MOLECULAR ECOLOGY

Molecular Ecology (2010) **19**, 3271–3284 doi: 10.1111/j.1365-294X.2010.04708.x

Molecular evolution of shattering loci in U.S. weedy rice

CARRIE S. THURBER,* MICHAEL REAGON,* BRIANA L. GROSS,† KENNETH M. OLSEN,† YULIN JIA‡ and ANA L. CAICEDO*

Biology Department, University of Massachusetts, Amherst, MA 01003, USA, †Department of Biology, Washington University, St Louis, MO 63130, USA, ‡USDA-ARS Dale Bumpers National Rice Research Center, Stuttgart, AR 72160, USA

Abstract

Cultivated rice fields worldwide are plagued with weedy rice, a conspecific weed of cultivated rice (*Oryza sativa* L.). The persistence of weedy rice has been attributed, in part, to its ability to shatter (disperse) seed prior to crop harvesting. In the United States, separately evolved weedy rice groups have been shown to share genomic identity with exotic domesticated cultivars. Here, we investigate the shattering phenotype in a collection of U.S. weedy rice accessions, as well as wild and cultivated relatives. We find that all U.S. weedy rice groups shatter seeds easily, despite multiple origins, and in contrast to a decrease in shattering ability seen in cultivated groups. We assessed allelic identity and diversity at the major shattering locus, *sh4*, in weedy rice; we find that all cultivated and weedy rice, regardless of population, share similar haplotypes at *sh4*, and all contain a single derived mutation associated with decreased seed shattering. Our data constitute the strongest evidence to date of an evolution of weeds from domesticated backgrounds. The combination of a shared cultivar *sh4* allele and a highly shattering phenotype, suggests that U.S. weedy rice have re-acquired the shattering trait after divergence from their progenitors through alternative genetic mechanisms.

Keywords: abscission, candidate gene, *Oryza sativa*, red rice, seed dispersal

Received 5 November 2009; revision received 11 February 2010; accepted 19 February 2010

Introduction

Invasive weeds that colonize agricultural fields cost millions of dollars in crop losses and weed control measures every year. Many of these agricultural weeds share similar fitness-related traits that make them highly competitive with crop species. For example, rapid growth, deep roots, high seed production and increased seed dispersal allow weeds to acquire more resources, as well as to produce more offspring (Basu *et al.* 2004). Efficient seed dispersal, in particular, may be a trait crucial to weed fitness. By increasing seed dispersal via 'shattering' or scattering their seeds, weeds can increase their presence in the seed bank and spread into new areas (Harlan & DeWet 1965). Plants that shatter their seeds within agricultural fields can often avoid collection by farmers, and subsequent seed consumption/destruction, thus persisting within fields. Additionally, shattering at maturity is sometimes necessary to retain sufficient seed moisture for dormancy, a trait favoured in agricultural weeds for winter survival and germination during the cropping season (Gu *et al.* 2005a; b; Delouche *et al.* 2007).

Most wild cereals, including wild relatives of rice, wheat and barley, have brittle, easily shed (shattering) seeds. Cultivated cereals, however, have undergone selection for reduction of shattering during the domestication process, to increase the amount of seed harvested by humans (Harlan & DeWet 1965). Reduced seed shattering is thought to be among the earliest and most important traits selected upon during grain domestication (Harlan 1992; Fuller *et al.* 2009). A reduction in seed shattering may have been favoured over complete nonshattering to minimize labour during harvest (Li *et al.* 2006; Sang & Ge 2007a). The shattering trait is thus under strong opposing selection in agricultural environments, with high levels of shattering favoured in invasive weeds and reduced shattering in cultivated crops.

Correspondence: Ana L. Caicedo,
E-mail: caicedo@bio.umass.edu

Weedy or red rice is a weedy type of rice (*Oryza sativa* L.) that invades cultivated rice fields and costs United States farmers millions of dollars each year (Burgos *et al.* 2008). Weedy rice is an aggressive competitor, decreasing yields and contaminating rice harvests with off-colour, brittle grains (Burgos *et al.* 2006; Cao *et al.* 2006). The appearance of weedy rice has been associated with a transition to direct seeding, and it is present worldwide, wherever rice is cultivated (Bres-Patry *et al.* 2001; Olsen *et al.* 2007). Although morphologically diverse, a suite of possible weediness-enhancing traits tends to characterize weedy rice in the field; these include the presence of red pericarps (bran), high levels of dormancy, and high levels of seed shattering (Vaughan *et al.* 2001; Gealy *et al.* 2003; Delouche *et al.* 2007). Several of these traits are also found in the wild ancestor of cultivated rice, *Oryza rufipogon*, and other wild *Oryza* relatives, but weedy rice differs from truly wild species in its adaptation to the agroecosystem and presence of some traits characterizing cultivated rice (e.g. high selfing rate, Delouche *et al.* 2007).

There are multiple efforts underway to understand the worldwide origins of weedy rice groups. Hypotheses range from invasion of wild *Oryza* relatives, to hybridization among wild and cultivated groups, or de-domestication of cultivated rice varieties (Bres-Patry *et al.* 2001; Gealy 2005a). In the United States, weedy rice is prevalent in the rice growing regions of the southern Mississippi basin (Gealy 2005b). No *Oryza* species is native to the U.S., and the evolutionary origin of U.S. weedy rice has been a source of debate since it was first documented in the 1840s (Delouche *et al.* 2007). Previous assessments of genetic diversity have determined that several populations of morphologically divergent weedy rice are present in the U.S. (Gealy *et al.* 2002; Londo & Schaal 2007; Reagon *et al.* 2010). The main populations of U.S. weedy rice, designated after their most common grain morphology, include the straw-hulled (SH) group, characterized by straw-coloured hulls, high yielding panicles and lack of awns, and the black-hulled awned (BHA) group, characterized by its greater height, black hulls and long awns (Gealy *et al.* 2002). The BHA group is subdivided into two genetically distinct subpopulations, BHA1 and BHA2 (Reagon *et al.* 2010). A third group (BRH), characterized by brown hulls, is most likely a result of hybridization between the SH and BHA groups (Reagon *et al.* 2010).

Studies have shown that U.S. weedy rice shares most of its genome with Asian cultivated rice (Londo & Schaal 2007; Reagon *et al.* 2010). Interestingly, U.S. weedy rice does not share a recent evolutionary origin with cultivars grown in the U.S., which belong to the *tropical japonica* variety group, though there is evidence for limited hybridization (Gealy *et al.* 2009; Reagon *et al.* 2010). Instead, studies suggest that SH weeds are most closely related to *indica*, a cultivated rice variety typical of lowland tropical regions, while the BHA groups share a closer relationship with *aus*, a rapidly maturing, photoperiod insensitive rice variety from Bangladesh and Northeastern India. However, neither of these crop varieties has been cultivated in the southern U.S. Moreover, though patterns of genome-wide variation suggest that weedy rice is not directly descended from wild rice (Gealy *et al.* 2009; Reagon *et al.* 2010), questions about possible contributions of wild rice to U.S. weedy rice evolution remain.

Recently, candidate genes underlying some domestication-related traits have begun to be identified in cultivated rice (e.g.: Fan *et al.* 2006; Gu *et al.* 2008; Xing *et al.* 2008). Because these traits often differ between cultivated rice and wild/weedy relatives, candidate genes have opened up new sources of potential information about the evolution of weediness-enhancing traits. Combined with information about genome-wide patterns of polymorphism, candidate genes may help provide a complete picture of the evolutionary origin of weedy rice groups. A recent investigation into a pericarp colour candidate gene, *Rc*, revealed that U.S. weedy rice groups carry alleles distinct from those in sampled cultivated or wild rice groups (Gross *et al.* 2010). Although genomic data suggests that U.S. weedy rice originated from cultivated rice varieties, *Rc* data suggests that weeds are not direct descendants of cultivated rice (Gross *et al.* 2010; Reagon *et al.* 2010). However, because different key traits may have been selected at different stages of the domestication process (Purugganan & Fuller 2009), weedy rice alleles at important domestication loci may tell complementary stories about the origins of weedy rice.

As a trait crucial to modern cultivation and harvesting practices, there has been great interest in discerning the genetic basis of seed shattering in rice. To date, two quantitative trait loci (QTL) of large effect have been cloned, *qsh1* and *sh4/SHA1*, each explaining over 70% of the variation in their respective crosses. The *qsh1* locus is a homeodomain gene, similar to *Arabidopsis thaliana* REPLUMLESS, which was isolated in a cross between two *O. sativa* varieties, *aus* and *temperate japonica*, that differ in their shattering propensity (Konishi *et al.* 2006). A single nucleotide substitution in the regulatory region of the gene decreases the shattering ability in a subset of cultivated *temperate japonica* rice (Konishi *et al.* 2006; Zhang *et al.* 2009).

