New Phytologist

Selection on grain shattering genes and rates of rice domestication

Lin-Bin Zhang[1], Qihui Zhu[1], Zhi-Qiang Wu[1], Jeffrey Ross-Ibarra[2], Brandon S. Gaut[2], Song Ge[1] and Tao Sang[3]

[1]State Key Laboratory of Systematic and Evolutionary Botany, Institute of Botany, Chinese Academy of Sciences, Beijing, 100093, China; [2]Department of Ecology and Evolutionary Biology, University of California, Irvine, CA 92697, USA; [3]Department of Plant Biology, Michigan State University, East Lansing, MI 48824, USA

Author for correspondence:
Song Ge
Tel: +86 10 62836097
Email: gesong@ibcas.ac.cn

Received: *15 May 2009*
Accepted: *19 June 2009*

New Phytologist (2009) **184**: 708–720
doi: 10.1111/j.1469-8137.2009.02984.x

Key words: artificial selection, domestication rate, nucleotide diversity, *Oryza*, *qSH1*, *sh4*.

Summary

- Molecular cloning of major quantitative trait loci (QTLs) responsible for the reduction of rice grain shattering, a hallmark of cereal domestication, provided opportunities for in-depth investigation of domestication processes.
- Here, we studied nucleotide variation at the shattering loci, *sh4* and *qSH1*, for cultivated rice, *Oryza sativa* ssp. *indica* and *Oryza sativa* ssp. *japonica*, and the wild progenitors, *Oryza nivara* and*Oryza rufipogon*.
- The nonshattering *sh4* allele was fixed in all rice cultivars, with levels of sequence polymorphism significantly reduced in both *indica* and *japonica* cultivars relative to the wild progenitors. The *sh4* phylogeny together with the neutrality tests and coalescent simulations suggested that *sh4* had a single origin and was fixed by artificial selection during the domestication of rice. Selection on *qSH1* was not detected in *indica* and remained unclear in *japonica*.
- Selection on *sh4* could be strong enough to have driven its fixation in a population of cultivated rice within a period of *c*. 100 yr. The slow fixation of the nonshattering phenotype observed at the archeological sites might be a result of relatively weak selection on mutations other than *sh4* in early rice cultivation. The fixation of *sh4* could have been achieved later through strong selection for the optimal phenotype.

Introduction

Cereal crops such as maize, rice and wheat were domesticated from wild grasses between 7000 and 10 000 yr ago (Doebley *et al.*, 2006). In all cereal crops, the transition to domestication included a dramatic reduction in grain shattering (Harlan, 1992). The mature grains of wild grasses dissociate easily from the panicle to ensure successful seed dispersal. However, grain shattering makes harvest difficult. Archeological evidence suggests that before domestication, gatherers collected immature grains by cutting off panicles before the onset of shattering (Tanno & Willcox, 2006). The development of nonshattering cultivars substantially improved grain yield and is considered to be the hallmark of cereal domestication.

Asian cultivated rice (*Oryza sativa*) was domesticated from a pair of closely related wild species, *Oryza nivara* and *Oryza rufipogon*, distributed from southeastern Asia to India. Two major types of rice cultivars, *O. sativa* subspecies *indica* and subspecies *japonica*, were domesticated independently from diverged wild populations with distinct genomic backgrounds (Kovach *et al.*, 2007; Sang & Ge, 2007a,b; Sweeney & McCouch, 2007; Vaughan *et al.*, 2008). Genetic analyses of grain shattering between cultivars and wild progenitors have identified a major quantitative trait locus (QTL), *sh4*, which accounts for 69% of phenotypic variation between *indica* rice and the wild annual species *O. nivara* (Li *et al.*, 2006a). Positional cloning of *sh4* has shown that a single nonsynonymous substitution in an Myb3 DNA-binding domain leads to incomplete development and partial function of the abscission zone, resulting in a reduction in grain shattering (Li *et al.*, 2006b). Another QTL, *qSH1*, which explains 69% of the shattering difference between *indica* and temperate *japonica* cultivars, has also been cloned (Konishi *et al.*, 2006). The causative mutation is a nucleotide substitution approx. 12 kb upstream of a homeobox gene. The mutation substantially reduces gene expression and disrupts the development of the abscission zone.

For each shattering gene, the detection of major QTLs with a single underlying causative mutation suggests the possibility of rapid phenotypic evolution. That is, nonshattering alleles may have been fixed quickly under strong artificial selection;

such a scenario has intuitive appeal given the importance of nonshattering for grain yield. This hypothesis, however, does not seem to agree with recent archeological evidence that suggests rice domestication was a relatively slow process. Archaeological observation from the Lower Yangtze region of China suggests that fixation of the nonshattering phenotype of rice could have taken two to three millennia (Fuller *et al.*, 2009). This is concordant with archeological data that indicates the development of nonshattering wheat and barley was also a gradual process that lasted well over a millennium (Tanno & Willcox, 2006; Fuller, 2007). Preliminary genetic data thus seem to be at odds with archaeological evidence concerning the rate of rice domestication.

The rate and timing of domestication continues to be a topic of broad interest and debate. The shattering phenotype is integral to the process of domestication, and thus provides a unique opportunity to synthesize genetic and archaeological evidence to advance our knowledge on the rate of domestication. Here, we investigate nucleotide variation of *sh4* and *qSH1* using samples representing all major subdivisions of rice cultivars and populations of wild progenitors collected from the entire geographic range of potential rice domestication. In contrast to previous surveys of genetic diversity at *sh4* and *qSH1* (Lin *et al.*, 2007; Onishi *et al.*, 2007), the sampling strategy here allows us to estimate, for the first time, the extent and strength of artificial selection at the shattering loci. Given the strength of selection, we also estimate the time required to fix *sh4* in rice cultivars. The results yield new insights into the dynamic processes of rice domestication, especially with regard to the potential rate of rice domestication.

Materials and Methods

Plant material

We sampled 31 accessions of cultivated rice (*O. sativa* L.), including 17 accessions of *O. sativa* ssp. *indica* and 14 accessions of *O. sativa* ssp. *japonica*. This sample represents all five recognized subgroups within *O. sativa* (Garris *et al.*, 2005). A total of 44 accessions of the wild progenitors were sampled, including 28 accessions of *O. rufipogon* and 16 accessions of *O. nivara* (Table 1). These collections covered a wide distributional range of the wild species from southeastern Asia to India, including all potential areas for rice domestication (see the Supporting Information, Fig. S1). Of the samples, 30 cultivars and 28 wild accessions were previously studied by Zhu *et al.* (2007). *Oryza barthii* was used as an outgroup for phylogenetic analyses. Seeds of the sampled accessions were germinated and DNA was isolated from seedlings.

DNA sequencing

A *c.* 2700-bp region of the *sh4* gene was amplified using PCR primers InF 5′-ACCAGCTCAACTCAACAACG and BR 5′-GTGTCAAAGCCTTAATGTGC. This region includes the entire first exon and *c.* 1900 bp of 5′ upstream region of the gene. For *qSH1*, we amplified a *c.* 1600-bp region surrounding the previously reported functional single-nucleotide polymorphism (SNP) using primers q_rF1 5′-GGCTACGAGCGATAAATCTG and q_rR2 5′-GCTACAGGCTCCTACTATAC. The PCR and internal sequencing primers were designed based on rice genome sequences (GenBank accession AP008210 and AL606619 for *sh4* and AB071330 and AB071331 for *qSH1*). The *sh4* region was amplified with *Takara LA Taq* DNA polymerase with GC buffer I (Takara, Shiga, Japan). The *qSH1* region was supplemented with Ex*Taq* DNA polymerase (Takara). The PCR products were purified with either a Pharmacia purification kit (Amersham Pharmacia Biotech) or a Dinggou purification kit (Dingguo, Beijing, China). For cultivated rice, purified PCR products were sequenced directly on both strands. For all individuals of wild species *O. rufipogon* and *O. nivara*, PCR products were cloned into p*GEM* T-easy vectors (Promega). At least six clones were sequenced initially to obtain both alleles from the heterozygous individuals at a locus. Singletons were confirmed by reamplification and resequencing. All sequences were deposited into GenBank, with accession numbers EU999788–EU999948.

