population size.

lutionary histories of distinct loci (31). Indeed, incomplete lineage sorting leads to incongruent gene trees when multiple loci are analyzed independently, which has been recently shown in the genus *Oryza* (32, 33), but phylogenetic studies of rice have still largely relied on data concatenation across multiple loci under the assumption that the predominant phylogenetic information inherent in the data will swamp out any conflicting signal (9, 15, 17–19).

Conflicting gene trees because of coalescent stochasticity have been a problem for species delimitation, but statistical methods that combine population genetics with phylogenetics now allow for a more accurate inference of recent evolutionary history (reviewed in ref. 30). Phylogenetic analyses using the multispecies coalescent (MSC) are able to detect signals of species differentiation even before their gene trees are reciprocally monophyletic (29, 30, 34, 35).

In this study, we provide evidence for a single domestication of rice. We identified SNPs through direct resequencing of >250 kb of sequence from 630 gene fragments across three rice chromosomes in wild and cultivated rice accessions. We then use two methods to infer the evolutionary history of domesticated Asian rice—a diffusion-based approach to demographic modeling on the SNP data (36) and a Bayesian evolutionary approach to phylogenetic analysis, implementing the multispecies coalescent (35) using previously published phylogenetic datasets. Both approaches support a single origin for rice, and we are also able to estimate a date for the domestication of rice consistent with the previously published archaeological studies.

## Results

**Selective Sweeps in Three Rice Chromosomes.** We resequenced portions of 630 genes on rice chromosomes 8, 10, and 12 at ~100-kb intervals in multiple accessions (Table S1) of *O. sativa indica* ( $n = 20$ ) and tropical *japonica* ( $n = 16$ ) as well as *O. rufipogon* ( $n = 20$ ) and a single accession of *O. nivara*, *O. barthii*, and *O. meridionalis* (data may be downloaded from [http://puruggananlab.bio.nyu.edu/Rice\\_data/](http://puruggananlab.bio.nyu.edu/Rice_data/)). We obtained  $255.9 \pm 2.09$  kb of sequence data for each accession. A total of 2,800 SNPs in *indica*, 2,070 SNPs in tropical *japonica*, and 7,274 SNPs in *O. rufipogon* were identified. As expected, the mean silent site nucleotide diversity ( $\pi$ ) was lower in *O. sativa* ( $\pi = 0.0037$  for *indica* and 0.0028 for tropical *japonica*) compared with *O. rufipogon* ( $\pi = 0.0079$ ) across all three chromosomes.

The classification of accessions was confirmed based on results of a STRUCTURE (37) population stratification analysis (Fig. S1). These results suggest that our sample had four ancestral populations ( $K = 4$ ), which corresponds to *O. sativa* ssp. *indica*, *O. sativa* ssp. tropical *japonica*, and two *O. rufipogon* clusters. There was only a marginal difference, however, in the likelihood

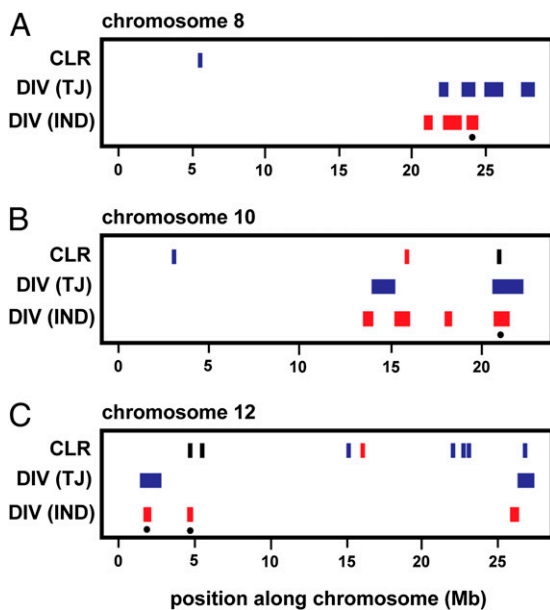
values between  $K = 3$  and  $K = 4$ , which is associated with the splitting of *O. rufipogon* into two subpopulations.

