2018

The Role of Standing Variation in the Evolution of Weedines Traits in South Asian Weedy Rice (Oryza spp.)

Zhongyung Huang  
*University of Massachusetts Amherst*

Shannon Kelly  
*University of Massachusetts Amherst*

Rika Matsuo  
*University of Massachusetts Amherst*

Lin-Feng Li  
*Washington University in St. Louis*

Yaling Li  
*Washington University in St. Louis*

See next page for additional authors

Follow this and additional works at: https://scholarworks.umass.edu/biology_faculty_pubs

Part of the Biology Commons

Recommended Citation  
Huang, Zhongyung; Kelly, Shannon; Matsuo, Rika; Li, Lin-Feng; Li, Yaling; Olsen, Kenneth M.; Jia, Yulin; and Caicedo, Ana L., "The Role of Standing Variation in the Evolution of Weedines Traits in South Asian Weedy Rice (Oryza spp.)" (2018). *G3: Genes, Genomes, Genetics*. 632.  
https://doi.org/10.1534/g3.118.200605

This Article is brought to you for free and open access by the Biology at ScholarWorks@UMass Amherst. It has been accepted for inclusion in Biology Department Faculty Publication Series by an authorized administrator of ScholarWorks@UMass Amherst. For more information, please contact scholarworks@library.umass.edu.
The Role of Standing Variation in the Evolution of Weediness Traits in South Asian Weedy Rice (Oryza spp.)

Zhongyun Huang,*1 Shannon Kelly,*2 Rika Matsuo,*3 Lin-Feng Li,†4 Yaling Li,†5 Kenneth M. Olsen,‡ Yulin Jia,* and Ana L. Caicedo,*6

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

ORCID IDs: 0000-0002-7716-3604 (Z.H.); 0000-0003-2820-3187 (L.-F.L.); 0000-0002-4711-5353 (Y.L.); 0000-0002-8338-3638 (K.M.O.); 0000-0002-2132-0498 (Y.J.); 0000-0002-0378-6374 (A.L.C.)

ABSTRACT Weedy rice (Oryza spp.) is a problematic weed of cultivated rice (O. sativa) around the world. Recent studies have established multiple independent evolutionary origins of weedy rice, raising questions about the traits and genes that are essential for the evolution of this weed. Among world regions, South Asia stands out due to the heterogeneity of its weedy rice populations, which can be traced to at least three origins: two through de-domestication from distinct cultivated rice varieties, and one from local wild rice (O. rufipogon/O. nivara). Here we examine five traits considered typical of or advantageous to weedy rice in weedy, cultivated and wild rice samples from South Asia. We establish that convergence among all three weed groups occurs for easy seed shattering, red pericarp color, and compact plant architecture, suggesting that these traits are essential for weed success in the South Asian agricultural environment. A high degree of convergence for black hull color is also seen among weeds with wild ancestors and weeds evolved from the aus cultivated rice group. We also examine polymorphism in five known domestication candidate genes, and find that Rc and Bh4 are associated with weed seed pericarp color and hull color, respectively, and weedy alleles segregate in the ancestral populations, as do alleles for the seed dormancy-linked gene Sdr4. The presence of a domestication related allele at the seed shattering locus, sh4, in weedy rice populations with cultivated ancestry supports a de-domestication origin for these weedy groups, and raises questions about the reacquisition of the shattering trait in these weedy populations. Our characterization of weedy rice phenotypes in South Asia and their associated candidate genes contribute to the emerging understanding of the mechanisms by which weedy rice evolves worldwide, suggesting that standing ancestral variation is often the source of weedy traits in independently evolved groups, and highlighting the reservoir of genetic variation that is present in cultivated varieties as well as in wild rice, and its potential for phenotypic evolution.

KEYWORDS red rice weed evolution candidate genes seed shattering seed dormancy Genomic Prediction GenPred Shared Data Resources

Agricultural weeds, defined as unwanted plants growing in agricultural environments (Monaco et al. 2002), are responsible for ~30% annual reduction in crop productivity worldwide (Oerke 2006). This negative impact of weeds can be attributed to various traits, which allow them to adapt and be competitive in the agricultural environment. Examples of such traits in diverse species can include rapid growth, efficient seed dispersal and seed dormancy (Baker 1965), and can be considered to be part of an agricultural weed syndrome (Vigueira et al. 2013b). Weed syndrome traits have been found to evolve repeatedly and independently in different weed lineages (Vigueira et al. 2013a; Kane and Rieseberg 2008; Cui et al. 2016; Huang et al. 2017), and are examples of what is known as parallel or convergent evolution (Arendt and Reznick 2008).

Weedy rice (Oryza spp.), an aggressive unwanted weed of cultivated rice (O. sativa), is a type of agricultural weed displaying parallel evolution of weediness traits (Ziska et al. 2015). Weedy rice occurs worldwide, and studies have established multiple independent evolutionary origins of this weed from non-weedy backgrounds (Reagon et al. 2010; Huang et al. 2017; Vigueira et al., 2017). Many typical traits of weedy rice resemble reversals to wild phenotypes. Despite diverse genetic backgrounds, red pericarp is a common feature of weedy rice, leading to the common name “red rice” (Gross et al. 2010; Cui et al. 2016). The fact that red
pericarp is common likely reflects selection for dormancy in weedy rice seeds, since the gene which regulates proanthocyanidin pigment synthesis also pleiotropically regulates ABA regulated seed dormancy (Gu et al. 2011; Cui et al. 2016), and seed dormancy has been detected in several populations (Gianinetto and Cohn 2008; Ziska et al. 2015). Ease of seed shattering or dispersal has also been reported for several populations (Thurber et al. 2010; Zhu et al. 2012). The wealth of knowledge of candidate genes underlying many morphological traits in rice has recently made it possible to examine the origins of weedy traits in weedy rice populations (Gross et al. 2010; Vigueira et al. 2013b), as well as to examine the extent to which the same genes underlie convergent traits across weedy rice populations.