Sequence analyses and neutrality tests

Sequence alignments were performed using CLUSTALX (Thompson *et al.*, 1997) and were refined manually. Genetic variation of each gene region was estimated with average pairwise differences per base pair between sequences (π) (Nei & Li, 1979) and Watterson's estimate (θ_w) based on the number of segregating sites (Watterson, 1975) using DnaSP version 4.10 (Rozas *et al.*, 2003). Under the standard neutral model for an autosomal gene from a random-mating population of a constant size, both π and θ_w are expected to be equal to $4N_e\mu$, where N_e represents the effective population size and μ represents the mutation rate per generation per site (Watterson, 1975).

Selection on the shattering genes was tested using multiple methods. Departure from neutrality was tested at segregating nucleotide sites with Tajima's *D* and Fu and Li's *D** and *F** using the program DNASP (Rozas *et al.*, 2003). Tajima's *D* measures the difference between the mean pairwise differences (π) and Watterson's estimator (θ_w) (Tajima, 1989). Fu and Li's *D** and *F** test the discrepancy between the number of polymorphic sites in external branches, or polymorphisms unique to a sequence, and number of polymorphic sites in internal phylogenetic branches, or polymorphisms shared by the sequences (Fu & Li, 1993).

We also performed maximum likelihood Hudson–Kreitman–Aguade (MLHKA) test (Wright & Charlesworth, 2004). The maximum likelihood ratio test assesses departure from neutrality at a focal locus compared with neutral standards. The degree of increase or decrease of polymorphism caused by selection is measured as *k*. Here seven neutrally evolving rice

New Phytologist

Table 1 Plant materials sampled in this study and information of geographic origin and functional single-nucleotide polymorphism (SNP)

Taxon	Accession[a]	Country of origin	Code[b]	Cultivar subgroups and collection localities of wild species[c]	Functional SNP[d]	
					sh4	*qSH1*
Oryza sativa ssp. *indica*	10594	Indonesia	ind-IDN	NA	T	G
$n_{sh4} = 17$	11713	Sri Lanka	ind-LKA1	NA	T	G
$n_{qSH1} = 17$	11724	Sri Lanka	ind-LKA2	NA	T	G
	30416	Philippine	ind-PHL	*indica*	T	G
	71503	Malaysia	ind-MYS	*indica*	T	G
	74716	India	ind-IND	NA	T	G
	RA4870	Vietnam	ind-VNM	*indica*	T	G
	RA4890	Iran	ind-IRN	*aus*	T	G
	RA4892	Madagascar	ind-MDG	NA	T	G
	RA4921	Thailand	ind-THA	*indica*	T	G
	RA4939	Bhutan	ind-AUS2	*aus*	T	G
	RA4966	Brazil	ind-BRA	*indica*	T	G
	RA4976	Afghanistan	ind-AFG	*aus*	T	G
	RA5020	Japan	ind-JPN	*indica*	T	G
	RA5321	Bangladesh	ind-AUS3	*aus*	T	G
	ZH249	China	ind-CHN	NA	T	G
	CL16	China	ind-CHN1	NA	T	G
O. sativa ssp. *japonica*	Nipponbare	Japan	jap-JPN	Temperate *japonica*	T	T
$n_{sh4} = 13$	11169	Philippines	jap-PHL	Temperate *japonica*	T	G
$n_{qSH1} = 14$	69317	France	jap-FRA	NA	T	G
	70932	Pakistan	jap-PAK	NA	T	G
	Au70140	Greece	jap-GRC	NA	T	G
	Au8134	Vietnam	jap-VNM	NA	T	G
	RA4875	USA	jap-USA2	Tropical *japonica*	T	G
	RA4913	India	jap-IND	Aromatic	T	G
	RA4950	Brazil	jap-BRA	Aromatic	T	G
	RA4972	South Korea	jap-KOR	Temperate *japonica*	T	T
	RA5123	USA	jap-USA1	Temperate *japonica*	T	T
	RA5363	Australia	jap-AUS	Temperate *japonica*	T	G
	RA5364	Nigeria	jap-NRA	Tropical *japonica*	T	G
	Taizhong 65	China	jap-CHN	NA	–	G
Oryza rufipogon	80505	India	ruf_IND1	NA	G	G
$n_{sh4} = 25$	80506	India	ruf_IND2	NA	G	G
$n_{qSH1} = 19$	80534	India	ruf_IND3	19/82	G	G
	80542	India	ruf-IND	19/81	G	G
	80774	Philippines	ruf-PHL	NA	G	G
	103423	Sri Lanka	ruf-LKA	8/81	G	G
	104311	Thailand	ruf-THA1	NA	G	G
	105494	Myanmar	ruf-MMR	NA	G	G
	105720	Cambodia	ruf-KHM	11/104	G	G
	105942	Thailand	ruf-THA	6/101	G	G
	105958	Indonesia	ruf-IDN	–7/106	G	G
	105960	Bangladesh	ruf-BGD	21/92	G	G
	106161	Laos	ruf-LAO	18/102	G	G
	106453	Indonesia	ruf-IDN1	4/104	G	–
	0112	Jiangxi, China	ruf-CHN	28/116	G	–
	0114	Jiangxi, China	ruf-CHN7	28/116	G	–
	0318	Jiangxi, China	ruf-CHN2	28/116	–	G
	2507	Guangdong, China	ruf-CHN11	22/112	G	–
	2510	Guangdong, China	ruf-CHN3	22/112	–	G
	5213	Hainan, China	ruf-CHN4	20/110	G	G
	6102	Guangxi, China	ruf-CHN5	22/109	–	G
	6109	Guangxi, China	ruf-CHN8	22/109	G	–
	6110	Guangxi, China	ruf-CHN9	22/109	G	–
	6113	Guangxi, China	ruf-CHN10	22/109	G	–
	6803	Yunnan, China	ruf-CHN6	24/102	G	G
	Yuan1–10	Yunnan, China	ruf-CHN12	23/103	G	–
	Yuan3–9	Yunnan, China	ruf-CHN13	23/103	G	–
	VOC4	Nepal	ruf-NPL	NA	G	G

Table 1 *continued*

Taxon	Accession[a]	Country of origin	Code[b]	Cultivar subgroups and collection localities of wild species[c]	Functional SNP[d]	
					sh4	*qSH1*
Oryza nivara	80470	India	niv-IND3	NA	G	G
$n_{sh4} = 16$	93191	Nepal	niv-NPL1	28/81	G	–
$n_{qSH1} = 13$	101967	India	niv-IND4	19/82	G	G
	103407	Sri Lanka	niv-LKA1	NA	G/T	G
	103416	Sri Lanka	niv-LKA2	8/80	G	G
	104687	India	niv-IND1	25/73	G	G
	105319	India	niv-IND2	10/76	G	G
	105391	Thailand	niv-THA	15/100	G	G
	105431	Sri Lanka	niv-LKA	6/81	G	G
	105705	Nepal	niv-NPL	28/81	G	G
	105734	Cambodia	niv-KHM1	11/104	G	–
	105742	Cambodia	niv-KHM	12/104	G/T	G
	106061	India	niv-IND	22/85	G	G
	106148	Laos	niv-LAO	18/102	G	G
	106155	Laos	niv-LAO1	18/102	G	–
	106345	Myanmar	niv-MMR	17/96	G	G
Oryza barthii	106234	Mali	bar-MLI		G	G

[a]Accession numbers preceded by RA were provided by Dr Susan McCouch at Cornell University; ZH249, VOC4, CL16, Yuan1–10, Yuan3–9 and the accessions with four digits were collected by the authors; all the other accessions were obtained from the Genetic Resources Center of the International Rice Research Institute (IRRI) at Los Banos, the Philippines.
[b]Individual accession is abbreviated by first three letters of the taxon name followed by the code of its origin of country.
[c]The cultivar subgroup classification followed Garris *et al.* (2005). The collection localities (latitude/longitude) of wild accessions are obtained from the IRRI. NA indicates the classification or locality information is not available for an accession.
[d]Nucleotide sequence at the functional SNP of *sh4* (Li *et al.*, 2006a) and *qSH1* (Konishi *et al.*, 2006). Dashes indicate that we were unable to amplify the gene from this accession using the primers.

genes (*Adh1*, *GBSSII*, *Ks1*, *Lhs1*, *Os0053*, *SSII1*, *TFIIAγ-1*; Zhu *et al.*, 2007) were used as reference sequences and *O. barthii* was used as the outgroup for the MLHKA tests. The test was performed using the program MLHKA provided by S. I. Wright (http://www.yorku.ca/stephenw/). With different random number seeds, Markov chain lengths of 100 000 were run in the MLHKA tests to specify each gene in each accession, and at least three independent runs per model were performed to assess the maximum likelihood values.

