Received: 2 May 2023 | Revised: 2 September 2023 | Accepted: 6 October 2023

DOI: 10.1111/mec.17172

ORIGINAL ARTICLE

MOLECULAR ECOLOGY WILEY

A derived weedy rice × ancestral cultivar cross identifies evolutionarily relevant weediness QTLs

Xiang Li[1] | Shulin Zhang[2] | Daniel Lowey[1] | Carter Hissam[1] | Josh Clevenger[3] | Sherin Perera[1] | Yulin Jia[4] | Ana L. Caicedo[1]

[1]Plant Biology Graduate Program and Department of Biology, University of Massachusetts Amherst, Amherst, Massachusetts, USA

[2]College of Biology and Food Engineering, Innovation and Practice Base for Postdoctors, Anyang Institute of Technology, Anyang, China

[3]HudsonAlpha Institute of Biotechnology, Huntsville, Alabama, USA

[4]United States Department of Agriculture-Agricultural Research Service, Dale Bumpers National Rice Research Center, Stuttgart, Arkansas, USA

Correspondence
Ana L. Caicedo, Plant Biology Graduate Program and Department of Biology, University of Massachusetts Amherst, Amherst 01003, MA, USA.
Email: caicedo@umass.edu

Funding information
National Science Foundation, Grant/Award Number: IOS-1947609

Handling Editor: Richard J Abbott

Abstract

Weedy rice (*Oryza* spp.) is a weedy relative of the cultivated rice that competes with the crop and causes significant production loss. The BHA (blackhull awned) US weedy rice group has evolved from *aus* cultivated rice and differs from its ancestors in several important weediness traits, including flowering time, plant height and seed shattering. Prior attempts to determine the genetic basis of weediness traits in plants using linkage mapping approaches have not often considered weed origins. However, the timing of divergence between crossed parents can affect the detection of quantitative trait loci (QTL) relevant to the evolution of weediness. Here, we used a QTL-seq approach that combines bulked segregant analysis and high-throughput whole genome resequencing to map the three important weediness traits in an F_2 population derived from a cross between BHA weedy rice with an ancestral *aus* cultivar. We compared these QTLs with those previously detected in a cross of BHA with a more distantly related crop, *indica*. We identified multiple QTLs that overlapped with regions under selection during the evolution of weedy BHA rice and some candidate genes possibly underlying the evolution weediness traits in BHA. We showed that QTLs detected with ancestor–descendant crosses are more likely to be involved in the evolution of weediness traits than those detected from crosses of more diverged taxa.

KEYWORDS
evolutionarily relevant QTL, flowering time, plant height, QTL mapping, seed shattering, weedy rice

1 | INTRODUCTION

Agricultural and environmental weeds cause significant ecological, economic and social problems in natural and managed ecosystems (Bagavathiannan et al., 2019). As one of the principal causes of crop loss, agricultural weeds lead to around 30% reduction annually in crop production worldwide (Oerke, 2006). The ability of weeds to infest and persist in crop fields—that is, their adaptability—is the major reason for their impact on agriculture and for their huge cost to society. Understanding how weedy plants arise and evolve can provide us with new insights into plant adaptation and help us devise management tactics to prevent their effect on crop production systems.

Cultivated rice (*Oryza sativa*) is one of the world's most important crops, and half of the world population consumes it on a daily basis (Bin Rahman & Zhang, 2022). Weedy rice (*Oryza* spp.), a weedy type of rice, competitively invades cultivated rice fields and causes rice production loss all around the world (Ziska et al., 2015). One weedy rice plant per square metre can cause more than 200 kg ha^{-1} yield loss (Burgos et al., 2006), and the lost yield in the U.S. alone could feed 12 million people per year (Durand-Morat et al., 2018). Weedy rice has independently evolved from cultivated rice groups multiple times, most often through a process of de-domestication (Hoyos et al., 2020; Huang et al., 2017; Imaizumi et al., 2021; Li et al., 2022; Qiu et al., 2017,

2020; Reagon et al., 2010). Because of its similarity to the crop, weedy rice is concealed in crop fields and difficult to remove at the vegetative growth stage (Wu et al., 2022). Weedy rice has also evolved traits instrumental in its competitiveness that distinguish it from cultivated rice. A suite of traits known as the 'agricultural weed syndrome' (Vigueira et al., 2013), including rapid growth, high nutrient use efficiency, seed dormancy, seed dispersal and herbicide resistance, have been reported in different weedy rice populations (Chauhan & Johnson, 2011; Gu et al., 2005; Sales et al., 2008; Shivrain et al., 2010; Thurber et al., 2010). For example, some weedy rice will grow taller than adjacent rice cultivars (Reagon et al., 2011; Sun et al., 2019), which can help the weeds access more light and heat (Falster & Westoby, 2003). Some weedy rice populations flower earlier than the crops they invade, a strategy that allows them to escape being collected by humans (Ellstrand et al., 2010). Flowering simultaneously as the crop can also be beneficial to weeds in some circumstances, as a way to avoid conspicuousness in the fields or to achieve dispersal by contaminating seed stocks (Ellstrand et al., 2010). And seed dispersal, known as seed shattering in the grasses, is a trait that most consistently differentiates weedy from cultivated rice (Huang et al., 2021; Thurber et al., 2010; Ziska et al., 2015) and helps weedy rice disperse and increases its presence in the seed bank.

To date, the loci underlying most weediness traits that led to de-domestication of cultivated rice and increased competitiveness of weedy rice have not been identified. In the U.S., the BHA group, a major weedy rice population characterized by having blackhull awned grain, has evolved from *aus* cultivars (Li et al., 2017; Reagon et al., 2010), a geographically constrained rice variety, centred in the Indian subcontinent (Zhou et al., 2022), and highly tolerant to environmental stresses, such as drought and heat (Casartelli et al., 2018; Civáň et al., 2015). A second major U.S. weedy population, known as the SH group, tends to have strawhull awnless grain and has evolved from *indica* varieties of cultivated rice (Li et al., 2017; Reagon et al., 2010) that are widely geographically distributed in tropical and subtropical regions (Zhou et al., 2022). Previously, quantitative trait locus (QTL) mapping in populations derived from crosses between BHA×*indica* and SH×*indica* have been used to detect the genetic bases of multiple weediness traits in BHA and SH (Qi et al., 2015; Thurber et al., 2013). *aus* and *indica* cultivars are closely related but genetically distinct populations, with *aus* proposed to have originated either as a separate domestication event from *indica*, or through divergence after a shared domestication event (Choi et al., 2017; Civáň et al., 2015; He et al., 2021; Moner et al., 2020). QTLs detected through a BHA×*indica* cross will thus capture mutations arising through the entire history of divergence of the *indica* and *aus* cultivar groups, including any improvement/adaptation to local environments and also the de-domestication divergence of BHA from *aus* (Figure 1). In contrast, mapping with a BHA weedy rice×ancestral *aus* cultivar cross will capture QTLs more likely to be specific only to the de-domestication process (Figure 1). The choice of parents in mapping populations can thus impact our understanding of how weediness evolves. Because BHA weeds shatter more and grow taller than both *indica* and *aus* cultivars, this similar pattern of phenotypic differentiation between BHA and the two crops makes this an excellent system for studying the impact of parental evolutionary relatedness on QTL mapping for the same traits.

We have generated a new mapping population from a cross of BHA weedy rice×*aus* cultivar to better understand the de-domestication process in this weedy rice group and how it differs from de-domestication in other independently derived weedy rice. We focussed on three weediness traits, flowering time, plant height and seed shattering, that differentiate BHA from its *aus* ancestors. The objectives of this study were to: (1) identify QTLs for these three important weediness traits; (2) determine which QTLs were likely under selection during the de-domestication of BHA; and (3) assess the effect of ancestry on QTL mapping by comparing the QTLs detected from BHA×*aus* and BHA×*indica* crosses.

2 | MATERIALS AND METHODS

2.1 | Plant material and trait measurements

We created a weed×crop mapping population by crossing one BHA weedy rice (10A) with one *aus* cultivar (Aus 196), with 10A as the pollen donor. Crosses were carried out at the Dale Bumpers National Rice Research Center in Stuttgart, AR, U.S. Aus 196 is a breeding/research material developed in Bangladesh and obtained from the International Rice Research Institute, while 10A is a BHA weedy rice collected in Arkansas, U.S. and maintained by the United States Department of Agriculture. F_2 were generated by allowing multiple F_1 plants to self-fertilize. Eighty F_2 individuals, randomly selected from F_1 progeny, and the two parents were grown until maturity in Conviron PGW36 growth chambers at UMass Amherst under the conditions of 11 h light (29°C) and 13 h dark (24°C).

Flowering time, plant height and seed shattering were measured in all grown F_2 plants and the two parents. Flowering time was determined as the days from sowing to the date when the first panicle had emerged halfway from the boot (i.e. heading date). Plant height was measured at flowering from the base of the plant at the soil to the tip of the tallest panicle, excluding awns. Seed shattering was measured 30 days after heading as breaking tensile strength (BTS), the force required to pull off seeds from the panicles, using a digital force gauge as in Thurber et al. (2010). Lower BTS values thus represent higher levels of seed shattering. For each plant, at least 10 mature seeds per panicle, from at least one panicle, were used to measure BTS.

2.2 | Pool selection and sequencing

Ten individuals with extreme phenotype (the lowest and the highest) among the 80 F_2 were selected to construct each

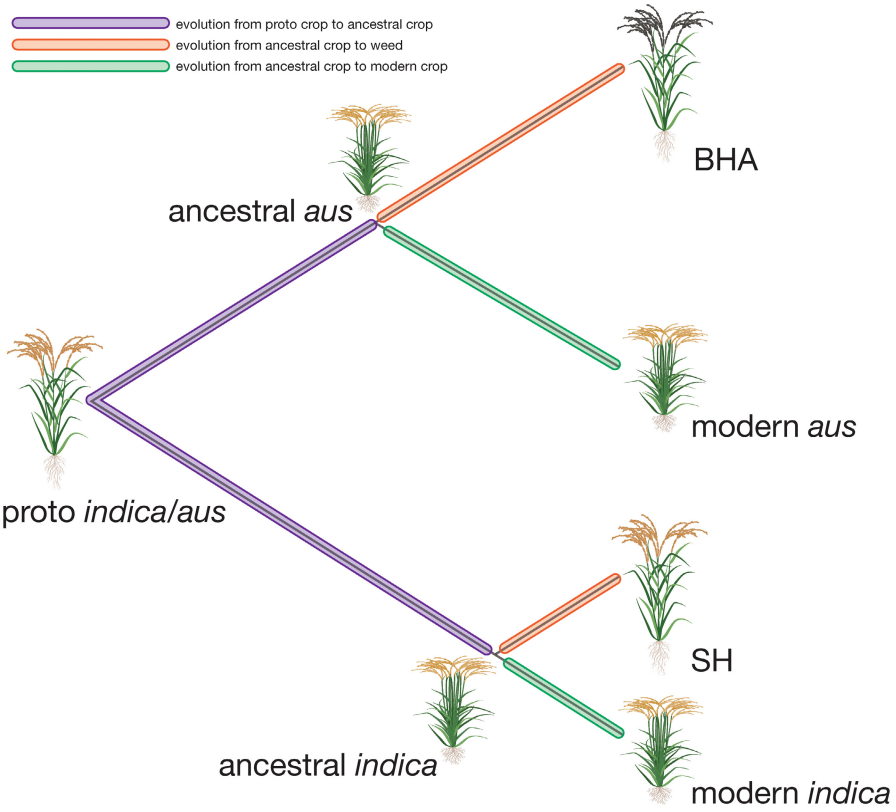

FIGURE 1 Illustration of the evolutionary history of the BHA and SH weedy rice groups and their respective ancestral cultivars, aus and indica. The purple bar represents the evolutionary divergence between ancestral aus and indica after diverging from proto indica/aus. The green bars indicate the further improvement or adaptation in modern crops relative to ancestral crops. The orange bars show the evolutionary divergence from ancestral crops to derived weeds. The divergence between BHA and aus occurred earlier than that of SH and indica. BHA, blackhull awned; SH, straw hulled. [Colour figure can be viewed at wileyonlinelibrary.com]