We recently characterized populations of weedy rice from South Asia (Huang et al. 2017), an area known for its great diversity of cultivated and wild Oryza (Civán et al. 2015; Garris et al. 2005; Huang et al. 2012; Zhu et al. 2007), that has also proven to contain a great diversity of weedy rice. Two weedy rice populations in this area trace their ancestry to cultivated rice varieties *aus* and *indica* through a process known as de-domestication (Huang et al. 2017). A third weedy population has evolved from local wild rice (the species complex known as *O. rufipogon/nivara*). Moreover, weeds presenting an admixture of all these backgrounds also occur (Huang et al. 2017). These South Asian weedy rice origins are also independent of the well-studied U.S. weedy rice populations de-domesticated from *aus* and *indica* varieties (Reagon et al. 2010; Huang et al. 2017). The diversity of weedy rice origins in a single world region raises questions about how each of these groups has acquired the traits that make them weedy.

To understand how weediness traits have evolved in each of the genetically distinct weedy populations of South Asia, we examine here the presence of iconic weedy rice traits in each of these populations. We additionally investigate the genetic mechanisms possibly underlying these traits by capitalizing on the extensive prior characterization of domestication trait candidate genes in cultivated rice, and deduce the evolutionary origins of alleles carried by weeds.

We first focus on pericarp pigmentation, as red pericarps have been proposed to be adaptive either by providing protection from abiotic or biotic stress (Ithal and Reddy 2004), and through association with seed pericarp color in weedy rice (Gross et al. 2010; Vigueira et al. 2013b), as well as to examine the extent to which the same genes underlie convergent traits across weedy rice populations.

We recently characterized populations of weedy rice from South Asia (Huang et al. 2017), an area known for its great diversity of cultivated and wild *Oryza* (Civán et al. 2015; Garris et al. 2005; Huang et al. 2012; Zhu et al. 2007), that has also proven to contain a great diversity of weedy rice. Two weedy rice populations in this area trace their ancestry to cultivated rice varieties *aus* and *indica* through a process known as de-domestication (Huang et al. 2017). A third weedy population has evolved from local wild rice (the species complex known as *O. rufipogon/nivara*). Moreover, weeds presenting an admixture of all these backgrounds also occur (Huang et al. 2017). These South Asian weedy rice origins are also independent of the well-studied U.S. weedy rice populations de-domesticated from *aus* and *indica* varieties (Reagon et al. 2010; Huang et al. 2017). The diversity of weedy rice origins in a single world region raises questions about how each of these groups has acquired the traits that make them weedy.

To understand how weediness traits have evolved in each of the genetically distinct weedy populations of South Asia, we examine here the presence of iconic weedy rice traits in each of these populations. We additionally investigate the genetic mechanisms possibly underlying these traits by capitalizing on the extensive prior characterization of domestication trait candidate genes in cultivated rice, and deduce the evolutionary origins of alleles carried by weeds.

We first focus on pericarp pigmentation, as red pericarps have been proposed to be adaptive either by providing protection from abiotic or biotic stress (Ithal and Reddy 2004), and through association with seed pericarp color in weedy rice (Gross et al. 2010; Vigueira et al. 2013b), as well as to examine the extent to which the same genes underlie convergent traits across weedy rice populations.

We recently characterized populations of weedy rice from South Asia (Huang et al. 2017), an area known for its great diversity of cultivated and wild *Oryza* (Civán et al. 2015; Garris et al. 2005; Huang et al. 2012; Zhu et al. 2007), that has also proven to contain a great diversity of weedy rice. Two weedy rice populations in this area trace their ancestry to cultivated rice varieties *aus* and *indica* through a process known as de-domestication (Huang et al. 2017). A third weedy population has evolved from local wild rice (the species complex known as *O. rufipogon/nivara*). Moreover, weeds presenting an admixture of all these backgrounds also occur (Huang et al. 2017). These South Asian weedy rice origins are also independent of the well-studied U.S. weedy rice populations de-domesticated from *aus* and *indica* varieties (Reagon et al. 2010; Huang et al. 2017). The diversity of weedy rice origins in a single world region raises questions about how each of these groups has acquired the traits that make them weedy.

We first focus on pericarp pigmentation, as red pericarps have been proposed to be adaptive either by providing protection from abiotic or biotic stress (Ithal and Reddy 2004), and through association with seed pericarp color in weedy rice (Gross et al. 2010; Vigueira et al. 2013b), as well as to examine the extent to which the same genes underlie convergent traits across weedy rice populations. In this study, we recorded seed pericarp color for 35 weedy rice accessions from South Asia; 2) investigate whether known candidate genes may explain the trait values found in each weedy rice population; and 3) determine whether weedy rice populations in South Asia have acquired weediness traits and alleles from standing variation in their ancestors, de novo mutation, or introgression from other sources.

Our goals were to: 1) determine which of these typical weedy traits characterize each of the known weedy rice populations occurring in South Asia; 2) investigate whether known candidate genes may explain the trait values found in each weedy rice population; and 3) determine whether weedy rice populations in South Asia have acquired weediness traits and alleles from standing variation in their ancestors, de novo mutation, or introgression from other sources.

**MATERIAL AND METHODS**

**Plant material and phenotypic measurements**

A panel of 101 *Oryza* accessions of types occurring in South Asia (including the countries Bangladesh, Myanmar, India, Nepal, Pakistan, and Sri Lanka) was used in this study. These included 35 South Asian weedy rice samples, for which predominant ancestries were previously characterized (Huang et al. 2017), comprising 15 *aus*-like, five *indica*-like, five wild-like and ten admixed weeds; admixed weeds are most often classified in the de-domestication origins of US weedy rice (Li et al. 2017). Our goals were to: 1) determine which of these typical weedy traits characterize each of the known weedy rice populations occurring in South Asia; 2) investigate whether known candidate genes may explain the trait values found in each weedy rice population; and 3) determine whether weedy rice populations in South Asia have acquired weediness traits and alleles from standing variation in their ancestors, de novo mutation, or introgression from other sources. We further discuss the evolution of weediness traits in South Asian in context of recent discoveries of weedy rice evolution in other world regions.