We used the chromosome scan data to first identify putative selective sweeps, regions of the genome that show evidence of recent positive selection, on all three chromosomes so that we could exclude these in our subsequent demographic analyses of rice origins (see below). We applied two different methods to identify regions with a genetic signature of a selective sweep within and between the domesticated species; these were based on (i) local reduction in nucleotide diversity and (ii) patterns of the multipopulation allele frequency spectrum (AFS; *Materials and Methods* and *SI Text*). We identified a total of 20 regions within the three chromosomes that had evidence for selective sweeps from at least one of these tests. We found that regions with reduced diversity were widespread in both *indica* and in particular, tropical *japonica* (Fig. 2, Fig. S2, and Table S2), consistent with the expectation of strong artificial selection during rice domestication. We were also able to find putative selective sweeps that were specific to *indica* or tropical *japonica* as well as regions with sweeps shared by both variety groups (Fig. 2 and Fig. S3).

**Demographic Inference.** We examined several demographic models for the origin and history of domesticated rice (Fig. 1) using a diffusion-based approach implemented in the program *∂a∂i* (36), which calculates the likelihood of a demographic model given an observed multipopulation AFS. The multipopulation AFS is the joint distribution of allele frequencies of diallelic variants from a genomic region sequenced in multiple individuals from each population. The program *∂a∂i* also computes the expected AFS under various demographic scenarios by numerically solving a multipopulation diffusion equation describing the effects of mutation, drift, and migration. Likelihoods of each model can be computed based on the product of the Poisson likelihoods for each entry of the AFS.

Based on 2,057 putatively neutral segregating sites (SI Text), we found that the single-origin models outperformed models of separate domestication of *indica* and tropical *japonica* (Table 1 and Fig. S4). The single-origin model provided a better fit to the data, even when putative sweep regions were excluded (Table 1). Each model has the same number of parameters, making the 22 log-likelihood unit difference between single- and double-founder models highly significant, even after correcting for linkage. We cannot, however, distinguish whether *indica* or tropical *japonica* was founded first under these scenarios. The maximum-likelihood parameter values inferred for both single-founder models (with and without sweeps) are presented in Table S3.

The parameters for population size and migration are very similar across all models, regardless of whether *indica* or tropical



**Fig. 2.** Summary of selective sweep mapping results. Candidate selective sweep regions identified in each test for selection for (A) chromosome 8, (B) chromosome 10, and (C) chromosome 12. Results for different subspecies are indicated by red for *indica* and blue for tropical *japonica*. CLR, composite likelihood ratio test based on the multipopulation allele frequency spectrum; DIV, diversity test based on regions of no variation; TJ, tropical *japonica*; IND, *indica*. Vertical bars in the CLR track correspond to single fragments, because it was calculated for each fragment separately. Colors in the track correspond to the respective population, with black bars indicating a sweep in both *indica* and tropical *japonica*. The four shared sweep regions are marked with a dot.

*japonica* was founded first. Although we present data for symmetric migration, we also tested the effect of four types of migration on the models to determine which one fits the AFS best (Table S4). These four models include symmetric migration, asymmetric migration, no migration from *O. rufipogon*, and no migration at all. We did not impose a founding bottleneck for *indica* and tropical *japonica* to simplify the parameter space when making these comparisons. Models with asymmetric migration have a slightly better fit, but the increase in log likelihood is not enough to justify the three additional migration parameters in the model.

Finally, a previous study suggests that selection could have genome-wide effects on the site frequency spectrum (18). To model the effects of selection, we reran our demographic models with weak positive selection (selection parameter = 1) occurring in all populations during the two- and three-population epochs. We found that, even after running the models with weak positive selection, the single-founder models still outperform the double-founder models (Table S5).

**Bayesian Reanalyses of Previously Published Phylogenetic Data.** Because the results of our demographic analysis challenge the currently accepted view that *indica* and *japonica* have independent origins, we examined the phylogeny of rice domestication. Previous phylogenetic studies concatenated data from multiple unlinked loci, a common practice but one that has been found to be problematic when dealing with recent speciation events (30, 31, 35). A Bayesian approach that accounts for gene tree heterogeneity while estimating the species tree can circumvent this issue, and this has been implemented in the program \*BEAST. Unfortunately, our resequenced gene fragments were too short (~500 bp) for \*BEAST to perform well, and instead, we reanalyzed previously published phylogenetic datasets that have been used to argue for the independent domestications of *indica* and *japonica* (15–17, 19, 38) (Table 2). We also used a recently published dataset that examined the evolution of the rice endosperm starch biosynthetic pathway (39).

