MOLECULAR ECOLOGY

Molecular Ecology (2014) **23**, 5003–5017 doi: 10.1111/mec.12922

Malaysian weedy rice shows its true stripes: wild *Oryza* and elite rice cultivars shape agricultural weed evolution in Southeast Asia

BENG-KAH SONG,*† TSE-SENG CHUAH,‡ SHEH MAY TAM§ and KENNETH M. OLSEN¶

*School of Science, Monash University Malaysia, 46150 Bandar Sunway, Selangor, Malaysia, †Genomics Facility, Monash University Malaysia, 46150 Bandar Sunway, Selangor, Malaysia, ‡Department of Agro-technology, Faculty of Agro-technology and Food Science, Universiti Malaysia Terengganu, 21030 Kuala Terengganu, Malaysia, §School of Biosciences, Taylor's University, 47500 Subang Jaya, Selangor, Malaysia, ¶Department of Biology, Washington University in St Louis, St. Louis, MO 63130, USA

Abstract

Weedy rice is a close relative of domesticated rice (*Oryza sativa*) that competes aggressively with the crop and limits rice productivity worldwide. Most genetic studies of weedy rice have focused on populations in regions where no reproductively compatible wild *Oryza* species occur (North America, Europe and northern Asia). Here, we examined the population genetics of weedy rice in Malaysia, where wild rice (*O. rufipogon*) can be found growing in close proximity to cultivated and weedy rice. Using 375 accessions and a combined analysis of 24 neutral SSR loci and two rice domestication genes (*sh4*, controlling seed shattering, and *Bh4*, controlling hull colour), we addressed the following questions: (i) What is the relationship of Malaysian weedy rice to domesticated and wild rice, and to weedy rice strains in the USA? (ii) To what extent does the presence of *O. rufipogon* influence the genetic and phenotypic diversity of Malaysian weeds? (iii) What do the distributions of *sh4* and *Bh4* alleles and associated phenotypes reveal about the origin and contemporary evolution of Malaysian weedy rice? Our results reveal the following: independent evolutionary origins for Malaysian weeds and US strains, despite their very close phenotypic resemblance; wild-to-weed gene flow in Malaysian weed populations, including apparent adaptive introgression of seed-shattering alleles; and a prominent role for modern Malaysian cultivars in the origin and recent proliferation of Malaysian weeds. These findings suggest that the genetic complexity and adaptability of weedy crop relatives can be profoundly influenced by proximity to reproductively compatible wild and domesticated populations.

Keywords: agricultural weeds, genetic introgression, hull colour, red rice, shattering, wild crop relatives

Received 19 June 2014; revision received 30 August 2014; accepted 3 September 2014

Introduction

Agricultural weeds pose a primary challenge to crop productivity worldwide, and their potential for rapid adaptation has made them appealing targets for evolutionary genomic studies (Stewart *et al.* 2009; Vigueira *et al.* 2013b). Weeds that are closely related to crop species have the potential for particularly complex evolutionary dynamics. Weedy crop relatives are often found exclusively in agricultural settings, where they have evolved to compete with crops through a combination of mimicry, rapid growth and reproduction, and persistence as dormant seed in crop soils (De Wet & Harlan 1975). In regions where crops are grown in

Correspondence: Beng-Kah Song, Fax: +603 55146099; E-mail: song.beng.kah@monash.edu and Kenneth M. Olsen, Fax: +1 314 935 4432; E-mail: kolsen@wustl.edu

proximity to related wild species, interfertile weeds may potentially acquire adaptive traits through both hybridization with cultivated plants and gene flow from wild populations. Despite their massive economic impact on worldwide agriculture, little is known about the origin of most weedy crop relatives or about the relative contributions of domesticated and wild populations in their ongoing evolution (Ellstrand *et al.* 2010; Vigueira *et al.* 2013b).

Weedy rice is a conspecific form of cultivated rice (*Oryza sativa* L.) that infests rice production areas worldwide. Also called red rice, it competes aggressively with cultivated rice, reducing yields in US rice fields by up to 80% and contaminating harvests with its dark-pigmented, undesirable grains (Estorninos *et al.* 2005; Gealy *et al.* 2012). Eradication efforts are hampered by the weed's close morphological and physiological similarity to its domesticated relative. A key distinguishing feature is the presence of highly shattering (disarticulating) inflorescences (Azmi & Karim 2008), a characteristic of wild *Oryza* species that was selected against during rice domestication (Zhang *et al.* 2009).

Studies of weedy rice populations worldwide using neutral markers have indicated a very close relationship to domesticated rice (Akasaka *et al.* 2009; Reagon *et al.* 2010; Grimm *et al.* 2013), suggesting an origin by dedomestication from feral crop strains. Domesticated rice is itself characterized by five major genetic subgroups: *indica* and *aus* (collectively referred to as the *indica* subspecies); and *tropical japonica*, *temperate japonica* and *aromatic* (comprising the *japonica* subspecies) (Kovach *et al.* 2007). In the southern USA, where weedy rice has been extensively studied, the two major weed forms, blackhull awned (BHA) and strawhull awnless (SH), appear to have originated in Asia from two different domesticated rice backgrounds – *aus* and *indica*, respectively – neither of which is cultivated in North America (Londo & Schaal 2007; Gealy *et al.* 2009; Reagon *et al.* 2010). US weeds show evidence of strong genetic bottlenecks consistent with accidental introductions into North America, possibly as seed contaminants of imported grain stocks (Reagon *et al.* 2010).

While genetic analyses suggest a very close relationship of weedy rice to domesticated rice, most studies to date have been conducted in world regions where there are no wild *Oryza* species present (e.g. North America, Korea, Japan, Europe and northern China) (Cao *et al.* 2006; Akasaka *et al.* 2009; Reagon *et al.* 2010; Chung & Park 2010; Zhu *et al.* 2012; but see also Pusadee *et al.* 2012; Zhang *et al.* 2012; Grimm *et al.* 2013). In regions where rice is cultivated in proximity to interfertile wild *Oryza* populations, gene flow and adaptive introgression could potentially play a major role in the weed's evolution. This may be especially likely in Southeast Asia, where populations of rice's wild progenitor, *Oryza rufipogon*, can be found growing at the borders of rice fields. For example, a recent study of Thai weedy rice has found evidence that local *O. rufipogon* populations may have contributed to the weed's genetic diversity (Pusadee *et al.* 2012). While weedy rice, like domesticated rice, is predominantly selfing, *O. rufipogon* is outcrossing, which would facilitate wild–weed gene flow. Moreover, while some *O. rufipogon* traits would most likely be maladaptive in an agricultural setting (e.g. prostrate growth, which would facilitate weed detection and reduce competitive ability in dense crop plantings), other wild traits such as seed dormancy and shattering would be highly advantageous in weed populations, as they allow for the establishment of a weed soil seed bank in crop fields. Selection for these weed-adaptive traits could thus favour hybridization and adaptive introgression from wild populations.

Among rice-growing regions of Southeast Asia, weed populations in Peninsular Malaysia present a particularly dynamic evolutionary system. Weedy rice was first reported in Malaysia in 1988 (Wahab & Suhaimi 1991), and it has emerged as a major problem only within the last two decades, as agricultural practices have shifted towards industrialization and direct seeding of crop fields by machine, which provides little opportunity for hand-weeding of fields. Industrialized farming in Peninsular Malaysia has also brought about the widespread planting of a relatively few 'elite' rice cultivars, which have largely displaced traditional Malaysian landraces in recent decades. These elite cultivars are derived from exotic *indica* germplasm developed at the International Rice Research Institute in the Philippines (Dalrymple 1986) and are genetically distinct from traditional Malaysian rice landraces, which comprise a mixture of *indica* and *japonica* varieties. Given their different genetic background, elite cultivars may also differ in their reproductive compatibility with weedy and/or wild populations; if highly interfertile, they could hybridize and potentially contribute to the weed's proliferation (e.g. Craig *et al.* 2014). Phenotypically, Malaysian weeds show far greater variation in both vegetative and reproductive traits than US strains; a recent morphological study identified more than 70 distinct morphotypes among Malaysian weeds (B. K. Song, unpublished observations; see also Fig. S1, Supporting information). This high phenotypic diversity suggests a level of population genetic complexity not present in other world regions.

In this study, we examine the origin and evolution of Peninsular Malaysian weedy rice in comparison with well-characterized weed populations in the USA. We use two complementary approaches, based on neutral markers and candidate genes. First, we assess the

genetic diversity and population structure using a panel of 24 polymorphic microsatellite (simple sequence repeat, or SSR) markers. These loci were originally identified and employed by Gealy *et al.* (2009) to examine the genetic structure of US weedy rice populations. In the present analysis, we combine the previously published North American data with Malaysian population samples, allowing a direct comparison of Malaysian and North American weed strains.