**Plant**

**material and**

**phenotypic measurements**

A panel of 101 *Oryza* accessions of types occurring in South Asia (including the countries Bangladesh, Myanmar, India, Nepal, Pakistan, and Sri Lanka) was used in this study. These included 35 South Asian weedy rice samples, for which predominant ancestries were previously characterized (Huang et al. 2017), comprising 15 *aus*-like, five *indica*-like, five wild-like and ten admixed weeds; admixed weeds are most often classified in the de-domestication origins of US weedy rice (Huang et al. 2017). The panel also included *Oryza* groups potentially contributing to the genetic make-up of weedy rice; these were 44 cultivar varieties (15 *aus*, 13 *indica*, 11 *japonica* and five admixed cultivars), 18 wild ancestors of cultivated Asian rice (*O. rufipogon/O. nivara*), and four outgroup species (two *O. meridionalis* and two *O. barthii*) (Table S1).

Plant DNA was extracted as described in Huang et al. (2017) and used for sequencing of candidate genes (see below). We recorded seed traits for all 101 accessions from the original seed storage, with pericarp color classified as red or white, hull color classified as black or straw and awn as present of absent (Table S1).
We planted a subset of 44 accessions, with three replicates per accession, in a randomized design in two Conviron PGW36 growth chambers, under 11 hr day length with 25° temperature until 30 days after flowering. Due to limited chamber availability, the growth experiment was carried out in two stages. The first batch was planted on Feb 15, 2013 and the second on June 28, 2013. Tiller number and tiller angle were recorded at flowering. We divided tiller angle into three categories, <30°, 30-60° and 60-90°, as determined by the maximum tiller angle of the accession among three replicates. Seed shattering was measured as breaking tensile strength, using the method described in Thurber et al. (2010), which measures the weight needed to detach seeds from panicles 30 days after flowering.

**DNA sequencing and genotyping of candidate genes**

Following the Agilent (Agilent, Santa Clara, CA) SureSelect Target Enrichment method (Girnke et al. 2009) instructions, we designed baits of 120 bp length targeting the five candidate genes’ open reading frame (ORF). Additional baits were designed to target other known gene ORFs and gene fragments (~1500 base pair (bp)) in upstream and downstream regions of each baited gene. A total of 128 loci were thus targeted and sequenced (Table S2). Baits were designed based on the rice reference genome (MSU 6.0 assembly; http://rice.plantbiology.msu.edu/). The SureSelect Target Enrichment method was applied to capture targeted genomic sequences in the Oryza panel, and Illumina Hi-seq 2500 sequencing was performed at the Whitehead Institute (Massachusetts Institute of Technology).

Raw sequencing reads were assessed for quality with FastQC software (http://www.bioinformatics.babraham.ac.uk/projects/fastqc/). The low quality reads were filtered using NGStoolkit (Patel and Jain 2012), keeping reads with depth >3 and mapping quality >30. Clean reads were mapped onto the MSU 6.0 reference genome using BWA (Li and Durbin 2009). SNP and indels were called with SAMtools (Li et al. 2009). A series of in-house Perl scripts were employed to convert the polymorphisms from variant call format (VCF) to FASTA format for sequence alignment. In the FASTA files, unmapped nucleotides were represented with N and low-quality nucleotides were represented with “?”. Due to limitations of the SureSelect method to accurately detect moderate and large insertion/deletion (InDel) polymorphisms, we confirmed the presence of InDels with cleaved amplified polymorphisms (CAPS; Konieczny and Ausubel 1993) for two cases: the 14 bp InDel in exon 6 of Rc and the 22-bp InDel in exon 3 of Bh4. WebMap (http://pga.mgh.harvard.edu/web_apps/web_map_/start/) was used to create a map from the gene sequence files taken from the MSU database and Genbank. Restriction enzyme site complexity was set to greater than or equal to 5 residues in order to find an enzyme that would cut in the target region. Primer3 (http://frodo.wi.mit.edu/primer3/) was used to choose primers that would amplify the target region of the sequence.

The **Rc** gene was amplified using the primer pair **Rc_4774_for** (Sequence: tcaattctgatttttttctca) and **Rc_5299_rev** (Sequence: aggctttgaggattgaggagt). The **Bh4** gene was amplified with the primer pair **Bh4_1_for** (Sequence: caatctggtgcataatcagaatgga) and **Bh4_1_rev** primers (Sequence: cccgaagatctgagtagcagg). The PCR products were set for one cycle of 95° for 3 min (RC) and 30 sec (Bh4), 40 cycles of melting, annealing, and extension (95° for 1 min, 55° (Rc) and 59° (Bh4) for 30 sec, 68° for 1 min) and one cycle of 72° for 10 min. PCR products were digested with 0.5 ul of 20 units/ul restriction enzyme BanI for Rc and ApalI for Bh4 by incubating at 37° for 60 min. The products of the digestion were analyzed using gel electrophoresis to identify insertion and deletion alleles in Rc and Bh4.

**RESULTS**

**Phenotypic traits characterizing South Asian weedy rice groups**

We had previously characterized several seed traits in a larger panel of South Asian weedy rice (Huang et al. 2017), and results for the subset of accessions characterized here confirm prior observations.
(Table 1, Table S1). Red pericarps are common in all groups of South Asian weedy rice (>60% of samples in weedy groups are red). This is in contrast to both progenitor cultivated rice groups, although they do show the presence of both pericarp types, but similar to wild rice (Table 1). Hull color varies among South Asian weedy rice, with aus-like, wild-like, and admixed weeds having primarily black hulls (>70%), while indica-like weeds have exclusively straw colored hulls (Table 1, Table S1). Cultivated groups tend to have straw colored hulls (>83% of our cultivated samples), and 83% of the wild rice samples we examined have black hulls.