extreme bulk for flowering time, plant height and seed shattering. DNA was extracted from approximately 1 g of fresh leaf tissue from each individual using Qiagen DNeasy Miniprep Kit (Qiagen, MD, U.S.) and was quantified using a fluorometric assay with a Qubit 2.0 Fluorometer (ThermoFisher Scientific, MA, U.S.). The concentration of DNA was diluted to 50 ng/μL and an equal amount of DNA (6 μL) from the 10 selected F_2 individuals for each bulk was mixed. The concentration and integrity of the mixed DNA (60 μL) for each bulk were confirmed by fluorometric assay and agarose gel, respectively. The bulks were labelled as FTL (flowering time low), FTH (flowering time high), PHL (plant height low), PHH (plant height high), SSL (seed shattering low) and SSH (seed shattering high). Bulked DNA pools were sent to Hudson Alpha, AL, U.S. for 100 base pair (bp) paired-end sequencing on an Illumina HiSeq platform. Raw reads were submitted to the NCBI Short Read Archive (SRA) under experiment PRJNA961674.

2.3 | Identifying QTL regions using QTLsurge

The raw reads for each bulk were cleaned with Trimmomatic (Bolger et al., 2014) using a sliding window trimming method with an average base quality set as at least 20. The clean reads were then mapped to the de novo assembled and well-annotated R498 *indica* rice reference genome (Du et al., 2017) using Burrows–Wheeler Aligner (BWA) with the method 'mem' (Li & Durbin, 2009). This reference genome was chosen because it is the currently available high-quality genome that is most closely related to *aus* crops and their derived weeds. The aligned genomes of the two extreme bulks for the same trait were used to call the genomic variants and calculate their frequencies for each bulk using bcftools (Li, 2011). Only single nucleotide polymorphisms (SNPs) were kept for the following analyses.

QTL mapping was carried out using a QTL-seq approach in QTLsurge (https://github.com/USDA-AR-GBRU/QTLsurge).To avoid the influence of high repeat regions or reads with poor mapping quality, SNPs were further filtered using the qtl. filter. vcf.pl script from QTLsurge with the parameters 'min_depth 5, max_depth 150, qual 30 and mq 40' and transformed to the QTLsurge format using vcf2freq.pl script. The SNP-index was defined for each base pair as the frequency of reads containing the reference sequence (reference allele) divided by the total reads covering the same genomic region in one bulk; ΔSNP-index was defined as the difference between SNP-indexes of the two bulks for a given trait, and the absolute value of ΔSNP-index

was used to estimate the possibility of the association of a SNP to a trait (Takagi et al., 2013). A SNP linked to a causal locus is expected to harbour alleles with different frequencies between the two bulks, leading to a high absolute ΔSNP-index, with fixation of different alleles in the two bulks yielding an absolute ΔSNP-index of 1. In contrast, nonrelevant loci should contain similar allele frequencies between the two bulks resulting in a low absolute ΔSNP-index. To visualize the changes in absolute ΔSNP-index along the genome, we set a window size of 200 SNPs with an overlap of 40 SNPs, calculated the average absolute ΔSNP-index within the window and made Manhattan plots using these average values in R Statistical Software (v 4.2.1; R Core Team, 2022). A window size of 200 SNPs was chosen due to its balance between marker density and the number of peaks and the overlap-to-window rate of 0.2 adhered to the guidance by the QTLsurge tutorial. The 97.5th percentile of absolute ΔSNP-index values calculated across the genome was used as the threshold to define significant QTLs. The windows with an average absolute ΔSNP-index greater than this threshold were considered as corresponding to QTLs.

2.4 | Identifying regions with signals of selection

A subset of *Oryza* genomes from Li et al. (2017) previously aligned to the MSU7 *japonica* Nipponbare genome (Kawahara et al., 2013), including 15 *aus* cultivar genomes and 20 BHA weedy rice genomes—one of which corresponded to the BHA crossing parent, were realigned to the R498 *indica* rice reference genome using the method mentioned above. The variants between these two populations were called through mpileup in samtools with the setting of '-q 30 -Q 20 -C 50 -ugst DP, SP' (Li, 2011). Genome-wide nucleotide diversity (π) was calculated for each population using VCFtools (Danecek et al., 2011) for nonoverlapping 100-kb windows across the whole genome. To identify regions of the genome with low amounts of genetic diversity in BHA, the nucleotide diversity ratio of *aus* cultivar to BHA weedy rice (π_{aus}/π_{BHA}) was calculated for each window. We further removed SNPs with minor allele frequency less than 5% using VCFtools (Danecek et al., 2011) and calculated the genetic differentiation (F_{ST}) between the BHA and *aus* populations for each nonoverlapping 100-kb window. Selection scan window size and approach was adopted from the analyses in Li et al. (2017), to ensure a comparable evaluation of the regions under selection when aligning the same population reads to different reference genomes. The windows with the top 5% of π ratios or top 5% of F_{ST} values were considered regions with signals of selection. Windows with selection signal were combined into a larger interval if the distances between them were 300 kb or less. Overlaps between significant QTLs and intervals under possible selection were determined manually. Regions under selection within 500 kb upstream or downstream of a QTL were considered to overlap with the QTL.

2.5 | Identifying candidate genes

To identify candidate genes for each weediness trait, we examined the SNPs with highest absolute ΔSNP-index within each QTL and referred to these as high ΔSNP-index SNPs. In our dataset, these high ΔSNP-index SNPs were either SNPs with an absolute ΔSNP-index greater than or equal to 0.75, or, if none were present, SNPs with the highest absolute ΔSNP-index within the QTL. For each of these high ΔSNP-index SNPs, we calculated the individual F_{ST} values from the BHA-*aus* population dataset using VCFtools (Danecek et al., 2011) and predicted the functional effects of these SNPs through SnpEff based on their genomic location and the known annotation of genes, transcripts and other genomic features (Cingolani et al., 2012). We prioritized our focus on the genes most seriously affected by these SNPs (e.g. missense mutations) and the genes affected by the SNPs with the highest ΔSNP-index of all. For a subset of candidate genes, we genotyped nearly 40 random F_2 individuals for one SNP near or inside the gene using a cleaved amplified polymorphic approach (CAPs) method, as in Thurber et al. (2010). The primers and restriction enzymes used to conduct CAPs are listed in Table S1. The phenotypes among individuals with different homozygous and heterozygous genotypes were evaluated using the Kruskal–Wallis (KW) test from the stats package (R Core Team, 2022) and followed by post hoc testing and plotting from the ggpubr package (Kassambara, 2022).

2.6 | Comparing the QTLs detected from different crosses

To compare our results to those obtained by prior QTL mapping using the BHA × *indica* cross (Qi et al., 2015; Thurber et al., 2013), we calculated the physical positions of these QTLs on the Nipponbare *japonica* reference genome based on their genetic positions using cMConverter (http://mapdisto.free.fr/cMconverter). Then, we converted the QTL coordinates from the Nipponbare *japonica* genome to the R498 *indica* rice genome through manual blasting and local alignments using Geneious 9.1.8 (https://www.geneious.com). For previously published QTLs with a confidence interval over 10 Mb (about ¼ of a chromosome) or without a confidence interval, the region ranging from 200 kb upstream to 200 kb downstream of the associated marker was defined as a QTL region. We also determined whether any of these previously identified QTLs were possibly under selection by examining the overlap between the QTLs and the selection regions detected in the above analyses. Since seed shattering is one of the most convergently evolved traits for weedy rice with different origins, we additionally compared the QTLs detected for seed shattering in BHA × *aus* mapping population, SH × *indica* mapping population, and two separate US weed × *japonica* mapping populations to assess degree of genetic convergence (Qi et al., 2015; Subudhi et al., 2014; Thurber et al., 2013). The same approach of converting QTL coordinates as above was employed.

3 | RESULTS

3.1 | Phenotypic variation

We contrasted phenotypes of the parents of a BHA weedy rice × *aus* cultivar cross for three quantitative traits: flowering time, plant height and seed shattering. The BHA weedy parent showed a later flowering time, greater plant height and a higher seed shattering level (lower BTS) than the *aus* cultivar parent (Figure 2a–c; Tables S2–S4), consistent with prior reports for the *aus* and BHA populations (Reagon et al., 2011; Thurber et al., 2010, 2014). Within the F_2 populations, nearly normal distributions were observed for flowering time and plant height (Figure 2a,b), but not for seed shattering, for which lower BTS values (corresponding to higher shattering) were much more prevalent (Figure 2c). This suggests different genetic architectures for these traits, with possible multiple genes and additive effects for flowering time and plant height, and perhaps with fewer genes or a dominance effect for seed shattering (Hartl & Clark, 1997). Interestingly, few F_2 plants reached the high BTS value of the *aus* parent, suggesting the alleles contributing to the high seed shattering in the BHA parent have dominant effects. For the other traits, transgressive variation was evident, suggesting that both the crop and weed parent carry alleles that can affect these traits in both directions (Hartl & Clark, 1997). For each trait, 10 F_2 individuals with extreme trait values were bulked (highlighted in Tables S2–S4). Flowering time bulks, designated as FTH and FTL, had an average of 112.2 and 70.3 days to heading, respectively (Figure 2d). Plant height bulks, designated as PHH and PHL, had average heights of 150.51 and 94.35 cm, respectively (Figure 2e). Seed shattering bulks, designated as SSH and SSL, had average mean BTS values of 0.39 and 9.07 gf, respectively (Figure 2f).