Second, because several key weediness traits correspond to well-characterized rice domestication genes, we can also directly examine the genetic underpinnings of weediness through candidate gene analyses. We focus on *sh4*, a major gene controlling the shattering phenotype in rice, and *Bh4*, a primary genetic determinant of hull colour. The *sh4* gene encodes a Myb3 transcription factor that functions in the formation of an abscission layer at the base of maturing caryopses, which allows grains to shatter when mature (Li *et al.* 2006). During rice domestication, a G-to-T reduced-function mutation in *sh4* exon 1 was a target of strong selection, leading to complete fixation of the reduced-shattering 'T' allele in all crop varieties surveyed to date (Li *et al.* 2006; Zhang *et al.* 2009). Interestingly, the 'T' allele has also been found to be completely fixed in all weedy rice samples examined to date, both in the USA and in worldwide (Thurber *et al.* 2010; Zhu *et al.* 2012); this is consistent with weed descent from domesticated germplasm (and also suggests that the 'T' allele alone is not sufficient to reduce shattering). However, studies to date have examined *sh4* variation primarily in weeds outside the geographical range of wild *Oryza* species. Detection of the wild 'G' allele in Malaysian weeds would serve as a strong indicator of wild-to-weed introgression in these Southeast Asian populations.

The other candidate gene, *Bh4*, encodes an amino acid transporter that, when functional, leads to the dark hull pigmentation characteristic of wild *Oryzas*. Selection for loss-of-function mutations has led to the non-pigmented (strawhull) phenotype characteristic of most cultivated rice varieties; a 22-bp frameshift deletion in exon 3 is the predominant crop allele, present in approximately 95% of strawhull varieties (Zhu *et al.* 2011). Unlike *sh4*, *Bh4* has been found to be polymorphic in weeds examined to date. Among US weed strains, strawhull weeds are fixed for the 22-bp deletion allele, while blackhull awned strains carry functional alleles that are closely related to alleles detected in *aus* rice (Vigueira *et al.* 2013a). The 22-bp deletion allele is also fixed in a phenotypically intermediate 'brownhull' US weed phenotype, indicating that *Bh4* alone cannot be the sole determinant of hull colour in weedy rice. In the context of Malaysian weedy rice evolution, allelic variation at *Bh4* could provide insight into the broad range of hull phenotypes present in these weeds (Fig. S1, Supporting information).

Through this combined analysis of SSRs and candidate genes, we address the following questions: (i) What is the relationship of Malaysian weedy rice to domesticated and wild rice, and to US weeds? Have Malaysian weeds, like US weeds, originated through dedomestication? If so, are Malaysian weeds derived from the same genetic backgrounds as US weedy rice? (ii) To what extent does the presence of *O. rufipogon* populations in Southeast Asia influence the evolution and genetic diversity of Malaysian weeds? Is there evidence of genetic introgression from wild populations into weeds? and (iii) What do the distributions of *sh4* and *Bh4* alleles and associated phenotypes reveal about the origin and contemporary evolution of Malaysian weedy rice? Our findings indicate an independent origin of Malaysian weeds from those in North America and a key role for both wild populations and recently introduced elite cultivars in the evolution of these Southeast Asian weed populations.

Materials and methods

Sampling and phenotypic classifications

A total of 207 seed samples, representing 17 populations of Peninsular Malaysian weedy rice distributed across six states, were collected in 2011 and 2012 (Fig. 1; Table S1, Supporting information). The sampled locations occur in three geographical regions within the major rice production area of Peninsular Malaysia (northwestern, northeastern and central-western), where infestation of weedy rice has been severe since the late 1980s (Wahab & Suhaimi 1991). Sampling was designed to maximize representation of the full range of grain phenotypes present in Malaysian weeds. Each accession corresponded to seeds harvested from a single mature panicle. Accessions were classified into six major morphotype groups, which together comprise 165 of the 207 samples. These include 36 strawhulled, awnless (SH), 17 strawhulled, awned (SHA), 68 brown-striped-hulled, awnless (BR), 28 brown-striped-hulled, awned (BRA), 5 blackhulled, awnless (BH) and 11 blackhulled, awned types (BHA) (Fig. S1, Supporting information). The remaining 42 accessions were determined to be phenotypically intermediate between strawhull and brown-striped hull with respect to hull coloration; these accessions were further subdivided as either intermediate-strawhulled-awnless types (mSH; 27 accessions) or intermediate-strawhulled-awned types (mSHA; 15 accessions) (Table S2, Supporting information).

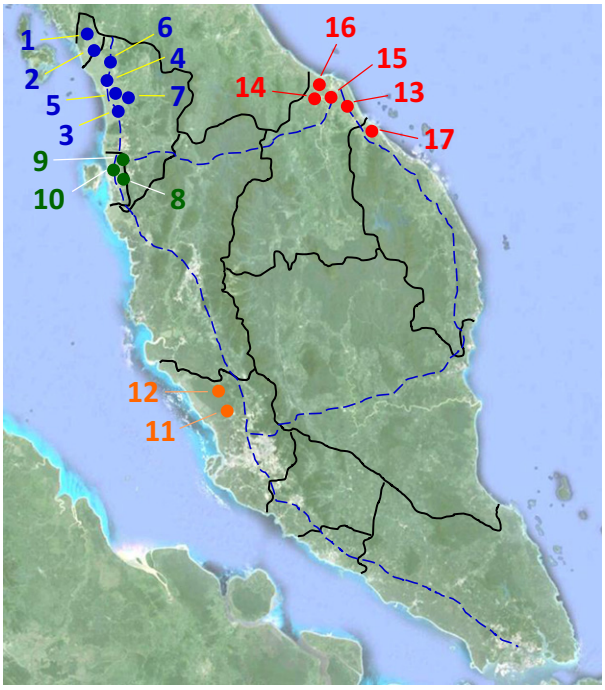

Fig. 1 Map of Peninsular Malaysia showing geographical collection localities for the weedy rice used in this study, colour-coded by sampling location (northwest: blue and green; central-west: orange; northeast: red). Blue dashed lines indicate major highways connecting rice sampling sites.

In addition to weedy rice samples, analyses also included 25 elite Malaysian cultivars collected from different rice planting areas (except for four that were provided by Rice Genebank, Seberang Perai, Malaysia) (designated CV-Mal); 17 Malaysian *O. rufipogon* accessions (OR_Mal); 18 *O. rufipogon* accessions representing other Southeast Asian countries (Thailand, Cambodia, India and Myanmar) (Oruf); and 16 Malaysian landrace accessions (Lr_Mal) (Table S1, Supporting information). Among the 17 wild Malaysian samples, 14 were generously provided from the Rice Genebank, Malaysian Agricultural Research & Development Institute (MARDI), Seberang Perai, and the other three were collected in the field. All wild rice and landrace accessions from outside of Malaysia were obtained from the International Rice Research Institute (Table S3, Supporting information). One plant per accession was grown in the greenhouses of either Washington University in St. Louis or Monash University Sunway Campus for DNA extraction.

DNA extraction and SSR analysis

Total genomic DNA was extracted from leaf tissue using a modified CTAB procedure (Doyle & Doyle 1990). Twenty-four microsatellite primer pairs (Table S4, Supporting information) were selected from a subset of the 31 used by Gealy *et al.* (2009) to characterize US weedy rice; loci were selected based on the evidence of polymorphism in the earlier study and marker dispersion across the rice genome. Loci were amplified in six multiplex PCR reactions, which include one hexaplex, one tetraplex, four triplex and one duplex PCR. Reactions were carried out in a volume of 8 μL containing 200 μM of each dNTPs, 20 mM Tris–HCl (pH 8.0), 0.2 μM of each primer (forward primers were labelled with either HEX, NED or 6FAM fluorescent dye), 2 mM $MgCl_2$, 0.2 unit Platinum *Taq* DNA polymerase (Invitrogen) and 20 ng genomic DNA. PCR amplifications were performed in a GeneAmp® PCR System 9700 (Applied Biosystems) according to the following PCR profile: 5 min at 94 °C for initial denaturation, followed by 25 cycles of 30 s at 94 °C, 30 s at 55 °C (except markers RM123, RM133 RM109, which required annealing temperature of 60 °C), and 1 min at 72 °C, and lastly 7 min at 72 °C for final extension.

All PCR products were electrophoresed on an ABI 3130xl capillary sequencer (Applied Biosystems) in the Biology Department of Washington University. Five previously genotyped accessions used in Gealy *et al.* (2009) (*indica* cv. Jasmine 85, *tropical japonica* cv. Cypress, *temperate japonica* cv. Nipponbare and US weedy rice accessions LA3 and TX4) were included in all genotyping runs; these reference standards were used for calibration to ensure consistency in assignment of allele sizes across Malaysian and US samples. SSR allele sizes were determined using GENEMAPPER 3.7 software (Applied Biosystems).