We additionally recorded the maximum tiller angle among replicates for each accession in our phenotyping study (Table S1). All South Asian weedy rice sub-groups have a relatively compact plant architecture, with angles <60° (Table 1, Table S1). This is similar to the cultivated aus group, but not as compact as indica, which usually has tiller angles <30°. As expected, wild rice has the most spread tillers, with the majority having angles between 60° and 90° (Table 1, Table S1). This result is consistent with previous observations on cultivars and wild rice (Jin et al. 2008).

We did not detect significant differences in tiller number among groups with the accession characterized for this study (P = 0.4), but trends confirmed prior significant observations from Huang et al. (2017), with higher tiller number occurring in wild rice and wild-like weeds (mean > 11), and lower numbers seen in cultivated aus and indica (mean < 7.71; Table 2). As previously reported (Huang et al. 2017), all weedy rice groups shatter easily, as does wild rice, and both cultivar-like weed groups shatter significantly more easily than their ancestors (P < 0.001; Table 2, Table S1).

Taken together, our results suggest convergence among South Asian weedy rice groups in ease of shattering, red pericarp color and compact plant architecture. Additionally, convergence occurs among aus-like, wild-like and admixed weeds for black hull color.

**Patterns of polymorphism in Rc and Bh4**

Rc is considered a rice improvement gene contributing to the prevalence of white pericarps in cultivated rice, with the recessive white pericarp phenotypes caused by at least two known loss-of-function alleles: a 14-bp deletion (rc) in exon 6 or a point mutation from C to A (Rc-s) (Sweeney et al. 2006). We first carried out a phylogenetic analysis of the Rc coding region based solely on SNP information, which verified a single origin for the white pericarp conferring Rc-s allele in aus cultivars (Figure 1). We mapped on to the tree the presence and absence of the 14-bp deletion, allowing us to distinguish Rc from rc alleles (Figure 1). Our data supports a single origin of the rc allele, which is prevalent in, but not exclusive to, cultivars. The two Rc haplotypes observed in the rc clade may have come about by recombination of alleles without the deletion into an rc SNP background. Our data also support the ancestral status of Rc alleles, and their prevalence in wild rice. Further examination of the association between Rc allele variation and pericarp color confirms that discrepancies between genotype and pericarp color are rare (six cases out of 95 accessions); most accessions with red pericarps carry the Rc allele and most accessions with white pericarps carry the rc or Rc-s allele across most Oryza groups (Table 3, Table S1). This supports prior observations that Rc is the major gene underlying pericarp color, although other genes may also have effects.

As expected due to the high frequency of red pericarps, South Asian weeds predominantly carry Rc alleles (Figure 1, Table 3). Furthermore, there is phylogenetic structure in Rc-type alleles, such that South Asian wild-like weeds most closely resemble wild rice Rc haplotypes, and aus-like weeds tend to possess alleles that are most closely related to haplotypes present in aus cultivars, be they Rc or Rc-s alleles. Indica-like and admixed weeds possess alleles falling in both the Rc and rc clades, but in all cases similar alleles from ancestral groups are also observed. Thus, for both weeds of wild and cultivated descent, the red pericarp trait in all South Asian weedy rice groups is likely inherited directly from ancestors where this trait is either common (wild rice) or rare but present (cultivated rice).

Like Rc, Bh4 is considered a rice improvement gene that contributes to the straw colored hull phenotype that is common, but not fixed, in cultivated rice (Zhu et al. 2011). A 22-bp frame-shift deletion in exon 3 has been shown to be the dominant functional polymorphism causing a straw hull phenotype (Zhu et al. 2011). Again we carried out a phylogenetic analysis based solely on SNP data in the ORF of the gene, and mapped the presence or absence of the deletion onto the tree (Figure 2). Phylogenetic SNP analysis broadly supports a single origin for Bh4 alleles carrying the 22 bp deletion, which are more common in crops, which tend to have straw-colored hulls, than in wild rice, which is predominantly black-hulled. A few deletion-carrying haplotypes occur outside the clade, as well as a few haplotypes without the deletion; recombination among alleles belonging to each clade could explain the rare lack of association between deletion status and Bh4 SNP background.

**Table 1 Categorical phenotypic traits in South Asian Oryza groups**

<table>
<thead>
<tr>
<th>Group</th>
<th>Pericarp</th>
<th>Hull</th>
<th>Tiller angle</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>red</td>
<td>white</td>
<td>black</td>
</tr>
<tr>
<td>Weedy rice</td>
<td>aus-like (15)</td>
<td>15 (100%)</td>
<td>0 (0%)</td>
</tr>
<tr>
<td></td>
<td>indica-like (5)</td>
<td>3 (60%)</td>
<td>2 (40%)</td>
</tr>
<tr>
<td></td>
<td>wild-like (5)</td>
<td>5 (100%)</td>
<td>0 (0%)</td>
</tr>
<tr>
<td></td>
<td>admixed (10)</td>
<td>8 (80%)</td>
<td>2 (20%)</td>
</tr>
<tr>
<td></td>
<td>aus (14)</td>
<td>5 (36%)</td>
<td>9 (64%)</td>
</tr>
<tr>
<td></td>
<td>indica (13)</td>
<td>3 (23%)</td>
<td>10 (77%)</td>
</tr>
<tr>
<td>South Asian wild rice</td>
<td>(18)</td>
<td>17 (94%)</td>
<td>1 (6%)</td>
</tr>
</tbody>
</table>

(Table 1, Table S1). Red pericarps are common in all groups of South Asian weedy rice (>60% of samples in weedy groups are red). This is in contrast to both progenitor cultivated rice groups, although they do show the presence of both pericarp types, but similar to wild rice (Table 1). Hull color varies among South Asian weedy rice, with aus-like, wild-like, and admixed weeds having primarily black hulls (>70%), while indica-like weeds have exclusively straw colored hulls (Table 1, Table S1). Cultivated groups tend to have straw colored hulls (>83% of our cultivated samples), and 83% of the wild rice samples we examined have black hulls.