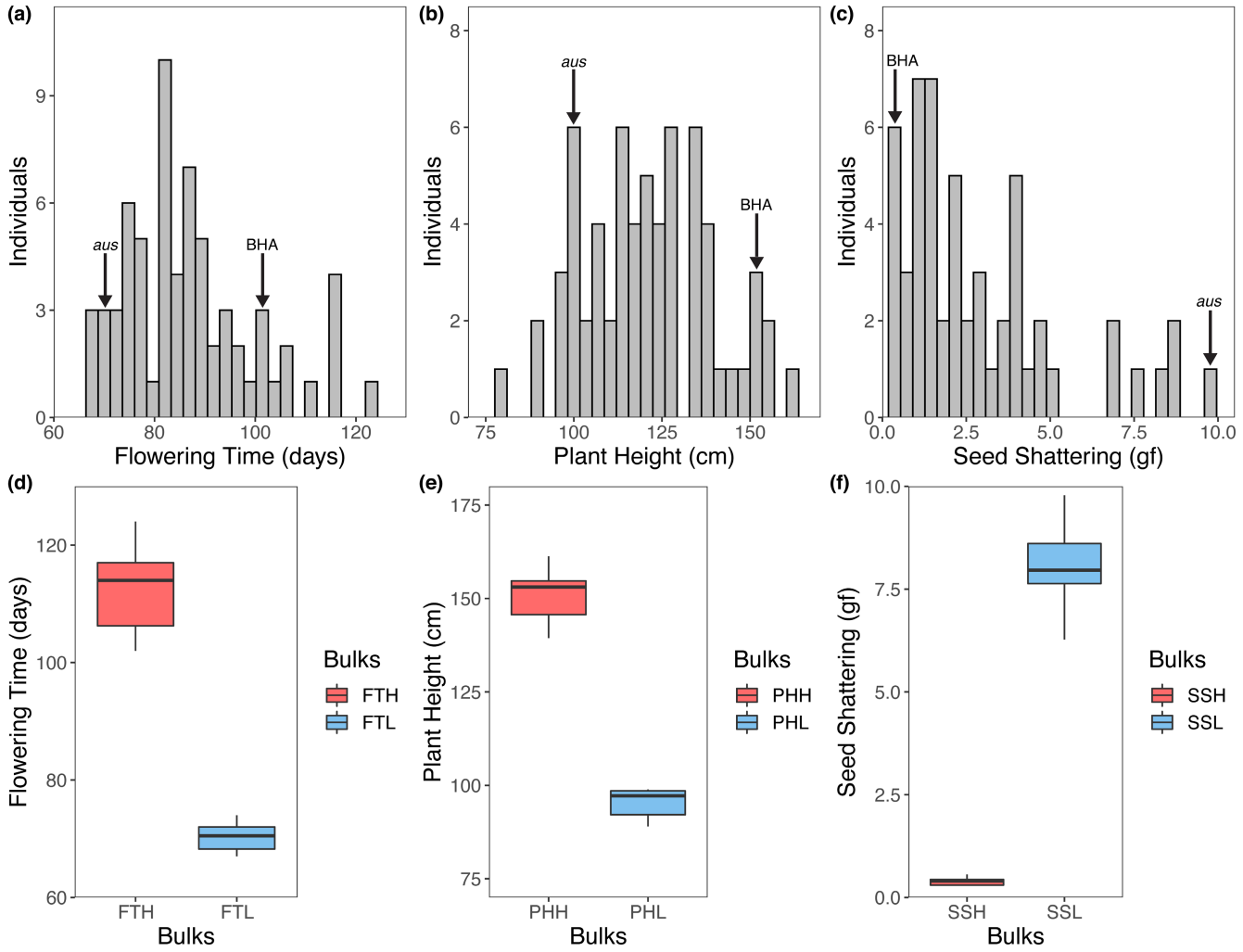

FIGURE 2 Phenotypic distributions of weedy traits in F_2 individuals and pairwise comparisons between extreme bulks for each trait. (a–c) The phenotypic distributions of F_2 individuals for flowering time, plant height and seed shattering, with the phenotypes of the BHA and aus parents indicated by arrows. (d-f) The pairwise comparisons between extreme bulks for each trait. FTH and FTL represent flowering time high and flowering time low bulks; PHH and PHL represent plant height high and plant height low bulks; and SSH and SSL indicate seed shattering high and seed shattering low bulks. Bulks with a phenotype more similar to weed groups are labelled in red, while those similar to cultivars are labelled in blue. BHA, blackhull awned. [Colour figure can be viewed at wileyonlinelibrary.com]

3.2 | QTL mapping and candidate gene identification

Among bulks, the average number of clean reads mapped to the R498 *indica* rice reference genome was 226,546,487. FTL had the lowest number of mapped clean reads at 139,789,206, and PHL had the highest number of mapped clean reads, with 281,391,018. On average, 3.5M SNPs were identified between each set of extreme bulks. After filtering out the SNPs with low mapping quality or extremely high sequencing depth, we kept 1,448,543 SNPs for the FTL-FTH bulks, 1,439,472 SNPs for the PHL-PHH bulks and 1,452,821 SNPs for the SSL-SSH bulks for the QTL-seq analyses. We plotted the average absolute ΔSNP-index between the respective high and low bulks for each window against the positions along the genome for each trait (Figure 3). The 97.5th percentile of the genome-wide absolute ΔSNP-index was 0.33 for flowering time, 0.30 for plant height and 0.38 for seed shattering. Using this as a threshold yielded 11, 12 and 5 QTLs for flowering time, plant height and seed shattering, respectively (Table 1; Figure 3). These QTLs spanned 738, 757 and 748 genes, respectively, and their gene IDs and annotations are listed in Tables S5–S7. Some QTLs were of small size, and no genes were included in the list. Below we focus on the genes affected by the high ΔSNP-index SNPs within each of these QTLs.

3.2.1 | Flowering time

Eleven flowering time QTLs occurred on chromosomes 1, 2, 6, 10 and 11 (Table 1; Figure 3). The QTL signals on chromosome 6 were particularly dominant, especially qHD6_1. Within these QTLs, 64 high ΔSNP-index SNPs were identified (Table S8). Most of these SNPs had modifier effects, the lowest putative impact category predicted by SnpEff, usually corresponding to noncoding variants or variants affecting noncoding regions (Cingolani et al., 2012; Table S8). Among the genes potentially affected by these SNPs, two contained missense mutations. One of these was unannotated while the other one codes for a transferase family protein, and a connection to flowering time was not immediately apparent (Table S8). Seven SNPs with extremely high ΔSNP-index, all in the qHD6_1 region, stood out. Of the four SNPs in the region from 9,702,496 to 9,702,574 bp, two had an absolute ΔSNP-index of 1, the highest possible value. No obvious candidate gene was evident in this region of qHD6_1, and all SNPs with a ΔSNP-index of 1 fell within intergenic regions (Table S8). Three SNPs in the region from 9,974,134 to 9,974,155 bp had an absolute ΔSNP-index greater than 0.93 (Table S8); in this region, about 2.4 kb downstream from these SNPs, we found a candidate gene for flowering time called *OsNF-YB9* (OsR498G0612149700.01; Table 2), a negative regulator of heading date. Overexpression of *OsNF-YB9* results in delayed heading and floral transition (Das et al., 2019). We also identified another candidate gene for flowering time in qHD6_1 named *DTH2* (OsR498G0612192000.01; Table 2), which was less than 1 kb upstream of four high ΔSNP-index SNPs with an absolute ΔSNP-index greater than 0.75 (two of them greater than 0.8; Table S8). *DTH2* is a QTL that promotes heading under long-day conditions (Wu et al., 2013). The well-known flowering time gene, *Heading date 1* (*Hd1*; Nemoto et al., 2016; Yano et al., 2000) also fell within qHD6_1 (Table S5); however, no high ΔSNP-index SNPs were observed near this gene. No obvious flowering time candidate genes were found in the other high-threshold flowering time QTLs.

We genotyped the two SNPs near the two flowering time candidate genes, *OsNF-YB9* and *DTH2*, in 39 random F_2 individuals (Tables S1, S9 and S10) and found a significant difference in flowering time between F_2 individuals with a homozygous BHA weed genotype and those with a homozygous *aus* crop genotype (Figure S1a,b). Genotypes at these two SNPs were completely correlated, even though positioned around 0.7 Mb apart (Table S9). This robust linkage disequilibrium would make identifying causal genes a challenge. Interestingly, in both instances, F_2 individuals with a homozygous BHA weed genotype flowered significantly earlier, opposite to the flowering pattern exhibited by the BHA parent. This indicates that late flowering alleles exist in the *aus* crop parent, and their action is likely modulated by genetic background, consistent with the transgressive segregation we observed. We examined the expression of both genes in the Rice Expression Database (Xia et al., 2017) and found that both genes are expressed in rice anthers, but *DTH2* was detected at higher levels in pistils and panicles, suggesting it may be a more compelling flowering time candidate gene.

3.2.2 | Plant height

Twelve QTLs located on chromosomes 1 and 7 were considered as plant height QTLs, with the strongest signals occurring on chromosome 7 (Table 1; Figure 3). These QTLs harboured 45 high ΔSNP-index SNPs (Table S11). Most of these SNPs also had effects as modifiers (Table S11). One missense SNP affected a gene annotated as cytochrome P450 72A1, but its connection to plant height was not obvious (Table S11). Two SNPs with an absolute ΔSNP-index equal or greater than 0.9 were found in qPH7_2, and one of these SNPs occurs around 10 kb upstream of the gene *OsGH3.8* (OsR498G0714682600.01; Table 2). This gene belongs to the GH3 family and codes an IAA-amido synthetase. Indole-3-acetic acid (IAA) induces the expression of expansins that promote plant growth, while IAA-amido synthetase prevents free IAA accumulation; thus, *OsGH3.8* is a negative regular for plant height, and overexpression of *OsGH3.8* causes retarded growth and development (Ding et al., 2008), making it a strong candidate for plant height. A third SNP with an absolute ΔSNP-index greater than 0.9 was in qPH7_3, but no known plant height candidate genes were found among the genes potentially affected (Table S11). Several genes annotated as putative gibberellin (GA) receptor GID1L2 (e.g. OsR498G0714845800.01; Table 2), possibly involved in the GA signalling pathway, could be candidate genes in qPH7_3 (Table S6). Although these genes did not contain any high ΔSNP-index SNPs (Table S11), multiple SNPs with relatively high ΔSNP-index were around 10 kb away from these GA receptor genes, for example, SNPs