Genetic diversity and population structure analyses

SSR genotype data were used to calculate observed allele number per sampling location (N), percentage of polymorphic loci (*P*), observed heterozygosity (H_o), expected heterozygosity (H_e), Shannon's diversity index (I), number of population- or morphotype-specific alleles (M_A) and average allelic richness per population over 24 loci (R_s); calculations were made using FSTAT version 2.9.3 (Goudet 2001). GENEALEX version 6.4 (Peakall & Smouse 2006) was used to perform a hierarchical analysis of molecular variance (AMOVA). Seed hull colour and presence of awns were used as designators for all morphotype-based analyses conducted in this study (see definition for the eight morphotype groups in the 'Sampling and phenotypic classifications' section of Materials and methods).

To infer the evolutionary origin(s) of Malaysian weedy rice and to compare its genetic composition with previously characterized US samples, population genetic analyses were conducted using a data set that

merged the Malaysian SSR data with those generated for the same 24 markers by Gealy *et al.* (2009). Samples from the 2009 study included 32 US weedy rice accessions, 21 *indica* crop varieties, 33 *japonica* crop varieties and 6 *O. nivara* accessions (where *O. nivara* is the taxonomic designation sometimes used for annual forms of *O. rufipogon*). The merged data set consisted of 375 accessions in total. The combined data set was analysed using the model-based Bayesian-clustering program STRUCTURE 2.3.3 (Pritchard *et al.* 2000) and principal coordinate analysis (PCoA). Pooled samples were also used for a morphological comparison between US and Malaysian weed samples.

For the STRUCTURE analysis, an admixture model was run five times for each value of K assumed genetic subpopulations (with K ranging from 1 to 14), using 100 000 iterations after a burn-in of 100 000 iterations. The ΔK *ad hoc* statistic of Evanno *et al.* (2005) was used to estimate the probability of best fit for each assumed K value. For comparison with candidate genes, a further analysis was conducted on Malaysian samples only (weedy rice, elite cultivars, landraces and *O. rufipogon*, total $N = 265$), with morphotype classifications used as population designators in the STRUCTURE runs.

PCoA was performed in GenAlEx 6 (Peakall & Smouse 2006) using a genetic distance matrix constructed from CS Chord distance values (Cavalli-Sforza & Edwards 1967). Eigenvalues were used to summarize and condense the variance among accessions to a limited number of dimensions, allowing for the identification of possible weedy rice clusters. Clustering patterns were also analysed as they related to morphological classifications of weed samples and geographical locations of sampling sites. To explicitly test for geographical isolation by distance among weedy rice sampling locations, a Mantel test was conducted using the program IBD (Bohonak 2002) to compare matrices of pairwise population CS Chord genetic distance vs. geographical distances. Statistical significance was assessed through 10 000 random permutations of the data.

Candidate gene analyses

Allelic variation at *sh4* was assessed by PCR and sequencing of a 669-bp region containing the exon 1G/T functional nucleotide polymorphism (FNP); the *sh4_004* primer pair designed by Thurber *et al.* (2010) was employed. PCR was performed in 20 μL reactions containing the following: 20 mM of Tris–HCl (pH 8.0), 50 mM of KCl, 1.5 mM of $MgCl_2$, 0.5 μM of each primer, 0.2 mM dNTP mix, 0.5 U of *Taq* DNA polymerase (Promega), 1 M betaine and 0.2 μg of genomic DNA. DNA amplifications were carried out with an initial denaturation at 95 °C for 5 min followed by 35 cycles of denaturation at 95 °C for 45 s, primer annealing at 60 °C for 45 s and primer extension for 45 s, with a final extension at 72 °C for 10 min. Amplified products were purified on agarose gels and sequenced directly with forward and reverse primers using Sanger sequencing (ABI PRISM BigDye Terminator Cycle Sequencing Reaction kit; PerkinElmer, USA). Sequence reactions were run on an ABI 3130 capillary sequencer at the Washington University Biology Departmental core facility. Contigs were aligned and checked for the exon 1 G/T FNP using BIOEDIT version 7.1.3 (Hall 2011).

To assess the relationship between *sh4* allelic variation and the shattering phenotype, the degree of shattering was measured in all 207 Malaysian weedy rice samples. The percentage of grain released from the panicle at maturity was determined and classified into five different categories: 1 (<1%); 3 (~3%); 5 (~15%); 7 (~35%); and 9 (>50%), following the standard protocol for rice descriptors (Bioversity International, IRRI & WARDA 2007). The relationship between degree of seed shattering and the *sh4* exon 1G/T FNP was determined by Spearman rank correlations carried out in SPSS version 16.0 (SPSS Inc. Chicago, IL).

For *Bh4*, accessions were scored for the presence or absence of the 22-bp frameshift deletion by scoring size variation in a PCR product containing this indel. The primer pair (forward primer *Bh4*-22F1, 5′-CAACCAGA TGCTAGTGATATGC-3′; reverse primer *Bh4*-22R1, 5′-AGGTTGAGCGTCACCTG-3′) amplifies a 92-bp or 114-bp product depending on whether an accession carries the loss-of-function mutation. PCR products were run on 2.5% agarose gels for scoring, with at least one accession of known genotype run per gel to assist in scoring.

Results

Genetic diversity of Malaysian samples

The 24 SSR loci were highly polymorphic in Malaysian samples, with a total of 265 alleles detected across all samples and an average of 11.8 alleles per locus. For Malaysian weedy rice accessions, genetic diversity was comparable among geographical regions; the lowest diversity was detected in a northwestern population (SPKB, $H_e = 0.250$, $R_S = 1.673$) and the highest in a central-western population (SLSB, $H_e = 0.483$, $R_S = 2.726$) (Table 1). Weedy rice populations were phenotypically variable in all regions sampled, with no clear geographical partitioning of morphotypes (Table S2, Supporting information). When weed accessions are grouped by morphotype rather than sampling location, diversity is highest in brown-striped-hull awned and intermediate-phenotype

Table 1 Genetic diversity in Peninsular Malaysian weedy, cultivated and wild rice samples based on 24 SSR loci. The first two letters of population abbreviations correspond to Malaysian states (see Fig. 1); full population names are indicated in Table S1 (Supporting information)

Region	Population	*N*	*P* (%)	H_o	H_e	R_s	M_A	*I*
Weedy rice								
Northwest	1 – PLPB	16	83.3	0.037	0.377	2.135	3	0.676
	2 – PLGH	8	75.0	0.010	0.379	2.318	2	0.643
	3 – KDGC	16	79.2	0.079	0.383	2.296	9	0.701
	4 – KDHK	15	66.7	0.014	0.299	1.859	1	0.496
	5 – KDKM	17	66.7	0.042	0.324	1.972	2	0.550
	6 – KDPD	19	91.7	0.066	0.404	2.343	8	0.741
	7 – KDSL	12	79.2	0.090	0.368	2.219	3	0.658
	8 – SPBM	13	91.7	0.035	0.411	2.417	2	0.749
	9 – SPKB	7	58.3	0.086	0.250	1.673	3	0.387
	10 – SPSD	9	70.8	0.106	0.301	1.937	0	0.506
	Pooled:	132	76.7	0.049	0.402	3.133		0.818
Central-west	11 – SLSB	15	95.8	0.043	0.483	2.726	10	0.909
	12 – SLSK	5	79.2	0.075	0.397	2.236	3	0.644
	Pooled:	20	87.5	0.077	0.469	3.689		0.923
Northeast	13 – KNBJ	7	62.5	0.124	0.312	2.102	3	0.531
	14 – KNMA	7	75.0	0.077	0.384	2.280	1	0.654
	15 – KNML	18	87.5	0.113	0.475	2.721	16	0.905
	16 – KNPB	13	87.5	0.052	0.371	2.233	9	0.662
	17 – TR	10	66.7	0.051	0.333	2.120	2	0.588
	Pooled:	54	75.8	0.086	0.408	3.535		0.807
	All weedy rice:	206	77.5	0.065	0.368	3.456		0.647
Cultivated and wild rice								
	Elite cultivars	25	70.8	0.033	0.085	1.351	4	0.169
	Landraces	16	95.8	0.011	0.620	3.952	34	1.249
	O. rufipogon	18	100.0	0.332	0.738	4.992	58	1.599

Abbreviations of genetic diversity parameters are as follows: *P*, percentage of polymorphic loci; H_o, observed heterozygosity; H_e, expected heterozygosity; R_s, average allelic richness per population over 24 loci; M_A, number of population-specific (private) alleles; *I*, Shannon diversity index.
N, number of accessions.

awnless accessions (BRA: H_e = 0.384, R_S = 2.679; mSH: H_e = 0.413, R_S = 2.588; Table S5, Supporting information) and lowest in strawhull awnless (SH) accessions (H_e = 0.231, R_S = 1.952). Collectively, the gene diversity of weed populations was approximately half that of the wild species, *O. rufipogon* (H_e = 0.368 and 0.738 for weedy and wild populations, respectively). Among cultivated rice samples, genetic diversity was very low in elite cultivars (H_e = 0.085), consistent with their narrow genetic background. Malaysian landraces, by contrast, showed a high level of diversity (H_e = 0.620), reflecting the fact that these comprise an assemblage of both *indica* and *japonica* rice.