Our \*BEAST analyses of two datasets by Zhu and Ge (15) and Londo et al. (16), each composed of less than five loci, resulted in several equivocal topologies, with at least 90% of the trees in the posterior set disputing the monophyly of *indica* and *japonica* (results not shown). The other four datasets, however, each with more than five loci, showed strong support for a single origin of domesticated rice when they were reanalyzed using \*BEAST (Table 2 and Fig. 3). We note that even the inclusion of *O. nivara*, which was suggested as an alternative ancestral species for *indica* (7, 9), still revealed a closer relationship between *indica* and *japonica* than each to any wild rice species (Fig. 3). Both datasets agree that *indica* and *japonica* are derived from one common ancestor, even if all five groups comprising Asian domesticated rice are represented in the Yu et al. (39) dataset and *O. rufipogon* population structure was considered (Fig. 3).

Applying a strict molecular clock of  $6.5 \times 10^{-9}$  substitutions/site per y for the grasses (40) resulted in an estimate for the mean time of the onset of domestication of 8,200 y before present (B.P.) [95% highest posterior density (HPD) = 4,400–12,100 y B.P.]. The estimate for the mean age for the *indica*–*japonica* split is 3,900 y B.P. (95% HPD = 1,700–6,600 y B.P.). Age estimates are higher when the mutation rate of  $3.8 \times 10^{-9}$  substitutions/site per y estimated from the chromosome scan data (SI Text) was applied. Using this molecular clock rate, we estimate the split of *O. sativa* from *O. rufipogon* commencing 13,500 y ago (95% HPD = 7,400–20,000 y B.P.) and the two domesticated rice variety groups splitting 6,700 y B.P. (95% HPD = 3,700–10,000 y B.P.).

## Discussion

The origin of Asian rice has long been a puzzle to biologists (22, 23), and over the last two decades, the multiple-origins domestication model proposing the independent domestication of *indica* and tropical *japonica* has gained support largely from molecular data analyzed by traditional phylogenetic methods (9, 15–19). These phylogenetic methods, however, can lead to heterogeneities in inferred gene tree topologies, particularly among

**Table 1. Single- and double-founder domestication models and their log likelihoods with and without selective sweep regions**

Type of model	Model	Log likelihood (with sweeps)	Log likelihood (no sweeps)
Single founder	<i>Japonica</i> from <i>indica</i>	–1,206.4	–1,038.7
Single founder	<i>Indica</i> from <i>japonica</i>	–1,206.3	–1,054.2
Double founder	<i>Indica</i> first	–1,232.1	–1,088.1
Double founder	<i>Japonica</i> first	–1,228.5	–1,088.0

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. Note that log-likelihood values from models with and without sweeps cannot be directly compared.



**Table 2. Results from \*BEAST reanalyses of previously published phylogenetic datasets**

Dataset	No. of loci	Total number of sites	Number of consensus topologies*	Single-origin support <sup>†</sup>	<i>Indica</i> and <i>japonica</i> not monophyletic
Zhu and Ge (15)	4	2,917	11	7.7%	90.1%
Londo et al. (16)	3	3,232	9	3.9%	92.9%
Tang et al. (17)	11	9,775	3	96.3%	3.4%
Zhu et al. (38)	10	8,031	3	100%	0%
Rakshit et al. (19)	22	19,789	1	100%	0%
Yu et al. (39)	5	19,780	3	100%	0%

Percentage values indicate the proportion of trees of a given topology in the posterior probability distribution of trees.

\*Of 19 possible topologies (including polytomies).

<sup>†</sup>*Indica* and tropical *japonica* monophyletic.

recently evolved species (30, 33). This ambiguity has prompted development of alternative phylogenetic inference methods, including those that use the multispecies coalescent (30, 35).

We have reassessed the phylogeny of domesticated rice using previously published datasets, five of which have been used to argue for a separate origin for *indica* and tropical *japonica* rice. Our study with the same data, reanalyzed in a multispecies coalescent framework, showed strong support for only a single origin of domesticated rice. Even the inclusion of *O. nivara* in two of the datasets that we have analyzed still revealed a closer relationship between *indica* and tropical *japonica* than each to any wild rice species. Two other datasets (15, 16) resulted in several equivocal topologies because of insufficient phylogenetic signal from too few loci (<5). This is not surprising, because simulations have shown that the probability of obtaining the correct species tree increases to 0.75 despite shallow tree depth when at least five loci are included (29, 41).