The *sh4* gene, encoding a nuclear transcription factor, was isolated from a cross between cultivated *O. sativa indica* and a wild species, *O. nivara*, and is involved in the degradation of the abscission layer between the

grain and the pedicel (Li *et al.* 2006; Lin *et al.* 2007). Highly shattering *O. nivara* possess very defined abscission layers, while nonshattering cultivated rice groups possess discontinuous abscission layers (Ji *et al.* 2006; Li *et al.* 2006). A single nonsynonymous substitution (G/T) in the second exon of *sh4* has been shown to lead to diminished DNA binding with the SH4 protein and incomplete development of the abscission layer in nonshattering rice (Li *et al.* 2006). Transgenic *japonica* plants expressing the wild *O. nivara* allele show a significantly increased ability to shatter (Li *et al.* 2006). Shattering QTL in the *sh4* genomic region have been consistently identified in studies involving other crosses between cultivated varieties and wild rice (Xiong *et al.* 1999; Cai & Morishima 2000).

Sh4 is considered the most significant shattering gene to have been selected upon during domestication (Li *et al.* 2006; Purugganan & Fuller 2009). Examination of *sh4* alleles has shown that all cultivated rice sampled to date shares the nonshattering T mutation, and most rice individuals share a common *sh4* haplotype, despite the fact that at least two separate domestication events gave rise to cultivated Asian rice (Li *et al.* 2006; Zhang *et al.* 2009). The sharing of a common *sh4* haplotype across divergent rice varieties has been attributed to a combination of introgression and strong positive selection (selective sweep) favouring a reduction in shattering in the crop during both domestication processes (Li *et al.* 2006; Sang & Ge 2007a,b; Zhang *et al.* 2009).

Here we assess patterns of polymorphism in weedy rice groups at the identified shattering genes and targeted flanking genomic regions, to determine the possible origin of the shattering phenotype in the U.S. weed and contribute to understanding of U.S. weedy rice evolution. The goals of the present study were to (1) assess levels of shattering in U.S. weedy rice groups, (2) determine the origin of U.S. weedy rice alleles at *qsh1* and *sh4* and (3) determine the role each locus may play in the shattering phenotype of weedy rice. We find that the shattering associated single nucleotide polymorphism (SNP) at *qsh1* has not played a role in the evolution of weedy rice, as all weeds, wild rice, and most cultivars share the ancestral allele at this locus. Moreover, although cultivated and weedy rice groups differ greatly in their shattering ability, all sampled weedy and domesticated accessions possess similar or identical alleles at the *sh4* locus, suggesting that the domestication-associated T substitution at *sh4* is not sufficient for loss of shattering. Our data supports a direct origin of U.S. weedy rice groups from domesticated ancestors, and implies that genetic changes at other loci must be responsible for the re-acquisition of the shattering trait during the weed's evolution.

Methods

Plant material

A phenotypically diverse sample of 58 weedy rice accessions, collected in the Southern U.S. rice belt, was generously supplied by David Gealy (USDA) (Table S1, Supporting information). An additional 87 samples of diverse *Oryza* species were included in the study as potential sources of weedy rice alleles. Cultivated rice accessions belong to five variety groups of Asian *O. sativa*: *indica* (9 samples), *aus* (7), *tropical japonica* (8), *temperate japonica* (4), and *aromatic* (3). Thirteen additional accessions of *tropical japonica* cultivars grown in the U.S were included. Other *Oryza* included geographically diverse samples of *O. rufipogon* (30), the wild ancestor of cultivated Asian rice, *O. nivara* (2), an annual plant that some consider an ecotype of *O. rufipogon* (Zhu & Ge 2005), *O. glumaepatula* (2), a wild rice from South America, *O. glaberrima* (4), cultivated African rice and *O. barthii* (2), the wild ancestor of domesticated African rice. *O. meridionalis*, a species native to Oceania, was included as an outgroup. All plants were grown for DNA extraction as described in Reagon *et al.* (2010).

Measurement of the shattering phenotype

A subset of 90 *Oryza* accessions, representing selfed progeny of plants grown for DNA extraction, was grown for phenotyping in a completely randomized block design in two Conviron PGW36 growth chambers at the University of Massachusetts Amherst (Table S1, Supporting information). Two seeds per accession, one per chamber (block), were planted in 4-inch pots and randomly assigned locations within a chamber. Watering and fertilizer schedules were the same in both chambers and plants were exposed to 12-h light/dark cycles. Upon heading, typically two to three months after germination, panicles were bagged to prevent pollen flow and loss of seeds. At 30 days after heading, panicles were tested for shattering using a digital force gauge (Imada, Northbrook, IL). Shattering was measured as Breaking Tensile Strength (BTS) (Konishi *et al.* 2006; Li *et al.* 2006), which is the amount of weight a seed can bear before releasing from the pedicel at the abscission layer. Briefly, panicles were suspended from a ring stand and an individual seed clipped with a small (~1 g) binder clip. Seeds that released at or prior to this point were recorded as zeros and considered highly shattering. For seeds remaining on the panicle, the force gauge was hooked onto the binder clip and the peak measurement upon grain removal was recorded. Preliminary trials revealed that considerable variation could occur within panicles of cultivated vari-

Table 1 Silent Site Nucleotide diversity per kb (Watterson's estimator nucleotide variation (θ_W), the average pairwise nucleotide diversity (θ_π) and Tajima's *D*) for wild *O. rufipogon*, cultivated *O. sativa* and weedy *O. sativa*

			Cultivated *Oryza sativa*					U.S. Weedy Rice		
		Oryza rufipogon	*indica*	*aus*	*tropical japonica*	*temperate japonica*	*aromatic*	SH	BHA1	BHA2
sh4 locus	θ_π	4	0	0.094	0.03	0	0.2	0	0.04	0.2
	θ_W	5	0	0.12	0.92	0	0.2	0	0.99	0.1
	Tajima's *D*	−0.85	N/A	−1.01	−1.16	N/A	0	N/A	−1.16	1.44
Flanking fragments										
sh4f_001	θ_π	2.2	2.4	0.79	0.26	0	0	0	0	0
	θ_W	2.4	2.1	1.1	0.77	0	0	0	0	0
	Tajima's *D*	−0.19	0.41	1.18	−1.16	N/A	N/A	N/A	N/A	N/A
sh4f_002	θ_π	7.3	0	1.7	0	0	1.5	0	2.3	2.6
	θ_W	1.2	0	1.8	0	0	1.5	0	1.4	1.7
	Tajima's *D*	−1.26	N/A	−1	N/A	N/A	0	N/A	1.8	1.79
sh4f_003	θ_π	2.2	0	0	0	0	0	0	0	0
	θ_W	4.4	0	0	0	0	0	0	0	0
	Tajima's *D*	−1.2	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A
sh4f_004	θ_π	1.8	0	0	0	4.1	0	0	0	0
	θ_W	3	0	0	0	3.3	0	0	0	0
	Tajima's *D*	−0.89	N/A	N/A	N/A	1.46	N/A	N/A	N/A	N/A
sh4f_005	θ_π	2.4	2.1	1.8	2	1.6	2.6	1.9	2	0
	θ_W	2.5	1.5	1.6	1.1	1.9	2.6	1.1	1.2	0
	Tajima's *D*	−0.13	1.17	0.56	1.57	−0.82	0	1.43	1.47	N/A
sh4f_006	θ_π	1.6	2	0	0.48	2.1	0	0	0	0
	θ_W	3.5	3	0	0.73	2.6	0	0	0	0
	Tajima's *D*	−1.34	−1.45	N/A	−0.62	−0.97	N/A	N/A	N/A	N/A

eties; thus, 25 randomly chosen seeds per plant were measured across two panicles and averages were calculated for each individual. Chamber effects on shattering were nonsignificant ($P > 0.15$), as determined by a Kruskal–Wallis nonparametric rank test, and were not considered in subsequent analyses.

DNA extraction, genotyping, and sequencing

DNA was extracted as described in Reagon *et al.* (2010). CAPs markers (Neff *et al.* 2002) were used to determine the *qsh1* allele in all individuals (Table S2, Supporting information). Variation at *sh4* was determined by DNA sequencing of the entire open reading frame, the promoter and a downstream region of the gene (Table S2, Supporting information). Additionally, six ~500 base pair (bp) regions of genes increasingly distant from the *sh4* locus (several kilobase pairs (kb) to several megabase pairs (Mb)) were sequenced spanning a region of 5.6 Mb (Table S2, Supporting information). Primers were generated using Primer3 (Rozen & Skaletsky 2000) based on the *O. sativa japonica* (var. Nipponbare) genome (TIGR v. 5 January, 2008). Initial PCR amplification and DNA sequencing was performed by Cogenics (Houston, TX) as described previously (Olsen *et al.* 2006; Caicedo *et al.* 2007). Additional PCR amplification was performed on a 500-bp region surrounding the loss-of-shattering associated SNP using LA Taq and GC rich buffer (TaKara) with added glycerol and DMSO. Sequence alignment, including base pair calls, quality score assignment and construction of contigs, was performed as described previously (Caicedo *et al.* 2007) using BioLign Version 2.09.1 (Tom Hall, NC State University). DNA sequences obtained for this study have been deposited in GenBank under accession nos GU220907–GU221904.