To further characterize selection on *sh4* and *qSH1* in the context of the demographic dynamics of rice domestication, we conducted a coalescent simulation (CS) as detailed in Tenaillon *et al.* (2004). This evaluates whether the domestication bottleneck alone could account for the reduction of nucleotide diversity of cultivated rice in comparison with its wild progenitors. In CS analyses, the statistic is a likelihood ratio of the best-fitting demographic model for the focal genes (*sh4* and *qSH1*) against an estimate of the demographic history of neutrally evolving genes, in this case *Adh1*, *GBSSII*, *Ks1*, *Lhs1*, *Os0053*, *SSII1* and *TFIIAγ-1*. We used parameter estimates for the seven neutral genes from Zhu *et al.* (2007). To estimate demographic parameters for *sh4* and *qSH1* we assumed that rice cultivation began 10 000 yr ago, and that the domestication bottleneck spanned 3000 generations, consistent with archaeological evidence (Khush, 1997). Given this model, we explored the bottleneck severity K (Wright *et al.*, 2005) over a grid of 25 values (0.0001, 0.0005, 0.001, 0.005, 0.01, 0.05, 0.1, 0.15, 0.2, 0.25, 0.3, 0.35, 0.4, 0.45, 0.5, 0.55, 0.6, 0.65, 0.7, 0.75, 0.8, 0.85, 0.9, 0.95, 1). The number of segregating sites (S) in cultivated rice was used as the goodness-of-fit statistic for all simulations. Results for *sh4* and *qSH1* were based on 10 000 simulations of the best-fitting coalescent model.

Phylogenetic analyses

The phylogeny of sequences was inferred by Neighbor-joining (NJ), maximum likelihood (ML), maximum parsimony (MP) and Bayesian inference (BI) methods. Neighbor-joining, ML and MP were conducted with PAUP 4.0b10 (Swofford, 2002). Bootstrap analyses were performed with 1000 replicates. Bayesian analyses were conducted using MRBAYES version 3.1 (Huelsenbeck & Ronquist, 2001; Ronquist & Huelsenbeck, 2003) with Markov chain Monte Carlo (MCMC) estimation of posterior probability distributions. Four chains of the MCMC were run each for 1 000 000 generations and were sampled. An initial 'burn-in' phase of 100 000 generations was discarded. We then saved one tree every 100 generations and constructed a majority-rule consensus tree of the 9000 trees sampled in

each run. The appropriate nucleotide substitution model for each data set was determined by MODELTEST 3.7 (Posada & Crandall, 1998). Optimization of the model and parameter search was made by using of Akaike information criterion (AIC, Sakamoto *et al.*, 1986).

Estimation of the time since the fixation of the *sh4* allele

We used two different approaches to estimate the time required to fix the nonshattering *sh4* allele in cultivated rice. Based on the initial frequency of nonshattering *sh4* allele ($p = 1/2N$, where N is the assumed population size) and selection coefficient (s), we first estimated the time required for *sh4* fixation (T_f) using equations from Kimura & Ohta (1969):

$$T_f = \frac{1}{s(1-e^{-2S})}\int_P^1 \frac{(e^{2S\xi}-1)(e^{-2S\xi}-e^{-2S})}{\xi(1-\xi)}d\xi + \frac{1-u(p)}{u(p)}\frac{1}{s(1-e^{-2S})}\int_0^P \frac{(e^{2S\xi}-1)(1-e^{-2S\xi})}{\xi(1-\xi)}d\xi$$

where $S = Ns$, and $u(p) = \frac{1-e^{-2Sp}}{1-e^{-2S}}$.

To calculate T_f, s was estimated using the relationship $d_{phys} = 0.01\ s/c$ (Kaplan *et al.*, 1989), where the physical distance of a selective sweep (d_{phys}) was obtained from the aligned genome sequences of rice chromosome 4 between an *indica* (GLA4) and a *japonica* (Nipponbare) cultivar and the local recombination rate c was calculated as described in Zhao *et al.* (2002).

Second, we used an approximate Bayesian method to investigate a simple model of a selective sweep in a bottlenecked, derived population (Thornton & Jensen, 2007). We first simulated a deterministic trajectory of the selected site forward in time, then a structured coalescent backward in time, conditional on the simulated allelic trajectory. For both *indica* and *japonica*, we simulated an ancestral population of size $N_A = N_1 + fN_1$ which splits at time t in the past into a modern wild population of size N_1 and a bottlenecked domesticated population of size fN_1 (see the Supporting Information, Fig. S2). At time *t-d* the bottlenecked population recovered to size N_1 (Fig. S2). We simulated a selective sweep in the bottlenecked populations starting shortly after the bottleneck began and ending at time *t-d*. We assumed $t = 10\,000$ yr and $N_1 = 250\,000$, in close agreement with archaeological data (Zhao & Piperno, 2000; Zong *et al.*, 2007) and diversity at neutral loci (Zhu *et al.*, 2007). Model parameters for each simulation were drawn from previous probability distributions, with $f \sim U(0.001, 0.5)$ and $s \sim U(0.0001, 1)$. The bottleneck duration d was set to be slightly longer than the expected duration of the selective sweep given f and s, using the equation $d = f \times \ln(2fN_1) \times (2[fN_1 + 0.5]s) + 10^{-5}$ (modified from Thornton & Jensen (2007) and Stephan *et al.* (1992)). The population mutation rate θ in each simulation was taken from previous estimates of nucleotide diversity (Zhu *et al.*, 2007), and the population recombination rate ρ was set equal to θ. We assumed that the population recombination rate, calculated as $\rho = 4N_1c$, where c is the recombination rate per bp per generation, was equal to θ. We simulated a locus of 1800 silent sites and a number of alleles for each population identical to the observed sample sizes at *sh4*.

For each simulated dataset we calculated the number of segregating sites (S) and nucleotide diversity (π) at silent sites in both domesticated and wild population. Datasets with values of all four summary statistics within 20% of the observed values at locus *sh4* were retained, and the selection coefficient (s), duration of the bottleneck (d), and severity of the bottleneck (f) were recorded. Recorded values of each parameter were used to estimate acceptance rates and form the posterior distribution.

Results

Molecular diversity of *sh4* and *qSH1*

In total we obtained *sh4* sequences from 30 accessions of cultivated rice and 41 accessions of the wild progenitors, and *qSH1* sequences from 31 accessions of rice cultivars and 32 accessions of the wild species (Tables 1, 2). These sequences were aligned with the outgroup, *O. barthii*, generating alignments of 2740 bp and 1575 bp for *sh4* and *qSH1*, respectively.

It has been shown that the G to T substitution of *sh4*, which results in an amino acid substitution from lysine to asparagine, was responsible for the reduction of grain shattering during the domestication of rice (Li *et al.*, 2006b). At the functional SNP site, all of the alleles from cultivars have a thymine (T) residue, whereas almost all of those from the wild accessions, including the outgroup, have a guanine (G) residue (Table 1). Two accessions of *O. nivara*, one from Sri Lanka (niv-LKA1) and the other from Cambodia (niv-KHM), are T/G heterozygotes, which is most likely the result of introgression of the nonshattering allele from cultivars into the wild populations (Li *et al.*, 2006b).

Phylogenetic analysis of the *sh4* dataset showed that all of the cultivars with T at the functional SNP site were grouped together with 65% bootstrap support (Fig. 1a). The alleles with T at the functional SNP site from the two heterozygous individuals of *O. nivara* were nested within cultivars. Thus, this monophyletic group of *sh4* sequences with T at the functional SNP site contains all nonshattering alleles derived from domestication. Within this group, *indica* and *japonica* cultivars are intermixed. Immediately outside of this clade, two wild accessions, an *O. nivara* from India and an *O. rufipogon* from Laos, form a sister group to the cultivated alleles with 74% bootstrap support. Both of the alleles in the sister group contain a G at the functional SNP site.