Observed heterozygosities (H_o) were low in weedy rice populations, with an average value of 0.065 across all populations (Table 1). This is consistent with a high selfing rate for the weeds, as has been observed for weedy rice in the USA (H_o = 0.02; see Table S5, Supporting information) and other world regions (e.g. Cao *et al.* 2006; Jiang *et al.* 2012). Similarly, low H_o values for cultivated rice (0.033 and 0.011 for elite cultivars and landraces, respectively) are consistent with the domesticate's high selfing rate. By comparison, observed heterozygosity was an order of magnitude higher for the outcrossing wild progenitor, *O. rufipogon* (H_o = 0.332). Compared with weedy rice in other world regions, overall genetic diversity in the Malaysian weeds (H_e = 0.368; Table 1) was higher than that of US strains (H_e = 0.270; reanalysed SSR data set of Gealy *et al.* 2009) and weeds in northeastern China (H_e = 0.31, Cao *et al.* 2006), but lower than values reported for weedy rice from northern Italy (H_e = 0.48, Jiang *et al.* 2012) and Thailand (H_e = 0.46; Pusadee *et al.* 2012).

Genetic differentiation among populations and morphotypes

Malaysian weed populations showed an overall low level of genetic differentiation among geographical sampling locations, with 96% of the total variance attributable to

within-population variation; only 2% of the total variation was distributed between the three geographical regions sampled (Table S6, Supporting information). All F_{ST} values between pairs of populations were < 0.10, and none of them were statistically significant after Bonferroni correction (Table S7, Supporting information). Nonetheless, a global estimate revealed statistically significant population differentiation (F_{ST} = 0.106, P < 0.001; Table S6, Supporting information). A test of geographical isolation by distance using a pairwise matrix of Chord distances among populations (Table S8, Supporting information) was also statistically significant, indicating a moderate correlation between genetic and geographical distances among populations (r = 0.425, P < 0.0034; Fig. S2, Supporting information).

Differentiation was high and significant when weed accessions were compared by morphotype. Nearly all pairwise F_{ST} comparisons between Malaysian weed morphotypes were significant at P < 0.05, and most remained statistically significant following Bonferroni correction (Table 2). Pairwise F_{ST} values ranged between 0.0073 and 0.494. Interestingly, a general trend is apparent across the weed morphotypes, with levels of genetic differentiation correlated with phenotypic divergence: strawhull awnless (SH) accessions (the weeds most closely resembling cultivated rice) fall at one end of this phenotypic gradient, and blackhull awned (BHA) accessions (most resembling *O. rufipogon*) fall at the other end (see Table 2, column 1). Another notable pattern is that US weed strains, while phenotypically very similar to Malaysian SH and BHA weeds, are genetically highly differentiated from them (F_{ST} = 0.700 between US and Malaysian SH weeds; F_{ST} = 0.637 between US and Malaysian BHA weeds; Table 2). This pattern suggests independent origins of these morphologically similar North American and Southeast Asian weed strains.

In a *STRUCTURE* analysis of merged Malaysian and US data, ΔK indicated maximum support for K = 2 genetic subpopulations (ΔK = 867.71; Fig. S3, Supporting information; Fig. 2a). However, considering the maximal ΔK at K = 2 may mostly be an artefact of extremely low likelihoods for K = 1 (Vigouroux *et al.* 2008), the second highest ΔK value at K = 5 was chosen (ΔK = 5.86; Fig. S3, Supporting information). Three of the subpopulations at K = 5 are present in Malaysian weeds, and two of these are present in morphotypes that span the phenotypic spectrum from SH through BH (Fig. 2a). In a pattern that parallels pairwise F_{ST} measures among morphotypes, the relative contribution of these two genetic subpopulations shows a gradient across morphotypes, with one genetic subpopulation predominating in lighter-hulled weeds (SH, SHA; coloured green in Fig. 2a) and the other in weeds with darker hull coloration (BR, BRA, BH; coloured brown in Fig. 2a) (see also membership coefficient values, Table S3, Supporting information). Among nonweed accessions, the green population component predominates in elite Malaysian cultivars, while the brown component is represented in *indica* Malaysian landraces and some *O. rufipogon* accessions. These distributions suggest a role for elite cultivars and either landraces or wild populations in the

Table 2 Pairwise F_{ST} among Malaysian weed morphotypes and other *Oryza* samples

	SH	SHA	mSH	mSHA	BR	BRA	BH	BHA	BHA-US	SH-US	CV_Mal	Lr_Mal	OR_Mal
SH	—												
SHA	**0.073***	—											
mSH	**0.165***	**0.095***	—										
mSHA	**0.247***	0.135	0.011	—									
BR	**0.380***	**0.3***	**0.093***	**0.088***	—								
BRA	**0.353***	**0.23***	**0.059***	0.021	0.045	—							
BH	**0.475***	**0.314***	0.106	0.078	0.025	0.036	—						
BHA	**0.494***	**0.356***	**0.351***	**0.352***	**0.4794***	**0.377***	0.456	—					
BHA-US	**0.637***	**0.512***	**0.45***	**0.4428***	**0.54***	**0.449***	**0.478***	**0.522***	—				
SH-US	**0.700***	**0.628***	**0.52***	**0.5469***	**0.567***	**0.509***	**0.663***	**0.698***	**0.599***	—			
CV_Mal	**0.058***	**0.182***	**0.304***	**0.4117***	**0.501***	**0.479***	**0.724***	**0.643***	**0.725***	**0.826***	—		
Lr_Mal	**0.437***	**0.28***	**0.228***	**0.2059***	**0.336***	**0.243***	0.209	**0.33***	**0.282***	**0.47***	**0.519***	—	
OR_Mal	**0.348***	**0.224***	**0.154***	**0.140 ***	**0.260***	**0.154***	0.154	**0.255***	**0.247***	**0.371***	**0.41***	**0.105***	—

SH, strawhulled awnless; SHA, strawhulled awned; mSH, intermediate-strawhulled-awnless; mSHA, intermediate-strawhulled-awned; BR, brown-striped-hulled awnless; BRA, brown-striped-hulled awned; BH, blackhulled awnless; BHA, blackhulled awned; BHA-US, US blackhulled awned; US-SH, US strawhulled awnless; CV_Mal, Malaysian elite cultivars; Lr_Mal, Malaysian landraces; OR_Mal, Malaysian *O. rufipogon*.

Bold font indicates significant differentiation at P < 0.05. Asterisks indicate significant differentiation after Bonferroni correction for multiple comparisons (adjusted P = 0.000641).