Data analysis

Summary statistics for the *sh4* locus and flanking genes for each population of interest were calculated as described previously (Caicedo *et al.* 2007). Statistics include Watterson's estimator nucleotide variation (θ_W), the average pairwise nucleotide diversity (θ_π) (Nei & Li 1979), and Tajima's *D* (Tajima 1989) for silent, synonymous, nonsynonymous and total sites (Table 1). Site type determination was based on annotations of the *O. sativa* genome (TIGR v. 5 January, 2008). Significance of Tajima's D values was tested using DNAsp (Rozas *et al.* 2003). Genealogical relationships among *sh4* alleles and flanking fragment alleles were determined with Maximum Parsimony (MP) and Neighbour Joining (NJ) analyses as implemented in MEGA 4 (Tamura *et al.* 2007). Both analyses considered pairwise deletion of

gaps/missing data. Distances were calculated using the Kimura 2-parameter model; branch bootstrap estimates were obtained from 1000 replicates. Heterozygotes were rare in our dataset, occurring occasionally only in *O. rufipogon*. When present, heterozygotes were phased using PHASE 2.1 prior to phylogenetic analyses (Stephens *et al.* 2001; Stephens & Scheet 2005), and no ambiguity was observed. For all loci, both NJ and MP trees produced similar results, so only the NJ trees are shown. Extended Haplotype Homozygosity (EHH) across the sampled genomic region containing *sh4* was calculated as described by (Sabeti *et al.* 2002), to test for extended linkage disequilibrium around the putatively selected mutation and assess the possibility of a selective sweep.

Results

The shattering phenotype in weedy, wild and cultivated rice

We recorded the degree of seed shattering for 90 accessions representing multiple groups of weedy, wild, and cultivated *Oryza* (Fig. 1, Table S1, Supporting information). Degree of shattering is a quantitative and highly variable trait (Ji *et al.* 2006). Our measurements revealed that some cultivated rice individuals show high variability in shattering within a single panicle (Table S1, Supporting information), with BTS for individual seeds occasionally varying by 10–200 g; however, extreme differences in BTS values, when present, occur for very few seeds within a panicle. In contrast, variation in shattering levels within panicles is much lower in weedy and wild rice accessions (Table S1, Supporting information). For all samples, mean and median shattering values are typically within 10 g.

Mean shattering differences among all measured *Oryza* accessions ranged widely, with values close to 0 g corresponding to a highly shattering phenotype, and values close to 100 g corresponding to complete nonshattering (Fig. 1, Table S1, Supporting information). In practice, BTS values of 5 g or less are considered shattering, as these seeds can be easily brushed off during measuring device attachment. Broad differences were observed across *Oryza* groups, and a Kruskal–Wallis test confirmed that variety has a significant effect on shattering levels ($P = 0.0013$).

Although lack of shattering is a hallmark of rice domestication, cultivated Asian rice varieties display a range of seed shattering phenotypes, with BTS values ranging from nearly 0 to 140 g (Fig. 1, Table S1, Supporting information). The *aus* group, in particular, shows a much narrower range of values (0–50 g), compared to *indica* (5–140 g) and *tropical japonica* (10–120 g) (Fig. 1). Additionally, one *indica* and one *aus* accession

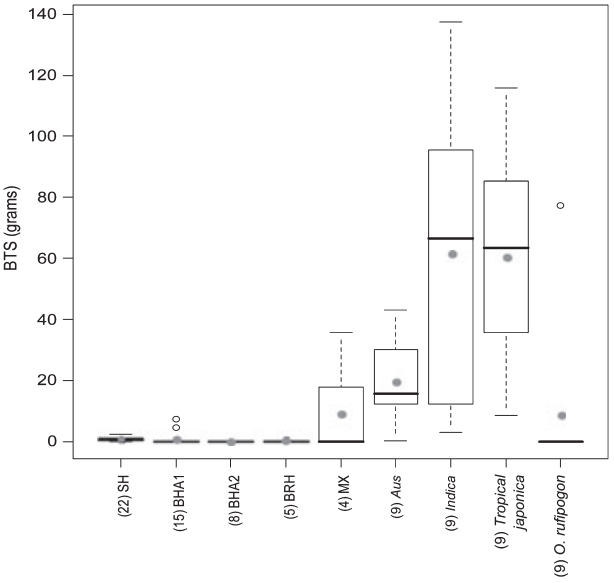

Fig. 1 Seed shattering phenotype in weedy, wild and cultivated *Oryza*. Distributions are of average accession BTS values for each *Oryza* group. The black line represents the median of each distribution, and the grey dot the mean; white dots represent outliers. Numbers in parenthesis correspond to sample sizes. Weedy rice groups are as follows: SH (straw-hulled), BHA1 and BHA2 (black hulled and awned), BRH (brown hulled) and MIX (mixed origin). Both *O. rufipogon* and *O. nivara* accessions have been grouped together under the heading *O. rufipogon*.

in our sample have average BTS values less than 5 g and may be considered shattering.

In contrast to cultivated rice, almost all of the wild Asian rice, *O. rufipogon* and *O. nivara* (Fig. 1), show BTS values of zero, indicating that the species are highly shattering. All weedy rice accessions, with the exception of a single individual (1B06, Table S1, Supporting information), show a propensity to shatter, registering BTS values very close to zero. Nonshattering weedy accession 1B06 has been shown to possibly have mixed ancestry (MX) (Reagon *et al.* 2010), and may have acquired additional nonshattering alleles through hybridization with cultivated rice. A single observed nonshattering *O. rufipogon* accession (2C04), on the other hand, does not resemble cultivated rice phenotypically or genetically (Reagon *et al.* 2010), suggesting that the nonshattering phenotype is not due to introgression from the crop.

Diversity at the qsh1 locus

We genotyped *Oryza* accessions at the *qsh1* locus, to determine whether the previously identified mutation (Konishi *et al.* 2006) might play a role in the shattering phenotype of weedy rice. All weeds and the majority of

rice cultivars were found to have the ancestral SNP, which also characterizes *O. rufipogon* and wild rice species, and is associated with higher levels of shattering (Table S1, Supporting information). Consistent with results from other research groups, we find that the nonshattering mutation is limited to two of our accessions belonging to the *temperate japonica* group (Table S1, Supporting information), and that the SNP is most likely not involved in variation in shattering levels outside of a small group within this cultivated variety (Konishi *et al.* 2006; Zhang *et al.* 2009).

The genealogy of sh4

To determine if the shattering locus, *sh4*, may underlie variation in shattering levels among cultivated and weedy rice, we sequenced the gene in a panel of 144 samples from weedy, cultivated and wild rice groups. The 3.9 kb of aligned sequence data includes the intron and both exons, plus 1040 bp of the promoter region upstream and 550 bp downstream of the *sh4* gene.

Relationships among haplotypes at the *sh4* locus (Fig. S1, Supporting information, Fig. 2) reveal a highly supported clade defined by the derived T mutation. As observed in previous research (Li *et al.* 2006; Zhang *et al.* 2009), all cultivated rice accessions sampled carry this mutation, which is associated with loss of shattering. Moreover, the majority of cultivated rice accessions share an identical haplotype across the 3.9 kb *sh4* region that we characterized. Three cultivars in our sample, one *aromatic*, one *tropical japonica* and one *aus*, differ from the common cultivated *sh4* haplotype by two, one and one nucleotide substitutions, respectively (Fig. S1, Supporting information, Fig. 2). These four SNPs have not been reported in other studies of the *sh4* locus to date, despite the detection of at least seven other low-frequency cultivated *sh4* haplotypes not detected here (Zhang *et al.* 2009). The two *aromatic* SNPs were the only ones found to occur in coding regions; one substitution alters amino acid 104 from a polar Serine to nonpolar Tryptophan, possibly resulting in the shattering phenotype in this individual (Fig. S2, Supporting information).

Eighteen *sh4* haplotypes were observed within wild *O. rufipogon* accessions. While the majority of the detected haplotypes are divergent from cultivated *sh4* alleles, six accessions carry an identical haplotype as the majority of cultivated rice, and two accessions carry haplotypes that differ by only one and three SNPs from this cultivated haplotype (Fig. S1, Supporting information, Fig. 2). Additionally, both *O. nivara* accessions sampled in this study have the same haplotype as the majority of cultivated rice (Fig. S1, Supporting information). These wild accessions were all found to shatter their seeds (BTS ~0 g, Table S1, Supporting information). The existence of shattering rice with the nonshattering T allele at *sh4* has not been previously reported (Li *et al.* 2006; Zhang *et al.* 2009), and indicates that the presence of this mutation alone is not sufficient to confer a reduction in shattering. Surprisingly, the single nonshattering *O. rufipogon* individual in our sample (2C04) does not carry the T mutation in *sh4*.

Contrary to our expectations, given their high propensity to shatter, all weedy rice accessions sampled carry the nonshattering associated T nucleotide in *sh4*. Moreover, the majority of weedy rice accessions, ~70%, have *sh4* haplotypes identical to the most common haplotype in cultivated rice. Four additional novel *sh4* haplotypes were detected in the 18 remaining accessions of weedy rice. Each of the four unique haplotypes differs from the main cultivated haplotype by a single SNP and is not shared with any cultivated or wild rice groups (Figs S1 and S2, Supporting information). Additionally, three of these SNPs are predicted to cause amino acid replacements and may have functional consequences.

