MOLECULAR ECOLOGY

Molecular Ecology (2011) **20**, 3743–3756 doi: 10.1111/j.1365-294X.2011.05216.x

The long and the short of it: *SD1* polymorphism and the evolution of growth trait divergence in U.S. weedy rice

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

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

Abstract

Growth-related traits, such as greater height, greater biomass, faster growth rate and early flowering, are thought to enhance competitiveness of agricultural weeds. However, weedy rice, a conspecific weed of cultivated rice (*Oryza sativa* L.), displays variation for growth traits. In the United States, separately evolved weedy rice groups have been shown to share genomic identity with exotic domesticated cultivars. Through a common garden experiment, we investigated whether growth trait divergence has occurred among U.S. weeds and their putative cultivated progenitors. We also determined polymorphism patterns in the growth candidate gene, *SD1*, to assess its possible role in the evolution of divergent phenotypes. We found considerable growth trait variation among weed groups, suggesting that growth trait convergence is not evident among weedy populations. Phenotypic divergence of weedy rice from cultivated ancestors is most apparent for flowering time. Introgression of a chromosomal block containing the *SD1* allele from *tropical japonica*, the predominant U.S. rice cultivar, was detected in one weedy rice population and is associated with a change in growth patterns in this group. This study demonstrates the role of introgressive hybridization in evolutionary divergence of an important weed.

Keywords: flowering time, introgression, *Oryza sativa*, plant height, red rice, *SD1* polymorphism

Received 9 April 2011; revision received 28 June 2011; accepted 3 July 2011

Introduction

Agricultural weeds are highly successful organisms, often possessing traits that enhance their ability to invade and persist in agricultural habitats. Conspecific weeds – those that have evolved within the same complex of species as domesticated plants – are frequently characterized by traits that differentiate them from their crop and wild relatives. Conspecific weeds are common; over half of the world's top 10 crops possess conspecific weedy forms (Ellstrand *et al.* 1999). Some traits typical of conspecific weeds are believed to enhance colonization ability, such as increased seed dispersal or dormancy. Other traits involve aspects of growth and development, such as faster growth rate, greater biomass and greater plant height, and may increase competitiveness (Baker 1974; Falster & Westoby 2003). There is great interest in understanding the evolutionary mechanisms by which weedy traits arise in conspecific weed species.

There are several mechanisms by which conspecific weed populations can evolve within crop–wild species complexes. Weed phenotypes may evolve *de novo*, through new mutations favoured by selection in weed groups. Alternatively, weedy traits may be inherited from standing variation in the founding progenitors. A third possible source of traits is introgressive hybridization between weeds and related wild/cultivated groups. Unlike demographic history, and similar to selection, introgression affects small portions of the genome (Scascitelli *et al.* 2010), and sampling of genomic fragments may miss introgressed regions (Lexer & Widmer 2008). As a result, the importance of introgression in weed evolution can be underestimated (Whitney *et al.* 2010),

Correspondence: Ana L. Caicedo, Fax: (413) 545 3243; E-mail: caicedo@bio.umass.edu

and its prevalence relative to other evolutionary processes is unknown.

One of the most notorious conspecific weeds is weedy rice (also called red rice), which invades cultivated rice (*Oryza sativa* L.) fields worldwide (Burgos *et al.* 2008). A weedy life history appears to have evolved multiple times around the world and from different sources in the *Oryza* complex (Olsen *et al.* 2007). In cultivated rice fields of the U.S., weedy rice has been among the most damaging weeds for more than 150 years, despite changes in agricultural practices and weed control efforts (Delouche *et al.* 2007). Currently, weedy rice infestations in the U.S. can reduce crop yields by up to 80% (Estorninos *et al.* 2005) and are estimated to lead to economic losses of about $50 million annually (Gealy *et al.* 2002).

Recent microsatellite and single-nucleotide polymorphism (SNP) data suggest that U.S. weedy rice evolved from Asian domesticated rice (Londo & Schaal 2007; Reagon *et al.* 2010). U.S. weedy rice populations contain several morphotypes based on grain and hull characteristics, and weedy individuals with the most frequently observed strawhull morphology (SH) share a similar genetic background with *indica* rice cultivars (Londo & Schaal 2007; Reagon *et al.* 2010). Weed strains with predominantly blackhulls and awns (BHA) tend to cluster with *aus* cultivars, a variety of rice grown in the northern Indian subcontinent (Londo & Schaal 2007; Reagon *et al.* 2010), and are structured into two genetically distinct groups (BHA1 and BHA2) (Reagon *et al.* 2010). Surprisingly, neither *indica* nor *aus* varieties have ever been commercially cultivated in the U.S. Cultivated rice in the southern U.S. belongs to the *tropical japonica* group, which is genetically distinct from *indica* and *aus* (Garris *et al.* 2005; Caicedo *et al.* 2007). No evidence has been found to support weedy rice evolving directly from local U.S. *tropical japonica* cultivars, and past assertions of contributions of the wild ancestor of cultivated Asian rice, *O. rufipogon*, are not supported by genome-wide SNP data (Reagon *et al.* 2010).

U.S. weedy rice strains are almost universally characterized by dark-pigmented pericarps, seed shattering and seed dormancy (Burgos *et al.* 2006; Delouche *et al.* 2007). These last two traits are believed to be conducive to weediness by promoting persistence in agricultural fields. In contrast, U.S. weeds vary in traits related to plant growth, such as height, growth rate, degree of tillering and flowering date (Pantone & Baker 1991; Estorninos *et al.* 2002; Delouche *et al.* 2007; Shivrain *et al.* 2010). This is surprising, as certain growth strategies are believed to better enhance competitiveness in weeds. For example, growing quickly and flowering early may ensure weed seed dispersal prior to crop harvest (e.g. Sahli *et al.* 2008); increased tillering is associated with greater competitiveness in rice (e.g. Estorninos *et al.* 2002). The presence of phenotypic diversity in weedy rice growth traits suggests that multiple competitive strategies may lead to success in the U.S. agricultural environment. Alternatively, convergence among independently evolved weed groups may be occurring for some key weedy growth traits.

Little is known about how the various growth-related weed phenotypes have arisen in U.S. weedy rice, and the evolutionary origin of U.S. weeds from cultivated ancestors further confounds the origin of weedy traits. U.S. weed groups are phenotypically distinct from the *tropical japonica* crop they invade. However, as the putative ancestral groups of U.S. weeds are not grown locally, it is not clear whether growth phenotypic divergence has occurred between weeds and their domesticated progenitors. Indeed, recently, the need for common garden experiments to demonstrate phenotypic divergence between conspecific weeds and their progenitors has been stressed (Ellstrand *et al.* 2010). Both major weedy rice groups show little genetic divergence from their respective progenitors and low levels of nucleotide diversity (Reagon *et al.* 2010). These genetic patterns of diversity are in contrast to the extensive morphological diversity found in weed populations, underscoring our lack of understanding of weedy trait evolution in these groups.

Clues about the origin and evolution of weedy traits may be obtained from examination of suitable candidate genes. For growth-related traits in weedy rice, one such locus is the vegetative growth candidate gene, *SD1* (also known as *GA20ox2*). *SD1* is most famous as the locus exploited in the breeding of high-yielding semi-dwarf rice cultivars during the 'green revolution' (Hedden 2003). A semi-dwarf mutant found in the Taiwanese *indica* landrace 'Dee Gee Woo Gen' was used to develop short-stature cultivars during the 1960s. A recessive allele, semi-dwarf 1 (*sd1*), caused by a 383-bp deletion in *SD1*, is primarily responsible for the reduction in height observed in most semi-dwarfs (Monna *et al.* 2002; Sasaki *et al.* 2002; Spielmeyer *et al.* 2002). Located on the long arm of chromosome 1, *SD1* encodes a GA20-oxidase, a critical enzyme involved in the final steps of gibberellin (GA) biosynthesis.