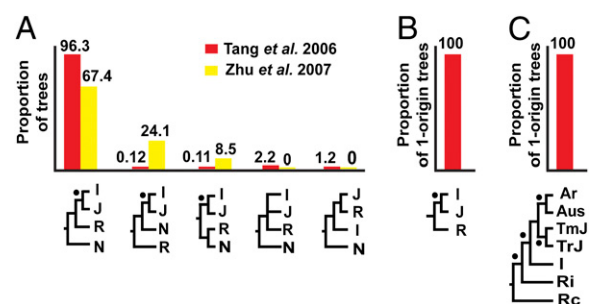
Previously detected population structure in *O. rufipogon* (16) violates one of the assumptions of the multispecies coalescent model. Accounting for this when we analyzed the data by Yu et al. (39), however, still showed *indica* and *japonica* as more closely related. Neither was affiliated with any wild rice group (Indian/Indochinese or Chinese), which would be expected if they were independently domesticated. There also seems to be phylogenetic support for the Indian/Indochinese *O. rufipogon* population as directly ancestral to domesticated rice. A larger sampling, however, will be necessary before a specific population can be identified as the ancestor of *O. sativa*, and there is also the possibility that such an ancestral population may be extinct.

The finding that domesticated rice has a single origin was also supported by demographic modeling using resequencing data of 630 nuclear loci from rice chromosomes 8, 10, and 12. The presence of selective sweeps shared by *indica* and tropical *japonica* may bias our inferences, because these shared sweeps could be related to domestication, and the same or very similar haplotypes may be fixed in the domesticated varieties (either from a single-origin event or parallel evolution or through post-domestication hybridization). There are, in fact, four putative cases of shared sweeps on these three chromosomes (Fig. 2), and these colocalize with known domestication quantitative trait loci (QTL) involved in traits for panicle length, plant height, days to heading, grain weight, and grain number (Fig. S2). Shared alleles between *indica* and *japonica* among known domestication genes have also been reported previously, most notably, red pericarp *rc* (42), nonshattering *sh4* (25, 43), and plant architecture *progl* (26) loci.

These and other shared sweeps were initially thought to arise from introgression between variety groups as a result of post-domestication hybridization (3, 22). Aside from the *rc*, *sh4*, and *progl* domestication genes, shared alleles have also been observed between *indica* and *japonica* at diversification genes that contribute to phenotypic diversity between rice cultivars. These

diversification genes include *BADH2* fragrance gene (44), the *sd1* semidwarfing gene (45), the *Pi-ta* disease resistance locus (45), the starch biosynthetic gene *Wx* (45), and the *GS3* grain length gene (46). In most of these cases, however, only a minor fraction of varieties carry the introgressed allele; this is in contrast to domestication alleles shared between *indica* and *japonica*, which are at or close to fixation in both variety groups (25, 26, 42, 43). Indeed, the alleles at some of these diversification loci seem to have been introgressed more recently as a result of recent breeding efforts (45).

In general, most rice cultivars surveyed in a genome-wide SNP assay show very little evidence level of introgression (45). Nevertheless, introgression between *japonica* and *indica* may obscure evidence for multiple origins of domesticated rice. Eckert and Carstens (41), however, have shown that coalescent-based methods of phylogenetic inference are still robust, despite moderate amounts of historical gene flow. Moreover, we find strong support for a single origin in our demographic modeling, even when putative selective sweep regions, including those shared between domesticated rice groups, are eliminated. Finally, our modeling took gene flow/migration into account in the demographic inference, and a single-origin model was still favored. Together, our results indicate that many of the shared selective sweeps observed among rice domestication genes (as opposed to diversification loci)



**Fig. 3.** Results from \*BEAST analyses of phylogenetic datasets. Nodes supported with high posterior probability ( $\geq 95\%$ ) by all datasets are shown with a dot. I, J, R, and N represent *indica*, tropical *japonica*, *O. rufipogon*, and *O. nivara*, respectively. The numbers above the bars in the graphs indicate the percentage of trees in the posterior probability distribution with a given topology or that support a single origin of rice. (A) Alternative species trees and their proportions in the posterior distribution resulting from analyses of the Tang et al. (17) (red) and Zhu et al. (38) (yellow) data. (B) Only one well-supported species tree was recovered from the Rakshit et al. (19) data. (C) All trees in the posterior distribution, despite inclusion of the five main cultivar groups [aromatics (Ar), Aus, temperate *japonica* (TmJ), tropical *japonica* (TrJ), and *indica* (I)] in an analysis of the Yu et al. (39) dataset, support a single origin for domesticated taxa. Ri and Rc indicate *O. rufipogon* from India/Indochina and China, respectively.

arise not from introgression between already domesticated *indica* and *japonica* but instead, reflect the single origin of this cultivated crop species.