Genealogy of the sh4 genomic region

To further elucidate the possible origin of *sh4* alleles in weedy rice, we examined phylogenetic relationships at loci increasingly distant from *sh4* in both the 5′ and 3′ directions in the genome. Six ~500 bp loci were chosen for analysis, spaced 7.9 kb, 600 kb and 1.2 Mb from *sh4* on the 5′ side of the gene and 300 kb, 1.1 Mb and 2.4 Mb from *sh4* on the 3′ side of the gene (Table S2, Supporting information). Further exploration on the 5′ side of *sh4* was not carried out, as the final fragment is within 50 kb of the telomere and only one other gene exists within this region. Two additional loci downstream of *sh4*, sts_040 and sts_021, examined in a previous study (Reagon *et al.* 2010), were also included in our analyses. The furthest locus, sts_021, is 7.9 Mb away from the centromere; thus, our sampling encompasses over two-thirds of the chromosome arm containing *sh4* (~16 Mb). Phylogenies of the eight selected loci surrounding *sh4* were produced to visualize changes in relationship of weedy, wild and cultivated alleles with distance from the *sh4* locus (Fig. 2). Because of their likely hybrid origin and rarity in U.S. rice fields (Reagon *et al.* 2010), BRH and MX groups were excluded from these phylogenetic analyses.

The resolution of relationships among *Oryza* groups varies greatly according to the diversity at each locus (Fig. 2). Because multiple sources of evidence support a minimum of two separate rice domestication events (e.g. Sang & Ge 2007a; Vaughan *et al.* 2008), we examined the *sh4* genomic region to determine at what point cultivated groups began to separate into distinct clades.

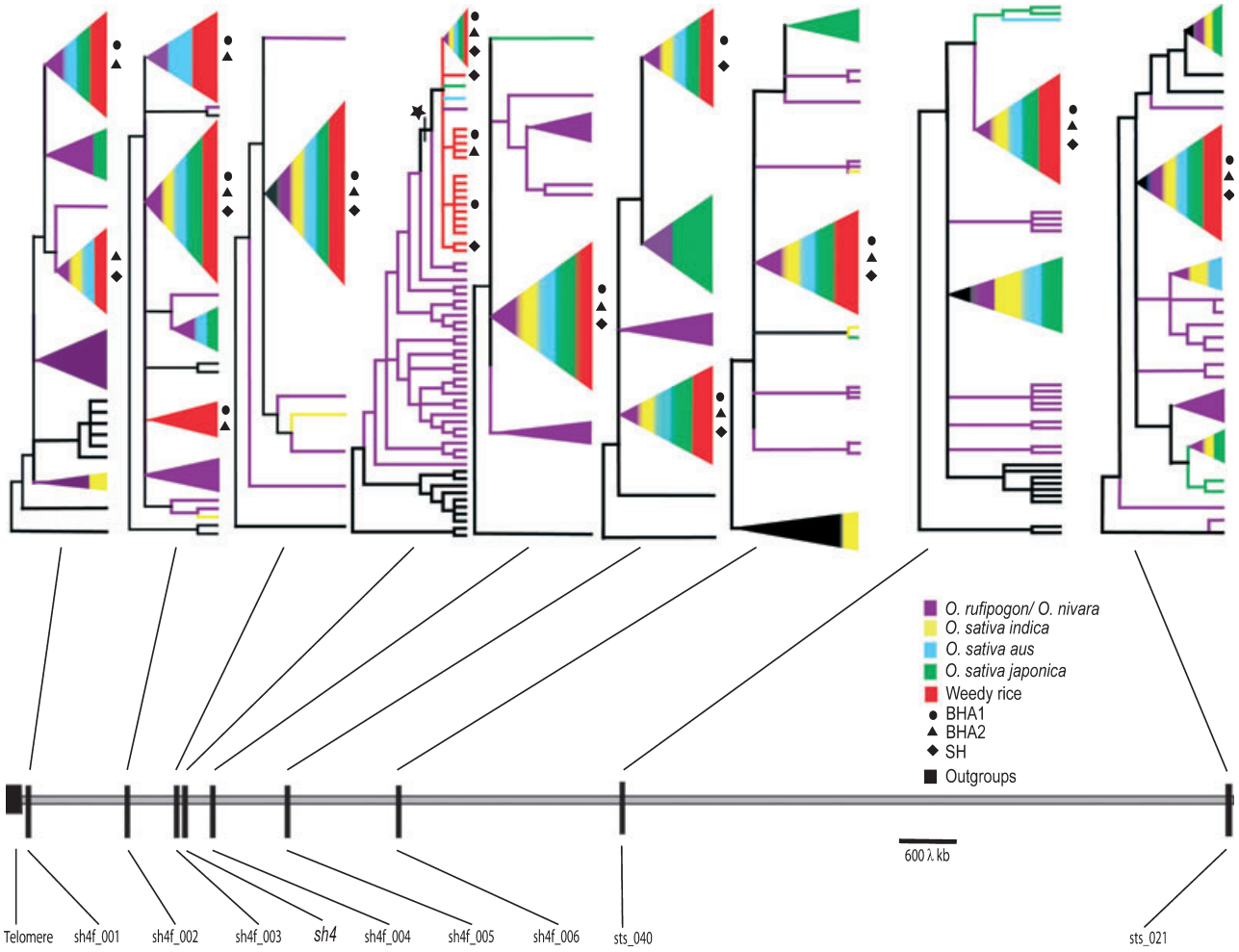

Fig. 2 Phylogenies of flanking regions surround *sh4*. Neighbour joining trees for each of eight ~500 bp regions at varying distances from the *sh4* locus. Diagram is to scale. Only branches with bootstrap values over 50% are shown. The star on the *sh4* locus tree denotes the T substitution associated with loss of shattering. For clarity, all *tropical japonica*, *temperate japonica* and *aromatic* rice have been grouped under the *japonica* heading and coloured green. Additionally, all weed groups have been coloured red, but the main groups are distinguishable via icons placed to the right of each tree.

Similar to *sh4*, most cultivated rice individuals share a single haplotype in the two closest flanking fragments sampled (sh4f_003 and sh4f_004; Fig. 2). This is consistent with hitchhiking of linked regions during selection on *sh4*; however, these loci are also highly conserved within all *Oryza* (Table 1). In the region upstream of *sh4*, multiple clades of domesticated rice appear ~600 kb (fragment sh4f_002), primarily due to diverse haplotypes in the *aus* and *japonica* groups. This trend continues 1.1 Mb upstream (fragment sh4f_001), but a clear separation into the two domesticated clades (*japonica* vs. *aus* and *indica*) is not seen. Downstream of *sh4*, greater haplotype diversity among cultivars is evident in fragment sh4f_005, ~1.1 Mb away and the remaining fragments. However, unlike many STS fragments previously examined (Caicedo *et al.* 2007), strong divergence of the two domesticated clades is not observed, with haplotype sharing evident among cultivated varieties in the sampled regions. This suggests that the effect of positive selection on *sh4* during rice domestication is evident throughout the genomic region sampled (see below).

In most fragments flanking *sh4*, weedy rice groups share haplotypes with cultivated rice varieties (Fig. 2). As expected, weedy groups tend to share haplotypes with their putative ancestors; thus, the majority of SH weeds group with *indica* cultivars (e.g. fragment sh4f_001), and the majority of BHA1 and BHA2 weeds group with *aus* cultivars. However, novel weed haplotypes were also observed in some fragments sampled; for example, some BHA1 and BHA2 weeds (11 accessions) share an identical haplotype in fragment sh4f_002 not seen in any other *Oryza* group. Moreover, in nearly every clade containing both weeds and

cultivars, some wild *Oryza*, principally *O. rufipogon* or *O. nivara*, is also present (Fig. 2).

Because a simple look at genealogical relationships within individual fragments in the *sh4* genomic region does not immediately reveal the source of weedy *sh4* alleles, we examined concatenated SNP haplotypes across the region (Fig. 3). Within 6.2 Mb (up to sts_040) surrounding *sh4*, 13 SH weed accessions are identical to a single *indica* accession (2B02), and seven SH weeds and a single BHA1 weed are identical to three *indica* cultivars. Additionally, two BHA1 and four BHA2 accessions are identical to a single *aus* accession (3A06). When the region 14 Mb away from *sh4* is included (sts_021) only the weeds identical to the *aus* accession remain grouped, indicating a breakdown of the other associations due to recombination. The lack of extended haplotype sharing between weeds and *tropical japonica*, suggests that weeds cannot have acquired *sh4* alleles through introgression with the local crop.