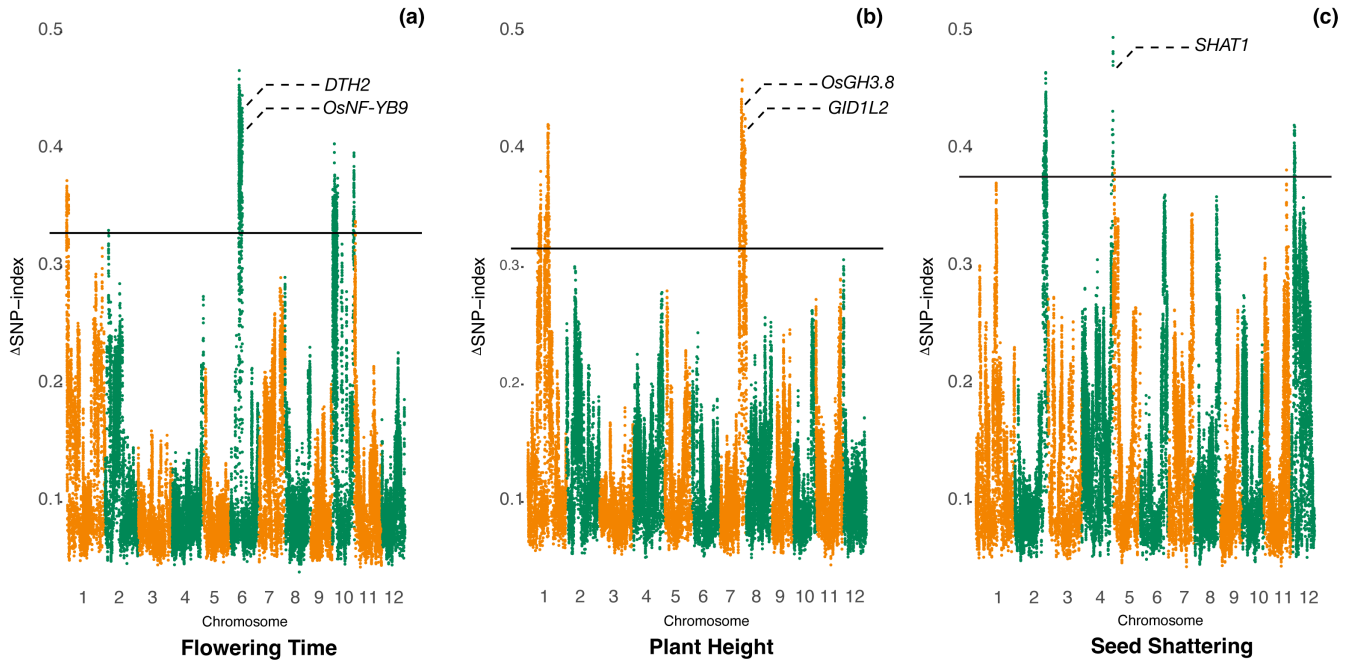

FIGURE 3 Manhattan plots of the average absolute ΔSNP-index between the two respective extreme bulks within sliding windows along the rice genome. (a–c) Manhattan plots of the average absolute ΔSNP-index within sliding windows between the two respective extreme bulks for the traits of flowering time, plant height and seed shattering along the rice genome. The window size is 200 SNPs with an overlap of 40 SNPs. The black line represents the threshold of the 97.5th percentile of the genome-wide absolute ΔSNP-index. The candidate genes identified for each trait were pointed to the corresponding Quantitative Trait Loci (QTL) peaks with dashed arrows. [Colour figure can be viewed at wileyonlinelibrary.com]

27,456,440 and 27,451,048 with absolute ΔSNP-indexes of 0.72 and 0.64, respectively.

We attempted to design CAP markers for genotyping the high ΔSNP-index SNPs close to *OsGH3.8* and the GA receptor genes to confirm the phenotype associations but failed. We examined our genome sequence data for the BHA parent to determine the identity of alleles predominant in each plant height bulk. For SNPs close to *OsGH3.8*, the BHA parent carried reference alleles, which were associated with shorter height, despite it being the taller parent (Table S9); this is again consistent with the transgressive segregation observed in the mapping population and suggests that the effect of this QTL is background dependent. For high ΔSNP-index SNPs close to GA receptor genes, the same situation was observed (Table S9). The proximity between QTLs qPH7_2 and qPH7_3 suggests these results are driven by high linkage disequilibrium between these regions, making it hard to identify the most likely causal locus.

3.2.3 | Seed shattering

Five QTLs for seed shattering were found on chromosomes 2, 4, 5, 11 and 12 (Table 1; Figure 3). For these QTLs, 30 high ΔSNP-index SNPs were identified, and the genes predicted to be potentially affected by these SNPs are listed in Table S12. One gene, *OsMGD*, a synthase for monogalactosyldiacylglycerol (MGDC), a polar lipid in the thylakoid membrane of plants (Hui et al., 2022), was predicted to carry a missense mutation but had no apparent connection to seed shattering (Table S12). Four highest ΔSNP-index SNPs were detected: one on chromosome 2 and three on chromosome 12, and the genes affected by these SNPs are highlighted in Table S12. However, no known seed shattering genes were noted in this list and the genes on chromosome 12 are not well annotated, suggesting that substantial more work is needed to identify potential candidates in these regions.

3.3 | Signals of selection and interactions with identified QTLs

QTL mapping detects loci that underlie trait differences between the two parental individuals, but these loci may or may not have been instrumental in the evolution of weediness at a population level. To discern whether any of our detected QTLs played an adaptive role in the evolution of the entire BHA group, we determined which QTLs overlapped with selective sweeps. Prior efforts to detect selective sweeps between *aus* and BHA had aligned the sequence data to the reference genome of a *japonica* cultivar (Kawahara et al., 2013), which is distantly related to the lineage that includes *aus* and BHA (Choi et al., 2017; Fuller et al., 2009, 2010; Gross & Zhao, 2014). We realigned 35 genomes from Li et al. (2017), 15 *aus* and 20 BHA (including our BHA mapping parent, but not the *aus* parent for which whole genome sequence was not available), to the R498 *indica* reference genome (Du et al., 2017) on average, 61.5 M reads were successfully mapped per sample.

TABLE 1 QTLs for flowering time, plant height and seed shattering detected using a cross of BHA×*aus*.

Phenotype	Chromosome	QTL	Physical range (bp)	
Flowering time	1	qHD1_1	24,799	468,345
		qHD1_2	2,240,489	2,354,841
	2	qHD2	4,001,775	4,028,588
	6	qHD6_1	9,250,219	11,903,051
		qHD6_2	12,185,732	12,697,917
		qHD6_3	13,350,694	13,502,404
	10	**qHD10_1**	**626,642**	**5,380,199**
		qHD10_2	6,211,127	6,216,812
		qHD10_3	**24,202,767**	**24,322,818**
		qHD10_4	**24,746,786**	**25,123,568**
	11	qHD11	1,065,280	1,094,337
Plant height	1	**qPH1_1**	**11,590,007**	**11,592,856**
		qPH1_2	**12,809,806**	**13,392,938**
		qPH1_3	14,378,114	14,903,956
		qPH1_4	**19,553,128**	**19,809,406**
		qPH1_5	**21,243,832**	**21,956,589**
		qPH1_6	**22,901,892**	**23,704,909**
		qPH1_7	**24,358,720**	**24,504,943**
	7	qPH7_1	22,322,762	22,888,318
		qPH7_2	**24,506,739**	**25,533,547**
		qPH7_3	26,433,407	27,984,964
		qPH7_4	28,579,279	28,906,309
		qPH7_5	29,833,641	29,868,402
Seed shattering	2	**qSH2**	**32,653,521**	**36,284,375**
	4	**qSH4**[a]	**34,966,466**	**35,793,877**
	5	**qSH5**	**1,518,830**	**1,539,044**
	11	qSH11	25,722,001	25,730,429
	12	**qSH12**	**3,061,735**	**4,198,161**

Note: QTLs that overlap with regions of selective sweeps are shown in bold.

[a]qSH4 does not overlap with selective sweeps directly, but there are two selective sweeps near this QTL.

We detected 5,920,505 SNPs among these weedy and cultivated genomes. A higher average nucleotide diversity was observed in the *aus* population (0.00363) than in the BHA population (0.00212), consistent with prior findings of a bottleneck in the evolution of the BHA group (Li et al., 2017). The average π ratio between *aus* to BHA across the genome was 2.74, and the 5% cut-off for a region under selection was set to 7.13 (Table S13). The average genomic F_{ST} between these two populations was moderate across the genome at 0.193, and the 5% cut-off corresponded to an F_{ST} of 0.409 (Table S14). The selection regions identified by these methods were largely consistent, and we thus considered all regions identified by either method (Tables S13 and S14). All selective sweeps identified in our prior study (Li et al., 2017) were detected in our newly aligned data, and we discovered additional regions putatively under selection. For example, regions on chromosomes 2 and 5 were exclusively revealed when we aligned the genomes to the *indica* reference genome, suggesting that choice of reference genome can impact the power of sweep detection. We examined which QTLs overlapped with regions putatively under selection for each of our traits. For all traits, we found more than one QTL occurring in putatively selected regions (bold in Table 1; Table S15). Below we go over the results obtained for each one of our traits.

3.3.1 | Flowering time

Three flowering time QTLs on chromosome 10 overlapped with selected regions (Table 1; Table S15). No known flowering time candidate genes were identified in these QTLs as we mentioned above (Table S8). No signal of selection was detected for the dominant QTLs on chromosome 6 (Table S15); thus, the candidate genes identified in these QTLs (*OsNF-YB9* and *DTH2*) may not have evolved under selection in the BHA group, while unidentified candidates on chromosome 10 most likely contributed to BHA adaptation. However, we did observe that SNPs near *DTH2* had different frequencies in the crop and weed populations. For example, for the SNP genotyped for

TABLE 2 Possible candidate genes for flowering time, plant height and seed shattering.

Trait	QTL	Gene name	R498 gene ID	Start	End	Gene annotation
Flowering time	qHD6_1	*OsNF-YB9*	OsR498G0612149700.01	9,976,607	9,977,653	Histone-like transcription factor and archaeal histone
	qHD6_1	*DTH2*	OsR498G0612192000.01	10,671,044	10,677,736	CCT/B-box zinc finger protein
Plant height	qPH7_2	*OsGH3.8*	OsR498G0714682600.01	24,842,656	24,845,085	OsGH3.8—Probable indole-3-acetic acid-amido synthetase
	qPH7_3	*GID1L2*[a]	OsR498G0714845800.01	27,423,213	27,426,415	Gibberellin receptor GID1L2, putative
Seed shattering	qSH4_1	*SHAT1*	OsR498G0409385900.01	33,346,944	33,350,643	AP2 domain-containing protein

[a]Multiple genes are annotated as *GID1L2* in the same region, but only one gene is used here as a representative.

DTH2, the reference allele frequency was 0.35 in the BHA population and 0.95 in the *aus* (Table S9), suggesting significant differentiation between the two populations at this locus, which may perhaps have been driven by genetic drift rather than selection.

3.3.2 | Plant height

Seven plant height QTLs overlapped with selective sweeps (Table 1; Table S15). One such region was qPH7_2, containing the candidate gene, *OsGH3.8*. It is worth noting that the SNPs located within or close to this gene, including those displaying notably high ΔSNP-index values, were fixed in the BHA population whereas both genotypes were present in the *aus* population (Table S9). These findings suggest that evolution of plant height in BHA population probably has occurred, in part, through selection on standing variation of this QTL. No other known plant height candidate genes were identified in the other QTLs overlapping with selection sweeps, indicating a need for more work to determine the loci most contributing to the weed adaptation via greater plant height in the BHA. The series of GA receptor genes located in qPH7_3 did not display any selection signal.