Gibberellins are an important family of plant hormones that strongly influence growth and development, including traits associated with weed success (e.g. germination, rapid emergence, stem elongation, tillering, plant height, flowering, fruit development and final biomass) (Reid 1993; Hooley 1994; Ross *et al.* 1997; Richards *et al.* 2001). *SD1* is strongly expressed in rice leaf blades, stems and flowers, suggesting that multiple aspects of growth may be influenced by mutations in this gene (Ashikari *et al.* 2002; Sasaki *et al.* 2002). Both

natural and induced *SD1* mutations with phenotypic effects occur in cultivated rice (Ashikari *et al.* 2002; Monna *et al.* 2002; Spielmeyer *et al.* 2002; Nagano *et al.* 2005). *SD1* is thus a particularly ideal candidate gene for growth-related phenotypic divergence in cultivated and weedy rice. Indeed, semi-dwarf weedy rice strains in Japan carrying the green revolution *sd1* allele have recently been detected (Kawasaki *et al.* 2009).

In this study, we use a common garden approach to determine whether the vegetative growth phenotypes characterizing U.S. weedy rice have evolved since weed divergence from cultivated ancestral groups. Concurrently, we assess whether any convergence in growth trait strategies is evident in U.S. weedy rice populations. Finally, we take advantage of the known links between GA and plant growth traits, and between GA synthesis and the *SD1*gene, to infer possible mechanisms driving the evolution of growth-related weedy traits in U.S. weedy rice. Our results show that phenotypic divergence from cultivated ancestors has occurred in weedy rice evolution primarily for flowering time and that growth trait convergence is not evident among weedy rice populations. We further show that introgressive hybridization has played a previously unrecognized role in the evolution of growth-related traits in U.S. weedy rice.

Methods

Plant material

We used the same panel of 144 *Oryza* accessions described in Thurber *et al.* (2010) for sequencing (Table S1, Supporting information). This panel included 58 accessions of U.S. weedy rice, representative of the geographical and morphological diversity present in seed stocks maintained by the USDA-ARS Dale Bumpers National Rice Research Center in Stuttgart, Arkansas. Weedy accessions were classified according to morphology and genome-wide SNP variation (Reagon *et al.* 2010) as follows: 24 strawhull samples, 15 and 9 from each blackhull awned group (BHA1 and BHA2, respectively), five samples with predominantly brownhull morphology that seem to have originated from hybridization between SH and BHA weeds (BRH) and five samples that are predominantly an admixture of weeds and local *tropical japonica* crops (MX groups).

Further *Oryza* samples in the panel comprise both landrace and modern accessions from five genetically distinct variety groups of cultivated *O. sativa* that originated in Asia: *indica* (9), *aus* (6), *tropical japonica* (7), *temperate japonica* (8) and *aromatic* (3), as well as an additional 13 *tropical japonica* cultivars representative of southern U.S. rice cultivation (Dilday 1990). Together, *indica* and *aus* varieties compose the *indica* subspecies of domesticated Asian rice, and *tropical japonica*, *temperate japonica* and *aromatic* varieties compose the genetically distinct *japonica* subspecies (Garris *et al.* 2005; Londo *et al.* 2006; Caicedo *et al.* 2007). The panel also included accessions of the wild progenitor of cultivated rice, *O. rufipogon* (30), and its annual ecotype, *O. nivara* (2). Other closely related *Oryza* species were examined as potential contributors of weedy traits and as outgroups, including two accessions of wild South American rice (*O. glumaepatula*), four accessions of African domesticated rice (*O. glaberrima*), three of its wild progenitor (*O. barthii*) and two of Australian wild rice (*O. meridionalis*). A subset of all accessions was used for phenotyping growth traits. Further details on plant materials are provided in Table S1 (Supporting Information) and in Reagon *et al.* (2010).

Common garden and growth measurements

We used the same plants as Thurber *et al.* (2010) to measure growth traits that have been shown to be affected by endogenous GA levels in cultivated rice and are therefore probably affected by *SD1*. Briefly, a subset of 90 *Oryza* accessions was grown in a completely randomized block design in two Conviron PGW36 growth chambers at the University of Massachusetts Amherst (Table S1, Supporting Information). Experimental plants were selected prior to the availability of *SD1* sequence data, and excluded *Oryza* groups previously shown not to have contributed to the genomic background of weedy groups (Reagon *et al.* 2010). Cultivation conditions were as described in Thurber *et al.* (2010). We measured plant height at 10 days postemergence, plant height at the time of flowering, number of days to flowering and the total number of tillers. Plant height was measured from the soil surface to the tip of the longest flag leaf at 10 days. At flowering, plant height was measured from the soil surface to the base of the panicle. The number of days to flowering was determined by when the first panicle emerged ~50% from the sheath (once at least 10 florets have emerged). We calculated an 'emergence growth rate' (plant height at 10 days/10) to characterize early growth and 'average growth rate' (height at flowering/days to flowering) to determine whether differences in overall growth rate occur among *Oryza* groups.

To visualize morphological divergence between weedy and putative parental populations as well as among weedy groups, we used several principal component analyses (PCA) to decompose the multiple trait measurements into two primary axes of variation, which were plotted to show differentiation among the various groups. PCA were performed using plant height at

flowering, number of tillers, number of days to flowering, emergence growth rate and average growth rate. Raw data were first scaled to have a unit variance and centred on the column means. All calculations were performed using pcaMethods package in R, using an option that corrects for missing data (R Core Development Team 2007). To further assess the degree to which ancestral and weedy populations can be differentiated by growth traits, we used linear discriminant analysis (LDA) on the same group combinations as the PCA, with the LDA function of the MASS package in R. Canonical structure coefficients were generated with the ade4 package in R (Chessel *et al.* 2004) to determine the relationship between each variable and each discriminant function.

To determine whether individual traits differed significantly between weedy and putative ancestors and among weedy groups, we used a series of one-way Kruskal–Wallis tests, one for each trait. Nonparametric tests were chosen because at least one of our traits (tiller number) is unlikely to have been sampled from a normal distribution, some of our samples sizes were small and sample sizes were unbalanced. Chamber effects were not detected, so trait means were calculated for each accession prior to further analyses. For each significant Kruskal–Wallis test ($P < 0.05$), groups with significant differences were identified post-hoc with Mann–Whitney tests using the Benjamini and Hochberg (BH) correction for multiple comparisons (Benjamini & Hochberg 1995). Statistical tests were carried out with the kruskal function of the agricolea package and the kruskal.test function of the stats package in R (R Core Development Team 2007).

DNA extraction and sequencing

DNA was extracted from approximately 1 g of fresh leaf material from one plant per accession in a panel of 144 weedy, wild and cultivated *Oryza* as in Reagon *et al.* (2010). A larger genotyping panel was used to include outgroup species and assess the frequency of weedy alleles in all rice cultivar groups and related wild species. Using Primer3 (Rozen & Skaletsky 2000), we designed overlapping PCR primers from the *O. sativa japonica* (var. Nipponbare) genome (TIGR v. 5 January, 2007) to cover the entire open reading frame of *SD1*, as well as ~1500 base pairs (bp) upstream of the start codon and ~1000 bp downstream of the end of the gene. An additional six ~500 bp regions of genes at increasing distance (~150 kb to ~2 Mb) from *SD1* were also sequenced, spanning a region of ~ 4.1 Mb of the long arm of chromosome 1. A description of flanking gene locations and primers used can be found in Supporting Information Table S2 (Supporting information). PCR amplification and DNA sequencing were performed by Cogenics (Beckman Coulter Genomics) as described in Caicedo *et al.* (2007).

Sequence alignment and editing were carried out with BioLign version 2.09.1 (Tom Hall, NC, State University). New DNA sequences obtained for this study were deposited in GenBank under accession numbers JN541407–JN542382. We also added *SD1* sequences available on GenBank to check for polymorphisms in *Oryza* found in earlier studies.

Genetic diversity and divergence

Levels and patterns of genetic diversity in *SD1* and six flanking loci were compared among weedy and other *Oryza* groups, using θ_W and θ_π calculated for silent sites with the perl scripts used in Caicedo *et al.* (2007). Tajima's D values and their significance were calculated using DNAsp version 5 (Librado & Rozas 2009). Groups were defined based on the clustering analysis of Reagon *et al.* (2010), using 48 ~500-bp sequence-tagged site (STS) fragments, randomly distributed across the genome.