We also examined concatenated SNP haplotypes for *O. nivara* or *O. rufipogon* accessions sharing the common domesticated *sh4* haplotype. The seven SH and single BHA1 accessions that share extended haplotypes with the three *indica* cultivars, are identical to a single *O. nivara* (2F01) and a single *O. rufipogon* (2C09) across a 6.2-Mb region (Fig. 3). Once the region 14 Mb away is added, these two wild accessions no longer group with the weeds yet still group with two *indica*s. Of the remaining wild accessions, a single *O. rufipogon* (2D06) is identical to a single *indica* (3A11) accession, but no other possess haplotypes identical to weeds or cultivars across the *sh4* genomic region. The greater sharing of extended haplotypes between weeds and cultivars than between weeds and wild rice strongly suggest that weedy rice populations have inherited the derived *sh4* T substitution from domesticated ancestors.

The impact of a selective sweep in the sh4 genomic region

The ubiquity of the derived *sh4* T substitution among cultivated rice accessions, stemming from multiple domestication events, suggests that *sh4* has been subjected to strong positive selection during the domestication process (Vaughan *et al.* 2008; Zhang *et al.* 2009). To determine how positive selection on *sh4* in cultivated rice has affected *sh4* diversity in weedy rice groups, we assessed levels of genetic diversity at the sampled regions. As expected, silent site nucleotide diversity at *sh4* is very low in cultivated rice (Fig. S3, Supporting information, Table 1). Values for *indica*, *aus* and *tropical japonica*, the three rice varieties most likely to have contributed to weedy rice, are all over an order of magnitude smaller than genome-wide averages estimated from a set of 111 STS loci (1.9, 1.9 and 1.6 per kb, respectively) (Caicedo *et al.* 2007). A recent study reported higher levels of *sh4* variation in cultivated groups, but still well below genome averages (Zhang *et al.* 2009). Conversely, *sh4* nucleotide diversity in *O. rufipogon* (Table 1) is close to the genomic average

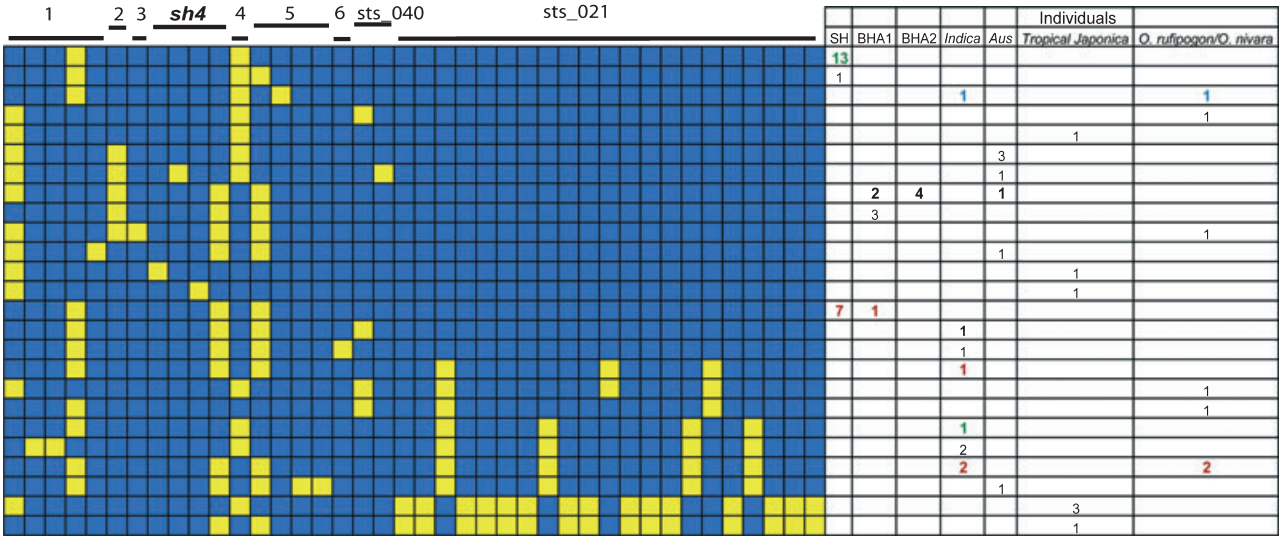

Fig. 3 Graphical view of concatenated *sh4* haplotypes. Haplotypes across the genomic region surrounding *sh4* are shown for the 90 individuals (wild, weedy, and cultivated) that share the common *sh4* haplotype containing the T SNP. The numbers across the top represent flanking regions (1–6 = sh4f_001–_006). Yellow squares represent SNPs found in at least one haplotype. A tally of individuals from each cultivated, weedy, or wild group is shown to the right. Bolded colours of accession counts indicate haplotypes that are identical across a 6.2-Mb region (up to sts_040) containing *sh4*.

(~5.2 per kb) (Caicedo *et al.* 2007) and in line with the diversity seen in Zhang *et al.* (2009).

The three main groups of U.S. weedy rice also show a reduction in nucleotide diversity at *sh4*, but the level of reduction differs among groups. Silent site nucleotide diversity values for SH, BHA1 and BHA2 range from 0 to 0.2 per kb (Table 1), while their genome wide averages based on 48 STS loci are 0.692, 0.829 and 0.657, respectively (Reagon *et al.* 2010). In general, the reduction in diversity at *sh4* compared to genomic values in weedy rice groups is less drastic than in cultivated rice, perhaps due to the genome-wide low levels of diversity associated with the bottlenecks giving rise to weedy groups (Reagon *et al.* 2010). Surprisingly, the BHA2 group showed only a mild decrease in diversity at *sh4* and a positive Tajima's D (Table 1), consistent with the presence of two moderate frequency haplotypes.

In cultivated and most weedy rice groups, there is also a decrease in diversity, to differing degrees, in genes flanking *sh4* (Table 1). The majority of loci sampled show diversity below the genome average within all cultivars. The *indica*, *aus* and *tropical japonica* groups have negligible amounts of diversity in fragments sampled up to 1.2 Mb on 5′ side (sh4f_002) and 1.1 Mb on 3′ side (sh4f_005) (Table 1), consistent with a selective sweep in the region. However, these fragments also show low levels of diversity in *O. rufipogon,* in line with overall reduced diversity previously reported on this arm of chromosome 4 (Mather *et al.* 2007). Remarkably high levels of diversity are evident in the furthest locus sampled from the *sh4* gene, sts_021, which shows a particularly drastic increase in diversity in *indica* and *tropical japonica* varieties.

Most loci sampled in the *sh4* genomic region show no diversity in the three major weedy rice populations, consistent with the proposed bottlenecks at founding (Reagon *et al.* 2010). Remarkably, however, some fragments in the *sh4* genomic region display higher levels of diversity in weedy groups than their putative progenitors (Table 1, Fig. S3, Supporting information). In particular, the BHA2 group is highly diverse at *sh4* and locus sh4f_002; because some BHA2 haplotypes at these loci are not found in other cultivated or wild *Oryza* groups sampled (Fig. 2, Fig. S1, Supporting information), high diversity levels may be due to inheritance from diverse unidentified ancestors, or new mutations since the origin of the weedy group.

To further assess the genomic impact of selection on the *sh4* T substitution in cultivated rice and subsequent inheritance in weedy rice, we determined the breakdown of linkage disequilibrium (LD) across the *sh4* region using the Extended Haplotype Homozygosity (EHH) analysis (Sabeti *et al.* 2002). As expected, homozygosity breaks down most quickly for the *O. rufipogon* group possessing the ancestral G substitution in *sh4*, within 100 bases of the SNP (Fig. 4A). For *O. rufipogon* accessions containing the derived T substitution, breakdown occurs more slowly, consistent with its derived status. For both groups homozygosity is at or near zero within 1.1 Mb downstream of the mutation.

In contrast to wild rice, and indicative of strong positive selection on *sh4*, cultivated rice groups all have more extensive haplotype homozygosity throughout the examined genomic region (Fig. 4B). Particularly noteworthy is the fact that *indica* shows no breakdown of homozygosity within *sh4*, although the *aus* and *tropical japonica* groups do. No group reached an EHH value of zero upstream of *sh4* within the region sampled; however, downstream of the gene, *tropical japonica* is the first group to reach a homozygosity value of zero. These patterns of LD suggest that *sh4* originated in the ancestors of *tropical japonica* and subsequently introgressed into *indica*, where there may have been less time for recombination to lead to breakdown of LD.

Homozygosity patterns for weed groups in the *sh4* genomic region are similar to those of the cultivars above but show a much slower breakdown of LD overall (Fig. 4C). Unlike cultivated rice, however, all weedy groups possess unique SNPs within *sh4*. This accounts for the initial breakdown of homozygosity within the gene. The high levels of homozygosity observed for weedy groups are consistent with inheritance of *sh4* alleles from ancestors with low levels of diversity and high levels of LD within the *sh4* genomic region.

Discussion

The loss of shattering as a seed dispersal mechanism is a key domestication trait, distinguishing cultivated cereals from their wild relatives. Seed shattering is also a trait associated with weed fitness, with increased levels of seed dispersal likely favoured in weeds infesting agricultural systems (Harlan & DeWet 1965). Recent advances dissecting the genetic basis of seed shattering variation in cultivated and wild rice (Konishi *et al.* 2006; Li *et al.* 2006; Lin *et al.* 2007) offer a unique opportunity to assess the evolution of this fitness-related trait in populations of weedy rice.

