JOURNAL OF Evolutionary Biology

doi: 10.1111/jeb.12171

The role of *Bh4* in parallel evolution of hull colour in domesticated and weedy rice

C. C. VIGUEIRA, W. LI & K. M. OLSEN

Department of Biology, Washington University, St. Louis, MO, USA

Keywords:

de-domestication;
Oryza glaberrima;
Oryza sativa;
parallel evolution;
weedy rice.

Abstract

The two independent domestication events in the genus *Oryza* that led to African and Asian rice offer an extremely useful system for studying the genetic basis of parallel evolution. This system is also characterized by parallel de-domestication events, with two genetically distinct weedy rice biotypes in the US derived from the Asian domesticate. One important trait that has been altered by rice domestication and de-domestication is hull colour. The wild progenitors of the two cultivated rice species have predominantly black-coloured hulls, as does one of the two U.S. weed biotypes; both cultivated species and one of the US weedy biotypes are characterized by straw-coloured hulls. Using *Black hull 4* (*Bh4*) as a hull colour candidate gene, we examined DNA sequence variation at this locus to study the parallel evolution of hull colour variation in the domesticated and weedy rice system. We find that independent *Bh4*-coding mutations have arisen in African and Asian rice that are correlated with the straw hull phenotype, suggesting that the same gene is responsible for parallel trait evolution. For the U.S. weeds, *Bh4* haplotype sequences support current hypotheses on the phylogenetic relationship between the two biotypes and domesticated Asian rice; straw hull weeds are most similar to *indica* crops, and black hull weeds are most similar to *aus* crops. Tests for selection indicate that Asian crops and straw hull weeds deviate from neutrality at this gene, suggesting possible selection on *Bh4* during both rice domestication and de-domestication.

Introduction

Understanding the molecular basis of parallel evolution, the occurrence of the same derived trait in different lineages, can provide fundamental insights into the process of adaptive evolution. Studies in coat colour patterning in animals have revealed that mutations in the melanocortin-1 receptor (*Mc1r*) gene have led to the repeated evolution of lighter colour in taxa as diverged as mice (Hoekstra *et al.*, 2006), mammoths (Römpler *et al.*, 2006) and lizards (Rosenblum *et al.*, 2010). Conversely, within a single species, there can be different genes that underlie repeated trait evolution, as seen in the convergence of light coat colour in different populations of beach mice (Steiner *et al.*, 2009). Domesticated plants provide one of the most useful study systems for examining this process of parallel evolution because many crop species are characterized by a shared suite of traits, termed the 'domestication syndrome' (Hammer, 1984), that are selectively favoured during the domestication process; these include traits involved in easier harvest (e.g. reduced seed dispersal or shattering, uniform and more erect plant architecture, uniform flowering and seed maturation) and desirability for human consumption (e.g. larger fruits or seeds and decreased plant defences).

Recent studies have revealed both similarities and differences among crop species in the underlying genetic targets of selection for domestication traits. For example, in sorghum, reduced shattering arises through several loss-of-function mutations at *Sh1*, encoding a putative YABBY transcription factor (Lin *et al.*, 2012); orthologs of this same gene appear to have been targeted during selection for reduced shattering in maize, rice and foxtail millet (Paterson *et al.*, 1995; Lin *et al.*, 2012). On the other hand, the rice *Sh1* ortholog controls

Correspondence: Kenneth M. Olsen, Department of Biology, Washington University, St. Louis, MO 63130, USA.
Tel.: +314 935 7013; fax: +314 935 4432; e-mail: kolsen@wustl.edu

only a minor-effect shattering QTL, and a different shattering gene, *Sh4*, appears to have been the major target of selection during rice domestication (Li *et al.*, 2006).

Rice, which is the most important food crop for most of the world's population, serves as a particularly effective system to study the genetics of parallel evolution during domestication. Within the genus *Oryza,* there have been at least two independent domestication events, corresponding to the globally cultivated Asian rice (*O. sativa*) and the minor crop African rice (*O. glaberrima*) (Fig. 1). Asian rice was the first crop genome to be completely sequenced and mapped (Goff *et al.*, 2002; Yu *et al.*, 2002), allowing for genomics tools to be utilized in discovering the genetic basis of important domestication traits. In addition, because African and Asian rice are phylogenetically closely related within the *Oryza* genus (both have AA genomes; Morishima *et al.*, 1992), comparative genomics approaches can be readily applied to study the parallel evolution of domestication traits in these crop species.

Asian rice was domesticated from its wild progenitor, *O. rufipogon*, starting about 7000 years ago in southern Asia (Fuller *et al.*, 2009). There is controversy in the literature pertaining to the number of independent domestication events within Asian rice (Londo *et al.*, 2006; Molina *et al.*, 2011; Yang *et al.*, 2012); the two major *O. sativa* subspecies, *indica* and *japonica,* are phenotypically distinct and derived from genetically diverged ancestral gene pools (Caicedo *et al.*, 2007), potentially reflecting two independent domestications.

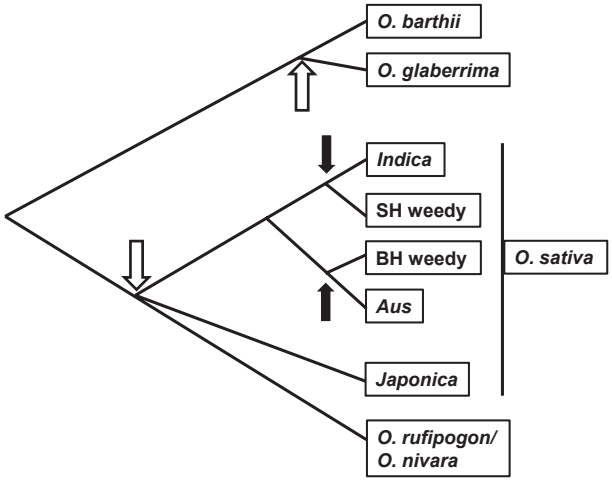

Fig. 1 Schematic phylogeny of relationships between rice species included in this study. White arrows indicate two separate domestication events: African rice (*Oryza glaberrima*) from its wild progenitor (*Oryza barthii*) and Asian rice (*Oryza sativa*) from its wild progenitor (*Oryza rufipogon/Oryza nivara*). Black arrows indicate de-domestication events leading to United States weedy rice biotypes: straw-hulled weedy rice (SH) from *indica* crops and black-hulled weedy rice (BH) from *aus* crops.

African cultivated rice (*O. glaberrima*) was domesticated from the African wild species, *O. barthii,* about 3500 years ago in the Niger River Delta and is only cultivated in West Africa (Li *et al.*, 2011). Some of the same domestication traits have been selectively favoured in both Asian and African rice (e.g. straw hulls, larger seeds and white pericarps); however, African rice is highly variable and often possesses traits characteristic of wild *Oryzas* (Sarla & Swamy, 2005).

Not only has rice undergone multiple domestication events, but it has also undergone multiple de-domestications, leading to the evolution of weedy rice. Weedy crop relatives have evolved in a number of crop species (reviewed by Vigueira *et al.*, 2013) and are characterized by the re-emergence of many of the progenitor-like phenotypes that were selected against during domestication, including seed shattering, seed dormancy and asynchronous reproduction. In the context of parallel evolution, de-domestication makes it possible to examine the genetic mechanisms accounting for both domestication traits as well as the phenotypic reversions to more progenitor-like characteristics.