Previous studies have estimated the divergence time between *indica* and *japonica* at ~86–440 ky ago (14, 15, 17, 21), long predating the domestication of rice. These estimates were derived from application of a molecular clock to divergence estimates between pairs of sequences from *O. sativa* ssp. *japonica* and *O. sativa* ssp. *indica*, and the inferred divergence time was interpreted as evidence that they were derived independently from diverged source populations of *O. rufipogon*. This interpretation of divergence time estimates is not warranted, however, and there is no need to invoke the existence of a deeply structured source population to account for the ancient coalescent time. In the case of recent population divergence, the common ancestor of a pair of alleles drawn at random from each population will have existed long before the appearance of the populations themselves. Therefore, the inferred coalescent times of *indica* and *japonica* alleles may greatly exceed the time since domestication, even if the alleles were derived from a single panmictic progenitor population. This failure to consider ancestral variation has led to erroneous inferences in a number of contexts (47) and seems to have unduly influenced perceptions about rice origins.

When ancestral variation is taken into account with the multispecies coalescent, the timing of divergence between *O. sativa* and *O. rufipogon* and between *indica* and *japonica* is found to be much more recent. The exact divergence time estimate is dependent on which molecular clock rate that we use. If we use an estimate for nucleotide substitution rates in the grasses (40), we find a divergence time between *O. rufipogon* and *O. sativa* at ~8,200 y ago and between tropical *japonica* and *indica* at ~3,900 y ago. If we apply the molecular clock rate estimated from the chromosome scan data (*SI Text*), we obtain an earlier mean date of domestication for rice (13,500 y B.P.). The former molecular estimates are in remarkable agreement with archaeological estimates for the onset of rice domestication in the Yangtze Valley (~8,000–9,000 y ago) and the expansion of *indica* rice in South Asia (~4,000 y ago) (3, 28), and even the latter date still falls within the upper boundary of archaeological dating estimates of rice phytoliths collected from the lower Yangtze (4).

Although our analyses are consistent with a single origin of rice, one possibility is that both *indica* and *japonica* originated from highly differentiated *O. rufipogon* gene pools that were not sampled by both us and the other previously published phylogenetic studies. We think this is unlikely, because to obtain our results, both of these gene pools (one for *indica* and another for *japonica*) must not be represented in our sampling. Moreover, if these gene pools existed, they would have split from each other at about the time of rice domestication to be consistent with our estimates of the timing of the *indica/japonica* split.

Archaeological studies have been interpreted as corroborating phylogenetic evidence for multiple origins of rice. Two centers of rice domestication—the Yangtze River Valley of China and the Ganges in India—have been identified based on the discovery of nonshattering rice spikelet bases in archaeological sites from these regions (3). The oldest archaeological evidence for rice domestication comes from the Yangtze Valley, where *japonica* or a *japonica*-like domesticated rice seems to have been present as early as ~8,000–9,000 y ago (2–4, 28). Rice domestication in the Ganges has also been observed, but rice here seems to be a minor crop (or was gathered wild) and only substantially grew in importance about 4,000 y after rice domestication in the Yangtze Valley (3).

It has been suggested that ancient peoples may have brought *japonica* westward along the Silk Road with other crops such as millet, apricots, and peaches (3). Hybridization of this *japonica* with local proto-*indica* cultivars along with active selection could have rapidly led to the rise and expansion of present day *O. sativa*

ssp. *indica* (3). Other models suggest hybridization of a single domesticate with locally differentiated *O. rufipogon* populations, leading to the present day *indica* and tropical *japonica* (48, 49). Although they differ in details, these models are consistent with a single origin of domesticated rice in the Yangtze Valley followed by spread and hybridization of this original domesticate that eventually led to *indica* (48).