We additionally recorded the maximum tiller angle among replicates for each accession in our phenotyping study (Table S1). All South Asian weedy rice sub-groups have a relatively compact plant architecture, with angles <60° (Table 1, Table S1). This is similar to the cultivated aus group, but not as compact as indica, which usually has tiller angles <30°. As expected, wild rice has the most spread tillers, with the majority having angles between 60° and 90° (Table 1, Table S1). This result is consistent with previous observations on cultivars and wild rice (Jin et al. 2008).

We did not detect significant differences in tiller number among groups with the accession characterized for this study (P = 0.4), but trends confirmed prior significant observations from Huang et al. (2017), with higher tiller number occurring in wild rice and wild-like weeds (mean > 11), and lower numbers seen in cultivated aus and indica (mean < 7.71; Table 2). As previously reported (Huang et al. 2017), all weedy rice groups shatter easily, as does wild rice, and both cultivar-like weed groups shatter significantly more easily than their ancestors (P < 0.001; Table 2, Table S1).

Taken together, our results suggest convergence among South Asian weedy rice groups in ease of shattering, red pericarp color and compact plant architecture. Additionally, convergence occurs among aus-like, wild-like and admixed weeds for black hull color.

**Table 2 Quantitative phenotypic traits in Asian Oryza groups. Numbers in brackets beside means are standard deviations**

<table>
<thead>
<tr>
<th>Group</th>
<th>Tiller number</th>
<th>Shattering</th>
</tr>
</thead>
<tbody>
<tr>
<td>South Asian weed (25)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>aus-like (8)</td>
<td>8.44(3.35)</td>
<td>2.14(5.79)</td>
</tr>
<tr>
<td>indica-like (5)</td>
<td>11.16(6.68)</td>
<td>17.46(16.70)</td>
</tr>
<tr>
<td>wild-like (5)</td>
<td>11.50(4.14)</td>
<td>8.36(19.05)</td>
</tr>
<tr>
<td>admixed (7)</td>
<td>7.57(3.95)</td>
<td>14.70(16.58)</td>
</tr>
<tr>
<td>aus (7)</td>
<td>7.71(1.50)</td>
<td>25.61(17.92)</td>
</tr>
<tr>
<td>indica (6)</td>
<td>7.00(2.97)</td>
<td>35.96(6.49)</td>
</tr>
<tr>
<td>South Asian wild rice (6)</td>
<td>11.44(7.55)</td>
<td>2.47(5.51)</td>
</tr>
</tbody>
</table>

Kuskal-Wallis 0.4040086 9.42E-08

1Chamber effect detected, only data from Chamber 1 plants were used in the analysis.

2No chamber effect detected.
Across *Oryza* groups, there is a strong association between hull color and the presence or absence of the Bh4 deletion, with 75% of samples matching expected phenotype based on deletion status, but several exceptions were also observed (Table S3, Figure 2). For example, three straw-hulled wild rice accessions and eight straw-hulled cultivars in our panel also have wild type alleles (Table S1, Table S3, Figure 2). This indicates that other loci can also affect hull color, as previously suggested (Kuriyama and Kudo 1967). In a previous survey of Bh4 sequences, two other loss-of-function mutations were detected leading to straw hulls: a 1-bp frame shift deletion in exon 1 and a SNP in exon 3 that causes an early stop codon (Zhu et al. 2011), but neither was observed in our panel.

South Asian weeds of different origins carry Bh4 haplotypes that primarily cluster with haplotypes of their ancestors (Figure 2). For example, wild-like weed alleles cluster together in the tree and seem to originate from a few closely related wild alleles. Aus-like and indica-like weeds with haplotypes carrying the 22 bp deletion cluster with various aus and indica haplotypes, and aus-like weeds lacking the...
Patterns of polymorphism in sh4
A single nonsynonymous G to T substitution in the first exon of sh4 has been shown to lead to incomplete development of an abscission layer in the seed-pedicel junction in cultivated rice, and thus loss of seed shattering during domestication (Li et al. 2006). Our phylogenetic analysis of the sh4 gene sequences obtained confirms the ancestral status of the G wild-type allele, and the prevalence of the derived T allele that is associated with loss of shattering in cultivated rice (Figure 3). South Asian weeds varied in the type of sh4 alleles they possess. Wild-like weeds all bear the ancestral G allele, and aus-like and indica-like weeds (with only two exceptions) have the derived T allele, indicating that all weeds groups have inherited their sh4 alleles from their respective ancestors. Given the near fixation of the T sh4 allele during rice domestication (Li et al. 2006), these results support the origin of indica-like and aus-like weeds in South Asia through de-domestication (Huang et al. 2017).

Patterns of polymorphism in Sdr4
The Sdr4 locus has been implicated in the seed dormancy trait, with Sdr4-k alleles being common in wild rice (O. rufipogon) and present in some indica cultivar varieties, and Sdr4-n alleles conferring reduced dormancy in almost all japonica and some indica cultivated rice (Sugimoto et al. 2010). We used SNP information to construct a phylogenetic tree of Sdr4, and overlay information of allele type, based on the ten SNPs and one 3 bp indels that differentiate the reported Sdr4-k and Sdr4-k haplotypes, as well as the two SNPs and one 3 bp indel that differentiate Sdr4-k from its closely related, but rarer, Sdr4-k’ allele (Sugimoto et al. 2010). We detected many variations of Sdr4, with multiple haplotype that could not be classified as either Sdr4-k or Sdr4-n (Figure 4, Table S1). For cultivars, consistent with previous reports (Sugimoto et al. 2010), japonica accessions only have the Sdr4-n allele, while Sdr4-n, Sdr4-k, Sdr4-k’ alleles, as well as various other haplotypes occur in aus and indica accessions.