To infer genealogical relationships at *SD1*, we performed a neighbour-joining (NJ) analysis using MEGA 4 (Tamura *et al.* 2007), considering the promoter, ORF and 3′ region. We considered pairwise deletion of gaps/missing data and calculated distances using the Kimura 2-parameter model. Heterozygotes were rare in our data set, occurring occasionally only in *O. rufipogon* and were included in the analysis as ambiguity codes. A few individuals were removed in generating the final tree owing to extensive missing data.

Using the 48 STS fragments from Reagon *et al.* (2010) as a reference set, we tested whether observed patterns of intraspecific polymorphism and interspecific divergence for *SD1* were consistent with neutrality in the SH and BHA1 groups with a maximum-likelihood multilocus HKA test (Wright & Charlesworth 2004). This method compares the likelihoods of a model that assumes all loci sampled are evolving neutrally with a model where one or more loci are assumed to have been the targets of selection. We compared two models, one where we assumed all loci in the reference set and *SD1* evolved neutrally and the other where *SD1* was assumed to be the target of selection. We used silent site polymorphisms for all tests, and *O. barthii* as the outgroup, because of sequencing failures in the more distant *O. meridionalis*. Four STS fragments missing *O. barthii* sequence were not used. The mean of three independent runs, each initiated using a different random seed and run for 1 million iterations, was used for likelihood ratio testing to identify the best fitting model. To determine whether individual traits differed

between *Oryza* groups carrying different *SD1* haplotypes, we used Kruskal–Wallis tests for each trait, as described previously.

Results

Phenotypic variation

The mean and standard deviation for morphological traits calculated for cultivated rice and the four largest weedy groups are given in Table 1, and the various PCA of putative cultivated ancestors and weedy rice groups are shown in Fig. 1. For all PCA, the first two principal components (PCs) explained >68% of the variation observed. LDA revealed that the predictive value of the five growth traits was high, exceeding 64% in all group combinations evaluated (Table S3, Supporting Information).

The putative progenitor populations of weedy rice, *aus* and *indica*, were moderately divergent and distinguishable by PCA (Fig. 1a). On average, *aus* were observed to grow more and more quickly than *indica* (Table 1), and both PCA and LDA suggested important roles for emergence and average growth rate in distinguishing the two groups (Tables S3 and S4, Supporting Information). However, confidence intervals overlap for most traits between the two cultivar groups, and difference among groups was not significant.

For most traits measured, weedy rice groups displayed a broad range in variation, exceeding or similar in magnitude to that of putative progenitor *O. sativa* populations (Table 1, Fig. 1b,c). However, we found evidence for the evolution of a novel phenotype in growth/life history of the SH weedy group compared with *indica* (Fig. 1b, Table 1). In particular, SH individuals have a faster average growth rate and flower significantly earlier than their putative *indica* progenitor (Table 1); these variables are also the greatest contributors to PC1 and PC2, respectively, and the ones with the greatest predictive value between groups in the LDA (Tables S3 and S4, Supporting Information). Confidence intervals for all other traits overlapped between SH and *indica*, and differences were not significant (Table 1).

There was considerable overlap in the 50% concentration ellipses between BHA1, BHA2 and their putative progenitor, *aus*, indicative of less growth trait divergence between these weed–progenitor pairs (Fig. 1c). Average and emergence growth rate were the greatest contributors to PC1 and PC2, respectively, but these traits displayed no significant differences among weed–progenitor groups (Table S4, Supporting Information; Table 1). However, BHA1 individuals were found to flower significantly later than *aus*, consistent with the large predictive value of this variable in LDA, and BHA2 individuals were significantly taller than *aus* (Table S3, Supporting Information; Table 1).

Patterns of growth between the major weedy rice groups were substantially divergent, with nearly all BHA and SH individuals falling on opposite sides of the first PC axis (Fig. 1d). The degree of morphological divergence between SH and BHA is more distinct (based on confidence ellipses) than that observed between their respective progenitor populations (Fig. 1a,d). The greatest contributors to the variation explained by PC1 and PC2 were average growth rate and flowering time, respectively (Table S4, Supporting Information). Flowering time was also identified by LDA as the variable with the largest predictive value (Table S3, Supporting Information). We observed that

Table 1 Means and standard deviations (in parenthesis) of growth traits measured by group

	Height at flowering (cm)	Emergence growth rate (cm/day)	Average growth rate (cm/day)	Days to flowering	Tiller number
Cultivated rice					
*indica**	75^{abc}† (23)	2.6^{a} (0.8)	0.67^{bc} (0.17)	113^{ab} (20)	6.5^{a} (3.8)
aus	80^{b} (13)	3.4^{a} (0.7)	0.81^{abc} (0.18)	100^{b} (9)	8.2^{a} (1.6)
japonica‡	76^{b} (15)	2.8^{a} (0.5)	0.76^{abc} (0.26)	109^{ab} (30)	5.8^{a} (3.7)
Weedy rice					
SH	69^{bc} (14)	2.7^{a} (0.7)	0.97^{a} (0.25)	73^{c} (11)	7.3^{a} (3.8)
BHA1	85^{b} (25)	3.1^{a} (0.7)	0.68^{b} (0.22)	126^{a} (16)	6.6^{a} (2.3)
BHA2	99^{a} (19)	3.2^{a} (0.9)	0.96^{ac} (0.36)	111^{ab} (26)	6.9^{a} (2.4)
BRH	59^{c} (11)	2.7^{a} (0.8)	0.73^{abc} (0.20)	84^{bc} (17)	9.9^{a} (4.7)

*Includes one cultivar with green revolution deletion (mean and standard deviation does not change with inclusion).
†Means with the same letter are not significantly different ($P < 0.05$), as determined by Mann–Whitney tests using the BH correction for multiple comparisons.
‡Includes *tropical japonica* from Asia (6) and the U.S. (3) and 1 *temperate japonica*.

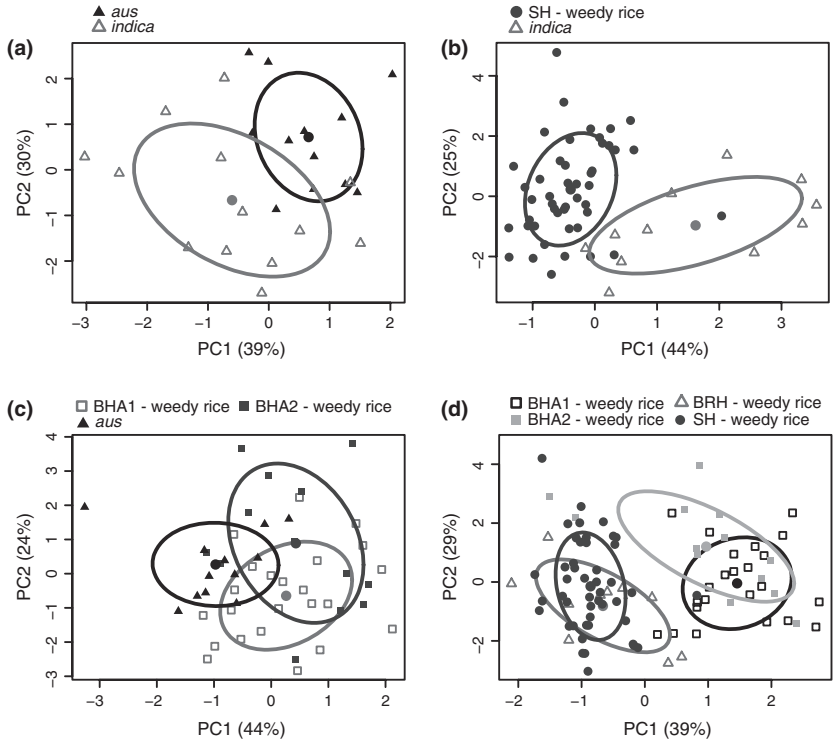

Fig. 1 Scatterplots of the first two principal components calculated from standardized morphological data. In each plot, data points are labelled based on genetically defined groups; the 50% confidence ellipses for each group are shown, and numbers in parentheses represent the amount of variation explained by each principal component. Four separate principal component analyses were carried out involving the following groups: (a) ancestral cultivated populations; (b) SH (strawhull) weedy rice and putative ancestral cultivated group, *indica*; (c) BHA1 and BHA2 (black hull awned) weedy rice groups, and putative ancestral cultivated group, *aus*; (d) all weedy groups.