Weedy rice (also called red rice) is an interfertile form of *O. sativa* that invades cultivated rice fields in the United States (US) and across the world, contaminating rice harvests with its dark, unpalatable grains and reducing yields by up to 80% (Estorninos *et al.*, 2005). Recent microsatellite and DNA sequence data indicate that US weedy rice is composed of two genetically diverged groups that are closely related to cultivated Asian rice within the *indica* subspecies (Londo & Schaal, 2007; Reagon *et al.*, 2010). The two weedy rice groups have several phenotypic differences, but are most easily discerned by their hull colour and are thus commonly referred to as black hull (BH) and straw hull (SH) weedy rice. A less common brown hull (BrH) phenotype appears to reflect rare hybridization events between BH and SH weed strains (Reagon *et al.*, 2010). BH weeds are closely related to *aus* cultivated rice varieties, a subgroup of *indica* rice grown in north-eastern India and Bangladesh, whereas SH weeds share a closer relationship with *indica* varieties *sensu stricto*, the other major subgroup within the *indica* subspecies (Londo & Schaal, 2007; Reagon *et al.*, 2010; see also Garris *et al.*, 2005) (Fig. 1). Given that *indica* and *aus* varieties have never been cultivated commercially in the US, these patterns suggest that the US weeds evolved through two de-domestication events in Asia and were subsequently introduced into the US, possibly as contaminants of seed stocks. Consistent with this hypothesis, coalescent simulations place the divergence of the weed strains from their putative domesticated progenitors well before rice cultivation began in the US (Reagon *et al.*, 2010).

In addition to distinguishing the major US weed strains, hull colour is also a trait that differs between wild and domesticated rice. All wild *Oryza* species are

characterized by the presence of a black hull, whereas most Asian rice cultivars and many African cultivars have undergone selection for straw hulls. The selective pressures regulating hull colour are not well understood. One hypothesis is that black hulls provide effective camouflage for wild *Oryzas* from bird predation, as the shattering seeds are likely to land in dark mud substrates, whereas the nonshattering grains of domesticated rice are better camouflaged if they have straw hulls that blend in with dried leaves (Zhu *et al.*, 2011). Straw hulls might also have been favoured during domestication simply because they can serve as a visual marker to distinguish cultivated plants from nearby wild populations.

Two or three genes are thought to control the black hull phenotype (Kuriyama & Kudo, 1967); however, the biochemical basis of hull pigmentation remains uncharacterized. One recently described major-effect gene responsible for black hull vs. straw hull phenotype is *Black hull 4* (*Bh4*) on rice chromosome IV, which encodes an amino acid transporter that is only expressed in maturing hulls (Zhu *et al.*, 2011). In a survey of *Bh4* sequence variation of 29 cultivated *O. sativa* accessions and 23 wild *O. rufipogon* accessions (including annual forms that are sometimes classified as a separate species, *O. nivara*), Zhu *et al.* (2011) documented three different loss-of-function mutations that generate the straw hull phenotype in domesticated Asian rice. In a larger PCR-genotyping screen of 433 straw hull landraces from China, they found that about 95% of sampled accessions carry one of these mutations, a 22-bp frameshift deletion in exon 3; about 4% carry a 1-bp frameshift deletion in exon 1. A single straw hulled *aus* cultivar, Kasalath, was found to have a SNP in exon 3 that results in an early stop codon. No loss-of-function mutations were found in *Bh4* in about 1% of the straw-hulled varieties surveyed, indicating that the straw hull phenotype can arise through genetic mechanisms other than those altering the *Bh4* protein-coding sequence; these might involve *Bh4 cis*-regulatory changes or mutations at other genes altogether. When testing for selection on *Bh4* using the maximum-likelihood Hudson–Kreitman–Aguade (MLHKA) test, performed in reference to seven neutral genes, Zhu *et al.* found significant deviations from neutrality for cultivated rice, but not for wild rice. This supports the hypothesis of a selective sweep on *Bh4* during Asian rice domestication. Zhu and colleagues were able to rescue the black hull phenotype in an *indica* plant carrying the common loss-of-function allele by expressing a transgenic copy of the functional gene; however, they were not able to rescue the black hull phenotype in a *japonica* variety, providing further evidence that one or more genes other than *Bh4* play a role in the hull colour phenotype.

Given the parallel domestication events in rice, with selection on a similar suite of domestication traits (white arrows in Fig. 1), and the parallel de-domestication events (black arrows in Fig. 1) with selection for traits that are often more like those of wild species, rice provides a tremendously useful system in which to study the genetic basis for parallel evolution. Using *Bh4* as a candidate gene for the molecular basis of parallel evolution of hull colour, our goal in this study was to determine this gene's role in hull colour evolution in US weedy rice and African rice. We find that *Bh4* sequences show reduced nucleotide diversity and significant deviations from neutral evolution in US weedy rice, consistent with selection at the *Bh4* locus during the process of Asian rice de-domestication. *Bh4* sequences also provide support for the hypothesis of separate origins of SH and BH weedy rice from *indica* and *aus* rice, respectively. In African rice, we identify three independent deletions in *Bh4* exons that are associated with the straw hull phenotype, indicating that this gene has been a target of selection in both the African and Asian crop species. As is the case with Asian rice, some straw-hulled African accessions lack any detectable loss-of-function mutations at *Bh4*, once again pointing to genetic mechanisms other than *Bh4*-coding variation as a contributor to the straw hull phenotype.

Materials and methods

Sampling and phenotyping

Samples for this study were selected from a panel of *Oryza* accessions used in previous studies of US weedy rice (Reagon *et al.*, 2010; Gross *et al.*, 2010a,b) and germplasm made available through the International Rice Research Institute (IRRI). Accessions included in the study are listed in Table 1 (accession numbers can be found in Table S1). Previously sampled accessions included 41 weedy accessions collected from US rice fields that were grouped by hull colour (10 black hull (BH), 27 straw hull (SH) and 4 brown hull (BrH)); 34 accessions representing the five genetically distinct subgroups within cultivated *O. sativa* (Garris *et al.*, 2005) (4 *aus*, 3 *aromatic*, 5 *indica*, 18 *tropical japonica* and 4 *temperate japonica*); 19 accessions of the Asian rice progenitor, *O. rufipogon* (including two accessions of the annual form, *O. nivara*); and one accession of *O. glumaepatula*, a South American AA genome species that was used as an outgroup. African *Oryza* accessions were obtained from IRRI and included 14 African domesticated rice varieties (*O. glaberrima*) and 12 accessions of the wild progenitor (*O. barthii*). Hull colour was determined post-harvest based on the International Rice GeneBank phenotyping protocol (http://cropgenebank.sgrp.cgiar.org). Plant DNA was extracted from approximately 1 g of fresh leaf tissue following a modified CTAB protocol as described previously (Gross *et al.*, 2009). Purified DNA was then diluted to 2 ng μL^{-1} in TE buffer for PCR.

Table 1 *Oryza* accessions used in this study. Hull colour and deletions found in *Bh4* are indicated.