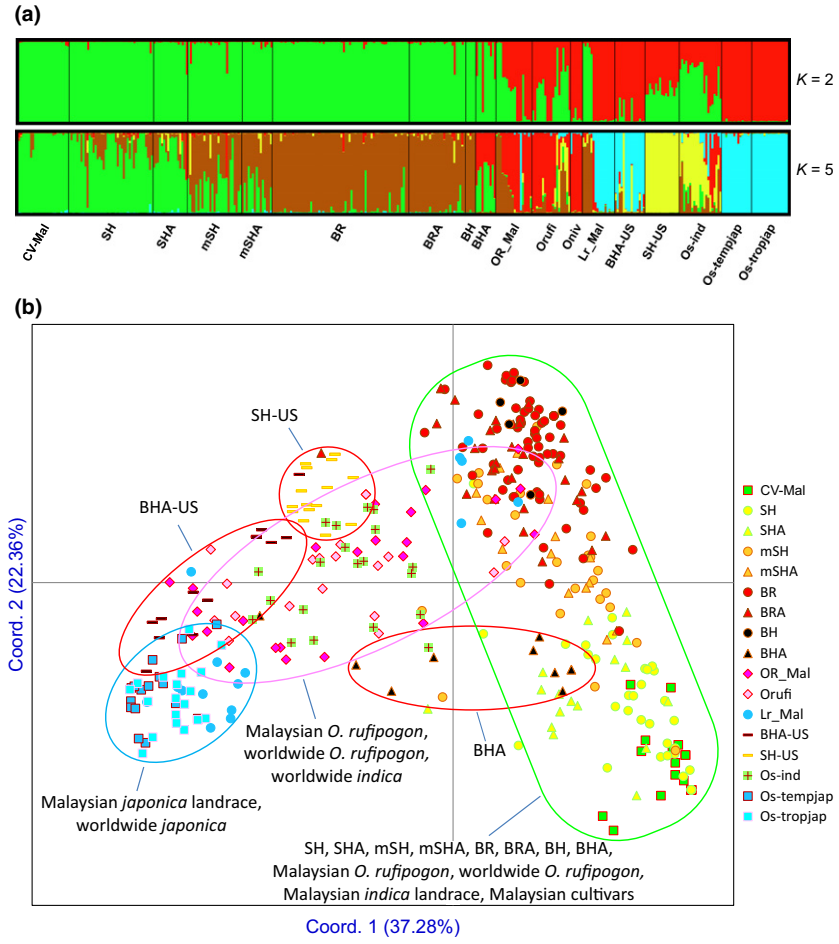

Fig. 2 (a) *STRUCTURE* analysis of the Malaysian and US weedy rice samples categorized by weed morphotype (K = 2 and 5). Abbreviations are as follows: CV-Mal, Malaysian *Oryza sativa* elite cultivars; OR_Mal, Malaysian *O. rufipogon*; Orufi, *O. rufipogon* from other countries; Oniv, *O. nivara* from other countries; Lr_Mal, Malaysian landrace rice; BHA-US, US blackhull awned weedy rice; SH-US, US strawhull awnless weedy rice; Os-ind, *O. sativa indica*; Os-tempjap, *O. sativa temperate japonica*; Os-tropjap, *O. sativa tropical japonica*. (b) Principal coordinate analysis (PCoA) showing genetic distances among Malaysian rice samples and worldwide *Oryza* collections. SH, strawhull awnless Malaysian weeds; SHA, strawhull awned Malaysian weeds; BR, brown-striped hull, awnless Malaysian weeds; BRA, brown-striped hull, awned Malaysian weeds; BH, blackhull awnless Malaysian weeds; BHA, blackhull awned Malaysian weeds; mSH, morphologically intermediate weed form between SH and BR; mSHA, morphologically intermediate weed form between SHA and BRA; SH-US, US strawhull awnless weeds; BHA-US, US blackhull awned weeds; CV-Mal, Malaysian *O. sativa* elite cultivars; OR_Mal, Malaysian *O. rufipogon*; Oruf, *O. rufipogon* accession collected from other Asian countries. Names of other sampled *Oryza* species are indicated within the figure.

evolution of Malaysian weeds. The BHA weed morphotype is unique in that accessions appear to be a genetic admixture of a subpopulation present in *O. rufipogon* (coloured red in Fig. 2a) and the subpopulation characteristic of strawhull weeds and elite cultivars. This could indicate an origin for BHA weeds through hybridization of *O. rufipogon* with SH weeds and/or elite lines. Consistent with F_{ST}-based analyses, SH and BHA weeds in the USA are genetically distinct from all Malaysian weeds in the *STRUCTURE* analysis, again suggesting independent origins of the SH and BHA strains found in the two world regions.

PCoA clustering results are congruent with the *STRUCTURE* analysis. In a PCoA using all Malaysian and US samples, Malaysian weeds in the SH-to-BH morphotype spectrum are arrayed along a gradient: SH morphotypes are closely grouped with elite cultivars, and BR, BRA and BH weeds are grouped at the other end of the gradient with *indica* Malaysian landraces and some *O. rufipogon* accessions (Fig. 2b). The morphologically intermediate mSH and mSHA weeds fall midway along the gradient. BHA weeds, which in the *STRUCTURE* analysis appear to be genetic admixtures between *O. rufipogon* and cultivated or weedy rice, show a distribution

consistent with that result; accessions are arrayed between wild accessions and a cluster containing SH weeds and elite cultivars (Fig. 2b). A second PCoA, performed using Malaysian samples only, suggests that the BHA weeds are closer to SH weeds than to elite cultivars (Fig. S4, Supporting information), possibly indicating a more direct role for the former in BHA weed evolution. As in the F_{ST} and *STRUCTURE* analyses, US weed strains are clearly genetically distinct from all Malaysian accessions (Fig. 2b).

Distributions of functional mutations in sh4 and Bh4 genes

In contrast to previous surveys of *sh4* variation in weedy rice worldwide (Thurber *et al.* 2010; Zhu *et al.* 2012), weed samples in Malaysia were not fixed for the reduced-shattering 'T' allele that is fixed in all domesticated rice. Of 178 accessions surveyed, 104 (58.4%) were T homozygotes, while 63 (35.4%) were homozygous for the ancestral 'G' allele characteristic of wild *Oryzas* and 11 (6.2%) were G/T heterozygotes (Table S9, Supporting information). This distribution suggests that, unlike previously examined weedy rice populations, genetic variation in domesticated rice cannot alone account for the variation present in weed strains. Instead, wild *O. rufipogon* populations have apparently played a role in their evolution.

When *sh4* functional variation is examined across morphotype classes, a gradient in allele frequencies is apparent, with the phenotypic gradient from SH to BHA morphotypes closely matched by an increasing frequency of the ancestral G allele (Fig. 3a, b). This gradient closely parallels the genetic contribution of the 'brown' subpopulation identified through *STRUCTURE* analysis. As noted above, the brown genetic subpopulation is present in both *O. rufipogon* and *indica* Malaysian landraces. Given that Malaysian landraces are

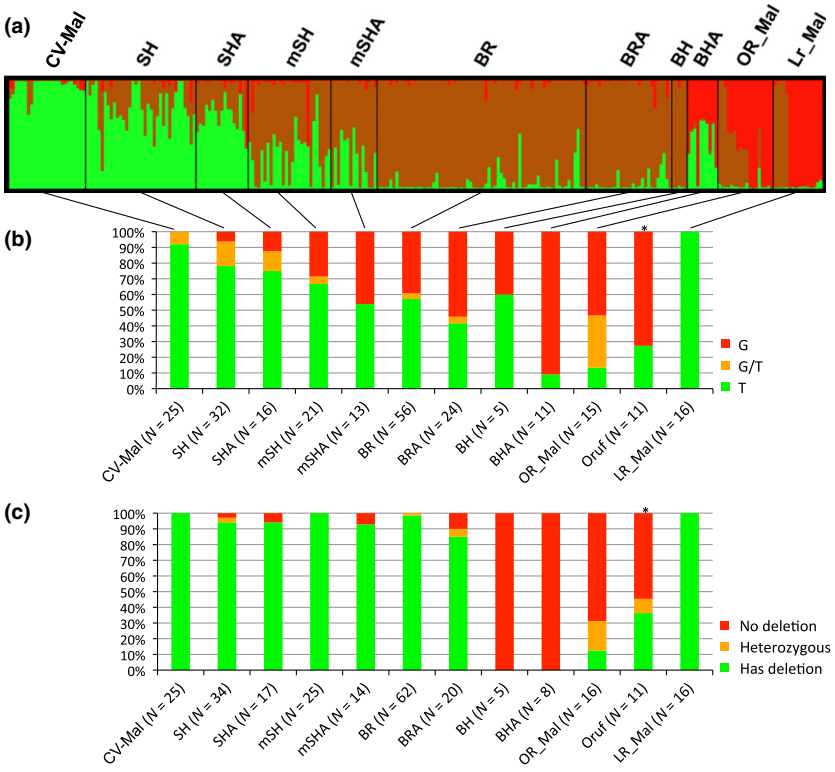

Fig. 3 Association of population structure with genotypic variation at the *sh4* and *Bh4* functional nucleotide polymorphisms. (a) *STRUCTURE* analysis of the Malaysian rice samples categorized by morphotypes ($K = 3$). (b) Percentage of rice individuals in each of phenotypic category possessing G, G/T and T genotypes at the *sh4* FNP. (c) Percentage of rice individuals in each of phenotypic category possessing, lacking or heterozygous for the 22-bp *Bh4* deletion. Numbers of accessions tested are indicated with labels (N). Data for individual accessions are listed in Table S5 (Supporting information). CV_Mal, Malaysian *O. sativa* elite cultivars; SH, strawhull awnless Malaysian weeds; SHA, strawhull awned Malaysian weeds; mSH, morphologically intermediate weed form between SH and BR; mSHA, morphologically intermediate weed form between SHA and BRA; BR, brown-striped hull, awnless Malaysian weeds; BRA, brown-striped hull, awned Malaysian weeds; BH, blackhull awnless Malaysian weeds; BHA, blackhull awned Malaysian weeds; OR_Mal, *O. rufipogon* accession collected from Malaysia; Oruf, *O. rufipogon* accession from other Asian countries (marked with asterisk * in Fig. 3b and Fig. 3c; *STRUCTURE* analysis results of these non-Malaysian samples are not included in Fig. 3a); Lr_Mal, Malaysian landrace rice.

completely fixed for the T domestication allele at *sh4* (Fig. 3b; Table S9, Supporting information) – and therefore could not be contributors of the G allele – these parallel distributions are most consistent with a direct role for *O. rufipogon* in the evolution of Malaysian weedy rice. Curiously, two elite cultivar accessions (MR219-TRJH01 and MR219-TRJH02) were found to be G/T heterozygotes; this could suggest recent admixture with wild populations, although no obvious signature of introgression is apparent in the STRUCTURE analysis (Fig. 2a).