The question of the origin of domesticated rice (or any domesticated species) is a complex problem, because human activity may have eroded genetic signatures that hamper attempts to reconstruct the evolutionary history of these recent human-associated species. Demographic factors, such as rampant admixture, compounded by the effects of prolonged bottleneck during the process of domestication may obscure genetic evidence for domestication models, including those that indicate multiple origins for cultivated species (50). It is clear, however, from our study that incomplete lineage sorting during the coalescent process can explain previous phylogenetic conclusions of the multiple origins of rice. As greater amounts of genome-wide data become available, it would be interesting to see if the results of our analysis are supported by these methods of reconstructing evolutionary history and demographic processes associated with the recent speciations characteristic of domestication.

Several other domesticated taxa, because of marked intraspecific phenotypic or genetic differentiation, also seem to have multiple evolutionary origins. Barley (51), grapes (52), and cucurbits (53) as well as livestock species such as sheep (54) and cattle (55) have been shown to have arisen more than one time, indicating that different cultures have reinvented these domesticated species several times rather than obtaining them through diffusion from other farming societies. Rice was also thought to be a clear example of a domesticated species with multiple origins (9, 15–19), suggesting that the Neolithic cultures of China and India separately led to the domestication of this cereal crop species. It now seems that rice, however, may have arisen only in one geographical region of Asia and that, from this single origin, we now find a food species with a wide geographical and cultural reach that has led to its becoming the major food crop for much of the world's population.

## Materials and Methods

**Resequencing of Gene Fragments on Three Rice Chromosomes.** Our resequencing panel consisted of 20 accessions of *O. rufipogon*, 36 landrace accessions of *O. sativa* (20 *indica* and 16 tropical *japonica*), and one each of *O. nivara*, *O. meridionalis*, and *O. barthii* obtained from the International Rice Research Institute and the US Department of Agriculture (*Table S1*). DNA was extracted, and ~500-bp gene fragments from protein-coding genes spaced at ~100-kb intervals on chromosomes 8, 10, and 12 were resequenced. Details of sequencing, population structure, and diversity analyses are in *SI Text*.

**Selective Sweep Mapping.** We used two different methods to map selective sweeps related to rice domestication from *O. rufipogon* based on local reductions in diversity as well as the multipopulation AFS used in the demographic inference. Details of these analyses are in *SI Text*.

**Demographic Inference.** We tested various single-origin and double-origin models (Fig. 1) using *∂a∂i* (36) (source code available on request). The single-origin models consisted of either *indica* from *japonica* or *japonica* from *indica* demographic scenarios (Fig. 1). The double-origin models can be categorized as either *indica* first or *japonica* first serial founder models, and they posit that each domesticate originated independently, one preceding the other (Fig. 1). We incorporated bottlenecks in all of our models for the founding of *indica* and tropical *japonica* populations. Details of the modeling are found in *SI Text*.

**\*BEAST Analyses of Published Rice Sequence Datasets.** Using the program Species Tree Ancestral Reconstruction/Bayesian Evolutionary Analysis by Sampling Trees (\*BEAST v1.6.1) (35), we reanalyzed six previously published phylogenetic datasets (15–17, 19, 38, 39). We should note that the study by

Tang et al. (17) chose atypical, highly divergent loci in their analyses. Moreover, the dataset by Yu et al. (39) included in our analysis contains possible selected loci, but none of these loci show selective sweeps shared by both *indica* and *japonica*. MrModeltest (56) was run on each locus to determine the best-fit nucleotide evolution models. A Yule prior, which assumes that lineages split at a constant rate (57), was specified for the species tree. Each dataset was analyzed independently in \*BEAST using a strict molecular clock for the reference locus based on the mean substitution neutral rate for grasses (0.0065 substitutions/site per Myr) (46),

from which the rate of the other genes was estimated. Other details on the phylogenetic analyses are in [SI Text](#).

**ACKNOWLEDGMENTS.** The authors would like to thank Dorian Fuller, Andrew Doust, and Joseph Heled for helpful discussions, Chris Smith for helping to develop the SNP quality control pipeline, and Xianfa Xie for helping in choice of some accessions. We would also like to thank Dennis Widjaja, Kelly Clemenza, Naeha Bhambra, Hannah Chaudry, and Sennia Gerard-Martinez for help in processing sequence data. This work was funded in part by the National Science Foundation Plant Genome Research Program.