Group	Hull colour	Deletions in *Bh4* (bp)	Number of accessions
US weedy rice, *Oryza sativa*			
SH	Straw	22	27
BrH	Brown	22	4
BH	Black	–	10
Asian domesticated variety groups, *O. sativa*			
aus	Straw	22	4
aromatic	Straw	22	3
indica	Straw	22	5
trop. & temp. japon.	Straw	22	21
	Straw	–	1
Asian wild rice			
Oryza rufipogon	Black	–	13
	Straw	22	3
	Brown	22	1
Oryza nivara	Black	–	1
	Brown	22	1
African domesticated rice, *O. glaberrima*			
	Straw	2	1
	Straw	3	2
	Straw	9	3
	Straw	–	3
	Black	–	5
African wild rice, *O. barthii*			
	Straw	9	1
	Straw	–	1
	Black	–	10
South American wild rice, *O. glumaepatula*			
	Black	–	1

PCR and DNA sequencing

Bh4 primers were selected from those used in a previous study (Zhu *et al.*, 2011) for PCR amplification of the entire coding and intronic regions of *Bh4*. Whenever possible, the full-length 2.5 kb-gene was amplified as a single amplicon. In cases of unsuccessful full amplification, internal primers were used to amplify smaller, overlapping regions that together span the gene (see Table S2). PCR amplifications were carried out in a GeneAmp® PCR System 9700. The profile was 2 min at 94 °C for denaturation, followed by 40 cycles of 30 s at 94 °C, 30 s at 50 °C and 3 min at 72 °C, and lastly 7 min at 72 °C for final extension. The reagents included 8.3 μL water, 2 μL 10× ExTaq buffer that contains 20 mM $MgCl_2$, 1.6 μL 2.5 mM dNTPs, 1 μL 20 μM of each forward and reverse primers, 5 μL 4 M betaine, 0.1 μL ExTaq DNA polymerase (Takara, Mountain View, CA, USA) and 1 μL of genomic DNA from the accessions – a total of 20 μL per reaction. Successful amplification of the complete *Bh4* gene was achieved in all accessions except the 10 BH weedy rice accessions. In the BH samples, we were unable to amplify a region corresponding to approximately 300 bp in the second intron, despite repeated efforts using multiple primer combinations. No differences in *Bh4* gene structure were detected on Southern blots that would explain a lack of amplification across this region (data not shown). The missing sequences in BH accessions were coded as missing data in subsequent DNA sequence alignments and analyses.

Big Dye Terminator sequencing was performed according to conventional methods and separated on an ABI 3130 capillary sequencer at the Washington University Biology Departmental core facility. When possible, direct sequencing of the PCR product was conducted. When cloning of PCR products was required, three clones for each product were sequenced and aligned. The consensus sequence of all three clones was used as the final sequence. Given that both domesticated and weedy rice are highly self-fertilizing, most of the sequenced accessions would be expected to be homozygous across the *Bh4* locus.

DNA sequence alignment and tree construction

Sequences were aligned and checked for quality using Phred and Phrap implemented in BioLign 4.0.6 (Hall, 2001). Base calls with Phred quality scores below 30 were checked and edited by eye. Finalized sequences were imported into BioEdit Ver. 7.1.3 (Hall, 2011), and published *Bh4* sequences including 29 cultivated *O. sativa* accessions and 23 wild *O. rufipogon* and *O. nivara* accessions were added into the alignment (Zhu *et al.*, 2011; GenBank accessions FQ377520 to FQ377566 and FQ377579 to FQ377583). GenBank accessions for newly sequenced samples are KC128691 to KC128808.

A maximum-likelihood (ML) tree was constructed for the *Bh4* gene using MEGA 5 (Tamura *et al.*, 2011), with the general time reversible (GTR) plus gamma nucleotide substitution model. The model was identified as most suitable by FindModel (available at http://www.hiv.lanl.gov/content/sequence/findmodel/findmodel.html), which is based on ModelTest (Posada & Crandall, 2001) and uses Akaike information criteria of 28 possible models. This tree was constructed from all available sequences of *Bh4*, including 52 sequences previously published (Zhu *et al.*, 2011) and 121 newly added sequences. Deletions in the sequence were excluded as they cannot be incorporated in the mutation model. One thousand bootstrap replicates were performed to evaluate the support for each branch on the ML tree.

Diversity analysis and tests for selection

Nei's average pairwise nucleotide diversity (π) and Watterson's estimator of theta (θ_W) were computed by DnaSP Ver.5 (Librado & Rozas, 2009) for both silent sites and all sites of individuals within each group. Tests for selection, including Tajima's D (Tajima, 1989) and Fu and Li's F (Fu, 1997) were calculated for each group

in DnaSP, with significance tested using a distribution of 10 000 coalescent simulations and the recombination parameter set to zero, which is consistent with the high rate of selfing in *Oryza*. Significance levels did not deviate substantially from those calculated with the recombination parameter set to the default value. Genetic distances between each group, assessed by pairwise haplotype frequency F_{ST} values, were calculated in Arlequin Ver.3.5 (Excoffier & Lischer, 2010).

To further determine whether *Bh4* sequences show signatures of selection, we compared π and θ_W values for *Bh4* with values calculated from a sample of unlinked coding gene regions from across the genome. The Asian domesticated, weedy and wild accessions used in the present study have been previously sequenced at 48 neutrally evolving sequence-tagged site (STS) loci (Reagon *et al.*, 2010); these STS loci served as the neutral reference markers for those accessions. A subset of 24 STS loci was used to test for selection on *Bh4* using the maximum-likelihood Hudson-Kreitman-Aguade test (MLHKA; Wright & Charlesworth, 2004). MLHKA was run six times from a random number seed for each group, three times including selection on *Bh4* in the model and three times including no selection on *Bh4* in the model. Significance was assessed by the likelihood ratio test, which is roughly equal to a Chi-squared distribution, with 1 degree of freedom.

For African wild and cultivated rice accessions, we used the average diversity across 14 coding genes previously analysed in a panel of African rice, including 20 accessions each of *O. barthii* and *O. glaberrima* (Li *et al.*, 2011), as a reference to compare diversity at *Bh4*. Because the accessions sequenced at the 14 coding genes differed from the accessions sequenced at *Bh4*, a MLHKA test was not run for the African samples.

Results

To determine how hull colour variation in African rice, Asian rice and US weedy rice is correlated with *Bh4* sequence variation, we analysed gene sequences using a total panel of 173 *Oryza* accessions, including 29 domesticated Asian accessions and 24 wild Asian accessions sequenced by Zhu *et al.* (2011). Newly added samples included 34 domesticated Asian rice accessions, 18 accessions representing its wild progenitor, *O. rufipogon/O. nivara*, 14 domesticated African accessions, 12 accessions of the African wild progenitor (*O. barthii*), 41 US weedy rice accessions and one wild South American accession (*O. glumaepatula*) as an outgroup.

The distribution of hull colour phenotypes from our samples is shown in Table 1. US weedy rice, which is traditionally categorized based on hull colour, was confirmed for hull colour phenotypes including straw hull (SH), black hull (BH) and brown hull (BrH). All sampled Asian crop varieties (34 accessions) had the straw hull phenotype, whereas African crop varieties (14 accessions) were variable, with 64% straw hull and 36% black hull. Asian wild rice (19 accessions) consisted of 74% black hull, 16% straw hull and 10% brown hull. African wild rice included 83% black hull accessions and 17% straw hull accessions. Given that the black hull phenotype is considered characteristic of nondomesticated *Oryzas*, the occurrence of the straw hull phenotype in some wild accessions suggests either past episodes of crop-to-wild genetic introgression or standing variation for the crop phenotype in the wild ancestor.