Weedy rice strains worldwide are typically highly shattering despite carrying the *sh4* reduced-shattering allele (Thurber *et al.* 2010; Zhu *et al.* 2012). To determine whether *sh4* variation in Malaysian weeds shows any correlation with the shattering phenotype, we assessed shattering levels as related to the G/T FNP. For weeds classified as having the lowest level of shattering, 93.3% were homozygous for the domestication allele; in contrast, less than half (49.2%) of the weeds classified as the most highly shattering were T homozygotes (Table S10, Supporting information). Across all five phenotypic classes, there was a weak but statistically significant positive correlation between degree of shattering and frequency of the ancestral G allele ($r = 0.387$, $P < 0.01$; Table S11, Supporting information). Thus, while not the sole determinant of the shattering phenotype, the *sh4* FNP appears to play some role in shattering variation among the weeds. Given that high levels of shattering are likely to be adaptive in weed populations, this phenotypic association suggests that adaptive introgression of the G allele from *O. rufipogon* could be selectively favoured in Malaysian weed populations.

Like *sh4*, functional variation at the *Bh4* hull colour gene shows a clear correlation with phenotypic variation in Malaysian weedy rice. Nearly all weeds classified as strawhull (SH, SHA) or intermediate strawhull (mSH, mSHA) were homozygous for the 22-bp frameshift deletion found in most domesticated rice (95.6%; Fig. 3c; Table S12, Supporting information). Weeds classified as black hull (BH, BHA) were completely fixed for the functional ancestral allele. *Bh4* heterozygosity was very low in all weed morphotypes (1.6% overall) and played no apparent role in hull phenotype variation across the strawhull-to-blackhull spectrum. Weeds classified as brown-striped hull (BR, BRA) were genotypically indistinguishable from strawhull weeds at *Bh4*, with 95% homozygous for the frameshift mutation. This finding parallels observations for *Bh4* variation in US weeds, where weeds classified as brownhull have also been reported to be fixed for the 22-bp deletion (Vigueira *et al.* 2013a). Thus, as with US weed strains, genetic factors other than this *Bh4* polymorphism apparently contribute to hull colour variation in Malaysian weed populations.

Discussion

As the primary weed of direct-seeded rice fields worldwide (Chauhan *et al.* 2013), weedy rice has commanded considerable attention from researchers seeking to understand its origin and ongoing evolution. While much of this research has focused on weed strains found in temperate regions that fall outside the viable habitat range of reproductively compatible wild species (e.g. North America, Europe, Japan, Korea and northern China) (Cao *et al.* 2006; Akasaka *et al.* 2009; Chung & Park 2010; Reagon *et al.* 2010; Zhu *et al.* 2012; Grimm *et al.* 2013), the present study has specifically examined weedy rice in a region where it has the potential to be influenced by contemporary gene flow from wild populations. We find that, in comparison with previously studied US weed strains, those in Malaysia are clearly shaped by genetic contributions from local wild *O. rufipogon* populations, with this pattern manifested both at neutral SSR loci and at the *sh4* and *Bh4* domestication genes (Figs 3 and 4). We find a clear correlation between a gradient in grain phenotypes in the Malaysian weeds, which range from very croplike SH strains to wildlike BHA strains, and the degree of apparent introgression from wild populations. Moreover, correlations between *O. rufipogon sh4* alleles and the shattering phenotype in weed populations point to a role for adaptive introgression in this process. Our data also provide evidence for independent evolutionary origins of SH and BHA weed strains in Malaysia from phenotypically similar SH and BHA strains in North America (Table 2 and Fig. 2), with an apparent role for Malaysian elite cultivars in SH weed origins. Below, we discuss these findings in the context of weedy rice evolution in Southeast Asia and broader implications for agricultural weed dynamics.

Contributors to Malaysian weedy rice evolution

Elite cultivars and landraces. Major changes in Malaysian rice-farming practices over the last four decades have likely contributed to the recent proliferation of the weedy rice populations examined here. It was not until the late 1970s that large-scale cultivation of elite Malaysian cultivars came into widespread practice. Direct-seeded, mechanized planting practices eliminated the opportunities for hand-weeding that occurred with traditional hand-transplanting of rice seedlings into fields. Coupled with cross-contamination of fields through shared farm machinery and replanting of fields with self-supplied, weed-contaminated seed stocks, these agricultural shifts strongly negatively impacted weed control efforts (Azmi & Baki 2002; Azmi & Karim 2008). Following initial reports of weedy rice in the late 1980s

(Wahab & Suhaimi 1991), weedy rice had become widespread across rice-growing regions of Peninsular Malaysia by the early 2000s (Azmi & Baki 2002; Azmi & Karim 2008). This rapid spread is reflected in the general absence of any strong geographical structuring of genetic variation across Peninsular Malaysia (Tables S6 and S7, Supporting information).

While Malaysian weeds show substantial genetic heterogeneity, they are clearly more closely related to *indica* rice (and some *O. rufipogon* accessions) than to *japonica* crop varieties (Fig. 2a, b). Malaysian SH weeds (the weed morphotype most similar to cultivated rice in grain characteristics) show a particularly close relationship to elite Malaysian cultivars (F_{ST} = 0.058; Table 2; see also Fig. 2a, b), suggesting a direct role for these modern cultivars in weedy rice evolution. Elite Malaysian cultivars were bred from exotic *indica* germplasm developed at the International Rice Research Institute (e.g. IR22, IR8) (Dalrymple 1986). Given that the SH weeds possess greater genetic diversity than the cultivars (H_e = 0.231 and 0.085, respectively; Table 1 and Table S5, Supporting information), it is reasonable to suggest that the modern cultivars have contributed to SH weed proliferation through a process of hybridization with weedy and/or wild populations, rather than solely through cultivar dedomestication. The relatively short time frame during which elite cultivars have been widely grown in Malaysia also argues against the evolution of weediness phenotypes directly from this narrow genetic stock.

Unlike elite cultivars, traditional Malaysian landraces have long been present in the areas examined in this study. *Indica* landraces share the 'brown' subpopulation component predominant in BR, BRA and BH weed strains (see Fig. 2) and so may have contributed to weedy rice evolution. However, as with elite cultivars and SH weeds, these landraces cannot be the sole ancestors of the weed morphotypes. Most notably, Malaysian landraces, like all domesticated rice, are fixed for the *sh4* reduced-shattering 'T' allele, whereas the weeds carry moderate frequencies of the ancestral 'G' allele found in *O. rufipogon* (Fig. 3b; Table S3, Supporting information) (Li *et al.* 2006; Zhang *et al.* 2009). Similarly, whereas all sampled landraces are fixed for the *Bh4* 22-bp deletion allele that is characteristic of strawhull domesticated rice, BH and BHA weeds carry functional *Bh4* alleles typical of wild *Oryza* species (Fig. 3c). Thus, while all Malaysian weeds show a close genetic similarity to domesticated rice (Fig. 2a), their origin is more complex than one of simple dedomestication from local crop varieties.

Oryza rufipogon. The clear role of *O. rufipogon* in Malaysian weed evolution detected in this study differs from previous findings, both for Malaysia and for other geographical regions. In Malaysia, two previous studies, based on 23 random amplified polymorphic DNA (RAPD) markers (Vaughan *et al.* 2003) and 10 SSR loci plus morphological characters (Zainudin *et al.* 2011), have examined relationships between weedy, cultivated and wild rice in two planting areas of Peninsular Malaysia (Selangor and Penang; see Fig. 1). Both studies supported a closer relationship of weedy rice to cultivated rice than to wild samples. Similarly, a study of weedy rice in China, including samples from Guangdong province where *O. rufipogon* occurs, detected no role for the wild species in local weed origins (Zhang *et al.* 2012). These findings stand in contrast to the very close relationship we observe between *O. rufipogon* and some BR, BRA and BH accessions based on SSR analyses (e.g. overlapping accessions in Fig. 2b; see also Table 2).

Similarly, the moderate-to-high frequency of the *sh4* ancestral 'G' allele we detect in Malaysian BR, BRA and BH and BHA weed morphotypes (Fig. 3b) stands in marked contrast to previous studies of *sh4* variation in weedy rice. Thurber *et al.* (2010) examined G/T allelic variation in 58 US weedy rice samples and found that these weeds all carry the 'T' allele fixed by selection during domestication. Zhu *et al.* (2012), in a worldwide survey of 165 weedy rice accessions (including samples from several countries where *O. rufipogon* occurs, e.g. India, Vietnam, Cambodia, Laos, Malaysia, Myanmar and Bangladesh), also found all weeds to be fixed for the domestication allele. It is possible that the very different pattern we observe in the present study is a reflection of the rapidly changing dynamics of weed evolution in Malaysia, where widespread weed proliferation has been so recent. Consistent with this hypothesis, *sh4* G/T heterozygote frequencies in Malaysian weedy rice accessions are relatively high for a predominantly selfing species (6.2%, N = 178; Table S3, Supporting information), suggesting recent descent from weed–wild hybrids. In Thailand, which, like Malaysia, has recently shifted to industrialized rice farming, SSR allele sharing between weedy rice and local *O. rufipogon* populations also points to recent wild-to-weed introgression associated with the weed's proliferation (Pusadee *et al.* 2012).