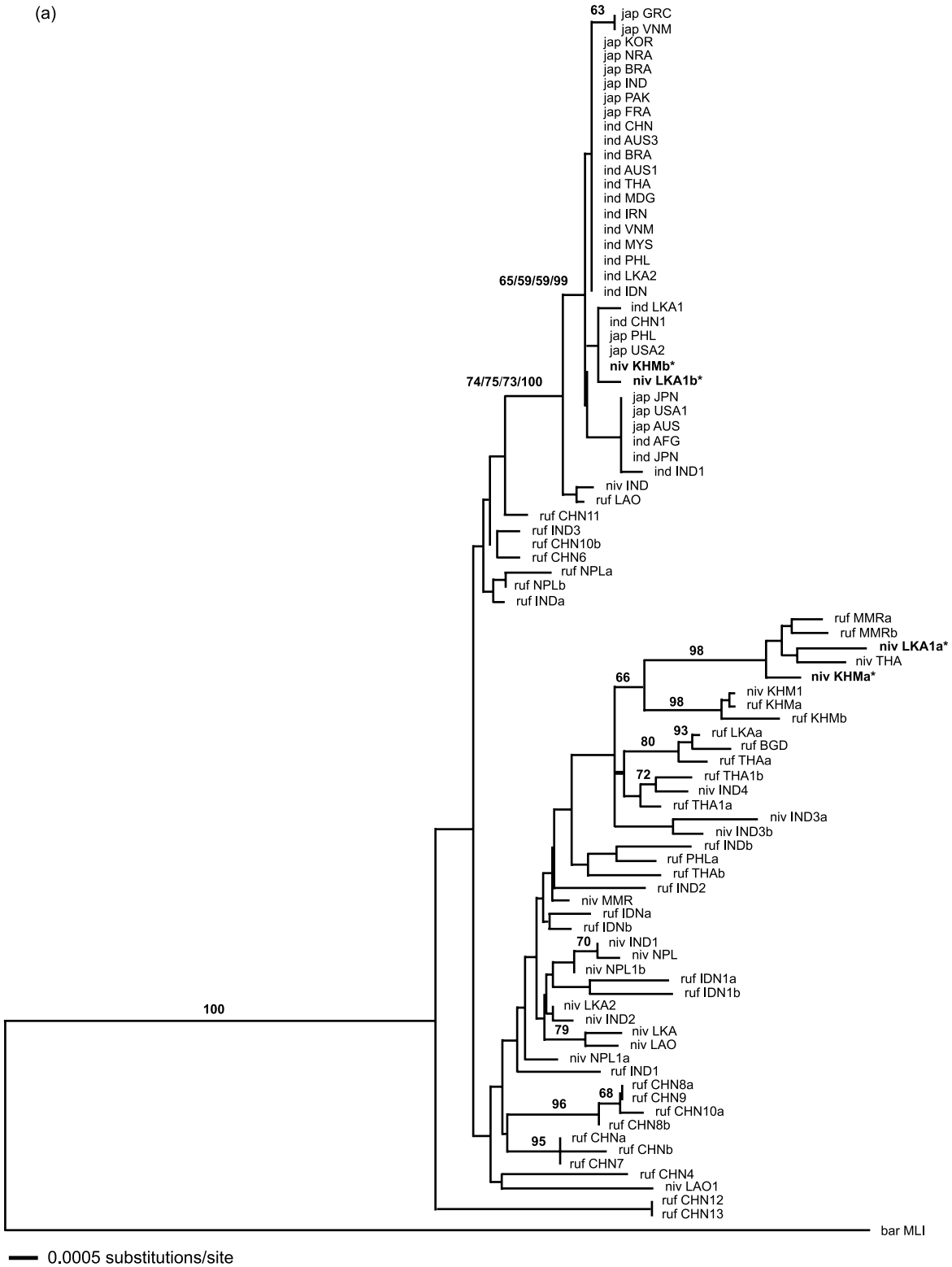

Fig. 1 The phylogeny of shattering genes. (a) Neighbor-joining (NJ) tree of *sh4*. (b) Neighbor-joining tree of *qSH1*. Maximum parsimony (ML), Maximum likelihood (MP) and Bayesian inference (BI) methods yielded essentially the same topology. Number-associated branches are NJ bootstrap values above 50%. Statistical support from the NJ, MP, ML and BI analyses are given at two clades most relevant to discussion. *, a and b following an accession name indicate two heterozygous alleles found from the same individual of this accession. **, Three *japonica* cultivars with nonshattering single-nucleotide polymorphism (SNP) at qSH1.

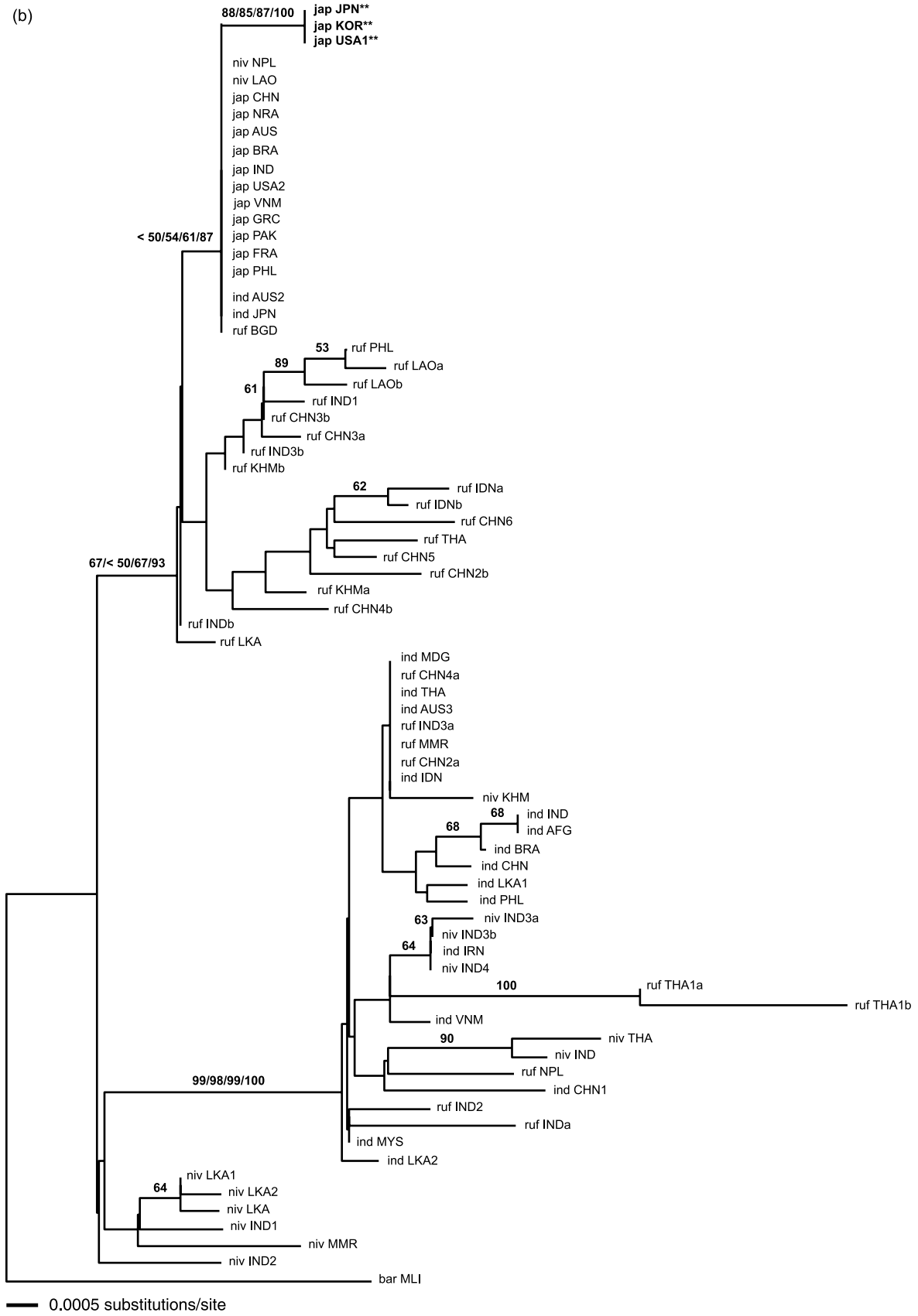

Fig. 1 continued

Table 2 Estimates of nucleotide variation

Locus	Taxon	N^a	L_T^b	H^c	S^d	π_T^e	θ_T^f	π_{sil}^g	θ_{sil}^h	R_M^i
sh4	*Oryza sativa* ssp. *indica*	34	2450.5	4	3	0.0003	0.0003	0.0004	0.0004	0
	O. sativa ssp. *japonica*	26	2450.3	3	3	0.0004	0.0003	0.0006	0.0004	0
	Oryza rufipogon	50	2454.1	32	64	0.0042	0.0060	0.0053	0.0080	8
	Oryza nivara	32	2449.6	20	45	0.0038	0.0046	0.0049	0.0060	8
qSH1	*O. sativa* ssp. *indica*	34	1542.0	12	21	0.0032	0.0033	0.0032	0.0033	3
	O. sativa ssp. *japonica*	28	1540.2	2	2	0.0005	0.0003	0.0005	0.0003	0
	O. rufipogon	38	1536.3	22	45	0.0054	0.0076	0.0054	0.0076	4
	O. nivara	26	1545.2	12	33	0.0057	0.0056	0.0057	0.0056	1