- Purugganan MD, Fuller DQ (2009) The nature of selection during plant domestication. *Nature* 457:843–848.
- Higham C, Lu TLD (1998) The origins and dispersal of rice cultivation. *Antiquity* 72: 867–877.
- Fuller DQ, et al. (2010) Convergence of genetics and archaeobotany in the entangled history of rice. *Archaeol Anthropol Sci* 2:115–131.
- Liu L, Lee G-A, Jiang L, Zhang J (2007) Evidence for the early beginning (c. 9000 cal. BP) of rice domestication in China: A response. *The Holocene* 17:1059–1068.
- International Rice Genome Sequencing Project 2005. The map-based sequence of the rice genome. *Nature* 436:793–800.
- Oka HI (1988) *Origin of Cultivated Rice* (Elsevier, Amsterdam).
- Li C, Zhou A, Sang T (2006) Genetic analysis of rice domestication syndrome with the wild annual species, *Oryza nivara*. *New Phytol* 170:185–193.
- Lu BR, Naredo MEB, Juliano AB, Jackson MT (2000) *Preliminary Studies on Taxonomy and Biosystematics of the AA Genome Oryza Species (Poaceae) in Grasses: Systematics and Evolution*, eds Jacobs SW, Everett LJ (CSIRO, Melbourne), pp 51–58.
- Cheng C, et al. (2003) Polyphyletic origin of cultivated rice: Based on the interspersed pattern of SINEs. *Mol Biol Evol* 20:67–75.
- Garris AJ, Tai TH, Coburn J, Kresovich S, McCouch S (2005) Genetic structure and diversity in *Oryza sativa* L. *Genetics* 169:1631–1638.
- Matsuo T, Futsuhara Y, Kikuchi F, Yamaguchi H (1997) *Science of the Rice Plant* (Food and Agriculture Policy Research Center, Tokyo).
- Second G (1982) Origin of the genic diversity of cultivated rice (*Oryza* spp.): Study of the polymorphism scored at 40 isoenzyme loci. *Jpn J Genet* 57:25–57.
- Nakano MA, Yoshimura A, Iwata N (1992) Phylogenetic study of cultivated rice and its wild relatives by RFLP. *Rice Genet Newsl* 9:132–134.
- Vitte C, Ishii T, Lamy F, Brar D, Panaud O (2004) Genomic paleontology provides evidence for two distinct origins of Asian rice (*Oryza sativa* L.). *Mol Genet Genomics* 272:504–511.
- Zhu QH, 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.
- 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.
- Tang T, et al. (2006) Genomic variation in rice: Genesis of highly polymorphic linkage blocks during domestication. *PLoS Genet* 2:e199.
- Caicedo AL, et al. (2007) Genome-wide patterns of nucleotide polymorphism in domesticated rice. *PLoS Genet* 3:1745–1756.
- Rakshit S, et al. (2007) Large-scale DNA polymorphism study of *Oryza sativa* and *O. rufipogon* reveals the origin and divergence of Asian rice. *Theor Appl Genet* 114: 731–743.
- Kumagai M, Wang L, Ueda S (2010) Genetic diversity and evolutionary relationships in genus *Oryza* revealed by using highly variable regions of chloroplast DNA. *Gene* 462: 44–51.
- Ma J, Bennetzen JL (2004) Rapid recent growth and divergence of rice nuclear genomes. *Proc Natl Acad Sci USA* 101:12404–12410.
- Sang T, Ge S (2007) Genetics and phylogenetics of rice domestication. *Curr Opin Genet Dev* 17:533–538.
- Sang T, Ge S (2007) The puzzle of rice domestication. *J Integr Plant Biol* 49:760–768.
- Oka H-I, Morishima H (1982) Phylogenetic differentiation of cultivated rice, XXIII. Potentiality of wild progenitors to evolve the *indica* and *japonica* types of rice cultivars. *Euphytica* 31:41–50.
- Li C, Zhou A, Sang T (2006) Rice domestication by reducing shattering. *Science* 311: 1936–1939.
- Tan L, et al. (2008) Control of a key transition from prostrate to erect growth in rice domestication. *Nat Genet* 40:1360–1364.
- 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.
- Fuller DQ, Qin L (2010) Declining oaks, increasing artistry, and cultivating rice: The environmental and social context of the emergence of farming in the Lower Yangtze Region. *Environ Archaeol* 15:139–159.
- Knowles LL, Carstens BC (2007) Delimiting species without monophyletic gene trees. *Syst Biol* 56:887–895.