Phylogeny of *Bh4*

A maximum-likelihood (ML) tree for *Bh4* was constructed using the entire 173 accession data set and is shown in Fig. 2. The purple triangle represents a collapsed clade (97 accessions) comprising all accessions characterized by either the 22 or 1 bp previously described loss-of-function deletions (92% and 8%, respectively); this clade includes all but two of the sampled cultivated *O. sativa* accessions, along with all SH and BrH weeds and five wild accessions (see also Table 1). Because deletions were not included in the tree construction, all sequences that had either the 22 or 1-bp deletion grouped as a single haplotype. One of the *O. sativa* cultivars that do not fall in this clade is a previously described *japonica* variety that has a functional *Bh4* gene and a black hull phenotype (Zhu *et al.*, 2011); the other is the previously described straw hull *aus* cultivar, Kasalath, which carries an exon 3 nucleotide substitution conferring a premature stop codon.

Among US weed strains, all straw hull (SH) and brown hull (BrH) accessions contain the 22-bp loss-of-function mutation characteristic of most cultivated Asian rice varieties (Table 1). Both of these types of weed strains share identical haplotypes with cultivated Asian rice, with the exception of three singleton SNPs (each unique to a single weed accession) in noncoding regions not found in Asian cultivars. All BH weed strains share an identical *Bh4* sequence haplotype, which, as expected, does not show evidence of any loss-of-function mutations in the coding region. These BH strains are closely grouped with the straw hull Kasalath *aus* accession (Fig. 2); the only coding sequence difference is the absence of the exon 3 loss-of-function SNP that characterizes the cultivar. Although Kasalath is diverged from other *aus* crops at the *Bh4* gene, *STRUCTURE* analysis across 48 STS loci indicates no genome-wide divergence of this accession from the *aus* group, nor any closer relationship with BH weeds (Reagon *et al.*, 2010).

The wild and domesticated African rice species, which are shown in blue in Fig. 2, group separately from Asian cultivated and weedy rice. Like Asian rice, African rice *Bh4* sequence haplotypes do not cluster by species; there are shared haplotypes between *O. glaberrima* and *O. barthii* for both straw-hulled and black-hulled accessions.

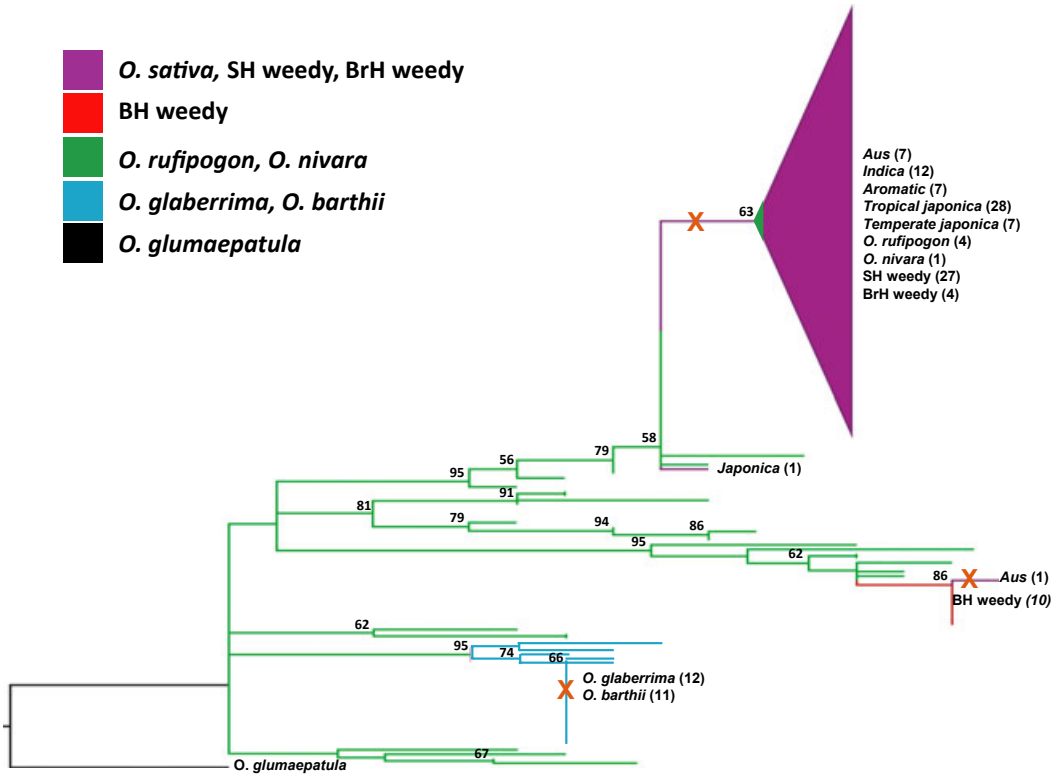

Fig. 2 Maximum-Likelihood (ML) tree of *Bh4* alleles. Colours of branches indicate grouping of accessions and numbers of individuals in a clade are recorded in parentheses. The purple triangle represents a collapsed clade containing the majority of Asian cultivars, SH and BrH weedy rice, and 5 Asian wild accessions. Orange X's represent the three locations on the tree where there is a transition from black hull to straw hull. Bootstrap support is reported for any branch of better than 50% support.

In straw-hulled individuals, there are three separate deletions in coding regions that could potentially cause loss of function of the *Bh4* gene. These include a 9-bp (three codon) in-frame deletion in exon 2, a 3-bp (1 codon) in-frame deletion in exon 2 and a 2-bp frameshift deletion in exon 3, which results in an early stop. The 9-bp deletion is shared in both *O. glaberrima* and *O. barthii*, whereas the other two deletions are unique in *O. glaberrima*. All African accessions that contain these exonic deletions in *Bh4* have the straw hull phenotype. In addition, there are two *O. glaberrima* accessions and one *O. barthii* accession that are straw hulled, but that do not have any loss-of-function mutations in the coding region of *Bh4*. This pattern suggests either a *cis*-regulatory change or another gene causing loss of black pigmentation in these accessions. The pairwise F_{ST} values between rice subgroups for *Bh4* broadly correspond to patterns observed in the phylogeny (Table S3).

Diversity of *Bh4*

Nucleotide diversity at *Bh4* is not reduced in *O. rufipogon*/*O. nivara* when compared to the average of 48 STS loci (Table 2, Fig. 3). In contrast, cultivated *indica* rice has a 10-fold reduction in diversity at *Bh4* when compared to the STS average value. The same pattern was seen for all cultivated Asian rice subgroups, except *aus* where diversity at *Bh4* is greater than the average across 48 STS loci (Table 2). However, if we remove one highly diverged *aus* cultivar, Kasalath, diversity estimates are reduced to zero for the *aus* varieties. This decrease in diversity is consistent with the near fixation of the 22-bp loss-of-function allele in cultivated Asian rice. Both SH and BH weedy rice have reductions in diversity at *Bh4* in comparison with the 48 STS average, indicating that selection may be acting on this gene. Interestingly, this effect is most pronounced in the BH weeds, which carry an apparently functional gene copy; all of these weeds are fixed for the same *Bh4* haplotype, potentially suggesting a recent sweep for this gene.

Unlike Asian wild rice, African wild rice (*O. barthii*) shows reductions in diversity at *Bh4* in comparison with the average across 14 unlinked loci (Table 2, Fig. 4). However, this reduction in diversity is lost when we examine only black hull *O. barthii*, where diversity at *Bh4* ($\pi_{\text{silent sites}} = 2.03$, $\theta_{\text{silent sites}} = 2.05$; Table S4) is very similar to the average across 14 unlinked loci ($\pi_{\text{silent sites}} = 2.5$, $\theta_{\text{silent sites}} = 2.4$). Because *O. barthii* and

Table 2 Nucleotide diversity per Kb.