We found that very few weeds have inherited the Sdr4-n allele despite its occurrence in all cultivated and wild groups. Instead, most weeds carry haplotypes that are more similar to Sdr4-k and k’. Unfortunately, for this candidate gene we had no phenotypic data. However, if the previous association between Sdr4 allele types and seed dormancy is taken at face value, our results suggest that most aus-like, indica-like, wild-like, and admixed South Asian weeds have some degree of seed dormancy conferred by the lack of Sdr4-n alleles. Moreover, although this trait may have been inherited from standing variation in cultivated and wild ancestors, it is significant that almost no weed derived from these groups has inherited the Sdr4-n variant that segregates in these groups, suggesting selection favoring Sdr4-k and similar alleles in South Asian weeds.
Genetic diversity analyses

If weediness candidate genes were targeted by selection during weed evolution, a decrease in nucleotide diversity should be evident at these loci. We compared nucleotide diversity and Tajima’s D values at silent sites in candidate genes with those obtained from all 128 targeted loci as a representation of genome-wide averages for groups of weeds and their putative ancestors. Mean genome-wide nucleotide diversity is highest in wild rice and in wild-like weeds (Table 4). The two cultivar groups and their derived weeds have similar levels of diversity, suggesting that these de-domesticated weed groups have not experienced strong bottlenecks upon evolving from cultivated ancestors.

Most candidate genes surveyed have levels of nucleotide diversity that are unremarkable compared to genomic averages, not exceeding two standard deviations from the mean. Sh4 stands out for its very low levels of diversity in aus, indica and indica-like groups, and this is consistent with selection for reduced shattering during the domestication process of cultivar groups (Li 2006; Konishi et al. 2006; Thurber et al. 2010). Low sh4 nucleotide diversity in indica-like weeds can be attributed to the lack of diversity at this gene in ancestral indica. Other candidate genes showing extremely low levels of nucleotide diversity are PROG1 and sh4 in the aus group, which suggests selection on some aspect of plant architecture in aus (Table 4).

Weedy groups derived from crops did not show bottleneck evidence on a genome-wide level (Table S5), but wild-like weeds on average possess about 60% of their ancestral diversity. Loss of diversity in

---

**Figure 2** Neighbor joining (NJ) tree of Bh4 haplotypes. The color range indicates the two major InDel alleles of Bh4; the inner ring indicates the hull color. The outer ring and the bullets indicate Oryza groups, as labeled in the key. Unmarked weeds are admixed according to results of Huang et al. (2017). Asterisks on branches indicate bootstrap values as labeled.
wild-like weeds is particularly evident in Rc and Bh4 suggesting possible selection on these pericarp and hull coloration genes in wild-like weeds. In contrast, crop-derived weeds show remarkable increases in diversity for PROG1 compared to their ancestors. This could be due to strong selection on cultivars for domesticated plant architecture, coupled with release of selection during de-domestication of the weeds. However, it is notable that weeds resembled crops in their predominant tiller angles (Table 1). A similar situation is seen for sh4 in aus-like weeds, where high diversity could be due to release from selection.

Genome-wide F$_{ST}$ values for the three weed-ancestor comparisons show that indica-like weeds have negligible amounts of differentiation from indica cultivars (F$_{ST}$ = 0.014), while differentiation between aus-like weeds and aus (F$_{ST}$ = 0.172) and wild-like weeds and wild rice (F$_{ST}$ = 0.165) is moderate (Table 5). Among candidate genes, Bh4 stands out by exceeding genome-wide F$_{ST}$ values by more than two standard deviations (SD) from the mean between aus-like weeds and aus. This is consistent with the higher frequency of black hulls observed in aus-like weeds compared to their cultivated ancestors and could indicate selection favoring black-hull Bh4 haplotypes in weeds (Table 1).
DISCUSSION

We had previously investigated the evolutionary origins of weedy rice in South Asia, an area harboring great cultivated and wild *Oryza* diversity (Huang et al. 2017). Weedy rice in this region has a highly heterogeneous genetic background, more so than many other world regions (Reagon et al. 2010; Vigueira et al. 2017; Li et al. 2017), with contributions from two cultivated rice varieties (*aus* and *indica*) and from the wild ancestor of domesticated rice. Despite these very diverse origins, our current study finds that weedy rice groups in South Asia share the presence of several iconic weediness traits, supporting the concept of an agricultural weed syndrome (Vigueira et al. 2013b), and helping define the suite of traits that are most likely to be adaptive in weedy rice.

Of the traits examined, the greatest convergence among all three weed groups in South Asia occurred for easy seed shattering, red pericarp color, and compact plant architecture (Table 1; Table 2; Figure 1; Figure 3; Figure S1; Figure 5), suggesting that these traits are advantageous in the South Asian agricultural environment. Convergence among the wild and cultivar derived weedy rice groups is striking, given that for some of these traits weeds do not resemble their ancestors. For example, red pericarps and shattering phenotypes do not predominate in cultivars but are common in *aus*-like and *indica*-like weeds; likewise, wild rice tends to have a spreading architecture, yet wild-like weeds tend to be more compact. Convergence for seed shattering and red pericarp also occurs with weed groups in other parts of the world such as the U.S. (Thurber et al. 2010; Gross et al. 2010; Figure 5).

**Figure 4** Neighbor joining (NJ) tree of *Sdr4* haplotypes. The color range indicates the three *Sdr4* haplotypes identified in Sugimoto et al. (2010). The ring and the bullets indicate *Oryza* groups, as labeled in the key. Unmarked weeds are admixed according to results of Huang et al. (2017). Asterisks on branches indicate bootstrap values as labeled.
and Malaysia (Song et al. 2014; Cui et al. 2016), supporting the idea that these are important traits in weedy rice evolution. Unfortunately, there is not enough information in the literature to gauge if compact plant architecture is also common in other weed groups. Our data suggest that, among the traits surveyed, seed dormancy potentially related to red pericarps, and seed shattering are the most essential weediness traits in weedy rice worldwide.