- Degnan JH, Rosenberg NA (2009) Gene tree discordance, phylogenetic inference and the multispecies coalescent. *Trends Ecol Evol* 24:332–340.
- Kubatko LS, Degnan JH (2007) Inconsistency of phylogenetic estimates from concatenated data under coalescence. *Syst Biol* 56:17–24.
- Zou XH, et al. (2008) Analysis of 142 genes resolves the rapid diversification of the rice genus. *Genome Biol* 9:R49.
- Cranston KA, Hurwitz B, Ware D, Stein L, Wing RA (2009) Species trees from highly incongruent gene trees in rice. *Syst Biol* 58:489–500.
- Liu L, Yu L, Kubatko L, Pearl DK, Edwards SV (2009) Coalescent methods for estimating phylogenetic trees. *Mol Phylogenet Evol* 53:320–328.
- Heled J, Drummond AJ (2010) Bayesian inference of species trees from multilocus data. *Mol Biol Evol* 27:570–580.
- Gutenkunst RN, Hernandez RD, Williamson SH, Bustamante CD (2009) Inferring the joint demographic history of multiple populations from multidimensional SNP frequency data. *PLoS Genet* 5:e1000695.
- 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.
- Zhu Q, Zheng X, Luo J, Gaut BS, Ge S (2007) Multilocus analysis of nucleotide variation of *Oryza sativa* and its wild relatives: Severe bottleneck during domestication of rice. *Mol Biol Evol* 24:875–888.
- 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.
- 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.
- Eckert AJ, Carstens BC (2008) Does gene flow destroy phylogenetic signal? The performance of three methods for estimating species phylogenies in the presence of gene flow. *Mol Phylogenet Evol* 49:832–842.
- Sweeney MT, et al. (2007) Global dissemination of a single mutation conferring white pericarp in rice. *PLoS Genet* 3:e133.
- Zhang LB, et al. (2009) Selection on grain shattering genes and rates of rice domestication. *New Phytol* 184:708–720.
- Kovach MJ, Calingacion MN, Fitzgerald MA, McCouch SR (2009) The origin and evolution of fragrance in rice (*Oryza sativa* L.). *Proc Natl Acad Sci USA* 106: 14444–14449.
- Zhao K, et al. (2010) Genomic diversity and introgression in *O. sativa* reveal the impact of domestication and breeding on the rice genome. *PLoS One* 5:e10780.
- Takano-Kai N, et al. (2009) Evolutionary history of GS3, a gene conferring grain length in rice. *Genetics* 182:1323–1334.
- Charlesworth D (2010) Don't forget the ancestral polymorphisms. *Heredity* 105: 509–510.
- Vaughan DA, Lu BR, Tomooka N (2008) The evolving story of rice evolution. *Plant Sci* 174:394–408.
- Ikehashi H (2009) Why are there *indica* type and *japonica* type in rice? History of the studies and a view for origin of two types. *Rice Sci* 16:1–13.
- Allaby RG, Fuller DQ, Brown TA (2008) The genetic expectations of a protracted model for the origins of domesticated crops. *Proc Natl Acad Sci USA* 105: 13982–13986.
- Morrell PL, Clegg MT (2007) Genetic evidence for a second domestication of barley (*Hordeum vulgare*) east of the Fertile Crescent. *Proc Natl Acad Sci USA* 104: 3289–3294.
- Arroyo-García R, et al. (2006) Multiple origins of cultivated grapevine (*Vitis vinifera* L. ssp. *sativa*) based on chloroplast DNA polymorphisms. *Mol Ecol* 15:3707–3714.
- Sanjurjo OI, Piperno DR, Andres TC, Wessel-Beaver L (2002) Phylogenetic relationships among domesticated and wild species of *Cucurbita* (Cucurbitaceae) inferred from a mitochondrial gene: Implications for crop plant evolution and areas of origin. *Proc Natl Acad Sci USA* 99:535–540.
- Pedrosa S, et al. (2005) Evidence of three maternal lineages in Near Eastern sheep supporting multiple domestication events. *Proc Biol Sci* 272:2211–2217.
- Loftus RT, MacHugh DE, Bradley DG, Sharp PM, Cunningham P (1994) Evidence for two independent domestications of cattle. *Proc Natl Acad Sci USA* 91:2757–2761.
- Nylander JAA (2004) *MrModeltest v2*. (Evolutionary Biology Centre, Uppsala University, Uppsala, Sweden).
- Drummond AJ, Rambaut A (2007) BEAST: Bayesian evolutionary analysis by sampling trees. *BMC Evol Biol* 7:214.