	Bh4 diversity				Genome-wide diversity†			
Group	π silent sites	π all sites	θ_w silent sites	θ_w all sites	π silent sites	π all sites	θ_w silent sites	θ_w all sites
SH weed	0.1	0.1	0.33	0.33	0.692	0.564	0.638	0.49
BrH weed	0	0	0	0	0.311	0.23	0.349	0.278
BH weed	0	0	0	0	0.829	0.75	0.933	0.827
Oryza sativa aus	4.36	3.17	6.72	4.9	1.57	1.19	1.09	1.4
O. sativa indica	0.33	0.22	0.22	0.14	2.18	1.61	2.22	1.65
O. sativa japonica	0.26	0.21	0.86	0.66	1.337	1.073	1.37	1.694
O. rufipogon/O. nivara	10.27	7.32	12.37	9.17	6.35	4.38	7.79	5.61
O. glaberrima	0.13	0.08	0.26	0.16	0.6	0.5	0.7	0.5
O. barthii	1.3	1.14	1.63	1.65	2.5	2.2	2.4	2.1

†Genome-wide diversity is averaged across 48 STS or 14 coding genes.

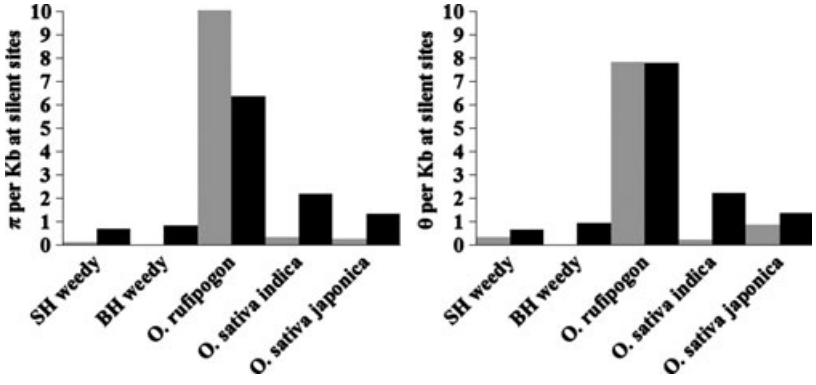

Fig. 3 Genetic diversity of SH and BH weedy rice, *Oryza rufipogon* and *Oryza sativa indica*. Diversity is measured by pairwise nucleotide diversity (π) on left and Watterson's theta (θ_W) on right at silent sites per 1000 bp. Grey bars represent *Bh4* and black bars represent average diversity across 48 sequence-tagged sites.

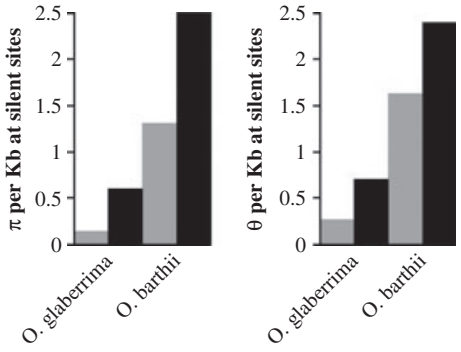

Fig. 4 Genetic diversity of *Oryza barthii* and *Oryza glaberrima*. Diversity is measured by pairwise nucleotide diversity (π) on left and Watterson's theta (θ_W) on right at silent sites per 1000 bp. Grey bars represent *Bh4* alone and black bars represent average diversity across 14 coding genes.

O. glaberrima may have high levels of ongoing gene flow (Li *et al.*, 2011), *Bh4* alleles that are selected on in the crop can be introgressed into the wild species, resulting in reduction in diversity seen when including all *O. barthii*. African cultivated rice (*O. glaberrima*) also has reductions in diversity at *Bh4* when compared to the 14 loci average. This difference becomes more pronounced when only the straw hull *O. glaberrima* are included (*Bh4* $\pi_{\text{silent sites}} = 0$, $\theta_{\text{silent sites}} = 0$; Table S4). It is important to note that the accessions included in the 14 unlinked loci data set are different than the accessions included in the *Bh4* data set, so there could be inherent differences in the genome-wide diversity between these two data sets.

Tests for selection

Tajima's D and Fu & Li's F were calculated in each group to determine whether signatures of selection are present in *Bh4* (Table 3). Groups for which these statistics could not be calculated due to lack of polymorphism included BH weedy rice, BrH weedy rice and straw hull African wild and cultivated rice. These groups have likely all undergone a very recent bottleneck and/or a selective sweep at this gene, resulting in the fixation of a single haplotype. SH weedy rice has significantly negative Tajima's D and Fu & Li's F values, also supporting the hypothesis of a recent bottleneck or selective sweep. The *japonica* and *aus* subspecies of *O. sativa* have significantly negative values for both statistics, consistent with selection for the 22-bp deletion haplotype. However, all segregating sites in the *aus*

Table 3 Tests for selection at *Bh4*.

Group	Number of accessions	Number of seg. sites	Tajima's D*	Fu & Li's F*	MLHK†
US weedy rice, *Oryza sativa*					
SH	27	3	−1.7333 (P = 0.04)	−2.68434 (P = 0.03)	0
BrH	4	0	N.A.	N.A.	N.T.
BH	10	0	N.A.	N.A.	0
Asian domesticated variety groups, *O. sativa*					
aus	8	31	−1.86695 (P = 0.00)	−2.24176 (P = 0.02)	0
indica	10	1	1.30268 (P = 0.93)	1.02064 (P = 0.93)	0
japonica	42	7	−2.03034 (P = 0.003)	−3.36633 (P = 0.005)	3.858 (P = 0.049)
Asian wild rice					
Oryza rufipogon	27	79	−0.42955 (P = 0.38)	−0.76844 (P = 0.27)	2.374 (P = 0.123)
Oryza nivara	15	68	−0.91150 (P = 0.19)	−0.92640 (P = 0.26)	N.T.
African domesticated rice, *O. glaberrima*					
Straw hull	9	0	N.A.	N.A.	N.T.
Black hull	5	1	−0.61237 (P = 0.62)	−0.47871 (P = 0.52)	N.T.
African wild rice, *O. barthii*					
Straw hull	2	0	N.A.	N.A.	N.T.
Black hull	10	13	−0.61495 (P = 0.28)	−0.79178 (P = 0.29)	N.T.

*Significance tested with 10,000 coalescent simulations.
†Likelihood ratio test results.

group are due to a single accession, Kasalath. All polymorphisms recorded in this group are therefore of low frequency because they only occur in this single phylogenetically diverged individual. When Kasalath was removed from the data set, these statistics could not be calculated due to a lack of polymorphism within the group. Asian wild rice and both black hull wild and domesticated African species also did not deviate from neutrality at *Bh4*.

MLHKA tests were used to test for deviation from neutrality at *Bh4* in comparison with a subset of 24 STS loci (Table 3). Both weedy rice biotypes had equal probabilities for a model of selection and a model of no selection on *Bh4* (MLHKA likelihood ratio test = 0). The same was true for *aus* and *indica* crops; this is in contrast to expectations of selection on this gene in all Asian cultivated rice and to what was found in the previous study (Zhu *et al.*, 2011). This difference in significance could be due to different loci used for comparison in our study and the previous analysis. The only test that was significant was in *japonica* crops (MLHKA likelihood ratio test = 3.858, P = 0.049), consistent with Tajima's D and Fu and Li's F. As expected, *O. rufipogon* did not have a signature of selection on *Bh4* (MLHKA likelihood ratio test = 2.374, P = 0.123).