BHA groups tended to be taller and have more tillers than SH, reflecting the trend seen between cultivated *aus* and *indica*, but these differences were not significant (Table 1). For flowering time, the trend is opposite that of progenitor groups, with the *indica*-derived SH flowering significantly earlier than *aus*-derived BHA groups. No other trait significantly differentiated the main weedy rice groups except BHA2 height compared with SH, and BHA1 average growth rate compared with SH. The BRH group, which is of putative SH–BHA hybrid origin, clustered with SH because of similar reduced heights and shorter time to flowering (Table 1, Fig. 1d).

Diversity in SD1 and flanking regions

We obtained sequence of sufficient quality for 139 individuals, with an alignment length of ~5400 bp for the *SD1* locus. For all of our analyses, we also added five sequences from GenBank (two *Oryza rufipogon*, and one *O. barthii*, *O. glaberrima* and *O. sativa indica*) that included most of the entire region sequenced in our alignment.

Diversity within groups was estimated for the *SD1* open reading frame (ORF), the 1545 bp promoter region upstream from the start codon (5′) and the 1299 bp following the stop codon (3′). For all *Oryza* groups except BHA2, the 5′ promoter region was more diverse than either the *SD1* ORF or the 3′ region (Table 2). Consistent with prior genetic surveys and the effects of a domestication bottleneck (e.g. Caicedo *et al*. 2007), we found that diversity in the entire *SD1* region was higher in *O. rufipogon* than for cultivated or weedy rice. While estimates of diversity for the *SD1* ORF did not differ substantially among cultivated groups and were similar to genome-wide averages estimated from 48 STS fragments used in Reagon *et al*. (2010), we found considerable variation in diversity estimates across weedy groups (Table 2). Notably, BHA1 diversity for the entire *SD1* region was considerably higher than genome-wide averages, and BHA1 was more diverse than its putative cultivated progenitor *aus* across this locus.

To assess patterns of diversity in the chromosomal region surrounding *SD1*, we examined six additional fragments spanning a region of ~2 Mb around *SD1* (Table S5, Supporting Information). Except for *tropical japonica* and BHA1, genetic diversity across most of the flanking loci was lower than observed within the *SD1*

Table 2 Values of θ_π, θ_w and Tajima's D for silent sites per 1000 bases at the *SD1* locus, and mean of 48 STS loci*

		Cultivated *O. sativa*			U.S. weedy rice			
	Oryza rufipogon	*indica*	*aus*	*tropical japonica*	SH	BHA1	BHA2	BRH
Gene†								
5′								
θ_π	9.63	4.51	3.81	5.59	0.72	7.50	0	6.19
θ_w	13.28	5.99	5.01	4.24	0.77	4.65	0	6.62
D	−0.96	−1.11	−1.48	1.26	−0.18	2.54	NA	−0.67
SD1-ORF								
θ_π	4.28	1.46	1.14	1.59	0.56	2.09	0.11	2.07
θ_w	6.74	1.89	1.63	1.45	0.48	1.21	0.17	2.23
D	−1.39	−1.02	−1.59	0.36	0.50	2.69	−1.05	−0.70
3′								
θ_π	4.46	0.48	0.52	1.17	0	2.39	0	0.91
θ_w	4.44	0.92	0.74	0.81	0	1.37	0	0.99
D	0.02	−1.60	−1.24	1.24	NA‡	2.45	NA	−0.71
Mean – 48 STS loci								
θ_π	6.35	2.18	1.57	1.34	0.69	0.83	0.66	0.31
θ_w	7.79	2.22	1.40	1.69	0.64	0.93	0.69	0.35
D	−0.61	0.04	0.14	−0.73	−0.40	−0.17	0.02	0.01

*Values for 48 STS loci from Reagon *et al.* 2010;.
†Statistics were calculated separately for the ~1500 bp upstream from the start codon, all exons and introns for *SD1* and ~1000 bp downstream from the stop codon (3′). Values of 0 indicate no genetic diversity.
‡NA not applicable. Tajima's D is undefined for monomorphic loci.

ORF and 5′ promoter region, though one fragment, sd13_002, 161 kbp downstream of *SD1*, was an outlier (Table S3, Supporting Information).

We tested whether patterns of *SD1* polymorphism in the two most commonly occurring weedy groups in the U.S. (SH and BHA1) are consistent with neutral expectations using Tajima's D. Tajima's D in the BHA1 group was highly positive and significantly different from zero when calculated for the *SD1* ORF alone and using the entire alignment, indicating the presence of moderate-frequency alleles (Table 2). Using the 48 STS fragments (Reagon *et al.* 2010) as a reference set, we also used the mlHKA test (Wright & Charlesworth 2004) to determine whether *SD1* patterns of diversity and divergence are consistent with neutrality in the two main weed groups. For the BHA1 group, the model that assumed selection at *SD1* performed significantly better than the neutral model (Table 3). The maximum-likelihood estimate for the selection parameter *k* was 11.6 for the BHA1 group, indicating that there is a large increase in polymorphism over neutral expectations at *SD1*.

Origins of SD1 alleles: introgression and founding effects

In the entire ~5400-bp *SD1* alignment, a total of 16 haplotypes were observed in weedy and cultivated *O. sativa* (Fig. 2). Haplotypes were assigned numbers

Table 3 Likelihood ratio tests of neutrality of silent polymorphisms for *SD1* and major weedy groups

Model	ln L*	Comparison	Likelihood ratio statistic†	*P*-value‡	*k*§
BHA1 – all neutral	−151.3				1
SH – all neutral	−128.7				1
BHA1 – *SD1*	−137.8	A vs. C	26.9	2.14×10^{-7}	11.6
SH – *SD1*	−128.3	B vs. D	0.89	0.34	0.94

*Mean from three independent runs with different random starting seeds and 500 000 iterations.
†Likelihood ratio statistic calculated using 2 ($\ln L_{selected} - \ln L_{all\ neutral}$).
‡*P*-value using χ^2 approximation with degrees of freedom = 1 for both comparisons.
§*k* measures the degree of loss or increase in diversity owing to selection.