Multiple populations of weedy rice with independent origins exist in the U.S. (Londo & Schaal 2007; Gealy *et al.* 2009; Reagon *et al.* 2010). Surveys of polymorphism have shown that the main populations of U.S. weedy rice share genetic backgrounds with, and are possibly descendants of, *indica* and *aus* cultivated rice varieties (Londo & Schaal 2007; Reagon *et al.* 2010). We have confirmed that all U.S. weedy rice populations are highly shattering (Fig. 1). The near complete lack of variability in this trait across weedy rice groups

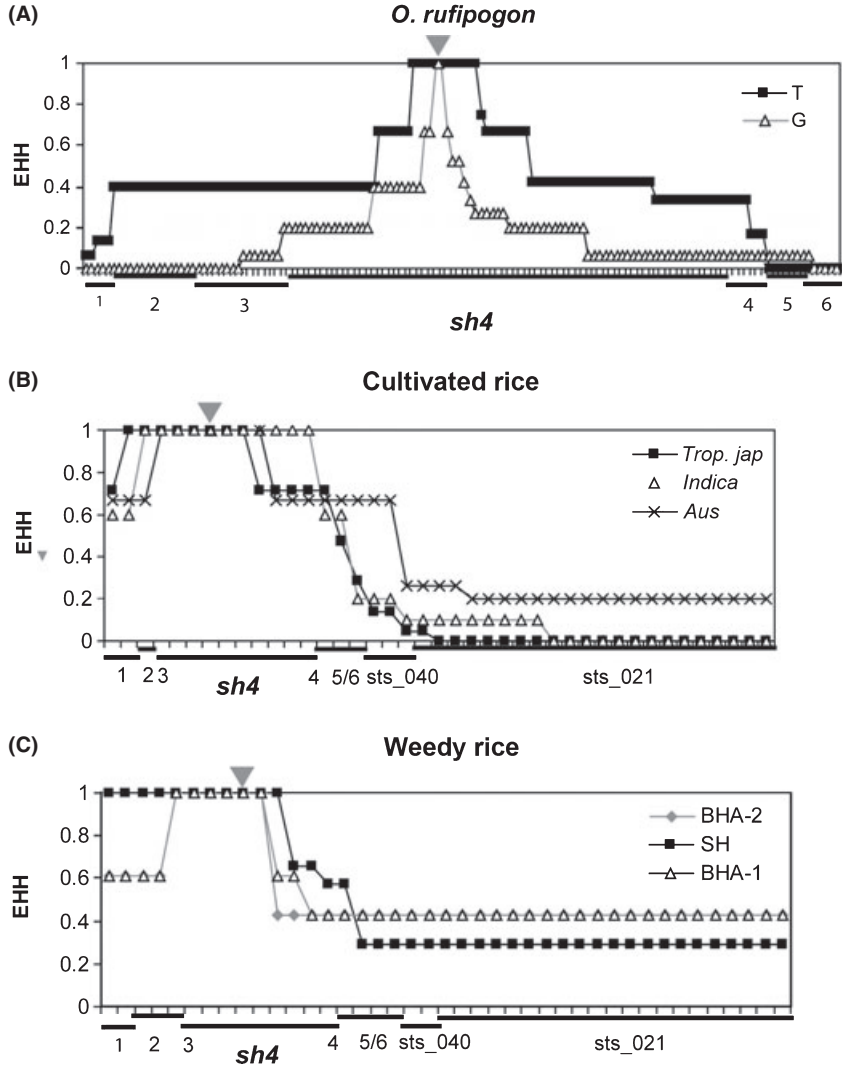

Fig. 4 Extended haplotype homozygosity surrounding *sh4*. EHH was performed on concatenated alignments containing the *sh4* gene and all eight flanking regions in order as they appear on the chromosome. Sts_040 and sts_021 were not included for *O. rufipogon* as haplotype homozygosity had already reached zero. The grey triangle atop each panel represents the location of the T mutation associated with loss of shattering in *sh4*. Numbers under black bars represent flanking regions (1–6 = sh4f_001–_006). (A) EHH for *O. rufipogon* groups possessing a T or a G at the SNP associated with shattering variation. (B) EHH results for three cultivated rice groups. (C) EHH results for the main U.S. weedy rice groups.

contrasts with the variance in shattering levels in cultivated rice varieties. The fact that all weedy rice shatters, despite separate origins of major weedy rice groups, suggests that shattering is a trait strongly selected for during weedy rice evolution. Coupled with genomic data indicating weedy rice origins from nonshattering ancestors, this pattern gives rise to questions about how weeds have acquired the shattering trait.

Environmental variation is known to affect the seed shattering trait in cultivated rice (Ji *et al.* 2006), and thus our shattering measurements could differ from those obtained under field conditions. However, extensive qualitative assessments of U.S. weedy rice in single and multiple U.S. rice fields report the U.S. weed as highly shattering (e.g. Noldin *et al.* 1999; Oard *et al.* 2000; Gealy *et al.* 2003; Delouche *et al.* 2007). Thus, our growth-chamber measurements of shattering levels in weedy rice seem consistent with observations in the weed's native environment. Likewise, multiple studies report wild rice as highly shattering in field conditions examined outside of the U.S. (e.g. Xiao *et al.* 1998; Cai & Morishima 2000), consistent with our results. Lastly, our measurements of U.S. cultivated *tropical japonica* varieties are consistent with low shattering levels of the crop in U.S. rice fields (Table S1, Supporting information). Thus, our measurements of shattering under

growth chamber conditions seem to accurately reflect known phenotypes of weedy and cultivated rice in U.S. fields.

To date, two loci of large effect have been shown to underlie the seed shattering trait in cultivated rice: *qsh1* and *sh4* (Konishi *et al.* 2006; Li *et al.* 2006; Lin *et al.* 2007). As reported by others (Konishi *et al.* 2006; Zhang *et al.* 2009), we found that the *qsh1* shattering associated SNP is only relevant to shattering variation within the cultivated *temperate japonica* group, where some individuals possess a derived mutation associated with extreme loss of shattering. All U.S. weedy rice individuals possess the ancestral allele that is common in all non*temperate japonica* cultivated and wild rice groups (Table S1, Supporting information).

In contrast to *qsh1*, *sh4* is considered to be a key gene under strong selection during rice domestication (Zhang *et al.* 2009). We found that all cultivated rice individuals examined are fixed for a T substitution in exon 1 of *sh4* (Fig. S1, Supporting information), which is associated with loss of shattering (Li *et al.* 2006). Moreover, consistent with prior observations (Li *et al.* 2006; Lin *et al.* 2007; Zhang *et al.* 2009), the majority of rice cultivars share an identical haplotype at *sh4*, suggesting a single origin of the nonshattering allele in domesticated rice. Surprisingly, despite their ability to shatter, our survey revealed that all U.S. weedy rice accessions carry the T substitution associated with nonshattering at *sh4*, and that most weeds share the common cultivated *sh4* haplotype (Fig. S1, Supporting information, Fig. 2). This demonstrates that the T substitution characteristic of cultivated *sh4* alleles is not sufficient for reduction of shattering in all genetic backgrounds.

Unequivocal determination of the ancestry of weedy rice from *sh4* sequence data is complicated by detection of the common cultivated *sh4* haplotype at low frequencies in wild rice accessions (6 out of 30 *O. rufipogon*). Three other surveys of *sh4* diversity, which have included *O. rufipogon* samples complementary to our own (>50), have not detected the common cultivated *sh4* haplotype in any *O. rufipogon* (Li *et al.* 2006; Lin *et al.* 2007; Zhang *et al.* 2009), which supports our conclusions regarding the rarity of this allele within wild rice. Interestingly, the wild rice accessions possessing the common cultivated *sh4* haplotype share at least 50% genomic identity with cultivated rice (Reagon *et al.* 2010), suggesting they may have acquired these alleles through introgression; however, intermediate crop-wild morphologies have not been observed for these accessions (e.g. height, tillering, hull colour, awns, etc.), and an ancestral existence of these alleles in wild rice cannot be completely ruled out.

We consider weedy inheritance of *sh4* alleles from wild ancestors unlikely for several reasons: (1) inheritance of the common cultivated *sh4* haplotype in the independently evolved SH, BHA1, and BHA2 weedy rice groups is more likely to have occurred from a group where the haplotype is nearly fixed (cultivated rice), than from one where it is rare; (2) for loci sampled across a 15.2-Mb genomic region surrounding *sh4*, clades containing SH weeds tend to contain at least one *indica* cultivar and clades containing BHA weeds tend to contain at least one *aus*, as expected from their genomic-inferred ancestry; (3) three distinct extended haplotypes across a 6.2-Mb genomic region containing *sh4* are shared among cultivated and weedy rice accessions, whereas a single extended haplotype is shared with wild rice (Figs 2 and 3).