Discussion

Dark hull colour is associated with wild *Oryza* species, and selection for straw-coloured hulls occurred as part of the domestication process in both Asian and African cultivated rice. Weedy rice in the US consists of two genetically distinct biotypes, which are easily distinguished by differences in hull colour; SH weedy rice shares the straw hull phenotype of cultivated rice, whereas BH weedy rice shares the dark hull phenotype of wild rice. We sequenced *Bh4*, a gene involved in hull colour in Asian rice, to determine whether the same gene is responsible for the parallel evolution of light hull colour in domesticated African rice and for the differences in hull colour found in the two US weedy rice biotypes.

Hull colour variation reflected in rice domestication and de-domestication

Light hull colour is almost universally fixed in cultivated Asian rice, supporting the hypothesis that this domestication trait has been selected for in both the *indica* and *japonica* subspecies of *O. sativa*. From our total panel of 62 Asian cultivars (34 from our study and 28 from Zhu *et al.*, 2011), only one was found to have a black hull. The black hull Chinese landrace, Lijiang, is a *japonica* variety. The *Bh4* sequence for Lijiang groups with *O. rufipogon* in the ML tree and has no loss-of-function mutations in its coding region. African cultivated rice, on the other hand, has much more variation in hull colour, potentially reflecting a lower selective pressure for this trait in African rice domestication; 9 of 14 accessions had straw hulls (64%). In a search of hull colour variation in African rice available on the International Rice Research Institute's (IRRI) germplasm database, we found that about 60% (501 of 839) of phenotyped *O. glaberrima* have straw-coloured hulls. Our random sample of 14 *O. glaberrima* is therefore reflective of hull colour variation in a much larger data set.

Both Asian and African wild rice (*O. rufipogon*/*O. nivara* and *O. barthii*) had variation in hull colour. In our

data set, 16% of Asian wild accessions and 17% of African wild accessions had straw hulls. This variation in the wild progenitors of both domesticated species has two potential mechanisms, introgression from the crop after domestication or standing variation in the wild species from which the crop alleles were selected (discussed below).

Weedy rice in the US also has hull colour variation. This pattern potentially supports a previously proposed hypothesis that these two weed biotypes, which also occur in rice fields in other world regions, reflect two different adaptive strategies: a crop mimic form (SH) and a more wild-like form (BH) (Federici *et al.*, 2001). However, it is also possible that the fixation of hull colour in each of the US weedy biotypes has no adaptive significance. SH weedy rice could be fixed for the straw hull phenotype simply because it was derived from *indica* crops that were already fixed for this trait (see genetic support below). BH weedy rice shares close similarity to one *aus* crop variety, Kasalath, which may have acquired the *Bh4* loss-of-function SNP after BH weedy rice had diverged from the *aus* crop lineage; this would suggest that BH weeds carry an ancestral black hull allele that was once characteristic of *aus* rice prior to selection for the straw hull phenotype. Similarly, haplotype variation at the pericarp colour gene *Rc* suggests that US weedy rice carries ancestral alleles that predate selection for white pericarps in the crop (Gross *et al.*, 2010a). However, genetic diversity in these groups points to a selective sweep in BH weeds (see below).

Molecular mechanisms for hull colour variation

Zhu *et al.* (2011) described mutations in *Bh4* that are responsible for the transition from black to straw hull in Asian cultivated rice. These included a very common 22-bp deletion, a less common 1-bp deletion and a SNP resulting in an early stop in a single *aus* accession, Kasalath. Our phylogeny is consistent with that of Zhu and colleagues for Asian wild and domesticated rice. However, our sampling includes some straw and brown hull wild accessions (4 *O. rufipogon* and 1 *O. nivara*) that have the 22-bp deletion haplotype and therefore group with *O. sativa* crops in the phylogeny. The presence of straw and brown hull wild rice could have two explanations. The first is that standing variation in hull colour is present in wild rice and that this is the source of loss-of-function alleles that were selected for during *O. sativa* domestication. The second is that introgression of alleles from the crop back into wild populations has occurred since domestication. Using the inferred ancestry assignment from InStruct runs on 48 STS loci (Reagon *et al.*, 2010), we found that all five wild accessions had some *indica* inferred ancestry, with values ranging from 0.114 to 0.842 (Table S5). This genome-wide signature of introgression from *indica* cultivated rice indicates that introgression is a more likely explanation for the presence of straw and brown hulls in Asian wild rice.

As found in Asian rice, haplotypes of *Bh4* were shared between both species of African rice. Shared alleles between these species are not uncommon, and there is no population structure found between *O. glaberrima* and *O. barthii* collected in the proposed centres of *O. glaberrima* domestication (Li *et al.*, 2011). Introgression and standing variation are both plausible explanations for these shared alleles. However, lack of genome-wide data in our samples impedes testing which of these is most likely.

Newly identified mutations in *Bh4*, including 9, 3 and 2-bp deletions, correlate with the straw hull phenotype in African wild and domesticated rice. Because all African rice sequences were clustered independent of Asian rice, we can conclude that introgression from Asian crops grown in Africa is not the source of *Bh4* alleles in *O. glaberrima* or *O. barthii*. Instead, there is evidence that selection for repeated trait evolution in African and Asian cultivated rice has resulted in independent mutations at the same gene. The same pattern has been found in selection at the *Rc* locus for white-coloured pericarps in African and Asian cultivated rice (Sweeney *et al.*, 2006; Gross *et al.*, 2010b). In both cases, the domestication trait results from loss-of-function mutations at these genes, suggesting that this pattern might occur predominantly with traits where deletions anywhere in the coding region can cause changes in phenotype. However, we might expect that other genes in the biochemical pathway could cause the same phenotype when lost. This could potentially be the case for the rice accessions (both African and Asian) that have a straw hull phenotype, but do not harbour any obvious loss-of-function mutations in the *Bh4* gene.

Previous studies have shown that weedy rice in the US is composed of two genetically and phenotypically distinct biotypes: SH and BH (Londo & Schaal, 2007; Reagon *et al.*, 2010). Our data support the evolutionary relationships proposed by these studies. SH weedy rice shares the 22-bp deletion haplotype for *Bh4* with several crop accessions including *indica* varieties, the putative progenitor to SH weedy rice. BH weedy rice shares an almost identical *Bh4* haplotype with at least one *aus* cultivar, the putative progenitor of BH weedy rice. In this case, BH weedy rice has a functional allele, whereas the straw-hulled *aus* contain one SNP difference from BH resulting in an early stop in exon three. Interestingly, this same *aus* cultivar, Kasalath, shares the haplotype for pericarp colour (*Rc*) with BH weedy rice in a previous study (Gross *et al.*, 2010a), supporting our findings that BH weedy rice is genetically most similar to this particular *aus* landrace among those included in our sampling. However, genome-wide variation across 48 STS loci indicates that Kasalath is not detectably

distinct from other *aus* crops, nor does this crop variety show differentially closer genetic relatedness to BH weeds (Reagon *et al.*, 2010).