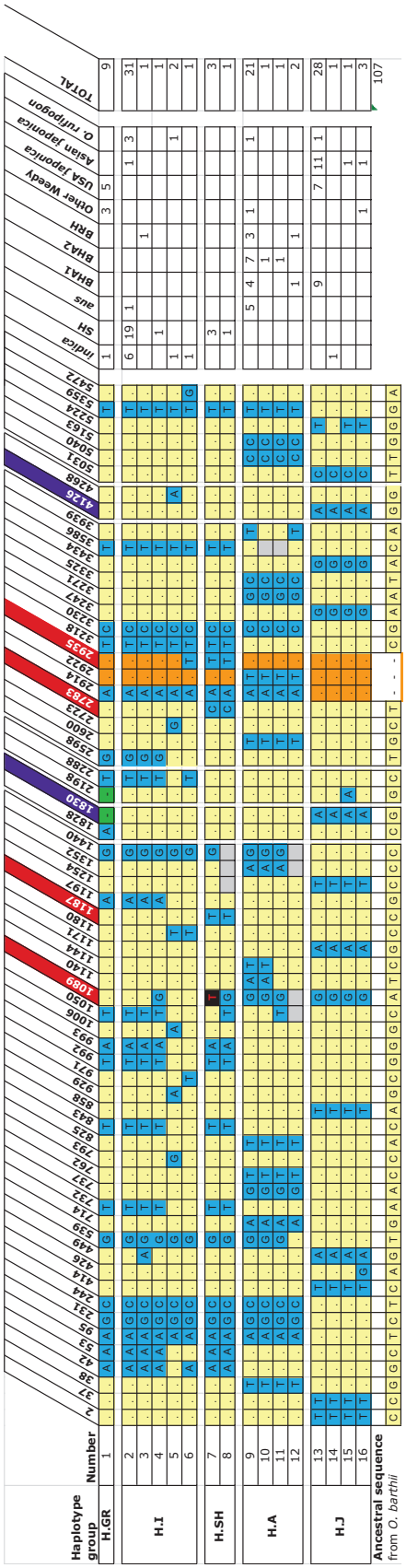

Fig. 2 Graphical view of *SD1* haplotypes. Haplotypes were constructed using the entire *SD1* open reading frame along with 1545 bp of promoter and 1299 bp downstream sequence. All SNPs in cultivated and weedy groups were used to construct haplotypes; low-frequency SNPs (<0.05) found only in wild *Oryza* were excluded. The only indel included in haplotype designation was the 'green revolution' functional deletion associated with semi-dwarfism in cultivated rice (shown in green). Haplotype groups and numbers are shown on the left. The location of SNPs from start of the alignment is given on top. Ancestral SNPs (based on *O. barthii* – last row) are marked in yellow. Three sites not found in *O. barthii* are shown with dashes; for these, yellow corresponds to the nucleotide found in Nipponbare. The four SNPs found in SH, but rare or absent in *indica*, are highlighted in red. The locations of the two nonsynonymous sites differentiating H.J from other haplotypes are highlighted in purple. The number of individuals carrying each haplotype within *Oryza* groups is shown in the table on the right.

and sorted into haplotype groups designated by letter combinations based on overall similarity of SNPs (e.g. H.I-2 indicates haplotype 2 in the H.I haplogroup).

Patterns of sequence diversity in *SD1* and other genomic regions are broadly consistent with expectations based on genome-wide surveys (Vaughan *et al*. 2001; Londo & Schaal 2007; Reagon *et al*. 2010) (Figs 2 and 3; Fig. S1, Supporting information). Similar to other loci, SH and *indica* cultivars mostly group together, as do BHA and *aus*. BRH individuals clustered with either BHA or SH as expected owing to their putative hybrid origin. We found that most weeds tended to share the same most frequent haplotype (MFH) with their cultivated progenitors (e.g. H.I-2 in SH and *indica*; Fig. 2). Most haplotypes that were unique to U.S. weedy rice were singletons and differed by only one or two SNPs from the MFH of their respective cultivated progenitor. One haplotype, H.SH-7 in three SH individuals, differed from the MFH of *indica* and SH by four SNPs; however, one SNP occurs in an *indica* sequence obtained from GenBank (AB213460), and overall similarity with the H.I (*indica*) haplogroup suggests that H.SH-7 is derived from *indica* (Fig. 2).

We found two exceptions to our broad expectations based on weed ancestry. Most notably, nine of fifteen BHA1 accessions clustered with *japonica* cultivars at the *SD1* locus, rather than their putative *aus* progenitors (Fig. S1, Supporting Information; Fig. 2). All of these individuals have the same haplotype (H.J-13) that is otherwise only found in cultivated *japonica* groups and two *O. rufipogon* accessions (Fig. 2). Four of these weedy accessions are from a single county in Arkansas, but the other five were collected in diverse locations within the southern rice-growing region (Table S2, Supporting Information). Presence of this allele accounts for

the high levels of *SD1* diversity and Tajima's D seen for BHA1. Other than two STS fragments within 2 Mb of *SD1*, no other locus was found in our reference data set of 48 STS loci where BHA1 shares a haplotype with *japonica* that is absent in *aus* (Reagon *et al.* 2010). Additionally, the BHA1 individuals with the H.J-13 haplotype also have the same cytoplasm cytotype common in *japonica* cultivars, but absent in *aus* accessions (Reagon *et al.* 2010). These data suggest that introgression between *japonica* and BHA weeds has occurred, most likely in the U.S., with *japonica* cultivars as the maternal parent.

The pattern of haplotype sharing observed in eight fragments flanking the *SD1* locus also supports introgression of *japonica* alleles as the most likely explanation for the H.J-13 haplotype found in BHA1 (Fig. 3). All BHA1 individuals with H.J-13 (Fig. 2) have the same extended haplotype unique to *japonica* at the five loci flanking *SD1* in the 5′ direction (Fig. 3), spanning ~2 Mbp and exceeding the extent of typical linkage disequilibrium in rice (Mather *et al.* 2007). This extended haplotype in the 5′ direction is not found in the single *indica* (H.J-14) in the H.J group. Although H.J-13 also occurred in one *O. rufipogon* accession, it is unlikely that its presence in BHA1 is because of the retention of shared ancestral polymorphism; this *O. rufipogon* accession (OR02; Table S1, Supporting Information) has been found to have a very similar genetic background as *temperate japonica* (Reagon *et al.* 2010) and likely acquired the extended 5′ haplotype itself through recent hybridization. Moreover, shared ancestral polymorphism with *O. rufipogon* would also be expected to occur in the *aus* group ancestral to BHA1, where the haplotype is not observed.

Another exception to the general trend of *SD1* weed clustering with putative ancestral haplotypes was the observation of the 383-bp deletion found in most modern semi-dwarf cultivars (i.e. the 'green revolution' deletion; H.GR-1 haplotype) in three of four weedy rice accessions previously identified as possible hybrids based on genome-wide patterns of diversity (Reagon *et al.* 2010) (Fig. 2; Table S1, Supporting Information). These weeds (MX) represent hybrids between *tropical japonica* and SH or BHA (Table S2, Supporting Information), indicating that the H.GR-1 haplotype has introgressed into the genetic backgrounds of both major weedy rice groups. Semi-dwarf U.S. *tropical japonica* cultivars have been grown for ~30 years (Dilday 1990), providing opportunities for local hybridizaton. Semi-dwarf cultivars have only been widely grown in Asia since the late 1960s (Khush 1997), making it unlikely that the deletion occurred in the founding populations of U.S. weedy rice. The size of the introgressed fragment in weedy individuals could not be determined because haplotypes at flanking fragments in these accessions resemble the *indica* genomic background in which the deletion arose (Fig. 3). As expected based on pedigree data, we also found the H.GR-1 *SD1* haplotype in

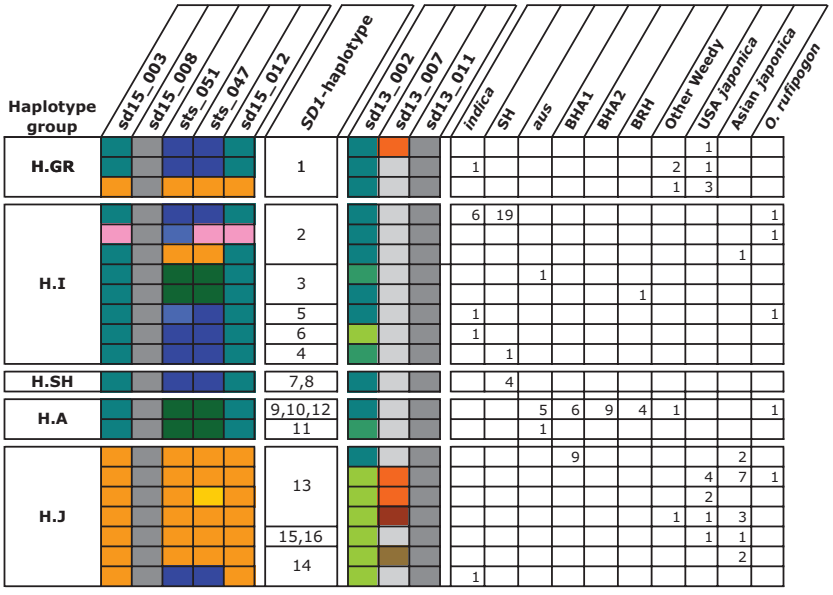

Fig. 3 Graphical view of extended haplotypes based on eight fragments flanking the *SD1* locus. Haplotype groups are arranged as in Fig. 2, and the central column (labelled *SD1* haplotypes) gives the haplotype number. Fragments include six fragments sequenced for this study, and two sequence-tagged site fragments from Reagon *et al.* (2010) found to be within 2 Mb of *SD1*. The flanking fragments are shown in the top row in the same order that they occur along chromosome 1 (5′–3′). Boxes with the same colour within a fragment (column) have the same haplotype. The table on the right gives the frequency of each haplotype within *Oryza* groups.

one *indica* and five U.S. *tropical japonica* semi-dwarf cultivars (Fig. 2; Table S1, Supporting Information).