Our observations that the ancestral *sh4* 'G' allele facilitates the shattering phenotype in weedy rice (Tables S10 and S11, Supporting information) suggest that introgression of this *O. rufipogon* allelic variation could be adaptive in weed populations, where a highly shattering, wildlike phenotype is generally considered to be advantageous for seed dispersal (Thurber *et al.* 2010). A number of other *O. rufipogon* predomestication traits with well-characterized candidate genes might also be

expected to undergo adaptive introgression into weedy rice populations. These include seed dormancy (conferred in part by ancestral, functional alleles at the *Rc* gene; Gu *et al.* 2011), open panicle architecture that facilitates freely shattering seed (conferred in part by ancestral, functional alleles of the *OsLG1* gene; Ishii *et al.* 2013) and the presence of awns that facilitate post-shattering seed dispersal (controlled in part by ancestral, functional alleles of the *An-1* gene; Luo *et al.* 2013). In contrast, other *O. rufipogon* traits would likely be maladaptive in agricultural fields and would be predicted to be selected against in weed populations. In particular, prostrate growth (conferred in part by ancestral, functional alleles at the *PROG1* gene; Jin *et al.* 2008; Tan *et al.* 2008) and perennial life history would both be expected to be strongly disadvantageous in the densely planted, seasonally harvested environment of rice production areas. It will be interesting to examine the fates of these potentially adaptive and maladaptive *O. rufipogon* alleles in Malaysian weed populations.

Weedy rice in Southeast Asia and other world regions

Our comparative analysis of Malaysian samples with previously analysed US weedy rice accessions permits several conclusions on the evolution of weedy rice in different world regions. Most strikingly, we find that the two common weed strains in southern US rice fields, SH and BHA morphotypes, have altogether different evolutionary origins from the phenotypically very similar strains found in Malaysia. US-SH strains, like those in Malaysia, are closely related to *indica* domesticated rice; however, they are derived from a genetically distinct subset of *indica* germplasm (Fig. 2a, b). Similarly, Malaysian BHA strains are altogether distinct from US strains, which evolved from *aus* rice, a domesticated variety not included in the present study. These findings contribute to a growing body of evidence for the independent, convergent evolution of weedy rice in different world regions. Besides the independent weed origins documented in the present study, studies in northeastern China, Korea and Japan indicate evidence for weed origins from *japonica* rice (Cao *et al.* 2006; Akasaka *et al.* 2009; Chung & Park 2010), a pattern not detected in either Malaysia (Figs 3 and 4) or the USA (Londo & Schaal 2007; Reagon *et al.* 2010).

Another key difference detected between US and Malaysian weedy rices is in the far greater genetic heterogeneity and phenotypic complexity of the Southeast Asian weeds. US weeds fall into two largely discrete phenotypic and genetic classes (SH, BHA), with only low levels of hybridization between them and little evidence that US crop varieties have contributed to their evolution (Reagon *et al.* 2010). Malaysian weeds, by contrast, are characterized by gradations in both phenotypic and genetic variations (Figs 3 and 4; Fig. S1, Supporting information). This greater complexity is likely a reflection of greater reproductive compatibility among Malaysian weed strains, as well as the hybridization discussed above between weeds and locally occurring wild and domesticated rice. To the extent that phenotypic and genetic heterogeneity can be taken as an indicator of adaptive flexibility, weed control efforts in Malaysia and other Southeast Asian countries will likely be especially challenging as weedy rice populations adapt to ongoing weed control efforts.

Conclusions

Crop domestication has long served as a model for understanding evolution (Darwin 1859; Larson *et al.* 2014), and weedy crop relatives are increasingly recognized as an important component of evolution in the agricultural context (Kane & Rieseberg 2008; Ellstrand *et al.* 2010; Vigueira *et al.* 2013b). The same capacity for rapid adaptation that makes weedy relatives so problematic for crop production also makes them especially interesting systems for exploring the mechanisms underlying their origin and ongoing evolution. In the case of weedy rice, our combined analysis of neutral markers and domestication genes has provided definitive evidence that wild *Oryza* populations contribute to the evolution of Malaysian weed strains, with introgression of weed-adaptive alleles a likely a component of this process. Moreover, we detect potentially troubling evidence of a major role for widely farmed, modern-bred elite cultivars in the evolution of these weeds. Taken together with our evidence for high genetic heterogeneity in Malaysian weeds, and of independent weed origins in different world regions, these findings suggest that current shifts towards industrialized rice farming worldwide will favour the continued proliferation of weedy rice. In the USA and elsewhere, weedy rice control efforts in the last decade have come to rely heavily on the use of herbicide-resistant crop varieties, a strategy that is now being adopted worldwide (Sudianto *et al.* 2013). We predict that as herbicide-resistant rice farming comes into widespread practice in Malaysia, the high reproductive compatibility of cultivars, weeds and wild populations will lead to the very rapid dissemination of resistance alleles into weedy and wild populations.

Acknowledgements

We especially thank Drs. David Gealy and Melissa Jia (Dale Bumpers National Rice Research Center, USA), for providing US weedy rice microsatellite data. We thank Linda Small,

Tonapha Pusadee, Pu Huang, Tina Neik Ting Xiang, Nicholas Kooyers, Kate Waselkov and Cindy Vigueira for technical assistance and computing support, and the Olsen lab group for helpful comments on the manuscript. We are grateful to the Washington University greenhouse staff for plant care support; to Site Noorzuraini (Rice Genebank, MARDI) for sharing of rice seeds; and to Dr. Azmi Man (MARDI), Abdullah Md Zain (University Malaysia Terengganu), Law Kok Choy (Sekinchan, Malaysia), Akmal Adilah Idris (University Malaysia Kelantan), the staff at Kemubu Agricultural Development Authority (KADA, Kelantan), Kawasan Pembangunan Pertanian Bersepadu Terengganu Utara (KETARA, Terengganu) and Muda Agricultural Development Authority (MADA, Kedah) for sharing information on weedy rice infestations and assisting in field sampling. Funding for this project was provided by Malaysian Ministry of Agriculture & Agro-Based Industry (Science Fund, 05-02-10-SF1001), the US National Science Foundation (Plant Genome Research Program, IOS-1032023) and the Monash University Malaysia (MUM) Tropical Medicine and Biology Multidisciplinary Platform.

References

Akasaka M, Ushiki J, Iwata H, Ishikawa R, Ishii T (2009) Genetic relationships and diversity of weedy rice (*Oryza sativa* L.) and cultivated rice varieties in Okayama Prefecture, Japan. *Breeding Science*, **59**, 401–409.

Azmi M, Baki BB (2002) Impact of continuous direct seeding culture on weed species diversity in the Malaysian rice ecosystem. In: *Proceedings of Regional Symposium on Environment and Natural Resources* (Kuala Lumpur, Malaysia 10–11 April 2002). Universiti Kebangsaan Malaysia, Bangi, 61–67.

Azmi M, Karim SMR (2008) *Weedy Rice – Biology, Ecology and Management*. Malaysian Agricultural Research and Development Institute, Kuala Lumpur.

Bioversity International, IRRI and WARDA (2007) *Descriptors for Wild and Cultivated Rice (Oryza spp.)*. Bioversity International, Rome, Italy; International Rice Research Institute, Los Banos, Philippines; WARDA, Africa Rice Center, Cotonou, Benin.

Bohonak JA (2002) IBD (isolation by distance): a program for analyses of isolation by distance. *Journal of Heredity*, **93**, 153–154.

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

Cavalli-Sforza LL, Edwards AWF (1967) Phylogenetic analysis: models and estimation procedures. *American Journal of Human Genetics*, **19**, 233–257.

Chauhan B, Singh K, Ladha J (2013) Weedy Rice: an emerging threat for direct-seeded rice production systems in India. *Journal of Rice Research*, **1**, 1–6.

Chung J-W, Park Y-J (2010) Population structure analysis reveals the maintenance of isolated sub-populations of weedy rice. *Weed Research*, **50**, 606–620.

Craig SM, Reagon M, Resnick LE, Caicedo AL (2014) Allele distributions at hybrid incompatibility loci facilitate the potential for gene flow between cultivated and weedy rice in the US. *PLoS ONE*, **9**, e86647.

Dalrymple DG (1986) *Development and Spread of High-Yielding Rice Varieties in Developing Countries*. Agency for International Development, Washington, District of Columbia.

Darwin C (1859) *The Origin of Species*. John Murray, London.

De Wet J, Harlan J (1975) Weeds and domesticates: evolution in the man-made habitat. *Economic Botany*, **29**, 99–107.

Doyle JJ, Doyle JL (1990) Isolation of plant DNA from fresh tissue. *Focus*, **12**, 13–15.

Ellstrand NC, Heredia SM, Leak-Garcia JA *et al*. (2010) Crops gone wild: evolution of weeds and invasives from domesticated ancestors. *Evolutionary Applications*, **3**, 494–504.

Estorninos LE Jr, Gealy DR, Talbert RE, Gbur EE (2005) Rice and red rice interference. I. Response of red rice (*Oryza sativa*) to sowing rates of tropical japonica and indica rice cultivars. *Weed Science*, **53**, 676–682.