Signatures of selection on *Bh4*

Asian cultivated rice shows reduced diversity at *Bh4* when compared to the average of 48 STS loci. This result is consistent for the *indica* and *japonica* subgroups. Tests for deviations from neutrality (Tajima's D and Fu and Li's F) were significant in *aus* and *japonica* varieties, supporting the hypothesis of positive selection at this gene during domestication. Very low levels of diversity within *indica* (a single segregating site) are likely to be the cause of a lack of statistical support for deviations from neutrality in this group. The increase in diversity for the *aus* subgroup is due to the highly diverged *Bh4* sequence from Kasalath. When this accession is removed from the data set, there is only a single shared haplotype among all remaining accessions. MLHKA tests were only significant in *japonica* crops, perhaps reflective of the domestication bottleneck impacting the rest of the genome and masking selection on this gene. We find no signatures of selection on *Bh4* from Asian wild-rice accessions. Diversity levels from *Bh4* are equal to or greater than that of the average of 48 STS. There were also no deviations from neutrality for Tajima's D, Fu and Li's F, and MLHKA tests. The lack of any signatures of selection in this group reflects the higher level of diversity as well as possible introgression of *Bh4* alleles from cultivated rice.

In African cultivated rice, there is reduced diversity in the *Bh4* gene in comparison with 14 coding genes from a previous study (Li *et al.*, 2011). However, we also see the same amount of diversity reduction in African wild rice, suggesting that this signature may not be due to selection from domestication. The increased percentage of straw hull accessions in African crops compared to African wild rice (64% and 17% respectively in our sample set) does suggest that this trait has been selectively favoured during African rice domestication. Rice farmers in Africa choose hull colour for their crop based on a variety of reasons including aesthetics and their own thoughts on likely bird predation (P. Richards, personal communication). However, selection is not the only explanation for the change in percentage; the domestication bottleneck could also be responsible. We could not calculate Tajima's D or Fu and Li's F in straw hull African rice due to the absence of segregating sites within these groups. Black hull African rice did not have significant deviations from neutrality at *Bh4*, so selection is likely not playing a role on maintaining black hulls in African wild rice.

SH and BH weedy rice also had reduced diversity at *Bh4* when compared to 48 STS loci that span the rice genome. This reduction in diversity, which is especially strong in BH weedy rice, suggests that hull colour may be an important trait in weedy rice. Because BH weedy rice had no genetic diversity at *Bh4*, which is suggestive of a recent sweep, we were not able to statistically test for deviations from neutrality. One hypothesis to explain this lack of diversity is similar to that put forth for dark hulls in wild rice; birds and other animals may not be able to distinguish the dark-hulled rice against the dark-coloured mud, eating more of the light-coloured grains. SH weedy rice could have a reduction in diversity in *Bh4* due to reduced diversity at this gene in its progenitors, and no selection on hull colour is taking place for SH weedy rice. We do find that diversity at *Bh4* is similar for both SH weeds and *indica* cultivars. In this case, the signature of selection, reduced diversity, may simply be an artefact of selection on this gene in the weed's progenitor. SH weedy rice may therefore only have a straw-coloured hull because that trait was fixed in its cultivated progenitor. SH weedy rice had significant deviations from neutrality for Fu and Li's F and Tajima's D, but was not significant in the MLHKA test. Interestingly, BrH weedy rice shares the 22-bp deletion of SH and cultivated Asian rice. We also find two wild Asian accessions that have a brown hull and the 22-bp deletion in *Bh4*. Given this observation, we can infer that at least one other gene is involved in hull colour variation in rice.

Parallel evolution

The repeated occurrence of loss-of-function mutations in *Bh4*, leading to a straw hull phenotype in both African and Asian crops, provides a new example of parallel evolution under domestication. These results are similar to our previous findings on the genetic basis for pericarp colour evolution in rice (Gross *et al.*, 2010a,b). Both African and Asian wild rice species and both US weedy rice biotypes have a dark-coloured (red) pericarp, although African and Asian cultivars often have a non-pigmented (white) pericarp that was selectively favoured during rice domestication (Sweeney *et al.*, 2006). A regulatory gene in proanthocyanidin synthesis (*Rc*) harbours loss-of-function mutations that are responsible for the transition of red to white pericarp in Asian cultivated rice (Sweeney *et al.*, 2006; Gross *et al.*, 2010a); different, independent loss-of-function mutations in this same gene are responsible for the white pericarp in African rice (Gross *et al.*, 2010b). *Rc* alleles from US weedy rice cluster with those from *indica* varieties of Asian cultivated rice, indicating that standing variation in the crop progenitors of weedy rice are responsible for this trait (Gross *et al.*, 2010a). This same pattern of parallel trait evolution in both domesticated species as well as standing variation in the progenitors of weedy rice has been found in *Bh4*. Other traits may not follow this same pattern; selection on seed shattering in wild, crop and weedy rices seems to have acted on different genes (Li *et al.*, 2006 and Thurber *et al.*, 2010).

Patterns of parallel trait evolution in rice do not differ substantially from patterns observed in natural populations, where we can find multiple genetic 'answers' to the similar selective pressure (reviewed by Elmer & Meyer, 2011). Selective pressures leading to rice domestication and subsequent de-domestication events are both recent and trackable, allowing this system to serve as a model for understanding the genomic basis for parallel evolution.

Acknowledgments

Funding for this research was provided by a grant from the NSF Plant Genome Research Program (IOS-1032023) to KMO. We thank Mike Dyer and other staff of the Washington University greenhouse for assistance in maintaining plant samples and Linda Small for her contributions to PCR amplification and sequencing. We are also grateful to members of the Olsen lab group, Beng-Kah Song (Monash University) and two anonymous reviewers who provided helpful constructive comments on an earlier version of this paper.

References

Caicedo, A.L., Williamson, S.H., Hernandez, R.D., Boyko, A., Fledel-Alon, A., York, T.L. *et al.* 2007. Genome-wide patterns of nucleotide polymorphism in domesticated rice. *PLoS Genet.* **3**: 1745–1756.

Elmer, K.R. & Meyer, A. 2011. Adaptation in the age of ecological genomics: insights from parallelism and convergence. *Trends Ecol. Evol.* **26**: 298–306.

Estorninos, L.E., Gealy, D.R., Gbur, E.E., Talbert, R.E. & McClelland, M.R. 2005. Rice and red rice interference. II. Rice response to population densities of three red rice (*Oryza sativa*) ecotypes. *Weed Sci.* **53**: 683–689.

Excoffier, L. & Lischer, H.E.L. 2010. Arlequin suite ver 3.5: a new series of programs to perform population genetics analyses under Linux and Windows. *Mol. Ecol. Resour.* **10**: 564–567.

Federici, M.T., Vaughan, D., Tomooka, N., Kaga, A., Wang, X.W., Doi, K. *et al.* 2001. Analysis of Uruguayan weedy rice genetic diversity using AFLP molecular markers. *Electron. J. Biotechnol.* **4**: 131–145.

Fu, Y.X. 1997. Statistical tests of neutrality of mutations against population growth, hitchhiking and background selection. *Genetics* **147**: 915–925.

Fuller, D.Q., Qin, L., Zheng, Y., Zhao, Z., Chen, X., Hosoya, L.A. *et al.* 2009. The domestication process and domestication rate in rice: spikelet bases from the Lower Yangtze. *Science* **323**: 1607–1610.

Garris, A.J., Tai, T.H., Coburn, J., Kresovich, S. & McCouch, S. 2005. Genetic structure and diversity in *Oryza sativa* L. *Genetics* **169**: 1621–1638.