Associations between SD1 haplotype and growth trait variation

To determine whether polymorphism at *SD1* is associated with observed phenotypic diversity among U.S. weedy rice groups and their progenitors, we compared patterns of growth across *SD1* haplotype groups (Table 4). Our phenotypic results suggest that the introgression of the *japonica* segment is associated with morphological diversity within BHA1. The nine BHA1 individuals with the H.J-13 haplotype were morphologically more similar to *japonica* with H.J-13, than to other BHA1 and BHA2, with whom they share a similar genomic background. Specifically, the H.J-13 BHA1 individuals were significantly shorter than BHA accessions in the H.A haplogroup (Table 4). The H.J BHA1 individuals also tended to be similar to *japonica* individuals in having slower emergence and average growth rates and earlier flowering than H.A BHA1 individuals, but these differences were not significant. We observed that the H.J haplogroup differs from other haplogroups at two nonsynonymous sites in *SD1* (Fig. 2).

No other clear correlation between *SD1* haplotype and phenotypic variation was observed. For example, BHA1 and BHA2 individuals flower later and are taller than *aus* with the same H.A haplogroup. Similarly, we did not find evidence for polymorphism in *SD1* affecting phenotypic divergence between SH and their putative *indica* progenitor or between SH individuals with different haplotypes (Table 4). No nonsynonymous SNPs were observed between SH and *indica* haplotypes (Fig. 2), indicating that the large differences in growth characteristics between these groups (Table 1) are not because of mutations in the *SD1*coding region.

Discussion

Phenotypic divergence for some growth traits characterizes U.S. weedy rice evolution

Traits increasing competitive ability – that is, the ability to deplete or efficiently use consumable resources (Radosevich *et al.* 1997; Dlugosch & Parker 2008) – are frequently observed in introduced agricultural weeds. Many traits that influence competitive outcomes are associated with plant growth (Falster & Westoby 2003); however, it is not known to what extent these traits differ between weeds and their nonweedy ancestors. U.S. weedy rice has been shown to be most closely related to, and likely descended from, domesticated *aus* and *indica* ancestors (Londo & Schaal 2007; Reagon *et al.* 2010). Although previous studies have documented considerable morphological differences between U.S.

Table 4 Means and standard deviations for growth traits for *Oryza* groups across *SD1* haplotype groups

	Height at flowering (cm)	Emergence growth rate (cm/day)	Average growth rate (cm/day)	Days to first flower	Tiller number*
H.A†					
BHA1	97^{ab}‡ (25)	3.5^{a} (0.6)	0.77^{ab} (0.25)	128^{a} (17)	6.3 (2.9)
BHA2	99^{a} (19)	3.2^{a} (0.9)	0.96^{ab} (0.36)	111^{abce} (26)	6.9 (2.9)
BRH§	57^{d} (11)	2.8^{a} (0.8)	0.74^{ab} (0.20)	79^{cd} (17)	10.8 (4.7)
aus¶	80^{abc} (14)	3.4^{a} (0.7)	0.81^{ab} (0.18)	100^{c} (9)	8.2 (1.6)
H.J					
BHA1	74^{cd} (20)	2.8^{a} (0.6)	0.60^{a} (0.15)	123^{ab} (16)	6. (1.8)
*japonica***	72^{c} (15)	2.8^{a} (0.4)	0.71^{ab} (0.26)	110^{abc} (30)	5.6 (3.7)
H.I					
SH	70^{cd} (23)	2.7^{a} (0.7)	0.97^{b} (0.17)	74^{d} (11)	7.6 (4.2)
indica††	77^{acd} (15)	2.7^{a} (0.7)	0.65^{ab} (0.27)	118^{abc} (20)	5.6 (3.8)
H.SH					
SH	67^{cd} (9)	2.5^{a} (0.8)	0.95^{ab} (0.18)	71^{de} (9)	6.4 (2.4)

*Tiller number Kruskal–Wallis test was not significant at $P < 0.05$.

†'H' designations refer to *SD1* haplotype groups as in Fig. 2.

‡Means with the same letter are not significantly different ($P < 0.05$) as determined by Mann–Whitney tests using the BH correction for multiple comparisons.

§Does not include BRH accession with H.I haplotype (see Table S1).

¶Does not include *aus* accession with H.I haplotype (see Table S1).

**Does not include *japonica* accessions with H.GR or H.I haplotypes (see Table S1).

††Does not include *indica* accessions with H.GR or H.J haplotypes (see Table S1).

weedy rice morphotypes (Gealy *et al.* 2006; Delouche *et al.* 2007; Shivrain *et al.* 2010), our study is the first to compare weedy groups of known genetic relatedness to each other and to their putative progenitors.

We find that weedy rice populations derived from different cultivated groups also have divergent patterns of growth (Fig. 1). Under our conditions, SH weeds flower significantly earlier than BHA groups (Table 1). SH weeds also tend to grow faster and remain shorter than at least one of our BHA populations (Table 1). Previous studies describing weedy rice morphology in the field also report that blackhull awned morphotypes are typically taller and flower later than strawhull types (Shivrain *et al.* 2010). This suggests that inferences from our data are relatively robust, and not because of effects of growth chamber conditions.

The observed variation in growth traits indicates that weedy rice groups in the U.S. have not converged to a single growth strategy. While it is hard to predict the exact fitness benefits of each of the observed alternative weedy rice growth traits, the temporal (e.g. historical changes in agricultural practices) and spatial heterogeneity in the U.S. rice agroecosystem (Burgos *et al.* 2008; Shivrain *et al.* 2010) may play a role in supporting these multiple strategies.

While differences in mean trait values were not always large, shifts in the distribution between weeds and their cultivated progenitor groups were also evident for some traits. Some shifts were in the same direction; both BHA and SH groups are slightly (though not significantly) taller than their progenitors. Other shifts are in the opposite direction: SH flowers significantly earlier than *indica*, whereas BHA groups flower significantly later than *aus* (Table 1). And some traits showed no shifts (emergence growth rate, tiller number), or only shifted in one weed–progenitor pair (SH grows significantly faster than *indica*) (Table 1). The greatest divergence between both weedy rice strains and their putative ancestral populations was for flowering time, suggesting either that founding individuals were highly divergent for this trait or that flowering time is more likely to come under selection in weedy plants. If selection has shaped flowering time in weedy rice, the selection is likely to be divergent among weedy groups, given the significant differences in flowering time between them.

The majority of weeds species investigated to date have found reduced levels of genetic diversity relative to their ancestors, consistent with the effects of strong bottlenecks (Dlugosch & Parker 2008), which may constrain the ability to evolve novel phenotypes. We did not find evidence, however, that the known demographic history of U.S. weedy rice has resulted in greater trait homogeneity within BHA groups (Fig. 1). These results indicate that considerable morphological variation can arise and/or persist in populations with little genetic diversity, adding to a growing list of invasive or weedy species in which this has been observed (Clements *et al.* 2004; Dlugosch & Parker 2008). In contrast, there does appear to be some correlation between low levels of genetic diversity and reductions in phenotypic diversity in SH strains. Clustering of SH individuals was much tighter, and, for most traits, SH had smaller confidence intervals than other weedy groups (Fig. 1). This reduction is consistent with prior simulations indicating that SH experienced a more intense bottleneck than BHA1 (Reagon *et al.* 2010). However, the overlap in confidence intervals between SH and *indica* suggests that the founding bottleneck had less dramatic effect on phenotypic variation than on sequence diversity.