Our identification of the main 'cultivated' *sh4* haplotype in all U.S. weedy rice groups constitutes the strongest evidence to date of an origin of these weeds from domesticated ancestors. If weeds inherited their *sh4* alleles from domesticated rice, two mechanisms could account for the novel SNPs carried by some weedy accessions at *sh4* and other sampled loci. The SNPs could have accumulated through mutation since divergence from cultivated ancestors, possibly aided by release from selection for nonshattering at *sh4*. Novel SNPs could also have been acquired through introgression with un-sampled wild and/or cultivated individuals.

The single origin of the *sh4* allele in cultivated rice is striking because a preponderance of evidence supports a minimum of two rice domestication events in different areas of Asia, one giving rise to the *indica* and *aus*, and another to the *japonicas* and *aromatic* group (see Sweeney & McCouch 2007). Several models have been proposed to account for this discrepancy (Lin *et al.* 2007; Sang & Ge 2007a; b). Recent evidence suggests that the *sh4* T mutation was first fixed in one set of cultivars, and quickly spread to independently domesticated rice groups via gene flow and selection (Zhang *et al.* 2009). The cultivated rice group in which the T substitution was initially fixed has not been identified, though some studies have suggested an origin in rice outside of China (Zhang *et al.* 2009). Haplotypes favoured by positive directional selection are expected to manifest an extended block of LD around the favoured mutation, and our survey of polymorphism in the *sh4* genomic region is consistent with strong selection on *sh4* in all cultivated rice groups prior to the evolution of weedy rice. Patterns of extended homozygosity in the region are also consistent with an origin of the *sh4* T mutation in ancestors of the *tropical japonica* group, with subsequent introgression into ancestors of *indica* (Fig. 4). Finer scale characterization of the *sh4* genomic region may be needed to rule out the effects of sampling stochasticity on the observed patterns.

The presence of 'nonshattering' *sh4* alleles in U.S. weedy rice despite their propensity to shatter (Fig. 1 and Fig. S1, Supporting information), implies that weedy groups must have re-acquired the shattering trait through the involvement of other, unidentified loci. These could be major loci that have not yet been identified within *Oryza*, or numerous loci with small effect that are thus difficult to detect. The ability to shatter despite having the T substitution is also present in some *O. rufipogon* and one *aus* cultivar. Alleles at genes facilitating shattering may have been acquired by weedy groups through *de novo* mutation, introgression from wild rice, or perhaps inherited from the few domesticated backgrounds that are able to produce BTS values at the lower end of the scale (Fig. 1). Whether divergent weedy rice groups have acquired the shattering traits through similar genetic mechanisms remains an open question. Ongoing fine scale characterization of the shattering trait via microscopy and BTS time-course evaluations across *Oryza* groups may help determine the likelihood of a shared genetic basis for shattering between wild and weedy rice. Ultimate identification of loci contributing to shattering in weedy rice may be facilitated by numerous QTL studies of this trait (Xiong *et al.* 1999; Cai & Morishima 2000; Thomson *et al.* 2003; Gu *et al.* 2005a; Ji *et al.* 2006; Onishi *et al.* 2007), including some involving crosses between Asian weeds and cultivated rice (Bres-Patry *et al.* 2001; Gu *et al.* 2005a). To shed further light on the genetic basis of shattering in U.S. weeds, we are currently generating mapping populations from U.S. weedy rice parents and their putative progenitors.

Assessments of genomic patterns of polymorphism have supported origin of U.S. weedy rice groups from two domesticated rice varieties, *indica* and *aus* (Londo & Schaal 2007; Reagon *et al.* 2010). In contrast, assessments of polymorphism at a candidate locus for pericarp colour, *Rc*, have revealed that alleles in weedy groups, which are exclusively red-pigmented, are not derived from alleles carried by the more common white pericarp cultivars (Gross *et al.* 2010). However, red pericarp cultivated rice varieties exist, implying that selection on *Rc* is likely to have been a feature during the development of modern cultivated varieties, rather than the early stages of rice domestication; thus, polymorphism at *Rc* suggests that U.S. weedy rice groups arose prior to the emergence of white-pericarp cultivated rice, perhaps from primitive red-pericarp domesticates (Gross *et al.* 2010). The *sh4* polymorphism data reported here further refines our understanding of the origin of U.S. weedy rice. All weed groups must have originated after the fixation of the nonshattering *sh4* allele in all cultivated rice groups. Thus, the progenitors of weedy rice must have been 'domesticated enough' to have undergone selection for reduced shattering. Future investigation of additional candidate domestication and weedy loci are likely to further contribute to our understanding of the evolutionary origins of this noxious weed.

Acknowledgements

We are grateful to D. Gealy for providing weedy rice accessions and S.R. McCouch for providing several accessions used in this study. Members of the Caicedo lab provided comments that much improved that manuscript. This project was funded by a grant from the U.S. National Science Foundation Plant Genome Research Programme (DBI-0638820) to A.L.C., K.M.O. and Y.J.

References

Basu C, Halfhill MD, Mueller TC, Stewart Jr CN (2004) Weed genomics: new tools to understand weed biology. *Trends in Plant Science*, **9**, 391–398.

Bres-Patry C, Lorieuz M, Clement G, Bangratz M, Ghesquiere A (2001) Heredity and genetic mapping of domestication-related traits in a temperate japonica weedy rice. *Theoretical and Applied Genetics*, **102**, 118–126.

Burgos NR, Norman RJ, Gealy DR, Black H (2006) Competitive N uptake between rice and weedy rice. *Field Crops Research*, **99**, 96–105.

Burgos NR, Norsworthy JK, Scott RC, Smith KL (2008) Red rice (*Oryza sativa*) status after 5 years of imidazolinone-resistant rice technology in Arkansas. *Weed Technology*, **22**, 200–208.

Cai HW, Morishima H (2000) Genomic regions affecting seed shattering and seed dormancy in rice. *Theoretical and Applied Genetics*, **100**, 840–846.

Caicedo AL, Williamson SH, Hernandez RD *et al.* (2007) Genome-wide patterns of nucleotide polymorphism in domesticated rice. *PLoS Genetics*, **3**, e163.

Cao Q, Lu B-R, Xia H *et al.* (2006) Genetic diversity and origin of weedy rice (*Orzya sativa f. spontanea*) populations found in North-eastern China revealed by simple sequence repeat (SSR) markers. *Annals of Botany*, **98**, 1241–1252.

Delouche JC, Burgos NR, Gealy DR *et al.* (2007) Seed shattering and dormancy in weedy rices. In: *Weedy Rices—Origin, Biology, Ecology and Control*, Chapter 4. Food and Agriculture Organization of the United Nations, Rome, Italy, **19**.

Fan C, Xing Y, Mao H *et al.* (2006) GS3, a major QTL for grain length and weight and minor QTL for grain width and thickness in rice, encodes a putative transmembrane protein. *Theoretical and Applied Genetics*, **112**, 1164–1171.

Fuller DQ, Qin L, Zheng Y *et al.* (2009) The domestication process and domestication rate in rice: spikelet bases from the lower yangtze. *Science*, **323**, 1607–1609.

Gealy D (2005a) Gene movement between rice (*Oryza sativa*) and weedy rice (*Oryza sativa*)—a U.S. temperate rice perspective. In: *Crop Ferality and Volunteerism* (ed. Gressel J), pp. 323–354. CRC Press, Boca Raton, FL.

Gealy DR (2005b) Gene movement between rice (*Oryza sativa*) and weedy rice (*Oryza sativa*)—a US temperate rice