3.3.3 | Seed shattering

Three seed shattering QTLs on chromosomes 2, 5 and 12 overlapped with selective sweeps (Table 1; Table S15). Another QTL, qSH4, did not directly overlap with selective sweeps, but two selective sweeps were found near it, at a distance less than 1 Mb (Table S15). Of all the QTLs detected for seed shattering, only the one on chromosome 7 was not close to any selection regions, indicating that loci underlying seed shattering may have often been under selection during weed evolution. As mentioned above, no known seed shattering candidate genes were detected in any QTLs (Table S12). However, one known shattering gene, *SHAT1* (OsR498G0409385900.01; Table 2) was near our qSH4 QTL and within a selective sweep (Table S15). This gene encodes an APETALA2 transcription factor and contributes to the development of the abscission zone, a structure that forms in the pedicel-floral interface and is important for shattering (Zhou et al., 2012). Interestingly, the major gene involved in the reduction of shattering during rice domestication, *Shattering4* (*SH4*; Li et al., 2006) was also located near qSH4 but not within any selective sweep and no high ΔSNP-index SNPs were found nearby.

We genotyped a SNP located in the 3′ UTR (untranslated region) of *SHAT1*, which did not have a very high ΔSNP-index value but was highly differentiated between the BHA and *aus* population sets (Table S9), in 37 randomly chosen F_2 individuals (Table S16). Notably, individuals with a homozygous BHA weed genotype frequently had low BTS values and high shattering levels while those with a homozygous *aus* crop genotype often exhibited high BTS values and low shattering levels (Table S16), although the differences were not statistically significant (Figure S1c). We took

advantage of unpublished expression data generated in our lab of the abscission zone in the BHA and *aus* parents at three developmental time points (Table S17); these revealed that *SHAT1* was expressed in the abscission zone in both of our mapping parents, confirming the expression pattern documented in earlier study from (Zhou et al., 2012). However, a higher expression of *SHAT1* occurred in the BHA parent compared with the *aus* parent, particularly during the heading stage (Table S17), suggesting the expression of this gene could be affected by cis-acting SNPs in the QTL we identified. We also noticed that SNPs near or within the *SHAT1* were fixed in the BHA population while segregating in the *aus* (Table S9). Together these results strongly support *SHAT1* likely contributing to adaptation via increased seed shattering in the BHA group. In contrast, expression of *SH4*, which occurred prior to heading, (as in Li et al., 2006), was comparable between the BHA and *aus* parents (Table S17), suggesting that *SH4* is unlikely account for the distinct seed shattering between them. Last, we note that unknown loci in chromosomes 2, 5 and 12 are also likely to have contributed to BHA adaptation via increased shattering.

3.4 | Comparisons of QTLs detected in different crosses

A higher number of QTLs were detected for the three weediness traits in our BHA×*aus* cross than in the prior BHA×*indica* cross (Qi et al., 2015; Thurber et al., 2013; Table 1; Table S18). Given the closer relationship between *aus* and BHA parents, which should lead to fewer accumulated genetic differences, this result is likely attributed to the power of the QTL-seq approach compared with classical linkage mapping.

Very few QTL from BHA×*indica* cross (Qi et al., 2015; Thurber et al., 2013) were found to overlap with the QTLs in the BHA×*aus* cross (Figure 4; Table S18). Flowering time and plant height had no overlapping QTL. Only two seed shattering QTLs from this study, qSH5 and qSH12 were close to minor effects QTLs, qSH5Bb and qSH12B, from the BHA×*indica* mapping population (Qi et al., 2015). We also discerned which of QTLs in BHA×*indica* cross were possibly involved in weediness adaptation by examining the overlap between the QTLs and our detected selective sweeps. Overlaps were observed in all three traits, and more frequently in seed shattering. For example, among the QTLs detected in (Qi et al., 2015), three out of six shattering QTLs overlapped with selection regions, but only one out of four flowering time QTLs overlapped (Table S18). Overall, fewer putatively weed adaptive QTLs were found in BHA×*indica* cross than in BHA×*aus* cross: 8 QTLs from (Qi et al., 2015; Thurber et al., 2013; Table S18) and 14 QTLs from our study (Table S15).

High seed shattering has likely independently evolved multiple times in different weedy rice populations and contributes to their survival and spread in the rice fields (Qi et al., 2015; Subudhi et al., 2014; Thurber et al., 2010, 2013). We compared the seed shattering QTLs from our BHA×*aus* cross with the prior SH×*indica* cross: only qSS12s in SH×*indica*, a minor effect QTL in that cross, overlapped with a QTL in the BHA×*aus* population (qSH12; Table S19). This QTL was also detected in the cross of BHA×*indica*, qSH12B, but only explaining a small portion of the total phenotype variation (PVE = 5.5%). Subudhi et al. (2014) also conducted two QTL studies for seed shattering in one US weedy rice of unknown ancestry through separate crosses with two *japonica* cultivars. One QTL on chromosome 4 (qSH4_Subudhi) consistently detected in both their crosses potentially overlapped with our qSH4 (Table S19) and was close to regions we detected as being under selection.

4 | DISCUSSION

The ability of weedy rice to evolve recurrently from different cultivated ancestral backgrounds is among the most startling examples of convergent or parallel evolution (Hoyos et al., 2020; Huang et al., 2017; Imaizumi et al., 2021; Li et al., 2022; Qiu et al., 2017, 2020; Reagon et al., 2010). Identifying the loci underlying these multiple weed de-domestications has important implications for our understanding of the genetic bases and repeatability of adaptation. As a major constraint on cultivated rice worldwide, finding the genes facilitating weedy rice evolution is also important for weed management and understanding which rice varieties are more likely to give rise to weedy descendants.

Our work and that of others have suggested that populations of weedy rice with different cultivated backgrounds have evolved weedy traits through different genetic mechanisms (Hoyos et al., 2020; Huang et al., 2017; Imaizumi et al., 2021; Li et al., 2022; Qiu et al., 2017, 2020). However, prior studies attempting to map weedy traits have not always taken ancestry into account (Bres-Patry et al., 2001; Gu et al., 2005; Qi et al., 2015; Subudhi et al., 2014; Thurber et al., 2013) leaving us with an incomplete understanding of the loci involved in the evolution of weediness traits. In the U.S., this has particularly impacted our understanding of the evolution of *aus*-derived BHA weedy rice, a group that has become an even bigger concern recently as it has been hybridizing with herbicide-resistant cultivated rice and creating novel herbicide-resistant weedy rice in the U.S. (Burgos et al., 2014; Wedger et al., 2022). By carrying out a QTL study involving a BHA×*aus* mapping population, we have identified QTLs more likely to be involved in adaptive evolution of three weediness traits and have explored the factors affecting the detection of evolutionarily relevant loci.

4.1 | Loci contributing to evolution of weediness in BHA weedy rice

Our study identified multiple QTLs for each of the weediness traits we examined (Table 1; Figure 3). Notably, we found that several of these QTLs overlapped with selective sweeps, particularly in

the case of the seed shattering QTL where three out of five QTLs showed a signal of selection and one QTL was very close to two selective sweeps (Table 1) and in the case of plant height where 7 of 12 detected QTLs overlapped with sweeps. These findings suggest that the evolution of weediness in rice has likely involved multiple loci, rather than a few mutations of large effect. This is consistent with the hypothesis that de-domestication has likely occurred primarily from sorting of standing variation present in cultivated rice rather than de novo mutations (Huang et al., 2017; Qiu et al., 2017).

The number of QTL detected was highest for plant height. BHA weedy rice tends to be taller than *aus* cultivars (Reagon et al., 2011), and in the U.S. rice fields, BHA weeds usually tower over commonly grown *tropical japonica* dwarf varieties. Increasing plant height can enhance the competitiveness of agricultural weeds through the most straightforward way of accessing more light and heat (Falster & Westoby, 2003; Liu et al., 2018), although there are also potential disadvantages such as the cost of construction and maintenance of the stem or the susceptibility to lodging (Falster & Westoby, 2003; Liu et al., 2018). Although more than half of these QTLs showed evidence of selection, we were unable to identify the most plausible candidates in most of these. Our most compelling plant height candidate gene, *OsGH3.8* is an IAA-amido synthetase preventing free IAA accumulation; IAA is the predominant auxin in plants and plays a significant role in plant growth and development (Vanneste & Friml, 2009); thus, *OsGH3.8* acts as a negative regulator for plant height (Ding et al., 2008). The overlap of this candidate gene with a selective sweep, coupled with the multiple SNPs proximal to and within this gene fixed in the BHA population (Table S9), suggests that this locus may have contributed to adaptation via increased plant height during the de-domestication of the BHA population, but that it is only one of several loci likely involved in the adaptation of the BHA group.

BHA weeds have evolved to flower later than their *aus* ancestors under day-neutral conditions, and this has led to near concurrent flowering with the *tropical japonica* cultivars that are their major competitors in the U.S. rice fields (Thurber et al., 2014). This flowering strategy can decrease conspicuousness of the weeds in the field to evade early removal attempts and increase the possibility of their seeds being unwittingly collected and planted (Thurber et al., 2014). In recent years, this has also led to facilitated hybridization with herbicide-resistant cultivars (Burgos et al., 2014; Wedger et al., 2022).

Mutations in *Hd1*, a major integrator of environmental signals as well as an important regulator of flowering time through controlling florigen's expression in rice (Nemoto et al., 2016; Yano et al., 2000), have been implicated in the evolution of daylength-insensitive cultivated rice during domestication and improvement (Takahashi et al., 2009). Mutations at *Hd1* have also been shown to explain the flowering time difference between SH and its cultivated ancestor *indica*, where SH has an ancestral functional allele while many *indica* cultivars have a nonfunctional allele (Qi et al., 2015; Thurber et al., 2014). Curiously, we detected a QTL encompassing *Hd1*, qHD6_1 (Table S5), but no surrounding SNPs were detected with high absolute ΔSNP-index. Prior studies have also shown that both *aus* cultivars and BHA weeds share nonfunctional allele at this locus, suggesting it cannot account for flowering time differences between them (Thurber et al., 2014).

Our two compelling flowering candidate genes control flowering time, *DTH2* and *OsNF-YB9*, bypassing *Hd1*. *DTH2* can affect heading through inducing florigens *Hd3a* and *RFT1* independent from the *Hd1* pathway (Wu et al., 2013). Another candidate gene, *OsNF-YB9* can alter the flowering time by interacting with *OsGI* (Izawa et al., 2011), a diurnally regulated protein that can control the expression of *Hd1* (Das et al., 2019). The significant phenotypic differences between the F_2 individuals with a homozygous BHA weed genotype and those with a homozygous *aus* crop genotype (Figure S1a,b) support that this locus affects the flowering time phenotype. However, the lack of a clear selection signal in this QTL suggests that these candidate loci were not instrumental to flowering time adaptation during de-domestication of BHA and that promising adaptive flowering time loci are more likely to be found on chromosome 10.