Introgression has contributed to growth trait divergence in weedy rice subgroups

Inferences from prior studies investigating the evolution of U.S. weedy rice make accounting for the occurrence of a morphologically diverse weedy *Oryza* difficult. No *Oryza* are native to North America, and all current weedy and cultivated populations are derived from recent (<400 years) introductions. Given this short time period, it is unlikely that novel mutations account for much of the phenotypic diversity observed within and among weedy populations, or between weedy rice and their progenitors. Our study joins an increasing number of studies that have found hybridization to be an important factor in promoting evolutionary divergence, particularly in recently founded weedy populations (e.g. Campbell *et al.* 2006; Kim *et al.* 2008; Gaskin & Kazmer 2009; Arnaud *et al.* 2010; Xia *et al.* 2011).

Owing to its drastic effect on plant height and its position in the GA pathway, *SD1* is a prime candidate for genes influencing weedy plant growth. We found some association of *SD1* haplotype with observed phenotypic diversity in *Oryza* groups. In particular, BHA1 individuals carrying an *SD1* haplotype common in *japonica* were shorter than BHA1 with *aus*-like haplotypes (Table 4).

Our sequencing results suggest that BHA1 individuals have acquired the H.J-13 *SD1* allele through introgression with the local *tropical japonica* crop (Figs 2 and 3). Based on identical haplotypes across the entire 2 Mb and the sharing of cytotypes, it is likely that the BHA1 H.J haplotypes are the result of a single hybridization event. Interestingly, although the signature of hybridization is not detectable genome-wide, four of the BHA1 accessions with H.J haplotypes had been identi-

fied as potential hybrids based on morphology (D. Gealy, personal communication). Owing to the extent of the introgressed region, we cannot determine whether patterns of growth in introgressed BHA1 individuals are because of *SD1* or another linked locus. Interestingly, however, the nonsynonymous SNPs characteristic of H.J haplotypes have recently been shown to lead to lower GA biosynthetic activity (Asano *et al.* 2011), which supports the contribution of this introgressed allele to phenotypic variation in BHA1.

Introgression is often discussed as a possible contributor to morphological diversity in U.S. weedy rice, but has not been documented to the degree shown here. Historically, a majority of studies have attributed phenotypic diversity and the success of weedy rice to hybridization with *Oryza rufipogon* (e.g. Delouche *et al.* 2007; Londo & Schaal 2007; Gealy *et al.* 2010). However, genome-wide assessment of nucleotide diversity (Reagon *et al.* 2010) and sequence data from candidate genes (Gross *et al.* 2010; Thurber *et al.* 2010) have not found evidence supporting hybridization between cultivated and wild rice as influencing phenotypic evolution in weedy rice. Patterns of shared polymorphism between weeds and *O. rufipogon* reported in most studies are more readily attributed to shared ancestral polymorphism than recent hybridization.

The emphasis historically given to hybridization also contrasts with recent genomic evidence showing almost nonexistent contributions of *japonica* cultivars to U.S. weedy rice (Vaughan *et al.* 2001; Londo & Schaal 2007; Gealy *et al.* 2010; Reagon *et al.* 2010). One previous study (Londo & Schaal 2007) has suggested *japonica* introgression at the ψ-Vatpase pseudogene (chromosome 5) in several of the same BHA1 accessions as this study, but other documented cases of introgression are rare. The low rates of hybridization suggested by genomic studies are consistent with the existence of significant crossing barriers between weedy groups and the local crop. However, our *SD1* data show that even rare hybridization events can result in stabilized introgressed lines. Notably, in contrast to expectations based on the fact that cultivated rice is harvested primarily for food, and hybrid seed from a crop mother plant is less likely to survive to reproduction, our data suggest that the maternal parent of the hybridization event was a cultivated *japonica*.

The HKA test and Tajima's D both suggested departures from neutral expectations in *SD1* in BHA1. Although the origin of one set of *SD1* alleles is through hybridization and introgression in BHA1, its maintenance at moderate frequency in the population is potentially indicative of adaptive introgression. Maintenance of several different moderate-frequency alleles suggests balancing or frequency-dependent selection, or selection mediated by habitat differences. Such signatures often characterize genes involved in defensive biotic interactions (e.g. Moeller & Tiffin 2008). The fitness benefit for a particular combination of growth characteristics will depend on competitive interactions between neighbouring plants; therefore, it is not unreasonable that genes affecting plant growth may have been targets of balancing selection during weedy rice evolution, regardless of their origin.

Conclusions

The ecological success of weedy species is probably influenced by growth traits. Our results show that the evolution of U.S. weedy rice has been marked by divergence and diversification of only certain growth traits. The main weedy rice groups have diverged from their putative cultivated ancestors primarily in flowering time. Likewise, weedy groups differ from each other in flowering time, but also for traits such as height and growth rate, suggesting that multiple life histories can lead to weed success. In at least one case, diversification of a weedy rice growth trait may have occurred through hybridization and introgression with the local crop. Our results also suggest that genetic diversity maintained in cultivated *O. sativa* is sufficient for the evolution of a weedy life history. Further studies combining neutral and putative candidate weedy genes may further elucidate the various evolutionary processes that have shaped morphological diversity in U.S. weedy rice.

Acknowledgements

We are grateful to D. Gealy for providing weedy rice accessions used in this study. This project was funded by a grant from the US NSF Plant Genome Research Program (DBI-0638820) to ALC, KMO and YJ. YJ is an employee of the USDA, which is an equal opportunity provider and employer.

References

Arnaud JF, Fenart S, Cordellier M, Cuguen J (2010) Populations of weedy crop-wild hybrid beets show contrasting variation in mating system and population genetic structure. *Evolutionary Applications*, **3**, 305–318.

Asano K, Yamasaki M, Takuno S *et al.* (2011) Artificial selection for a green revolution gene during japonica rice domestication. *Proceedings of the National Academy of Sciences*, **108**, 1034–1039.

Ashikari M, Sasaki A, Ueguchi-Tanaka M *et al.* (2002) Loss-of-function of a rice gibberellin biosynthetic gene, GA20 oxidase (GA20ox-2), led to the rice 'green revolution'. *Breeding Science*, **52**, 143–150.

Baker HG (1974) The evolution of weeds. *Annual Review of Ecology and Systematics*, **5**, 1–24.

Benjamini Y, Hochberg Y (1995) Controlling the false discovery rate: a practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society. Series B*, **57**, 289–300.

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

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

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

Campbell LG, Snow AA, Ridley CE (2006) Weed evolution after crop gene introgression: greater survival and fecundity of hybrids in a new environment. *Ecology Letters*, **9**, 1198–1209.

Chessel D, Dufour A-B, Thioulouse J (2004) The ade4 package – I – One-table methods. *R News*, **4**, 5–10.

Clements DR, DiTommaso A, Jordan N *et al.* (2004) Adaptability of plants invading North American cropland. *Agriculture, Ecosystems & Environment*, **104**, 379–398.

Delouche JC, Burgos NR, Gealy DR *et al.* (2007) Weedy rices – origin, biology, ecology and control. *FAO Plant Production and Protection Paper*, **188**, 1–144.

Dilday RH (1990) Contribution of ancestral lines in the development of new cultivars of rice. *Crop Science*, **30**, 905–911.

Dlugosch KM, Parker IM (2008) Founding events in species invasions: genetic variation, adaptive evolution, and the role of multiple introductions. *Molecular Ecology*, **17**, 431–449.

Ellstrand NC, Prentice HC, Hancock JF (1999) Gene flow and introgression from domesticated plants into their wild relatives. *Annual Review of Ecology and Systematics*, **30**, 539–563.

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, Gealy DR, Talbert RE (2002) Growth response of rice (*Oryza sativa*) and red rice (O-sativa) in a replacement series study. *Weed Technology*, **16**, 401–406.

Estorninos LE, 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.

Falster DS, Westoby M (2003) Plant height and evolutionary games. *Trends in Ecology & Evolution*, **18**, 337–343.