Goff, S.A., Ricke, D., Lan, T.-H., Presting, G., Wang, R., Dunn, M. *et al.* 2002. A draft sequence of the rice genome (*Oryza sativa* L. ssp. japonica). *Science* **296**: 92–100.

Gross, B.L., Skare, K.J. & Olsen, K.M. 2009. Novel Phr1 mutations and the evolution of phenol reaction variation in US weedy rice (*Oryza sativa*). *New Phytol.* **184**: 842–850.

Gross, B.L., Reagon, M., Hsu, S.-C., Caicedo, A.L., Jia, Y. & Olsen, K.M. 2010a. Seeing red: the origin of grain pigmentation in US weedy rice. *Mol. Ecol.* **19**: 3380–3393.

Gross, B.L., Steffen, F.T. & Olsen, K.M. 2010b. The molecular basis of white pericarps in African domesticated rice: novel mutations at the Rc gene. *J. Evol. Biol.* **23**: 2747–2753.

Hall, T. 2001. BioLign alignment and multiple contig editor. URL http://en.bio-soft.net/dna/BioLign.html.

Hall, T. 2011. BioEdit biological sequence alignment editor. URL http://www.mbio.ncsu.edu/BioEdit/bioedit.html.

Hammer, K. 1984. Das Domestikationssyndrom. *Kulturpflanze* **32**: 11–34.

Hoekstra, H.E., Hirschmann, R.J., Bundey, R., Insel, P. & Crossland, J.P. 2006. A single amino acid mutation contributes to adaptive beach mouse color pattern. *Science* **313**: 101–104.

Kuriyama, H. & Kudo, M. 1967. Complementary genes Ph and Bh controlling ripening black coloration of rice hulls and their geographical distribution. *Jpn. J. Breed.* **17**: 13–19.

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

Li, Z.-M., Zheng, X.-M. & Ge, S. 2011. Genetic diversity and domestication history of African rice (*Oryza glaberrima*) as inferred from multiple gene sequences. *Theor. Appl. Genet.* **123**: 21–31.

Librado, P. & Rozas, J. 2009. DnaSP v5: a software for comprehensive analysis of DNA polymorphism data. *Bioinformatics* **25**: 1451–1452.

Lin, Z., Li, X., Shannon, L.M., Yeh, C.-T., Wang, M.L., Bai, G. *et al.* 2012. Parallel domestication of the Shattering1 genes in cereals. *Nat. Genet.* **44**: 720–724.

Londo, J.P. & Schaal, B. 2007. Origins and population genetics of weedy red rice in the USA. *Mol. Ecol.* **16**: 4523–4535.

Londo, J.P., Chiang, Y.-C., Hung, K.-H., Chiang, T.-Y. & Schaal, B. 2006. Phylogeography of Asian wild rice, *Oryza rufipogon*, reveals multiple independent domestications of cultivated rice, *Oryza sativa*. *Proc. Natl. Acad. Sci.* **103**: 9578–9583.

Molina, J., Sikora, M., Garud, N., Flowers, J.M., Rubinstein, S., Reynolds, A. *et al.* 2011. Molecular evidence for a single evolutionary origin of domesticated rice. *Proc. Natl. Acad. Sci.* **108**: 8351–8356.

Morishima, H., Sano, Y. & Oka, H.-I. 1992. Evolutionary studies in cultivated rice and its wild relatives. In: *Oxford Surveys in Evolutionary Biology*, Vol. 8 (D. Futuyma & J. Antonovics, eds), pp. 135–184. Oxford University Press, Oxford.

Paterson, A.H., Lin, Y.R., Li, Z., Schertz, K.F., Doebley, J.F., Pinson, S.R. *et al.* 1995. Convergent domestication of cereal crops by independent mutations at corresponding genetic loci. *Science* **269**: 1714–1718.

Posada, D. & Crandall, K.A. 2001. Selecting the best-fit model of nucleotide substitution. *Syst. Biol.* **50**: 580–601.

Reagon, M., Thurber, C.S., Gross, B.L., Olsen, K.M., Jia, Y. & Caicedo, A.L. 2010. Genomic patterns of nucleotide diversity in divergent populations of U.S. weedy rice. *BMC Evol. Biol.* **10**: 180.

Römpler, H., Rohland, N., Lalueza-Fox, C., Willerslev, E., Kuznetsova, T., Rabeder, G. *et al.* 2006. Nuclear gene indicates coat-color polymorphism in mammoths. *Science* **313**: 62.

Rosenblum, E.B., Römpler, H., Schöneberg, T. & Hoekstra, H.E. 2010. Molecular and functional basis of phenotypic

convergence in white lizards at White Sands. *Proc. Natl. Acad. Sci.* **107**: 2113–2117.

Sarla, N. & Swamy, B.P.M. 2005. *Oryza glaberrima*: a source for the improvement of *Oryza sativa*. *Curr. Sci.* **89**: 955–963.

Steiner, C.C., Römpler, H., Boettger, L.M., Schöneberg, T. & Hoekstra, H.E. 2009. The genetic basis of phenotypic convergence in beach mice: similar pigment patterns but different genes. *Mol. Biol. Evol.* **26**: 35–45.

Sweeney, M.T., Thomson, M.J., Pfeil, B.E. & McCouch, S. 2006. Caught red-handed : Rc encodes a basic helix-loop-helix protein conditioning red pericarp in rice. *Plant Cell* **18**: 283–294.

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

Tamura, K., Peterson, D., Peterson, N., Stecher, G., Nei, M. & Kumar, S. 2011. MEGA5: molecular Evolutionary Genetics Analysis using Maximum Likelihood, Evolutionary Distance, and Maximum Parsimony Methods. *Mol. Biol. Evol.* **28**: 2731–2739.

Thurber, C.S., Reagon, M., Gross, B.L., Olsen, K.M., Jia, Y. & Caicedo, A.L. 2010. Molecular evolution of shattering loci in U.S. weedy rice. *Mol. Ecol.* **19**: 3271–3284.

Vigueira, C.C., Olsen, K.M. & Caicedo, A.L. 2013. The red queen in the corn: agricultural weeds as models for rapid adaptive evolution. *Heredity* **110**: 303–311.

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

Yang, C., Kawahara, Y., Mizuno, H., Wu, J., Matsumoto, T. & Itoh, T. 2012. Independent domestication of Asian rice followed by gene flow from japonica to indica. *Mol. Biol. Evol.* **29**: 1471–1479.

Yu, J., Hu, S., Wang, J., Wong, G.K., Li, S., Liu, B. *et al.* 2002. A draft sequence of the rice genome (*Oryza sativa* L. ssp indica). *Science* **296**: 79–92.

Zhu, B.F., Si, L., Wang, Z., Zhou, Y., Zhu, J., Shangguan, Y. *et al.* 2011. Genetic control of a transition from black to straw-white seed hull in rice domestication. *Plant Physiol.* **155**: 1301–1311.

Supporting information

Additional Supporting Information may be found in the online version of this article:

Table S1 Accession information.

Table S2 Primer sequences used to amplify *Bh4*.

Table S3 Population pairwise FSTs.

Table S4 Nucleotide diversity per Kb in African rice by hull color.

Table S5 *Straw hull and Brown hull wild rice* coefficents of ancestry inferred by InStruct.

Received 6 November 2012; revised 19 March 2013; accepted 28 March 2013