Seed shattering is an important trait that consistently distinguishes cultivated and weedy rice with high seed shattering being prevalent in weedy rice populations (Huang et al., 2021; Thurber et al., 2010). We identified seed shattering QTLs in multiple chromosomes (Table 1), indicating that there are numerous mutational targets that can affect this trait in rice. Interestingly, an obviously higher proportion of the QTLs for seed shattering (80%) overlapped with or close to regions under selection when compared to the QTLs for flowering time (27%) and plant height (58%; Table 1). This suggests that the loci contributing to seed shattering modification are likely subject to selection pressures in BHA. Despite the evolution of high shattering in weedy rice representing a reversal to a phenotype present in the wild ancestor of cultivated rice, several studies have shown that shattering in the U.S. weeds did not involve a reversal to wild alleles of the shattering domestication gene, *SH4* (Huang et al., 2018; Qi et al., 2015; Subudhi et al., 2014; Thurber et al., 2010, 2013). Although *SH4* occurred close to one of our QTL, qSH4, lack of differentiation at this gene between our crossing parents supported the involvement of a linked locusin explaining the difference in shattering ability between these weeds and crops, similar to (Subudhi et al., 2014). *SHAT1* (Zhou et al., 2012), the candidate gene we identified in the selective sweep near our qSH4 QTL, might contribute to the evolution of high seed shattering in BHA as evidenced by our genotype to phenotype correlation, the expression difference between the two parents and the observation of the fixed SNPs in the BHA population. Additionally, our study contributes further evidence for the role of distinct genetic mechanisms in explaining the convergence of high seed shattering in weedy rice populations with different origins (Huang et al., 2017; Qi et al., 2015; Thurber et al., 2013). When comparing the seed shattering QTLs between the BHA×*aus* and SH×*indica* crosses, there is only one overlapping QTL (qSH12; Qi et al., 2015). Similarly, comparison with results of US weed×*japonica* cultivar crosses (Subudhi et al., 2014) only revealed one overlapping QTL.

4.2 | The factors impacting the detection of evolutionarily relevant QTL

Our prior attempts (Qi et al., 2015; Thurber et al., 2013) to identify genes underlying weedy traits in BHA relied on a cross with an *indica* cultivar that is more distantly related to BHA than its *aus* progenitor group (Figure 1). Using an *aus* cultivar as the crop parent yielded novel QTLs, and very little overlap with QTLs detected in the BHA×*indica* cross (Figure 4), which is surprising given the closer relationship between *aus* and BHA. To some extent, the detection of novel QTLs in the BHA×*aus* cross may be due to the greater power of QTL-seq, which combines bulked segregant analysis and whole genome sequencing (Takagi et al., 2013). A similar pattern has been observed for one QTL with a large region detected with conventional QTL mapping, which was then subdivided into three smaller QTLs using QTL-seq (Wen et al., 2019). QTL-seq has other advantages to detect weediness loci, including being less time-consuming, more efficient and having higher accuracy. One major limitation to this approach is the difficulty in estimating the effect size for each QTL in explaining phenotypic variation (Sheng et al., 2021; Wen et al., 2019).

The diverging evolutionary history between *aus* and *indica* also plays a role in the QTLs we detected in this study. A comparative genomic analysis has suggested different selection profiles for the domestication of *aus* and *indica* in both position and magnitude of the selective sweeps (Civáň et al., 2015). Even within the genomic regions under selection in both groups, distinct sequence types seem to have been selected during domestication (Civáň et al., 2015). Additionally, compared with *indica*, *aus* has been hypothesized to have been under incomplete domestication selection, as some alleles associated with *indica* and *japonica* rice domestication are not found in the genome of *aus* (Zhao et al., 2018).

One example of the impact of cross on weediness QTL detection is the flowering time candidate gene *DTH8* (Qi et al., 2015), identified in the prior BHA×*indica* cross. A loss-of-function allele of *DTH8* can cause a reduced photoperiod sensitivity and earlier flowering and thus helped rice to adapt to the northern limit of rice cultivation (Fujino et al., 2013; Wei et al., 2010; Yan et al., 2011). Earlier flowering time in *indica* compared with BHA can be explained by a nonfunctional allele of *DTH8* present in the *indica* parent but not in the BHA parent (Qi et al., 2015). However, no signal of QTL for flowering time was found near this gene in our results, indicating that *DTH8* cannot explain the flowering time difference between *aus* and BHA. The same explanation applies to the other flowering candidate gene *DTH7* found in the BHA×*indica* cross (Gao et al., 2014; Qi et al., 2015) that is not located in any of the flowering time QTLs detected in our cross. Another example is *SD1* (Spielmeyer et al., 2002) occurring in a major plant height QTL in the BHA×*indica* cross (Thurber et al., 2013). The *indica* parent used in the cross is the semidwarf parent employed in breeding during the green revolution and has a disrupted *SD1* function, while BHA has a functional allele (Thurber et al., 2013). The *aus* parent in our study does not have this mutation and no QTL around this gene was identified, supporting its lack of relevance for BHA weed evolution. Because of the evolutionary distance between *indica* and *aus*, mutations in the above genes are likely to have contributed to *indica* domestication and not necessarily to weediness evolution of BHA. Our finding of greater QTL overlaps with selective sweeps also supports the BHA×*aus* cross as better for the detection of evolutionarily relevant weediness QTLs.

The choice of reference genomes can also have an impact both on mapping potential and on the detection of positive selection. For many years, rice sequences of all types have been aligned to

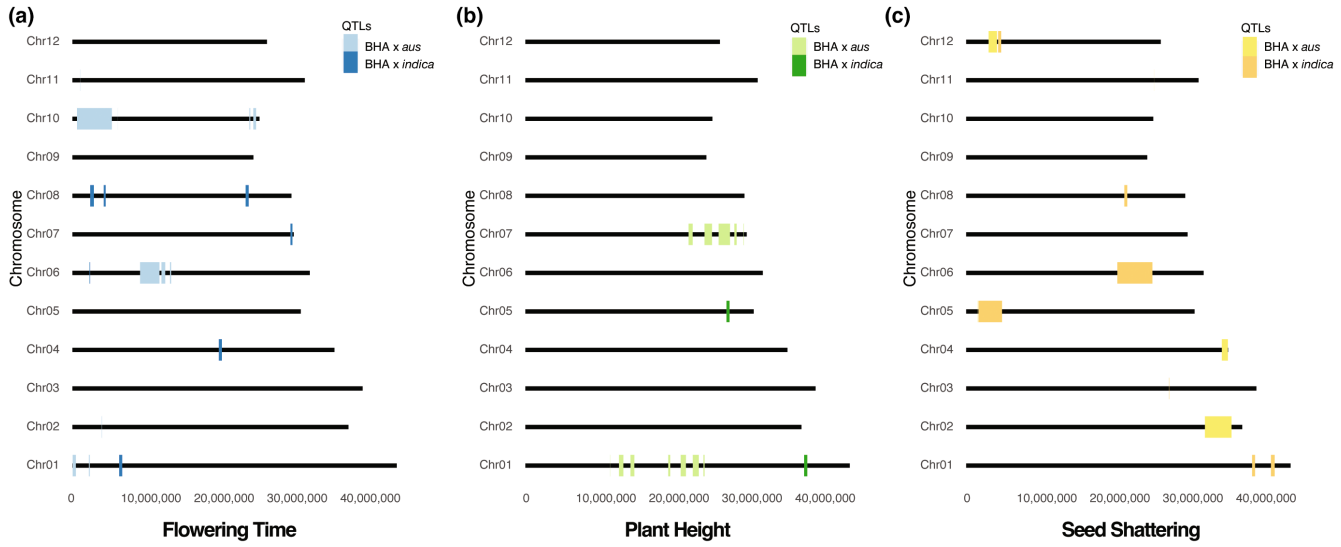

FIGURE 4 Distributions of the Quantitative Trait Loci (QTLs) identified in the crosses of BHA×aus and BHA×indica for the three weedy traits, flowering time, plant height and seed shattering along the chromosomes. (a) Flowering time; (b) plant height and (c) seed shattering. The coloured rectangles represent the QTLs along the chromosome. [Colour figure can be viewed at wileyonlinelibrary.com]

the Nipponbare genome (Kawahara et al., 2013) a *temperate japonica* variety which was the first rice genome sequenced. The advent of high-quality de novo assembled rice genomes (Du et al., 2017) now provides the opportunity to map reads to an *indica* reference. *indica* and *japonica* rice are highly differentiated (Du et al., 2017) and likely came from different domestication events (Choi et al., 2017; Fuller et al., 2009, 2010; Gross & Zhao, 2014). The *indica* reference differs markedly to the *japonica* in having a longer assembly and a large number of structure variations (Du et al., 2016). Both BHA and *aus* are more closely related to *indica* (Li et al., 2017). The effects of reference genome can be discerned in the mapping rate: we earlier attempted to align one bulked sequencing sample to the *japonica* reference genome, but only 65% reads were mapped; the mapping rate vastly increased to more than 90% when we aligned it to the *indica* reference. A poor mapping rate can impact the number of genetic markers for QTL mapping and thus affect QTL detection. We also found more selective sweeps when we realigned the BHA and *aus* population genomes to the *indica* reference, which may be explained by the correct alignment of previously unmapped or wrongly mapped reads. Wrongly mapped reads can even potentially inflate estimates of nucleotide diversity; lower nucleotide diversity was estimated for both the BHA and *aus* populations when we aligned their genomes to the *indica* reference genome than to the *japonica* reference genome. We hope our study will spark discussions about the impact of parental and reference genome choice on evolutionary inference and QTL detection.

AUTHOR CONTRIBUTIONS

A.L.C. designed the study. X.L., S.Z., S.P. and Y.J. performed the research. X.L. analysed the data. D.L., C.H. and J.C. helped with analyses. X.L. and A.L.C. wrote the manuscript. All authors read and approved the final manuscript.

ACKNOWLEDGEMENTS

This work was supported by NSF IOS-1947609 to A. Caicedo, K. Olsen and Y. Jia and Lotta Crabtree Fellowships in Production Agriculture to X. Li and Gilgut Scholarships to X. Li.

CONFLICT OF INTEREST STATEMENT

The authors declare no conflict of interest.

DATA AVAILABILITY STATEMENT

Raw reads of the six bulks were submitted to the NCBI (National Center for Biotechnology Information) Short Read Archive (SRA) under the project of RPJNA961674.

BENEFIT SHARING STATEMENT

A collaborative research effort was established between scientists from academic institutions and government agency. Results of the research have been shared with the broader scientific community. The research addresses an important topic for important questions in evolution, particularly regarding convergent evolution and the evolution of weediness.

ORCID

Xiang Li https://orcid.org/0000-0003-0257-6110
Ana L. Caicedo https://orcid.org/0000-0002-0378-6374

REFERENCES

Bagavathiannan, M. V., Graham, S., Ma, Z., Barney, J. N., Coutts, S. R., Caicedo, A. L., de Clerck-Floate, R., West, N. M., Blank, L., Metcalf, A. L., Lacoste, M., Moreno, C. R., Evans, J. A., Burke, I., & Beckie, H. (2019). Considering weed management as a social dilemma bridges individual and collective interests. *Nature Plants*, *5*(4), 343–351.