[a]Total number of sequences.
[b]Average length (bp) of the sequences from each species.
[c]Total number of haplotypes at an individual locus in each taxon.
[d]Total number of polymorphic sites.
[e]Average number of pairwise nucleotide difference per site calculated based on the total number of polymorphic sites.
[f]Watterson's estimator of θ per base pair calculated based on the total number of polymorphic sites.
[g]Average number of pairwise nucleotide differences per site calculated based on silent sites.
[h]Watterson's estimator of θ per base pair calculated based on silent sites.
[i]R_M, the minimum number of recombination events.

Table 3 Tests for neutrality and selection

Locus	Taxon	Tajima's *D*	Fu and Li's *D**	Fu and Li's *F**	MLHKA[a]	k^b
sh4	*Oryza sativa* ssp. *indica*	0.2452	0.9345	0.8510	2.60×10^{-6}	0.073
	O. sativas ssp. *japonica*	0.6963	0.9684	1.0301	2.13×10^{-5}	0.060
	O. rufipogon	–1.0419	0.0809	–0.3965	0.193	0.392
	O. nivara	–0.6242	1.1935	0.6964	0.042	0.237
qSH1	*O. sativa* ssp. *indica*	–0.1250	1.6582**	1.2688	0.182	2.080
	O. sativa ssp. *japonica*	0.7509	0.8154	0.9199	0.046	0.209
	O. rufipogon	–1.0477	–0.0001	–0.4281	0.826	0.805
	O. nivara	0.0086	1.5296**	1.2329	0.773	0.800

[a]Maximum likelihood Hudson–Kreitman–Aguadé (MLHKA) test was performed with *O. barthii* as the outgroup and seven neutrally evolving rice genes as the reference. *P*-values are reported.
[b]The maximum likelihood estimate of the selection parameter in MLHKA tests.
For Tajima's *D*, Fu and Li's *D**, and Fu and Li's *F**, only two significant ($P < 0.05$) test values were obtained, which are indicated by ** ($0.001 < P \leq 0.01$).

At the *qSH1* locus, three temperate *japonica* individuals were found homozygous T/T at the SNP site that confers the reduced shattering phenotype. The remaining accessions of cultivars, including two temperate *japonica* individuals, and all wild accessions, were homozygous G/G, conferring the shattering phenotype (Table 1; Fig. 1b). The three temperate *japonica* accessions with reduced shattering genotype form a monophyletic group with 88% bootstrap support. This group is nested in a clade that contains the rest of the *japonica* cultivars together with two *indica*, two *O. nivara* and one *O. rufipogon* accession. The remaining *indica* cultivars are intermixed with *O. nivara* and *O. rufipogon* in a well-supported clade.

With regard to genetic variation at the shattering loci, cultivars have lower levels of nucleotide diversity at both loci than the wild progenitors. However, the degree of reduction in nucleotide diversity differs between the two loci and between the cultivars. For *sh4*, the cultivars have the same level of nucleotide diversity ($\theta_{sil} = 0.0004$ for both *indica* and *japonica*), which is 15- to 20-times lower than that of *O. nivara* or *O. rufipogon* (Table 2). At the *qSH1* locus, however, the similar level of reduction in nucleotide diversity occurs only in *japonica*. The *indica* accessions contain about half of the nucleotide diversity of the wild species (Table 2).

Neutrality tests

We conducted several neutrality tests to determine whether the reduction in nucleotide diversity in rice cultivars could be caused by artificial selection. The MLHKA test, performed in reference to seven neutral genes (Zhu *et al.*, 2007), indicated that a selective sweep could be the primary cause of the severe reduction in nucleotide polymorphism in both *indica* and *japonica* cultivars at the *sh4* locus ($P < 0.0001$; Table 3). Because this could reflect selection on *sh4* in wild populations

Table 4 The *P*-values of the coalescent simulation test for selection using seven reference genes

Locus	*indica* vs wild	*japonica* vs wild	Cultivated vs wild
sh4	0.016	0.047	0.002
qSH1	0.939	0.395	0.930

(Yamasaki *et al.*, 2005), we also applied MLHKA to the *O. nivara* and *O. rufipogon* samples (Table 3); the test of neutrality is borderline significant ($P = 0.042$) for *O. nivara* but not significant for *O. rufipogon* ($P = 0.193$). The results thus indicate that there was strong artificial selection on *sh4* during rice domestication. At the *qSH1* locus, the MLHKA test was marginally significant in *japonica* ($P = 0.046$) but not significant for either *indica* or the wild species (Table 3).

Unlike the HKA test, Tajima's *D* and Fu and Li's D^* and F^* rejected neutrality in only two instances of 24 statistics obtained (Tajima, 1989; Fu & Li, 1993; Table 3). These were Fu and Li's D^* for *qSH1* of *indica* and *O. nivara* ($P < 0.01$). However, those nonsignificant statistics are likely to be an artifact of population demographics (Tajima, 1989; Nielsen, 2001), and do not necessarily support neutrality. After populations undergo subdivision and bottlenecks (Garris *et al.*, 2005; Zhu *et al.*, 2007) there is an excess of alleles of intermediate frequency that leads to higher Tajima's *D* values (Das *et al.*, 2004). In addition, theoretical work (Innan & Kim, 2004) has shown that the multilocus HKA test is more powerful in detecting artificial selection than the single-locus Tajima's *D* when polymorphisms is much reduced.

To determine whether the low nucleotide diversity at *sh4* and *qSH1* could be explained by a genetic bottleneck alone, we conducted the CS test in reference to seven neutrally evolving genes. We began by inferring a bottleneck model with the seven neutral genes, and then compared polymorphism in *sh4* and *qSH1* with the expectation of the model (Table 4). The results indicated that variation detected at *sh4* could not be explained by a domestication bottleneck alone for *indica* ($P = 0.016$), *japonica* ($P = 0.047$) or a combined sample from both cultivars ($P = 0.002$). By contrast, variation at *qSH1* was not significantly different from the expectation under a bottleneck scenario for *indica* ($P = 0.939$), *japonica* ($P = 0.395$) or the combined sample ($P = 0.930$). Together, the results of the MLHKA and CS tests suggest that *sh4* was fixed in cultivated rice under artificial selection. The extent to which *qSH1* was selected in *japonica* rice remains unclear.

Timing of *sh4* fixation under artificial selection

Taking advantage of the wealth of genomic information for rice chromosome 4, we estimated the time required to drive the nonshattering allele to fixation by artificial selection. We first calculated the selection coefficient (s) using the equation $d_{phys} = 0.01\ s/c$ (Kaplan *et al.*, 1989), where c is the recombination rate and d_{phys} is the physical distance between a selected and hitchhiking neutral site. Based on aligned genome sequences of rice chromosome 4 between an *indica* (GLA4) and a *japonica* (Nipponbare) cultivar, we observed hypervariable regions 21 kb upstream and 43 kb downstream of *sh4*. In this *c.* 66 kb region containing *sh4*, there were 34 SNPs between the *indica* and *japonica* cultivars. The SNP density of 0.52 SNPs per kb around the *sh4* locus thus is much lower than 3.5 SNPs per kb for the entire chromosome 4 (Feng *et al.*, 2002; B. Han, pers. comm.). This suggests that the selective sweep spanned a 66 kb region, thus signifying a 33 kb mean distance between the selected site and the farthest sweep site.