Garris AJ, Tai TH, Coburn J, Kresovich S, McCouch S (2005) Genetic structure and diversity in *Oryza sativa* L. *Genetics*, **169**, 1631–1638.

Gaskin JF, Kazmer DJ (2009) Introgression between invasive saltcedars (Tamarix chinensis and T-ramosissima) in the USA. *Biological Invasions*, **11**, 1121–1130.

Gealy DR, Tai TH, Sneller CH (2002) Identification of red rice, rice, and hybrid populations using microsatellite markers. *Weed Science*, **50**, 333–339.

Gealy DR, Yan WG, Rutger JN (2006) Red rice (*Oryza sativa*) plant types affect growth, coloration, and flowering characteristics of first- and second-generation crosses with rice. *Weed Technology*, **20**, 839–852.

Gealy DR, Agrama HA, Eizenga GC (2010) Exploring genetic and spatial structure of U.S. weedy red rice (*Oryza sativa* L.) in relation to rice relatives worldwide. *Weed Science*, **57**, 627–643.

Gross BL, Reagon M, Hsu S-C *et al.* (2010) Seeing red: the origin of grain pigmentation in US weedy rice. *Molecular Ecology*, **19**, 3380–3393.

Hedden P (2003) The genes of the green revolution. *Trends in Genetics*, **19**, 5–9.

Hooley R (1994) Gibberellin: perception, transduction and responses. *Plant Molecular Biology*, **26**, 1529–1555.

Kawasaki A, Imai K, Ushiki J, Ishii T, Ishikawa R (2009) Molecular constitution of weedy rice (*Oryza sativa* L.) found in Okayama prefecture, Japan. *Breeding Science*, **59**, 229–236.

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

Kim M, Cui M-L, Cubas P *et al.* (2008) Regulatory genes control a key morphological and ecological trait transferred between species. *Science*, **322**, 1116–1119.

Lexer C, Widmer A (2008) The genic view of plant speciation: recent progress and emerging questions. *Philosophical Transactions of the Royal Society B: Biological Sciences*, **363**, 3023.

Librado P, Rozas J (2009) DnaSP v5: a software for comprehensive analysis of DNA polymorphism data. *Bioinformatics* **25**, 1451–1452. DOI: 1410.1093/bioinformatics/btp1187.

Londo JP, Schaal BA (2007) Origins and population genetics of weedy red rice in the USA. *Molecular Ecology*, **16**, 4523–4535.

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

Mather KA, Caicedo AL, Polato NR *et al.* (2007) The extent of linkage disequilibrium in rice (*Oryza sativa* L.). *Genetics*, **177**, 2223–2232.

Moeller DA, Tiffin P (2008) Geographic variation in adaptation at the molecular level: a case study of plant immunity genes. *Evolution*, **62**, 3069–3081.

Monna L, Kitazawa N, Yoshino R *et al.* (2002) Positional cloning of rice semidwarfing gene, sd-1: rice ''green revolution gene'' encodes a mutant enzyme involved in gibberellin synthesis. *DNA Research*, **9**, 11.

Nagano H, Onishi K, Ogasawara M, Horiuchi Y, Sano Y (2005) Genealogy of the ''Green Revolution'' gene in rice. *Genes & Genetic Systems*, **80**, 351–356.

Olsen KM, Caicedo AL, Jia Y (2007) Evolutionary genomics of weedy rice in the U.S.A. *Journal of Integrative Plant Biology*, **49**, 811–816.

Pantone DJ, Baker JB (1991) Weed-crop competition models and response-surface analysis of red rice competition in cultivated rice – a review. *Crop Science*, **31**, 1105–1110.

R Core Development Team (2007) *R: A Language and Environment for Statistical Computing R Foundation for Statistical Computing*, Vienna, Austria.

Radosevich S, Holt J, Ghersa C (1997) *Weed Ecology: Implications for Management*, 2nd edn. John Wiley & Sons, Inc., New York.

Reagon M, Thurber CS, Gross BL *et al.* (2010) Genomic patterns of nucleotide diversity in divergent populations of invasive U.S. weedy rice. *BMC Evolutionary Biology*, **10**, 180.

Reid JB (1993) Plant hormone mutants. *Journal of Plant Growth Regulation*, **12**, 207–226.

Richards DE, King KE, Ait-ali T, Harberd NP (2001) How gibberellin regulates plant growth and development: a molecular genetic analysis of Gibberellin signaling. *Annual Review of Plant Biology*, **52**, 67–88.

Ross JJ, Murfet IC, Reid JB (1997) Gibberellin mutants. *Physiologia Plantarum*, **100**, 550–560.

Rozen S, Skaletsky HJ (2000) Primer3 on the WWW for general users and for biologist programmers. In: *Bioinformatics Methods and Protocols: Methods in Molecular Biology* (eds Krawetz S, Misener S). pp. 365–386, Humana Press, Totowa, NJ.

Sahli HF, Conner JK, Shaw FH, Howe S, Lale A (2008) Adaptive differentiation of quantitative traits in the globally distributed weed, wild radish (*Raphanus raphanistrum*). *Genetics*, **180**, 945–955.

Sasaki A, Ashikari M, Ueguchi-Tanaka M *et al.* (2002) Green revolution: a mutant gibberellin-synthesis gene in rice. *Nature*, **416**, 701–702.

Scascitelli M, Whitney KD, Randell RA *et al.* (2010) Genome scan of hybridizing sunflowers from Texas (*Helianthus annuus* and *H. debilis*) reveals asymmetric patterns of introgression and small islands of genomic differentiation. *Molecular Ecology*, **19**, 521–541.

Shivrain VK, Burgos NR, Scott RC *et al.* (2010) Diversity of weedy red rice (*Oryza sativa* L.) in Arkansas, USA in relation to weed management. *Crop Protection*, **29**, 721–730.

Spielmeyer W, Ellis MH, Chandler PM (2002) Semidwarf (sd-1), ''green revolution'' rice, contains a defective gibberellin 20-oxidase gene. *Proceedings of the National Academy of Sciences of the United States of America*, **99**, 9043.

Tamura K, Dudley J, Nei M, Kumar S (2007) MEGA4: molecular evolutionary genetics analysis (MEGA) software version 4.0. *Molecular Biology and Evolution*, **24**, 1596–1599.

Thurber CS, Reagon M, Olsen KM, Jia Y, Caicedo AL (2010) Molecular evolution of shattering loci in U.S. weedy rice. *Molecular Ecology*, **19**, 3271–3284.

Vaughan LK, Ottis BV, Prazak-Havey AM *et al.* (2001) Is all red rice found in commercial rice really *Oryza sativa*? *Weed Science*, **49**, 468–476.

Whitney KD, Randell RA, Rieseberg LH (2010) Adaptive introgression of abiotic tolerance traits in the sunflower *Helianthus annuus*. *New Phytologist*, **187**, 230–239.

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

Xia HB, Wang W, Xia H, Zhao W, Lu BR (2011) Conspecific crop-weed introgression influences evolution of weedy rice (*Oryza sativa* f. spontanea) across a geographical range. *PLoS ONE*, **6**, e16189.

The authors' research interests center on plant evolutionary genetics. This work represents part of a broad effort to understand the evolutionary origin of weedy rice in the U.S., and, more generally, how weediness evolves from within crop-wild species complexes.

Data accessibility

DNA sequences: GenBank accessions JN541407–JN542382.

Phenotype and genotype data uploaded as online supplemental material.

Supporting information

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

Table S1 Information of plant material used in the study

Table S2 Primers used in the study

Table S3 Structure coefficients and statistic of linear discriminant analyses

Table S4 Trait loadings on the first two principal components in each PCA

Table S5 Diversity statistics of *SD1* flanking regions

Fig. S1 Neighbor Joining tree of *SD1* haplotypes. Branches are colored by genetic subpopulation (Reagon *et al.* 2010), and the tips are labeled as in Supporting Information Table S2, which provides accession information. Numbers on branches correspond to bootstraps from 1000 replicates.

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