Bin Rahman, A. N. M. R., & Zhang, J. (2022). Trends in rice research: 2030 and beyond. *Food and Energy Security*, *12*, e390. https://doi.org/10.1002/fes3.390

Bolger, A. M., Lohse, M., & Usadel, B. (2014). Trimmomatic: A flexible trimmer for Illumina sequence data. *Bioinformatics*, *30*(15), 2114–2120.

Bres-Patry, C., Lorieux, M., Clément, G., Bangratz, M., & Ghesquière, A. (2001). Heredity and genetic mapping of domestication-related traits in a temperate japonica weedy rice. *Theoretical and Applied Genetics*, *102*(1), 118–126.

Burgos, N. R., Norman, R. J., Gealy, D. R., & Black, H. (2006). Competitive N uptake between rice and weedy rice. *Field Crops Research*, *99*(2), 96–105.

Burgos, N. R., Singh, V., Tseng, T. M., Black, H., Young, N. D., Huang, Z., Hyma, K. E., Gealy, D. R., & Caicedo, A. L. (2014). The impact of herbicide-resistant rice technology on phenotypic diversity and population structure of United States weedy rice. *Plant Physiology*, *166*(3), 1208–1220.

Casartelli, A., Riewe, D., Hubberten, H. M., Altmann, T., Hoefgen, R., & Heuer, S. (2018). Exploring traditional aus-type rice for metabolites conferring drought tolerance. *Rice*, *11*(1), 9.

Chauhan, B. S., & Johnson, D. E. (2011). Competitive interactions between weedy rice and cultivated rice as a function of added nitrogen and the level of competition. *Weed Biology and Management*, *11*(4), 202–209.

Choi, J. Y., Platts, A. E., Fuller, D. Q., Hsing, Y.-I., Wing, R. A., & Purugganan, M. D. (2017). The rice paradox: Multiple origins but single domestication in Asian rice. *Molecular Biology and Evolution*, *34*(4), 969–979.

Cingolani, P., Platts, A., Wang, L. L., Coon, M., Nguyen, T., Wang, L., Land, S. J., Lu, X., & Ruden, D. M. (2012). A program for annotating and predicting the effects of single nucleotide polymorphisms, SnpEff: SNPs in the genome of *Drosophila melanogaster* strain w1118; iso-2; iso-3. *Fly*, *6*(2), 80–92.

Civáň, P., Craig, H., Cox, C. J., & Brown, T. A. (2015). Three geographically separate domestications of Asian rice. *Nature Plants*, *1*, 15164.

Danecek, P., Auton, A., Abecasis, G., Albers, C. A., Banks, E., DePristo, M. A., Handsaker, R. E., Lunter, G., Marth, G. T., Sherry, S. T., McVean, G., Durbin, R., & 1000 Genomes Project Analysis Group. (2011). The variant call format and VCFtools. *Bioinformatics*, *27*(15), 2156–2158.

Das, S., Parida, S. K., Agarwal, P., & Tyagi, A. K. (2019). Transcription factor OsNF-YB9 regulates reproductive growth and development in rice. *Planta*, *250*(6), 1849–1865.

Ding, X., Cao, Y., Huang, L., Zhao, J., Xu, C., Li, X., & Wang, S. (2008). Activation of the indole-3-acetic acid–amido synthetase GH3-8 suppresses expansin expression and promotes salicylate- and jasmonate-independent basal immunity in Rice. *The Plant Cell*, *20*(1), 228–240.

Du, D., Gao, X., Geng, J., Li, Q., Li, L., Lv, Q., & Li, X. (2016). Identification of key proteins and networks related to grain development in wheat (*Triticum aestivum* L.) by comparative transcription and proteomic analysis of allelic variants in TaGW2-6A. *Frontiers in Plant Science*, *7*, 922.

Du, H., Yu, Y., Ma, Y., Gao, Q., Cao, Y., Chen, Z., Ma, B., Qi, M., Li, Y., Zhao, X., Wang, J., Liu, K., Qin, P., Yang, X., Zhu, L., Li, S., & Liang, C. (2017). Sequencing and de novo assembly of a near complete indica rice genome. *Nature Communications*, *8*, 15324.

Durand-Morat, A., Nalley, L. L., & Thoma, G. (2018). The implications of red rice on food security. *Global Food Security*, *18*, 62–75.

Ellstrand, N. C., Heredia, S. M., Leak-Garcia, J. A., Heraty, J. M., Burger, J. C., Yao, L., Nohzadeh-Malakshah, S., & Ridley, C. E. (2010). Crops gone wild: Evolution of weeds and invasives from domesticated ancestors. *Evolutionary Applications*, *3*(5–6), 494–504.

Falster, D. S., & Westoby, M. (2003). Plant height and evolutionary games. *Trends in Ecology & Evolution*, *18*(7), 337–343.

Fujino, K., Yamanouchi, U., & Yano, M. (2013). Roles of the Hd5 gene controlling heading date for adaptation to the northern limits of rice cultivation. *Theoretical and Applied Genetics*, *126*(3), 611–618.

Fuller, D. Q., Qin, L., Zheng, Y., Zhao, Z., Chen, X., Hosoya, L. A., & Sun, G.-P. (2009). The domestication process and domestication rate in rice: Spikelet bases from the lower Yangtze. *Science*, *323*(5921), 1607–1610.

Fuller, D. Q., Sato, Y.-I., Castillo, C., Qin, L., Weisskopf, A. R., Kingwell-Banham, E. J., Song, J., Ahn, S.-M., & van Etten, J. (2010). Consilience of genetics and archaeobotany in the entangled history of rice. *Archaeological and Anthropological Sciences*, *2*(2), 115–131.

Gao, H., Jin, M., Zheng, X.-M., Chen, J., Yuan, D., Xin, Y., Wang, M., Huang, D., Zhang, Z., Zhou, K., Sheng, P., Ma, J., Ma, W., Deng, H., Jiang, L., Liu, S., Wang, H., Wu, C., Yuan, L., & Wan, J. (2014). *Days to heading 7*, a major quantitative locus determining photoperiod sensitivity and regional adaptation in rice. *Proceedings of the National Academy of Sciences of the United States of America*, *111*(46), 16337–16342.

Gross, B. L., & Zhao, Z. (2014). Archaeological and genetic insights into the origins of domesticated rice. *Proceedings of the National Academy of Sciences of the United States of America*, *111*(17), 6190–6197.

Gu, X.-Y., Kianian, S. F., Hareland, G. A., Hoffer, B. L., & Foley, M. E. (2005). Genetic analysis of adaptive syndromes interrelated with seed dormancy in weedy rice (*Oryza sativa*). *Theoretical and Applied Genetics*, *110*(6), 1108–1118.

Hartl, D. L., & Clark, A. G. (1997). *Principles of population genetics*. Sinauer Associates.

He, W., Chen, C., Xiang, K., Wang, J., Zheng, P., Tembrock, L. R., Jin, D., & Wu, Z. (2021). The history and diversity of Rice domestication as resolved from 1464 complete plastid genomes. *Frontiers in Plant Science*, *12*, 781793.

Hoyos, V., Plaza, G., Li, X., & Caicedo, A. L. (2020). Something old, something new: Evolution of Colombian weedy rice (*Oryza* spp.) through de novo de-domestication, exotic gene flow, and hybridization. *Evolutionary Applications*, *13*(8), 1968–1983.

Huang, Z., Kelly, S., Matsuo, R., Li, L.-F., Li, Y., Olsen, K. M., Jia, Y., & Caicedo, A. L. (2018). The role of standing variation in the evolution of weedines traits in South Asian weedy rice (*Oryza* spp.). *G3 Genes|Genomes|Genetics*, *8*(11), 3679–3690.

Huang, Z., Rodriguez, J. A., Matsuo, R., & Caicedo, A. L. (2021). Assessing physiological and genetic evidence for evolution of shared weedy rice traits at the vegetative growth stage. *Frontiers in Agronomy*, *3*, 601414. https://doi.org/10.3389/fagro.2021.601414

Huang, Z., Young, N. D., Reagon, M., Hyma, K. E., Olsen, K. M., Jia, Y., & Caicedo, A. L. (2017). All roads lead to weediness: Patterns of genomic divergence reveal extensive recurrent weedy rice origins from South Asian Oryza. *Molecular Ecology*, *26*(12), 3151–3167.

Hui, L., Liu, D., Wang, Y., Li, S., Yin, L., & Wang, S. (2022). Overexpression of rice monogalactosyldiacylglycerol synthase OsMGD leads to enhanced salt tolerance in rice. *Agronomy*, *12*(3), 568.

Imaizumi, T., Ebana, K., Kawahara, Y., Muto, C., Kobayashi, H., Koarai, A., & Olsen, K. M. (2021). Genomic divergence during feralization reveals both conserved and distinct mechanisms of parallel weediness evolution. *Communications Biology*, *4*(1), 952.

Izawa, T., Mihara, M., Suzuki, Y., Gupta, M., Itoh, H., Nagano, A. J., Motoyama, R., Sawada, Y., Yano, M., Hirai, M. Y., Makino, A., & Nagamura, Y. (2011). Os-GIGANTEA confers robust diurnal rhythms on the global transcriptome of rice in the field. *The Plant Cell*, *23*(5), 1741–1755.

Kassambara, A. (2022). *ggpubr: "ggplot2" Based publication ready plots*. R package version 0.6.0. https://rpkgs.datanovia.com/ggpubr/

Kawahara, Y., de la Bastide, M., Hamilton, J. P., Kanamori, H., McCombie, W. R., Ouyang, S., Schwartz, D. C., Tanaka, T., Wu, J., Zhou, S., Childs, K. L., Davidson, R. M., Lin, H., Quesada-Ocampo, L., Vaillancourt, B., Sakai, H., Lee, S. S., Kim, J., Numa, H., … Matsumoto, T. (2013). Improvement of the *Oryza sativa* Nipponbare reference genome using next generation sequence and optical map data. *Rice*, *6*(1), 4.

Li, C., Zhou, A., & Sang, T. (2006). Rice domestication by reducing shattering. *Science*, *311*(5769), 1936–1939.

Li, H. (2011). A statistical framework for SNP calling, mutation discovery, association mapping and population genetical parameter estimation from sequencing data. *Bioinformatics*, *27*(21), 2987–2993.

Li, H., & Durbin, R. (2009). Fast and accurate short read alignment with Burrows–Wheeler transform. *Bioinformatics*, *25*(14), 1754–1760.

Li, L., Li, Y., Jia, Y., Caicedo, A. L., & Olsen, K. M. (2017). Signatures of adaptation in the weedy rice genome. *Nature Genetics*, *49*(5), 811–814.

Li, X., Zhang, S., Amaro-Blanco, I., Perera, S., Khandekar, N. S., Lowey, D., Osuna, M. D., & Caicedo, A. L. (2022). Multiple compensatory mutations contribute to the de-domestication of Iberian weedy rice. *Plants, People, Planet*, *4*, 499–510. https://doi.org/10.1002/ppp3.10286

Liu, F., Wang, P., Zhang, X., Li, X., Yan, X., Fu, D., & Wu, G. (2018). The genetic and molecular basis of crop height based on a rice model. *Planta*, *247*(1), 1–26.