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

Gealy DR, Agrama HA, Eizenga GC (2009) Exploring genetic and spatial structure of U.S. weedy red rice (*Oryza sativa*) in relation to rice relatives worldwide. *Weed Science*, **57**, 627–643.

Gealy DH, Agrama H, Jia MH (2012) Genetic analysis of atypical U.S. red rice phenotypes: indications of prior gene flow in rice fields? *Weed Science*, **60**, 451–461.

Goudet J (2001) *FSTAT*, a program to estimate and test gene diversities and fixation indices, version 2.9.3. Available from http://www.unil.ch/izea/softwares/fstat.html (Accessed November 6, 2011).

Grimm A, Fogliatto S, Nick P, Ferrero A, Vidotto F (2013) Microsatellite markers reveal multiple origins for Italian weedy rice. *Ecology and Evolution*, **3**, 4786–4798.

Gu X-Y, Foley ME, Horvath DP *et al*. (2011) Association between seed dormancy and pericarp color is controlled by a pleiotropic gene that regulates abscisic acid and flavonoid synthesis in weedy red rice. *Genetics*, **189**, 1515–1524.

Hall T (2011) BioEdit biological sequence alignment editor. Available from http://www.mbio.ncsu.edu/BioEdit/bioedit.html.

Ishii T, Numaguchi K, Miura K *et al*. (2013) *OsLG1* regulates a closed panicle trait in domesticated rice. *Nature Genetics*, **45**, 462–465, 465e1–2.

Jiang Z, Xia H, Basso B, Lu BR (2012) Introgression from cultivated rice influences genetic differentiation of weedy rice populations at a local spatial scale. *Theoretical and Applied Genetics*, **124**, 309–322.

Jin J, Huang W, Gao J-P *et al*. (2008) Genetic control of rice plant architecture under domestication. *Nature Genetics*, **40**, 1365–1369.

Kane NC, Rieseberg LH (2008) Genetics and evolution of weedy *Helianthus annuus* populations: adaptation of an agricultural weed. *Molecular Ecology*, **17**, 384–394.

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

Larson G, Piperno DR, Allaby RG *et al*. (2014) Current perspectives and the future of domestication studies. *Proceedings of the National Academy of Sciences United States of America*, **111**, 6139–6146.

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

Londo JP, Schaal BA (2007) Origins and population genetics of weedy red rice in the USA. *Molecular Ecology*, **16**, 4523–4535.
Luo J, Liu H, Zhou T *et al.* (2013) *An-1* encodes a basic helix-loop-helix protein that regulates awn development, grain size, and grain number in rice. *Plant Cell*, **25**, 3360–3376.
Peakall R, Smouse PE (2006) genalex 6: genetic analysis in Excel. Population genetic software for teaching and research. *Molecular Ecology Notes*, **6**, 288–295.
Pritchard JK, Stephens M, Donnelly P (2000) Inference of population structure using multilocus genotype data. *Genetics*, **155**, 945–959.
Pusadee T, Schaal BA, Rerkasem B, Jamjod S (2012) Population structure of the primary gene pool of *Oryza sativa* in Thailand. *Genetic Resources and Crop Evolution*, **60**, 335–353.
Reagon M, Thurber C, Gross B *et al.* (2010) Genomic patterns of nucleotide diversity in divergent populations of U.S. weedy rice. *BMC Evolutionary Biology*, **10**, 180.
Stewart CN, Tranel PJ, Horvath DP *et al.* (2009) Evolution of weediness and invasiveness: charting the course for weed genomics. *Weed Science*, **57**, 451–462.
Sudianto E, Beng-Kah S, Ting-Xiang N *et al.* (2013) Clearfield® rice: Its development, success, and key challenges on a global perspective. *Crop Protection*, **49**, 40–51.
Tan L, Li X, Liu F *et al.* (2008) Control of a key transition from prostrate to erect growth in rice domestication. *Nature Genetics*, **40**, 1360–1364.
Thurber CS, Reagon M, Gross BL *et al.* (2010) Molecular evolution of shattering loci in U.S. weedy rice. *Molecular Ecology*, **19**, 3271–3284.
Vaughan DA, Zain AM, Watanabe H, Okuno K (2003) Relationship between wild, weedy and cultivated rice in Malaysia. Rice Genetics Newsletter 12: 183. Available from: http://www.shigen.nig.ac.jp/rice/rgn/vol12/v12p183.html (Accessed July 23, 2013).
Vigouroux Y, Glaubitz JC, Matsuoka Y, Goodman MM, Jesus SG, Doebley J (2008) Population structure and genetic diversity of new world maize races assessed by DNA microsatellites. *American Journal of Botany*, **95**, 1240–1253.
Vigueira CC, Li W, Olsen KM (2013a) The role of *Bh4* in parallel evolution of hull colour in domesticated and weedy rice. *Journal of Evolutionary Biology*, **26**, 1738–1749.
Vigueira CC, Olsen KM, Caicedo AL (2013b) The red queen in the corn: agricultural weeds as models of rapid adaptive evolution. *Heredity*, **110**, 303–311.
Wahab A, Suhaimi O (1991) Padi angin: characteristics, adverse effects and methods of its eradication (in Malay). *Teknologi Padi*, **7**, 21–31.
Zainudin H, Azmi M, Ahmad SO (2011) Association of commercial rice varieties with weedy rice accessions (*Oryza sativa* complex) in Pulau Pinang's rice granary area. *Tropical Life Sciences Research*, **22**, 1–11.
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.
Zhang L, Dai W, Wu C, Song X, Qiang S (2012) Genetic diversity and origin of *Japonica*- and *Indica*-like rice biotypes of weedy rice in the Guangdong and Liaoning provinces of China. *Genetic Resources and Crop Evolution*, **59**, 399–410.
Zhu B-F, Si L, Wang Z *et al.* (2011) Genetic control of a transition from black to straw-white seed hull in rice domestication. *Plant Physiology*, **155**, 1301–1311.
Zhu Y, Ellstrand NC, Lu B-R (2012) Sequence polymorphisms in wild, weedy, and cultivated rice suggest seed-shattering locus *sh4* played a minor role in Asian rice domestication. *Ecology and Evolution*, **2**, 2106–2113.

B.K.S. conceived and coordinated the study, carried out the experiments, performed the computational analysis and interpretation of the data, supervised the work and drafted the manuscript. T.S.C. and S.M.T. helped in data analysis, field sampling and research supervision. K.M.O. edited the manuscript, contributed to the interpretation of the data, and supervised and coordinated the work. All authors read and approved the final manuscript.

Data accessibility

Microsatellite genotype data and *sh4* sequence alignments – Dryad doi http://dx.doi.org/10.5061/dryad.1m9q8

Sampling locations, morphological data, and *Bh4* and *sh4* FNP genotypes – online supporting information.

Supporting information

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

Fig. S1 Representatives of eight groups of weedy rice accessions classified according to seed morphological features: (a) strawhulled, awnless (SH); (b) strawhulled, awned (SHA); (c) intermediate-strawhulled-awnless (mSH); (d) intermediate-strawhulled-awned (mSHA); (e) brown-striped-hulled, awnless types (BR); (f) brown-striped-hulled, awned (BRA); (g) blackhulled, awnless (BH); and (h) blackhulled, awned (BHA) types.

Fig. S2 Isolation by distance among Peninsular Malaysian weedy rice populations.

Fig. S3 *STRUCTURE* analysis of the US and Malaysian rice samples categorized by morphotype.

Fig. S4 Principal coordinate analysis (PCoA) of Malaysian weedy, wild, landrace and cultivated rice accessions.

Table S1 Population code, location, number of sample, range of coordinates, coexisting rice varieties of the weedy rice populations and other rice samples used in this study.

Table S2 Distribution, morphotype code and sample number for weedy rice accessions phenotyped for hull color and awn presence.

Table S3 *Oryza* accession information, seed shattering phenotype, functional nucleotide polymorphism data for the *sh4 and Bh4* genes, and coefficients of ancestry inferred by *STRUCTURE*.

Table S4 List of multiplex PCR primers used in SSR genotyping.

Table S5 Genetic diversity in Peninsular Malaysian and US weedy rice samples grouped by morphotype.

Table S6 Analysis of molecular variance (AMOVA) and *F*-statistic values of 17 weedy rice populations from Peninsular Malaysia based on the 15 SSR loci.

Table S7 Pairwise population F_{ST} values generated by location-based clustering.

Table S8 Pairwise population C.S. Chord genetic distance (generated by location-based clustering) and geographical distance values.

Table S9 *Oryza* accessions and their associated genotypes at the *sh4* functional nucleotide polymorphism.

Table S10 Relationship between degree of seed shattering and genotype at the *sh4* FNP of the weedy rice accessions.

Table S11 Correlation coefficients (*r*) between degree of seed shattering and genotype at the *sh4* FNP of the weedy rice accessions.

Table S12 Numbers of weedy rice accessions and their associated genotypes at the *Bh4* functional nucleotide polymorphism.