Given the local recombination rate around the *sh4* locus of approx. 1.122×10^{-7}/bp (B. Han, Pers. Comm.), s is calculated as 0.37. Using the previously reported nucleotide diversity at silent sites ($\theta_{sil} = 0.0024$–0.00292) for *O. sativa* (Caicedo *et al.*, 2007; Zhu *et al.*, 2007) and a substitution rate of 5.9×10^{-9} substitutions per synonymous site per year for grasses (Gaut *et al.*, 1996; White & Doebley, 1999), we estimated that the effective population size ranged from 10 200–12 300. To be cautious, we used a much wider range of the effective population size of *O. sativa*, between 1000 and 100 000, for the calculation of fixation time. With estimates of s and N_e, we calculated the fixation time (T_f) of *sh4* according to Kimura & Ohta (1969). Assuming the initial allele frequency of $1/2N$, T_f for *sh4* was estimated to be between 76 and 112 yr.

Although our estimate of s was based on detailed genome sequence data, we still could not rule out potential bias associated with the calculation. Thus, a simulation method was also employed for estimating fixation time. Simulations incorporating a demographic model like that of Thornton & Jensen (2007) generate similar insights about the timing of the selective sweep. The joint posterior probability distributions of the bottleneck duration d and the selective coefficient s are shown in Fig. 2. The marginal posterior distributions for both parameters in both subspecies are leptokurtic, with the highest probability assigned to very low values but with long, relatively fat tails. The medians of the posterior distribution of f (*japonica* 0.034 and *indica* 0.043) and s (*japonica* 0.187 and *indica* 0.235) point to a dramatic reduction in population size and relatively strong selection. Under this model (see the Materials and Methods section), the duration of the bottleneck (d) is closely related to the length of time to fixation of the selective alleles. The median of the posterior distribution of d (*japonica* 118 yr, *indica* 97 yr) suggests that the fixation of *sh4* could have occurred in an extremely short period of time, which is concordant with the earlier calculation based on estimated s. Estimates of the duration based on the joint posterior distributions of s and f are similarly rapid, whether based on median values (*japonica* 86 yr and *indica* 72 yr) or the peak (*japonica* 308 yr and *indica* 280 yr) of the distribution. This suggests that the fixation of *sh4* could have occurred in a short period of time ranging from < 100 yr to only a few hundred years.

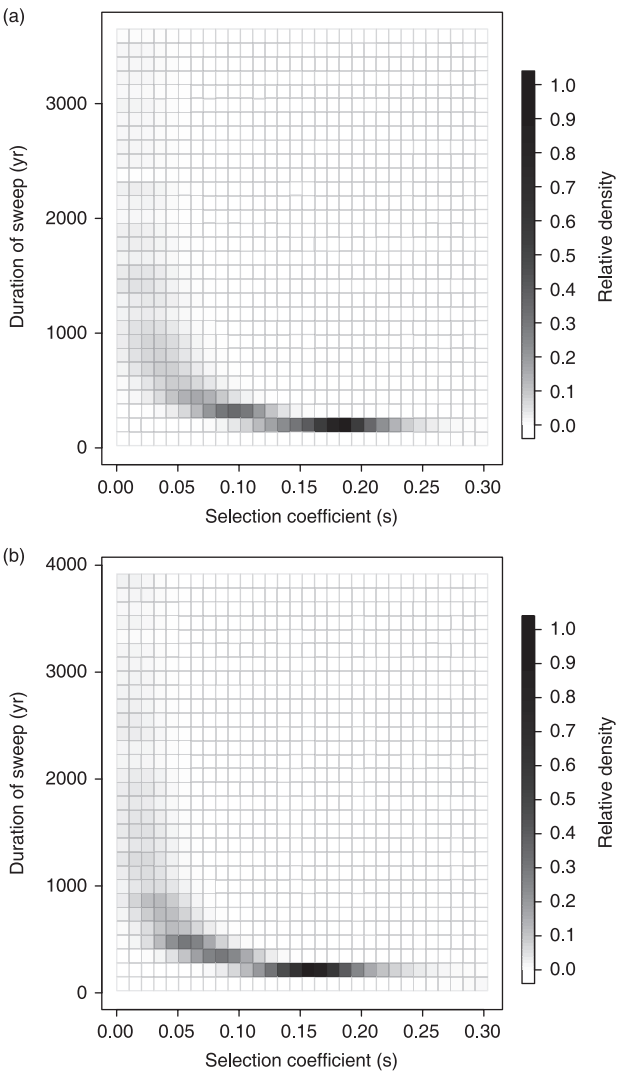

Fig. 2 Joint relative probability distribution of the selection coefficients (s) and the duration of sweep (*d*) for *sh4*. Darker squares indicate higher probability. (a) *Oryza sativa* ssp. *indica*; (b) *Oryza sativa* ssp. *japonica*.

Discussion

The question of whether a crop has had single or multiple origins has attracted considerable attention from biologists, archeologists and social scientists (Diamond, 2002; Doebley *et al.*, 2006; Zeder, 2006; Allaby *et al.*, 2008). Multiple origins are of interest because they imply that similar phenotypic changes occurred under parallel artificial selection imposed by people from different cultures and geographic regions. The inference and comparison of evolutionary history of genomes of domesticated species and domestication related genes provides an effective approach to study the origins of crops. Taking maize as an example, the phylogenies of genome-wide markers and a key domestication gene, *tb1*, are concordant in suggesting that there had been a single domestication event (Wang *et al.*, 1999; Matsuoka *et al.*, 2002; Clark *et al.*, 2004). For barley, phylogenetic analyses of neutral DNA markers and domestication genes have also been in agreement, but they suggest that there have been at least two independent domestications of the crop (Azhaguvel & Komatsuda, 2007; Komatsuda *et al.*, 2007; Morrell & Clegg, 2007).

In the case of rice domestication, previous phylogenetic analyses based on a variety of molecular markers have shown that *indica* and *japonica* cultivars fall into separate clades, thus supporting independent domestications of the two cultivars (Second, 1982; Cheng *et al.*, 2003; Zhu & Ge, 2005; Londo *et al.*, 2006). However, here we found that all rice cultivars form a single monophyletic group on the *sh4* phylogeny (Fig. 1a). Moreover, the sequence polymorphism at the *sh4* locus is significantly reduced in both cultivars relative to their wild progenitors, most likely as a result of artificial selection. With a sample size representing all subdivisions of rice cultivars and both wild progenitors across their distribution ranges, these phylogenetic and population genetic analyses strongly support a single origin of the nonshattering *sh4* allele and subsequent fixation in cultivated rice by artificial selection. The nonshattering allele at another locus, *qSH1*, is restricted to a portion of temperate *japonica* and does not appear to have been subjected to artificial selection in *indica* (Fig. 1b; Tables 2–4). The strength and extent of artificial selection imposed on *qSH1* in *japonica* rice has yet to be substantiated and may require studies of additional *japonica* cultivars.

The incongruence between genome-wide and domestication-gene phylogenies in rice may be explained by two mechanisms (Sang & Ge, 2007a,b). First, the earliest rice cultivar may have fixed a set of critical domestication alleles including *sh4*, which then spread into populations of the wild progenitors, *O. nivara* and *O. rufipogon*, through introgression. Under this model, the modern cultivars, *indica* and *japonica*, were derived from the hybrids between this early cultivar and the genetically divergent wild populations (the snowballing model; Sang & Ge, 2007a). Second, the domestication of rice may have started from divergent wild populations in different localities, giving rise to earlier cultivars possessing different sets of domestication alleles. The subsequent crosses among them may have allowed farmers to select the best set of domestication alleles, which are now fixed in modern cultivars (the combination model; Sang & Ge, 2007a). In either case, introgression, important in driving natural speciation (Rieseberg *et al.*, 2003), has played an essential role in the domestication of rice.

While incongruence between domestication gene and genome-wide phylogenies can be explained by models that invoke gene flow, the apparent contradiction between archaeological and genetic evidence for the timing of domestication requires additional consideration. Considering at first the functional consequence of the *sh4* mutation, the reduction in shattering is caused by the substitution of a neutral amino acid, asparagine, for a positively charged lysine in the predicted DNA-binding domain of *sh4*. The mutation weakens but

does not eliminate the function of *sh4*, and it leads to the incomplete development and partial function of the abscission zone that controls grain abscission (Li *et al.*, 2006b; T. Sang, unpublished). Thus, the mutation in *sh4* reduces shattering enough for mature grains so that they can be harvested with tools such as sickles. The partial function of *sh4* may have been maintained to allow grains to be separated and recovered from the straws by simple, traditional techniques such as beating panicles against a wood tub. Because of the superior phenotype, the mutation could, in theory, be quickly fixed in cultivars under strong artificial selection. Our population genetic analyses support this conjecture by suggesting that the fixation of the nonshattering *sh4* may have occurred in about 100 yr.