Moner, A. M., Furtado, A., & Henry, R. J. (2020). Two divergent chloroplast genome sequence clades captured in the domesticated rice gene pool may have significance for rice production. *BMC Plant Biology*, *20*(1), 472.

Nemoto, Y., Nonoue, Y., Yano, M., & Izawa, T. (2016). Hd1,a CONSTANS ortholog in rice, functions as an Ehd1 repressor through interaction with monocot-specific CCT-domain protein Ghd7. *The Plant Journal: For Cell and Molecular Biology*, *86*(3), 221–233.

Oerke, E.-C. (2006). Crop losses to pests. *The Journal of Agricultural Science*, *144*(1), 31–43.

Qi, X., Liu, Y., Vigueira, C. C., Young, N. D., Caicedo, A. L., Jia, Y., Gealy, D. R., & Olsen, K. M. (2015). More than one way to evolve a weed: Parallel evolution of US weedy rice through independent genetic mechanisms. *Molecular Ecology*, *24*(13), 3329–3344.

Qiu, J., Jia, L., Wu, D., Weng, X., Chen, L., Sun, J., Chen, M., Mao, L., Jiang, B., Ye, C., Turra, G. M., Guo, L., Ye, G., Zhu, Q.-H., Imaizumi, T., Song, B.-K., Scarabel, L., Merotto, A., Jr., Olsen, K. M., & Fan, L. (2020). Diverse genetic mechanisms underlie worldwide convergent rice feralization. *Genome Biology*, *21*(1), 70.

Qiu, J., Zhou, Y., Mao, L., Ye, C., Wang, W., Zhang, J., Yu, Y., Fu, F., Wang, Y., Qian, F., Qi, T., Wu, S., Sultana, M. H., Cao, Y.-N., Wang, Y., Timko, M. P., Ge, S., Fan, L., & Lu, Y. (2017). Genomic variation associated with local adaptation of weedy rice during de-domestication. *Nature Communications*, *8*(1), 1–12.

R Core Team. (2022). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing. https://www.R-project.org/

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

Reagon, M., Thurber, C. S., Olsen, K. M., Jia, Y., & Caicedo, A. L. (2011). The long and the short of it: SD1 polymorphism and the evolution

of growth trait divergence in U.S. weedy rice. *Molecular Ecology*, *20*(18), 3743–3756.

Sales, M. A., Shivrain, V. K., Burgos, N. R., & Kuk, Y. I. (2008). Amino acid substitutions in the acetolactate synthase gene of red rice (*Oryza sativa*) confer resistance to imazethapyr. *Weed Science*, *56*(4), 485–489.

Sheng, C., Song, S., Zhou, R., Li, D., Gao, Y., Cui, X., Tang, X., Zhang, Y., Tu, J., Zhang, X., & Wang, L. (2021). QTL-seq and transcriptome analysis disclose major QTL and candidate genes controlling leaf size in sesame (*Sesamum indicum* L.). *Frontiers in Plant Science*, *12*, 580846.

Shivrain, V. K., Burgos, N. R., Scott, R. C., Gbur, E. E., Estorninos, L. E., & McClelland, M. R. (2010). Diversity of weedy red rice (*Oryza sativa* L.) in Arkansas, U.S.A. in relation to weed management. *Crop Protection*, *29*(7), 721–730.

Spielmeyer, W., Ellis, M. H., & Chandler, P. M. (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*(13), 9043–9048.

Subudhi, P. K., Singh, P. K., DeLeon, T., Parco, A., Karan, R., Biradar, H., Cohn, M. A., & Sasaki, T. (2014). Mapping of seed shattering loci provides insights into origin of weedy rice and rice domestication. *The Journal of Heredity*, *105*(2), 276–287.

Sun, J., Ma, D., Tang, L., Zhao, M., Zhang, G., Wang, W., Song, J., Li, X., Liu, Z., Zhang, W., Xu, Q., Zhou, Y., Wu, J., Yamamoto, T., Dai, F., Lei, Y., Li, S., Zhou, G., Zheng, H., … Chen, W. (2019). Population genomic analysis and de novo assembly reveal the origin of weedy Rice as an evolutionary game. *Molecular Plant*, *12*(5), 632–647.

Takagi, H., Abe, A., Yoshida, K., Kosugi, S., Natsume, S., Mitsuoka, C., Uemura, A., Utsushi, H., Tamiru, M., Takuno, S., Innan, H., Cano, L. M., Kamoun, S., & Terauchi, R. (2013). QTL-seq: Rapid mapping of quantitative trait loci in rice by whole genome resequencing of DNA from two bulked populations. *The Plant Journal: For Cell and Molecular Biology*, *74*(1), 174–183.

Takahashi, Y., Teshima, K. M., Yokoi, S., Innan, H., & Shimamoto, K. (2009). Variations in Hd1 proteins, *Hd3a* promoters, and *Ehd1* expression levels contribute to diversity of flowering time in cultivated rice. *Proceedings of the National Academy of Sciences of the United States of America*, *106*(11), 4555–4560.

Thurber, C. S., Jia, M. H., Jia, Y., & Caicedo, A. L. (2013). Similar traits, different genes? Examining convergent evolution in related weedy rice populations. *Molecular Ecology*, *22*(3), 685–698.

Thurber, C. S., Reagon, M., Gross, B. L., Olsen, K. M., Jia, Y., & Caicedo, A. L. (2010). Molecular evolution of shattering loci in U.S. weedy rice. *Molecular Ecology*, *19*(16), 3271–3284.

Thurber, C. S., Reagon, M., Olsen, K. M., Jia, Y., & Caicedo, A. L. (2014). The evolution of flowering strategies in US weedy rice. *American Journal of Botany*, *101*(10), 1737–1747.

Vanneste, S., & Friml, J. (2009). Auxin: A trigger for change in plant development. *Cell*, *136*(6), 1005–1016.

Vigueira, C. C., Olsen, K. M., & Caicedo, A. L. (2013). The red queen in the corn: Agricultural weeds as models of rapid adaptive evolution. *Heredity*, *110*(4), 303–311.

Wedger, M. J., Roma-Burgos, N., & Olsen, K. M. (2022). Genomic revolution of US weedy rice in response to 21st century agricultural technologies. *Communications Biology*, *5*(1), 885.

Wei, X., Xu, J., Guo, H., Jiang, L., Chen, S., Yu, C., Zhou, Z., Hu, P., Zhai, H., & Wan, J. (2010). DTH8 suppresses flowering in rice, influencing plant height and yield potential simultaneously. *Plant Physiology*, *153*(4), 1747–1758.

Wen, J., Jiang, F., Weng, Y., Sun, M., Shi, X., Zhou, Y., Yu, L., & Wu, Z. (2019). Identification of heat-tolerance QTLs and high-temperature stress-responsive genes through conventional QTL mapping, QTL-seq and RNA-seq in tomato. *BMC Plant Biology*, *19*(1), 398.

Wu, D., Qiu, J., Sun, J., Song, B.-K., Olsen, K. M., & Fan, L. (2022). Weedy rice, a hidden gold mine in the paddy field. *Molecular Plant*, *15*(4), 566–568.

Wu, W., Zheng, X.-M., Lu, G., Zhong, Z., Gao, H., Chen, L., Wu, C., Wang, H. J., Wang, Q., Zhou, K., Wang, J.-L., Wu, F., Zhang, X., Guo, X., Cheng, Z., Lei, C., Lin, Q., Jiang, L., Wang, H., … Wan, J. (2013). Association of functional nucleotide polymorphisms at DTH2 with the northward expansion of rice cultivation in Asia. *Proceedings of the National Academy of Sciences of the United States of America*, *110*(8), 2775–2780.

Xia, L., Zou, D., Sang, J., Xu, X., Yin, H., Li, M., Wu, S., Hu, S., Hao, L., & Zhang, Z. (2017). Rice expression database (RED): An integrated RNA-Seq-derived gene expression database for rice. *Journal of Genetics and Genomics*, *44*(5), 235–241.

Yan, W. H., Wang, P., Chen, H. X., Zhou, H.-J., Li, Q. P., Wang, C. R., Ding, Z. H., Zhang, Y. S., Yu, S. B., Xing, Y. Z., & Zhang, Q. F. (2011). A major QTL, Ghd8, plays pleiotropic roles in regulating grain productivity, plant height, and heading date in rice. *Molecular Plant*, *4*(2), 319–330.

Yano, M., Katayose, Y., Ashikari, M., Yamanouchi, U., Monna, L., Fuse, T., Baba, T., Yamamoto, K., Umehara, Y., Nagamura, Y., & Sasaki, T. (2000). Hd1, a major photoperiod sensitivity quantitative trait locus in rice, is closely related to the Arabidopsis flowering time gene CONSTANS. *The Plant Cell*, *12*(12), 2473–2484.

Zhao, Q., Feng, Q., Lu, H., Li, Y., Wang, A., Tian, Q., Zhan, Q., Lu, Y., Zhang, L., Huang, T., Wang, Y., Fan, D., Zhao, Y., Wang, Z., Zhou, C., Chen, J., Zhu, C., Li, W., Weng, Q., … Huang, X. (2018). Pan-genome analysis highlights the extent of genomic variation in cultivated and wild rice. *Nature Genetics*, *50*(2), 278–284.

Zhou, J., Yang, Y., Lv, Y., Pu, Q., Li, J., Zhang, Y., Deng, X., Wang, M., Wang, J., & Tao, D. (2022). Interspecific hybridization is an important driving force for origin and diversification of Asian cultivated Rice *Oryza sativa* L. *Frontiers in Plant Science*, *13*, 932737.

Zhou, Y., Lu, D., Li, C., Luo, J., Zhu, B.-F., Zhu, J., Shangguan, Y., Wang, Z., Sang, T., Zhou, B., & Han, B. (2012). Genetic control of seed shattering in rice by the APETALA2 transcription factor shattering abortion1. *The Plant Cell*, *24*(3), 1034–1048.

Ziska, L. H., Gealy, D. R., Burgos, N., Caicedo, A. L., Gressel, J., Lawton-Rauh, A. L., Avila, L. A., Theisen, G., Norsworthy, J., Ferrero, A., Vidotto, F., Johnson, D. E., Ferreira, F. G., Marchesan, E., Menezes, V., Cohn, M. A., Linscombe, S., Carmona, L., Tang, R., & Merotto, A. (2015). Chapter three-weedy (red) rice: An emerging constraint to global rice production. In D. L. Sparks (Ed.), *Advances in Agronomy* (Vol. *129*, pp. 181–228). Academic Press.

SUPPORTING INFORMATION

Additional supporting information can be found online in the Supporting Information section at the end of this article.

How to cite this article: Li, X., Zhang, S., Lowey, D., Hissam, C., Clevenger, J., Perera, S., Jia, Y., & Caicedo, A. L. (2023). A derived weedy rice × ancestral cultivar cross identifies evolutionarily relevant weediness QTLs. *Molecular Ecology*, *32*, 5971–5985. https://doi.org/10.1111/mec.17172