Given separate origins, how have weedy groups in South Asia acquired convergent weedy traits? Our survey of PROG1 as a candidate gene for tiller angle is inconclusive, but our candidate genes for pericarp color and seed shattering suggest differing genetic mechanisms. Most weedy accessions with red pericarps carry wild-type Rc alleles that can explain the phenotype and were likely inherited from variation present in their ancestral groups (Figure 1). The similar nucleotide diversity in the Rc gene for all cultivar and weed groups, as well as the low pairwise Fst between weed and ancestor, also serve as evidence of the inheritance of ancestral alleles (Table 4; Table 5). The very occasional discrepancies between pericarp color and Rc alleles (Figure 1, Table 3) indicate that other genes or mutations may affect pericarp color, but the Rc genotype in weeds is mostly sufficient to explain weed phenotype. Consistent with origins for ancestral standing variation, there are no strong signatures of selection observed for this gene in weed groups (Table 3).

A similar situation to pericarp color and Rc may be occurring with seed dormancy and its candidate gene Sdr4. We unfortunately do not have seed dormancy phenotypic data for our weedy populations, but seed dormancy has often been considered a characteristic weedy trait (Sugimoto et al. 2010). The lack of dormancy-associated Sdr4 alleles in any weed population, while easily explained through inheritance from ancestral standing variation, suggests possible convergence in seed dormancy among all weed groups possibly mediated through the evolutionary advantage conferred by maintenance of the dormancy trait.

Convergence of seed shattering in South Asian weeds seems to have occurred in a different manner. In our South Asian panel, every wild-like weed has the same ancestral wild type sh4 allele as wild rice that has been shown to be associated with seed shattering (Li et al. 2006), thus sh4 inheritance from ancestors can explain the shattering phenotype in wild-like weeds. In contrast, almost all cultivar-like weeds have the derived sh4 allele common in crop rice (Table S4; Figure 3) that has been implicated in loss of shattering during domestication. Thus, although inheritance from cultivated ancestors explains the sh4 allele present in indica and aus-like weeds, this allele cannot explain the shattering phenotype. This is similar to what has been observed for US weedy rice (Thurber et al. 2010), where a novel genetic mechanism must account for seed shattering. Whether there is convergence in this novel mechanism between South Asian cultivar-derived weeds and any of the US groups is unknown, and we cannot yet draw conclusions about the roles of standing variation vs. new mutations in the evolution of shattering in these groups. The high levels of nucleotide diversity of aus-like weed in sh4 compared to aus cultivars, and the extremely negative Tajima’s D (Tables 4; Table S5), suggest that aus-like weed may be undergoing a release in selection for this gene.

While not common to all weedy groups in South Asia, a high degree of convergence for black hull color occurs among wild-like, admixed and aus-like weeds. This trait is also convergent with one group (BHA) of US weedy rice (Reagon et al. 2010). The fact that dark hulls and awns are present, but not prevalent in the aus ancestors of both BHA and aus-like South Asian weeds (Table S3), suggests there may be some adaptive value to these traits such as resistance to pests and predators. However, these cannot be considered essential weedy traits, as weedy rice groups that lack them nevertheless thrive around the world (Ziska et al. 2015).

### Table 4 Summary statistics on silent sites for candidate genes and all loci obtained with target-capture

<table>
<thead>
<tr>
<th>Locus</th>
<th>Summary statistics</th>
<th>aus</th>
<th>indica</th>
<th>aus-like weed</th>
<th>indica-like weed</th>
<th>wild-like weed</th>
<th>wild</th>
</tr>
</thead>
<tbody>
<tr>
<td>Rc</td>
<td>π</td>
<td>0.0025</td>
<td>0.0026</td>
<td>0.0021</td>
<td>0.0040</td>
<td>0.0029</td>
<td>0.0057</td>
</tr>
<tr>
<td></td>
<td>θw</td>
<td>0.0029</td>
<td>0.0033</td>
<td>0.0018</td>
<td>0.0039</td>
<td>0.0027</td>
<td>0.0089</td>
</tr>
<tr>
<td></td>
<td>Tajima’s D</td>
<td>-0.5243</td>
<td>-0.9636</td>
<td>0.6020</td>
<td>0.2043</td>
<td>0.4154</td>
<td>-1.5138</td>
</tr>
<tr>
<td>Bh4</td>
<td>π</td>
<td>0.0068</td>
<td>0.0092</td>
<td>0.0059</td>
<td>0.0096</td>
<td>0.0013</td>
<td>0.0141</td>
</tr>
<tr>
<td></td>
<td>θw</td>
<td>0.0074</td>
<td>0.0077</td>
<td>0.0074</td>
<td>0.0011</td>
<td>0.0015</td>
<td>0.0130</td>
</tr>
<tr>
<td></td>
<td>Tajima’s D</td>
<td>-0.2999</td>
<td>0.8084</td>
<td>-0.7779</td>
<td>-1.2185</td>
<td>-0.9726</td>
<td>0.3387</td>
</tr>
<tr>
<td>sh4</td>
<td>π</td>
<td>0.00035</td>
<td>0</td>
<td>0.0009</td>
<td>0</td>
<td>0.0090</td>
<td>0.0058</td>
</tr>
<tr>
<td></td>
<td>θw</td>
<td>0.00023</td>
<td>0</td>
<td>0.0021</td>
<td>0</td>
<td>0.0083</td>
<td>0.0083</td>
</tr>
<tr>
<td></td>
<td>Tajima’s D</td>
<td>1.1224</td>
<td>NA</td>
<td>-2.124</td>
<td>NA</td>
<td>0.6457</td>
<td>-1.2382</td>
</tr>
<tr>
<td>PROG1</td>
<td>π</td>
<td>0</td>
<td>0.0111</td>
<td>0.0010</td>
<td>0.0089</td>
<td>0.0089</td>
<td>0.0087</td>
</tr>
<tr>
<td></td>
<td>θw</td>
<td>0</td>
<td>0.0024</td>
<td>0.0023</td>
<td>0.0106</td>
<td>0.0071</td>
<td>0.0107</td>
</tr>
<tr>
<td></td>
<td>Tajima’s D</td>
<td>NA</td>
<td>-1.1491</td>
<td>-1.1594</td>
<td>-1.0485</td>
<td>1.4588</td>
<td>-0.6036</td>
</tr>
<tr>
<td>Sdr4</td>
<td>π</td>
<td>0.0018</td>
<td>0.0022</td>
<td>0.0018</td>
<td>0.0036</td>
<td>0.0036</td>
<td>0.0037</td>
</tr>
<tr>
<td></td>
<td>θw</td>
<td>0.0011</td>
<td>0.0023</td>
<td>0.0033</td>
<td>0.0034</td>
<td>0.0035</td>
<td>0.0063</td>
</tr>
<tr>
<td></td>
<td>Tajima’s D</td>
<td>1.3759</td>
<td>-0.1269</td>
<td>-1.3165</td>
<td>0.2431</td>
<td>0.2431</td>
<td>-1.3434</td>
</tr>
<tr>
<td>All 128 genes</td>
<td>π</td>
<td>0.0041</td>
<td>0.0040</td>
<td>0.0044</td>
<td>0.0048</td>
<td>0.0062</td>
<td>0.0090</td>
</tr>
<tr>
<td></td>
<td>θw</td>
<td>0.0040</td>
<td>0.0043</td>
<td>0.0044</td>
<td>0.0049</td>
<td>0.0060</td>
<td>0.0105</td>
</tr>
<tr>
<td></td>
<td>Tajima’s D</td>
<td>-0.0211</td>
<td>-0.2047</td>
<td>-0.0885</td>
<td>-0.1591</td>
<td>0.2740</td>
<td>-0.8180</td>
</tr>
</tbody>
</table>