perspective. *Proceedings of the National Academy of Sciences, USA*, **99**, 11256–11259.
Gealy DR, Tai TH, Sneller CH (2002) Identification of red rice, rice, and hybrid populations using microsatellite markers. *Weed Science*, **50**, 333–339.
Gealy DR, Mitten DH, Niel-Rutger J (2003) Gene flow between red rice (*Oryza sativa*) and herbicide-resistant rice (*O. sativa*): implications for weed management. *Weed Technology*, **17**, 627–645.
Gealy DR, Agrama HA, Eizenga GC (2009) Exploring Genetic and spatial structure of U.S. weedy red rice (*Oryza sativa* L.) in relation to rice relatives worldwide. *Weed Science*, **157**, 627–643.
Gross BL, Reagon M, Hsu S-C, Caicedo AL, Jia Y, Olsen KM (2010) Seeing red: the origin of grain pigmentation in US weedy rice. *Molecular Ecology*, **19**, doi: 10.1111/j.1365-294X.2010.04707.x
Gu X-Y, Kianian SF, Foley ME (2005a) Seed dormancy imposed by covering tissues interrelates to shattering and seed morphological characters in weedy rice. *Crop Science*, **45**, 948–955.
Gu X-Y, Kianian SF, Hareland GA, Hoffer BL, Foley ME (2005b) Genetic analysis of adaptive syndromes interrelated with seed dormancy in weedy rice (*Oryza sativa*). *Theoretical and Applied Genetics*, **110**, 1108–1118.
Gu X-Y, Turnipseed EB, Foley ME (2008) The qSD12 Locus controls offpring tissue-imposed seed dormancy in rice. *Genetics*, **179**, 2263–2273.
Harlan JR (1992) *Crops & Man*. American Society of Agronomy, Madison, WI.
Harlan JR, DeWet JM (1965) Some thoughts about weeds. *Economic Botany*, **19**, 16–24.
Hillis DM, Mable BK, Moritz C (1996) Applications of molecular systematics: the state of the field and a look to the future. In: *Molecular Systematics* (eds Hillis DM, Moritz C, Mable BK). Sinauer Associates, Inc., Sunderland, MA.
Ji H-S, Chu S-H, Jiang W *et al.* (2006) Characterization and mapping of a shattering mutant in rice that corresponds to a block of domestication genes. *Genetics*, **173**, 995–1005.
Konishi S, Izawa T, Lin SY *et al.* (2006) An SNP caused loss of seed shattering during rice domestication. *Science*, **312**, 1392–1396.
Li C, Zhou A, Sang T (2006) Rice domestication by reducing shattering. *Science*, **311**, 1936–1939.
Lin Z, Griffith ME, Li X *et al.* (2007) Origin of seed shattering in rice (*Oryza sativa L.*). *Planta*, **226**, 11–20.
Londo JP, Schaal BA (2007) Origins and population genetics of weedy red rice in the USA. *Molecular Ecology*, **16**, 4523–4535.
Mather KA, Caicedo AL, Polato NR *et al.* (2007) The extent of linkage disequilibrium in rice (*Oryza sativa* L.). *Genetics*, **177**, 2223–2232.
Neff MM, Turk E, Kalishman M (2002) Web-based primer design for single nucleotide polymorphism analysis. *Trends in Genetics*, **18**, 613–615.
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**, 5269–5283.
Noldin JA, Chandler JM, McCauley GN (1999) Red rice (*Oryza sativa*) biology. I. Characterization of red rice ecotypes. *Weed Technology*, **13**, 12–18.
Oard J, Cohn MA, Linscombe S, Gealy D, Gravois K (2000) Field evaluation of seed production, shattering, and dormancy in hybrid populations of transgenic rice (*Oryza sativa*) and the weed, red rice (*Oryza sativa*). *Plant Science*, **157**, 13–22.
Olsen KM, Caicedo AL, Polato N *et al.* (2006) Selection under domestication: evidence for a sweep in the rice Waxy genomic region. *Genetics*, **173**, 975–983.
Olsen KM, Caicedo AL, Jia Y (2007) Evolutionary genomics of weedy rice in the USA. *Journal of Integrative Plant Biology*, **49**, 811–816.
Onishi K, Takagi K, Kontani M, Tanaka T, Sano Y (2007) Different patterns of genealogical relationships found in the two major QTLs causing reduction of seed shattering during rice domestication. *Genome*, **50**, 757–766.
Purugganan MD, Fuller DQ (2009) The nature of selection during plant domestication. *Nature*, **457**, 843–848.
Reagon M, Thurber CS, Gross BL, Olsen KM, Jia Y, Caicedo AL (2010) Genomic patterns of nucleotide diversity in parallel populations of U.S. weedy rice. *BMC Evolutionary Biology*, **10**, 180.
Rozas J, Sanchez-DelBarrio JC, Messeguer X, Rozas R (2003) DnaSP, DNA polymorphism analyses by the coalescent and other methods. *Bioinformatics*, **19**, 2496–2497.
Rozen S, Skaletsky HJ (2000) Primer3 on the WWW for general users and for biologist programmers. In: *Bioinformatics Methods and Protocols: Methods in Molecular Biology* (ed. Krawetz SMS), pp. 365–386. Humana Press, Totowa, NJ.
Sabeti PC, Reich DE, Higgins JM *et al.* (2002) Detecting recent positive selection in the human genome from haloptype structure. *Nature*, **419**, 832–837.
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.
Stephens M, Scheet P (2005) Accounting for decay of linkage disequilibrium in haplotype inference and missing data imputation. *American Journal of Human Genetics*, **76**, 449–462.
Stephens M, Smith NJ, Donnelly P (2001) A new statistical method for haplotype reconstruction from population data. *American Journal of Human Genetics*, **68**, 978–989.
Sweeney M, McCouch S (2007) The complex history of the domestication of rice. *Annals of Botany*, **100**, 951–957.
Tajima F (1989) Statistical-method for testing the neutral mutation hypothesis by DNA polymorphism. *Genetics*, **123**, 585–595.
Tamura K, Dudley J, Nei M, Kumar S (2007) MEGA4: molecular evolutionary genetics analysis (MEGA) software version 4.0. *Molecular Biology and Evolution*, **24**, 1596–1599.
Thomson MJ, Tai TH, McClung AM *et al.* (2003) Mapping quantitative trait loci for yield, yield components and morphological traits in an advanced backcross population between *Oryza rufipogon* and the *Oryza sativa* cultivar Jefferson. *Theoretical and Applied Genetics*, **107**, 479–493.
Vaughan LK, Ottis BV, Prazak-Havey AM, Sneller C, Chandler JM, Park WD (2001) Is all red rice found in commercial rice really *Oryza sativa*? *Weed Science*, **49**, 468–476.
Vaughan DA, Lu B-R, Tomooka N (2008) The evolving story of rice evolution. *Plant Science*, **174**, 394–408.

Xiao J, Li J, Grandillo S *et al.* (1998) Identification of trait-improving quantitative trait loci alleles from a wild rice relative, *Oryza rufipogon*. *Genetics*, **150**, 899–909.

Xing YZ, Tang WJ, Xue WY, Xu CG, Zhang Q (2008) Fine mapping of a major quantitative trait loci, qSSP7, controlling the number of spikelets per panicle as a single Mendelian factor in rice. *Theoretical and Applied Genetics*, **116**, 789–796.

Xiong LZ, Liu KD, Dai XK, Xu CG, Zhang Q (1999) Identification of genetic factors controlling domestication-related traits of rice using an F2 population of a cross between Oryza sativa and O. rufipogon. *Theoretical and Applied Genetics*, **98**, 243–251.

Zhang L-B, Zhu Q, Wu Z-Q *et al.* (2009) Selection on grain shattering genes and rates of rice domestication. *New Phytologist*, **184**, 708–720.

Zhu QH, 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.

This work is part of C.T.'s PhD thesis on convergent evolution among weedy rice groups. M.R. and B.L.G are postdoctoral researchers interested in the evolutionary genetics of crop weeds. K.M.O. uses population genetics to study the genetics to study the genetic basis of adaption in wild, weedy and cultivated plants. Y.J. studies the evolution and functional genetics of pathogen resistance in cultivated rice and relatives. A.L.C.'s research involves population genomic approaches to understand the evolutionary dynamics of wild, weedy and cultivated species.

Supporting information

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

Table S1 List of accessions used for this study**.** Accessions are grouped by type (weed, wild or cultivar). Identification numbers as well as genotypes at *qsh1* and *sh4* are listed along with phenotypic values for seed shattering

Table S2 List of primer sequences and their location**.** Primers are grouped by gene (*sh4* and *qsh1*) as well as genetic vs. flanking region. Additionally, the enzyme used in the *qsh1* CAPS study is identified and the cut site is listed

Fig. S1 Neighbour Joining Tree of *sh4* haplotypes. Numbers below branches represent bootstrap support in percentages; only clades with over 50% support are labelled. The black star denotes the G to T substitution associated with loss of shattering in domesticated rice. Colour key at left of the tree identifies *Oryza* groups represented by the observed haplotypes. The *O. sativa* group contains accessions of the five recognized domesticated rice populations: *aus*, *indica*, *aromatic*, *tropical japonica* and *temperate japonica*. Labels on the right side the tree identify the number of individuals sharing a haplotype. A triangle is placed anywhere more than ten individuals share an identical haplotype. Four haplotypes unique to weedy rice are numbered (I, II, III, and IV) while haplotypes unique to *O. rufipogon* are not labelled or numbered. Three of the unique weedy rice haplotypes contain mutations that alter amino acids: Glutamic Acid to Lysine in exon 1 in haplotype II, Arginine to Leucine in exon 2 in haplotype III, and Arginine to Tryptophan in exon 2 in haplotype I.

Fig. S2 Graphical view of unique *sh4* haplotypes**.** The top haplotype represents the common shared cultivated haplotype found in 90 individuals from cultivated, weedy, and wild groups. Of the three unique cultivar haplotypes, only the *aromatic* individual (2B01) contains a nonsynonymous SNP. Four unique weedy haplotypes (I–IV) are displayed where three of the four contain nonsynonymous SNPs. Haplotype numbers match those of Fig. S1. Additionally, three wild individuals are shown. 2E01 and 2C03 contain the nonshattering T nucleotide plus additional coding and noncoding SNPs. 2C05 was chosen to represent one of the many wild haplotypes containing a shattering G nucleotide for comparison.

Fig. S3 Ratios of silent site nucleotide diversity at *sh4* and surrounding loci. The ratio of the average pairwise nucleotide diversity (θ_π) per kb is shown (A) for three cultivated groups (*indica*, *aus* and *tropical Japonica*) against wild *O. rufipogon* and (B) for the three major weed groups (SH, BHA1 and BHA2) against their putative progenitors (*Indica* or *Aus*). Overall diversity is low across the entire region in cultivated and weedy rice groups.

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 corresponding author for the article.