The genetic evidence for rapid fixation of *sh4* contrasts sharply with archeological evidence for the slow increase in the proportion of nonshattering rice grains grown in the Lower Yangtze region of China between 6900 and 6600 yr ago (Fuller *et al.*, 2009). To better understand this discrepancy, we need to recognize first that nonshattering includes two components, with one being advantageous and the other being disadvantageous. The advantage is to allow the harvest of mature grains at once without considerable grain loss. A tradeoff of the nonshattering phenotype is the added difficulty in grain recovery. If it takes too much labor and energy to separate grains from harvested panicles, such cultivars might have not been acceptable to farmers who preferred slightly immature grains, which were easy to thresh, over mature grains that were difficult to recover from straw. Thus, nonshattering is not the only important trait for domestication selection. Perhaps the balance between nonshattering and threshing determined the overall magnitude of benefit of mutations for cereal domestication (Sang, 2009). As a consequence, the frequency of the nonshattering phenotype may not have increased very quickly until farmers came across a mutation that balanced these two processes.

Although mutations that benefited both harvest and threshing practices were clearly advantageous, this type of mutation must be much less frequent than loss-of-function mutations, which could cause the complete loss of the abscission layer and lead to the over-reduction of shattering (Hu *et al.*, 1964; Ji *et al.*, 2006). Completely nonshattering cultivars may have arisen periodically during rice cultivation, but may not have become widely adopted by early farmers. A mutation of *sh4* that ideally balanced harvest and threshing would have allowed farmers to use simple harvesting and threshing techniques and tools to achieve higher yields. The selection for such a mutation would be very strong and could eventually drive its fixation in all cultivars given sufficient opportunities for introgression (Sang, 2009). This could also explain why, unlike some other domestication-related alleles such as *rc* that arose independently from different mutations (Sweeney *et al.*, 2006, 2007), *sh4* had a single origin and became fixed in all rice cultivars.

However, it is possible that rice had been cultivated or even partially domesticated for some time before the appearance and fixation of *sh4*. The slow increase in proportion of nonshattering grains at the Lower Yangtze archeological sites could be a result of relatively weak selection at nonshattering loci other than *sh4*, which could have conferred limited advantage of harvest efficiency, with a nonshattering phenotype that was either too weak or too strong. Under this hypothesis, it is not surprising to find relatively slow fixation of a crucial domestication trait such as nonshattering at some archeological sites of early crop domestication (Purugganan & Fuller, 2009). The hypothesis is consistent with our finding that *O. rufipogon* populations of China are not closely related to cultivars on the *sh4* phylogeny. Instead, an *O. nivara* from Indian and an *O. rufipogon* from Laos form the sister group of the cultivars, suggesting that *sh4* could have originated in areas outside of China, although reliable determination of the location for initial *sh4* fixation requires much more extensive sampling, especially in the southern Himalayas where *indica* domestication is suggested to have occurred (Londo *et al.*, 2006). Similar archeological studies in this region may also help paint a more complete picture of the history and process of the domestication of rice.

The possibility that domestication preceded the global fixation of *sh4* lends further support to the combination model of domestication (Sang & Ge, 2007a), which seems to be consistent with the growing evidence from both neutral markers and other domestication-related genes (Kovach *et al.*, 2007; Gao & Innan, 2008; Konishi *et al.*, 2008). Under this model, it must have taken time, possibly hundreds of years, for the nonshattering *sh4* allele to spread into areas where rice had already been cultivated. Thus, our estimated fixation time of around 100 yr may principally reflect the original selection event in the local populations but does not account for the subsequent distribution of the *sh4* allele throughout the globe. Thus, a highly beneficial domestication allele could quickly sweep through local populations under strong artificial selection, yet could not do the same in geographically isolated populations without human-mediated gene flow. To better understand the process and estimate the time of fixation of such alleles, population genetic models are needed that consider both the fixation of strongly selected alleles in location populations and migration of those alleles to other locales (Ross-Ibarra & Gaut, 2008).

Acknowledgements

We thank B. Han (Beijing Institute of Genomics, CAS) for sharing unpublished data, C. K. Wang and Y. F. Xie (SLU) for mathematical assistance, the International Rice Research Institute (Los Banos, Philippines) and S. McCouch (Cornell University) for providing some of the plant material, and anonymous reviewers for valuable comments on the earlier version of the manuscript. This work was supported by the National Basic Research Program of China (2007CB815704), National Natural Science Foundation of China (30121003, 30430030), and the grants from the Chinese Academy of

Sciences to S.G., and by a United States National Science Foundation grant (DBI 0321467) to B.S.G.

References

Allaby RG, Fuller DQ, Brown TA. 2008. The genetic expectations of a protracted model for the origins of domesticated crops. *Proceedings of the National Academy of Sciences, USA* **105**: 13982–13986.

Azhaguvel P, Komatsuda T. 2007. A phylogenetic analysis based on nucleotide sequence of a marker linked to the brittle rachis locus indicates a diphyletic origin of barley. *Annals of Botany* **100**: 1009–1015.

Caicedo A, Williamson S, Hernandez RD, Boyko A, Fledel-Alon A, York T, Polato N, Olsen K, Nielsen R, McCouch SR *et al.* 2007. Genome-wide patterns of nucleotide polymorphism in domesticated rice. *PLoS Genetics* **3**: e163.

Cheng C, Motohashi R, Tsuchimoto S, Fukuta Y, Ohtsubo H, Ohstubo E. 2003. Polyphyletic origin of cultivated rice: based on the interspersion pattern of SINEs. *Molecular Biology and Evolution* **20**: 67–75.

Clark RM, Linton E, Messing J, Doebley JF. 2004. Pattern of diversity in the genomic region near the maize domestication gene *tb1*. *Proceedings of the National Academy of Sciences, USA* **101**: 700–707.

Das A, Mohanty S, Stephan W. 2004. Inferring the population structure and demography of *Drosophila ananassae* from multilocus data. *Genetics* **168**: 1975–1985.

Diamond J. 2002. Evolution, consequences and future of plant and animal domestication. *Nature* **418**: 700–707.

Doebley JF, Gaut BS, Smith BD. 2006. The molecular genetics of crop domestication. *Cell* **127**: 1309–1321.

Feng Q, Zhang Y, Hao P, Wang S, Fu G, Huang Y, Li Y, Zhu J, Liu Y, Hu X *et al.* 2002. Sequence and analysis of rice chromosome 4. *Nature* **420**: 316–320.

Fu YX, Li WH. 1993. Statistical tests of neutrality of mutations. *Genetics* **133**: 693–709.

Fuller DQ. 2007. Contrasting patterns in crop domestication and domestication rates: recent archaeobotanical insights from the Old World. *Annals of Botany* **100**: 903–924.

Fuller DQ, Qin L, Zheng Y, Zhao Z, Chen X, Hosoya LA, Sun GP. 2009. The domestication process and domestication rate in rice: spikelet bases from the lower Yangtze. *Science* **323**: 1607–1610.

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.

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

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*. *Proceedings of the National Academy of Sciences, USA* **93**: 10274–10279.

Harlan JR. 1992. *Crops and man*, 2nd edn. Madison, WI, USA: American Society of Agronomy and Crop Science Society of America.

Hu CH, Kao KN, Chang CC. 1964. Histological and genetic studies on shedding and lodging habits of rice plants. *Botanical Bulletin of Academia Sinica* **5**: 170–180.

Huelsenbeck JP, Ronquist F. 2001. MrBayes: Bayesian inference of phylogeny. *Bioinformatics* **17**: 754–755.

Innan H, Kim Y. 2004. Pattern of polymorphism after strong artificial selection in a domestication event. *Proceedings of the National Academy of Sciences, USA* **101**: 10667–10672.

Ji H-S, Chu S-H, Jiang W, Cho Y-I, Hahn J-H, Eun M-Y, McCouch SR, Koh H-J. 2006. Characterization and mapping of a shattering mutant in rice that corresponds to a block of domestication genes. *Genetics* **173**: 995–1005.

Kaplan NL, Hudson RR, Langley CH. 1989. The 'hitchhiking effect' revisited. *Genetics* **123**: 887–899.

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

Kimura M, Ohta T. 1969. The average number of generations until fixation of a mutant gene in a finite population. *Genetics* **61**: 763–771.