Nucleotide diversity (π or θw) of individual gene <0.0005. Tajima’s D >2 or < -2.

### Table 5 Pairwise Fst between weed groups and ancestors for each of the 15 candidate genes (Hudson et al. 1992)

<table>
<thead>
<tr>
<th>Locus</th>
<th>aus-like vs. aus</th>
<th>indica-like vs. indica</th>
<th>wild-like vs. wild</th>
</tr>
</thead>
<tbody>
<tr>
<td>Rc</td>
<td>0.0123</td>
<td>0.0000</td>
<td>0.2917</td>
</tr>
<tr>
<td>Bh4</td>
<td>0.6362</td>
<td>0.0000</td>
<td>0.3067</td>
</tr>
<tr>
<td>Sh4</td>
<td>0.8884</td>
<td>NA</td>
<td>0.1047</td>
</tr>
<tr>
<td>PROG1</td>
<td>0.1364</td>
<td>0.0124</td>
<td>0.2072</td>
</tr>
<tr>
<td>Sdr4</td>
<td>0.2279</td>
<td>0.0000</td>
<td>0.1386</td>
</tr>
<tr>
<td>All 128 genes</td>
<td>0.1717</td>
<td>0.0142</td>
<td>0.1654</td>
</tr>
</tbody>
</table>
Song et al. 2014). Both wild-like and aus-like weeds likely inherited these phenotypes from variation present in ancestral groups. In the case of hull color, this is supported by Bh4 variation (Figure 2), with wild-like weeds carrying Bh4 alleles similar to wild rice, and aus-like weeds carrying alleles similar to the aus in our panel with no-deletion alleles. While the black hull phenotype is rare in the aus in our panel, a slightly higher frequency has been previously reported for a wider set of aus surveyed (Huang et al. 2017). Surprisingly, although aus-like weeds likely inherited Bh4 alleles from a small subset of aus cultivars with black hulls, both groups show similar nucleotide diversity at Bh4 (Table 4; Table S5), but very large differentiation as measured by FST (Table 5). This is compatible with a scenario whereby aus-like weeds were de-domesticated from aus cultivars early in the domestication process, before the majority of aus cultivars underwent selection for straw hulls with the 22-bp Bh4 deletion (Vigueira et al. 2013a).

Our characterization of weedy rice phenotypes in South Asia and the associated candidate genes contribute to the emerging understanding of the mechanisms by which weedy rice evolves worldwide. Whether evolving from crop ancestors, as inferred for BHA and SH weeds in the US and aus-like weeds and indica-like weeds in South Asia, or evolving from wild ancestors, such as the wild-like weeds of South Asia, standing variation present in the non-weedy ancestors often seems a sufficient explanation for the origin of weedy rice traits. Thus, for none of the genes surveyed, even those that underlie traits where weeds diverge from their ancestors, is introgression required to explain the alleles present in weedy rice (the only exception being the unknown genetic mechanism for shattering in cultivar-related weeds). The fact that cultivated rice varieties, which have been shown to have gone through domestication bottlenecks and loss of diversity (Caicedo et al. 2007), as well as wild rice, which is not a priori adapted to the agricultural environment, nonetheless harbor the necessary standing variation (i.e., allelic diversity) to give rise to troublesome weeds, may help explain the prevalence of the weedy rice problem worldwide. The possibility of continued repeated evolution of weedy rice presents an interesting conundrum for usual calls for maintenance of diversity in crop species.

ACKNOWLEDGMENTS

The authors thank the University of Massachusetts greenhouse staff and S. Perera for plant care support. We also thank the larger working group of the Plant and Microbial Innovation theme in the Institute for Applied Life Sciences for discussions. The United States Department of Agriculture (USDA) is an equal opportunity provider and employer. This work was supported by a grant from the US National Science Foundation Plant Genome Research Program (IOS-1032023) to A.L.C., K.M.O and Y.J.

LITERATURE CITED


Ithal, N., and A. R. Reddy, 2004 Rice flavonoid pathway genes, OsDfr and OsAns, are induced by dehydration, high salt and ABA, and contain stress responsive promoter elements that interact with the transcription activator, OsC1-MYB. Plant Sci. 166: 1505–1513. https://doi.org/10.1016/j.plantsci.2004.02.002