Komatsuda T, Pourkheirandish M, He C, Azhaguvel P, Kanamori H, Perovic D, Stein N, Graner A, Wicker T, Tagiri A *et al.* 2007. Six-rowed barley originated from a mutation in a homeodomain-leucine zipper I-class homeobox gene. *Proceedings of the National Academy of Sciences, USA* **104**: 1424–1429.

Konishi S, Ebana K, Izawa T. 2008. Inference of the japonica rice domestication process from the distribution of six functional nucleotide polymorphisms of domestication-related genes in various landraces and modern cultivars. *Plant Cell Physiology* **49**: 1283–1293.

Konishi S, Izawa T, Lin SY, Ebana K, Fukuta Y, Sasaki T, Yano M. 2006. An SNP caused loss of seed shattering during rice domestication. *Science* **312**: 1392–1396.

Kovach MJ, Sweeney MT, McCouch SR. 2007. New insights into the history of rice domestication. *Trends in Genetics* **23**: 578–587.

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

Li C, Zhou A, Sang T. 2006b. Genetic analysis of rice domestication syndrome with the wild annual species, *Oryza nivara*. *New Phytologist* **170**: 185–194.

Lin Z, Griffith M, Li X, Zhu Z, Tan L, Fu Y, Zhang W, Wang X, Xie D, Sun C. 2007. Origin of seed shattering in rice (*Oryza sativa* L.). *Planta* **226**: 11–20.

Londo JP, Chiang YC, Hung KH, Chiang TY, Schaal BA. 2006. Phylogeography of Asian wild rice, *Oryza rufipogon*, reveals multiple independent domestications of cultivated rice, *Oryza sativa*. *Proceedings of the National Academy of Sciences, USA* **103**: 9578–9583.

Matsuoka Y, Vigouroux Y, Goodman MM, Sanchez GJ, Buckler E, Doebley J. 2002. A single domestication for maize shown by multilocus microsatellite genotyping. *Proceedings of the National Academy of Sciences, USA* **99**: 6080–6084.

Morrell PL, Clegg MT. 2007. Genetic evidence for a second domestication of barley (*Hordeum vulgare*) east of the Fertile Crescent. *Proceedings of the National Academy of Sciences, USA* **104**: 3289–3294.

Nei M, Li WH. 1979. Mathematical model for studying genetic variation in terms of restriction endonucleases. *Proceedings of the National Academy of Sciences, USA* **76**: 5269–5273.

Nielsen R. 2001. Statistical tests of selective neutrality in the age of genomics. *Heredity* **86**: 641–647.

Onishi K, Takagi K, Kontani M, Tanaka T, Sano Y. 2007. Different patterns of genealogical relationships found in the two major QTL causing reduction of seed shattering during rice domestication. *Genome* **50**: 757–766.

Posada D, Crandall KA. 1998. Modeltest: testing the model of DNA substitution. *Bioinformatics* **14**: 817–818.

Purugganan MD, Fuller DQ. 2009. The nature of selection during plant domestication. *Nature* **457**: 843–848.

Rieseberg LH, Raymond O, Rosenthal DM, Lai Z, Livingstone K, Nakazato T, Durphy JL, Schwarzbach AE, Donovan LA, Lexer C. 2003. Major ecological transitions in wild sunflowers facilitated by hybridization. *Science* **301**: 1211–1216.

Ronquist F, Huelsenbeck JP. 2003. MrBayes 3: Bayesian phylogenetic inference under mixed models. *Bioinformatics* **19**: 1572–1574.

Ross-Ibarra J, Gaut BS. 2008. Multiple domestications do not appear monophyletic. *Proceedings of the National Academy of Sciences, USA* **105**: E105.

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

Sakamoto Y, Ishiguro M, Kitagawa G. 1986. *Akaike Information Criterion Statistics*. Tokyo, Japan/Dordrecht, the Netherlands: KTK Scientific Publishers/D. Reidel Publishing.

Sang T. 2009. Genes and mutations underlying domestication transitions in grasses. *Plant Physiology* **149**: 63–70.

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

Sang T, Ge S. 2007b. Genetics and phylogenetics of rice domestication. *Current Opinion in Genetics & Development* **17**: 533–538.

Second G. 1982. Origin of the genic diversity of cultivated rice (*Oryza* spp.): study of the polymorphism scored at 40 isozyme loci. *Japanese Journal of Genetics* **57**: 25–57.

Stephan W, Wiehe T, Lenz MW. 1992. The effect of strongly selected substitutions on neutral polymorphism – analytical results based on diffusion-theory. *Theoretical Population Biology* **41**: 237–254.

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

Sweeney MT, Thomson MJ, Cho YG, Park YJ, Williamson SH, Bustamante CD, McCouch SR. 2007. Global dissemination of a single mutation conferring white pericarp in rice. *PLoS Genetics* **3**: 1418–1425.

Sweeney MT, Thomson MJ, Pfeil BE, McCouch SR. 2006. Caught red-handed: Rc encodes a basic helix-loop-helix protein conditioning red pericarp in rice. *Plant Cell* **18**: 283–294.

Swofford DL. 2002. *PAUP*: phylogenetic analysis using parsimony (*and other methods), version 4.0*. Sunderland, MA, USA: Sinauer Associates.

Tajima F. 1989. Statistical method for testing the neutral mutation hypothesis by DNA polymorphism. *Genetics* **123**: 585–595.

Tanno KI, Willcox G. 2006. How fast was wild wheat domesticated? *Science* **311**: 1886.

Tenaillon MI, U'Ren J, Tenaillon O, Brandon SG. 2004. Selection versus demography: a multilocus investigation of the domestication process in maize. *Molecular Biology and Evolution* **21**: 1214–1225.

Thompson JD, Gibson TJ, Plewniak F, Jeanmougin F, Higgins DG. 1997. The Clustal X windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools. *Nucleic Acids Research* **25**: 4876–4882.

Thornton KR, Jensen JD. 2007. Controlling the false positive rate in multilocus genome scans for selection. *Genetics* **175**: 737–750.

Vaughan DA, Lu B-R, Tomooka N. 2008. The evolving story of rice evolution. *Plant Science* **174**: 394–408.

Wang RL, Stec A, Hey J, Lukens L, Doebley J. 1999. The limits of selection during maize domestication. *Nature* **398**: 236–239.

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

White S, Doebley J. 1999. The molecular evolution of *terminal ear1*, a regulatory gene in the genus Zea. *Genetics* **153**: 1455–1462.

Wright SI, Charlesworth B. 2004. The HKA test revisited: a maximum-likelihood-ratio test of the standard neutral model. *Genetics* **168**: 1071–1076.

Wright SI, Bi IV, Schroeder SG, Yamasaki M, Doebley JF, McMullen MD, Gaut BS. 2005. The effects of artificial selection on the maize genome. *Science* **308**: 1308–1310.

Yamasaki M, Tenaillon MI, Vroh Bi I, Schroeder SG, Sanchez-Villeda H, Doebley JF, Brandon SG, McMullen MD. 2005. A large-scale screen for artificial selection in maize identifies candidate agronomic loci for domestication and crop improvement. *Plant Cell* **17**: 2859–2872.

Zeder MA. 2006. Central questions in the domestication of plants and animals. *Evolutionary Anthropology: Issues, News, and Reviews* **15**: 105–117.

Zhao Z, Piperno DR. 2000. Late Pleistocene/Holocene environments in the middle Yangtze River Valley, China and rice (*Oryza sativa* L.) domestication: the phytolith evidence. *Geoarchaeology* **15**: 203–222.

Zhao Q, Zhang Y, Cheng Z, Chen M, Wang S, Feng Q, Huang Y, Li Y, Tang Y, Zhou B *et al.* 2002. A fine physical map of the rice chromosome 4. *Genome Research* **12**: 817–823.

Zhu Q, Ge S. 2005. Phylogenetic relationships among A-genome species of the genus *Oryza* revealed by intron sequences of four nuclear genes. *New Phytologist* **167**: 249–265.

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. *Molecular Biology and Evolution* **24**: 875–888.

Zong Y, Chen Z, Innes JB, Chen C, Wang Z, Wang H. 2007. Fire and flood management of coastal swamp enabled first rice paddy cultivation in east China. *Nature* **449**: 459–462.

Supporting Information

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

Fig. S1 Geographic locations from which the wild *Oryza* species were sampled in this study.

Fig. S2 Diagram of the coalescent model used in simulation.

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