Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

Predictive models and wild progenitor-derived introgression lines dissect the population dynamics of domestication syndrome traits associated with crop mimicry in weedy rice

Swayamsiddha Aswita Dhal
Central Rice Research Institute
Anilkumar Chandrappa
Central Rice Research Institute
Debashree Dalai
Central Rice Research Institute
Anowar Hossain
Central Rice Research Institute
Anarghya Brahmankodu Kumar
Central Rice Research Institute
Virender Kumar
International Rice Research Institute
Soham Ray
Indian Agricultural Research Institute
Kshirod Kumar Jena
International Rice Research Institute
Nimai Prasad Mandal
ICAR-CRRI-Central Rainfed Upland Rice Research Station
Divya Balakrishnan
Indian Institute of Rice Research
Kutubuddin A Molla
Central Rice Research Institute
Pratik Satya
Central Research Institute for Jute and Allied Fibres
Chinmay Kumar Pradhan
Utkal University
Trilochan Mohapatra
Protection of Plant Varieties and Farmers' Rights Authority
Dipti Ranjan Pani
ICAR-National Bureau of Plant Genetic Resources, Base Centre

Meera Kumari Kar
Central Rice Research Institute
Mridul Chakraborti
mridul.chakraborti@icar.org.in

Central Rice Research Institute

Research Article

Keywords: Synteny, Conspecific weeds, Rotational cultivation, Weedy rice evolution, Tissue pigmentation

Posted Date: April 24th, 2026

DOI: https://doi.org/10.21203/rs.3.rs-9327164/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

Additional Declarations: No competing interests reported.

Abstract

Weedy-rice, causing 15-90% yield loss of rice across continents, is a conspecific weed originating from natural outcrossing between *Oryza sativa* and its wild-progenitors. Bidirectional gene flow and phenotypic mimicry of cultivated ecotypes in early growth stages enable evasion of manual weeding, and facilitate rapid proliferation in cultivated ecosystems. Their sustainable management requires understanding of population dynamics in a crop-weed continuum which is challenging in field-populations due to genetic admixtures and multidimensional selection pressures. To resolve this, we generated bulked backcross-derived inbred lines (BILs) and chromosome-segment substitution lines (CSSLs) from *Oryza rufipogon* using syntenic microsatellite and SNP markers.Leaf-sheath pigmentation and short-term seed dormancy traits were fine-mapped and BILs mimicking weedy-rice populations were selected for awns, seed shattering, and dormancy, but differing in leaf-sheath pigmentation and plant height. A near-isogenic line (NIL) with purple leaf sheath was also developed to facilitate application of bidirectional selection pressure. Simulated models of population dynamics under different levels of gene flow (5 and 10%) and selection pressures were developed after baseline-surveys in natural populations. Using an experimental population of the weedy-derivatives and the recurrent parent or its purple-sheathed NIL at 40:60 ratio, population dynamics under different degree and directions of selection were characterized in field trials. While resurgence of weedy-derivatives was 59.20–64.27% in the absence of selection, it could be restricted to 16.27–19.60% by applying unidirectional selection pressure and further to 3.94–4.33% through bidirectional selection. These findings highlight the effectiveness of population-dynamics guided breeding and agronomic strategies to reduce weedy-rice load.

Key message

Synteny-guided chromosome segment substitution lines were developed from at reduced cost and efforts through an improved methodology and utilized to genetically dissect an evolution-guided strategy for weedy rice management.

Introduction

Weedy rice is one of the most serious threats to rice production causing 15–90% yield loss across continents (Roma-Burgos et al. 2021). Due to close genomic affinity and cross-compatibility with cultivated rice ecotypes, high seed-shattering, rich phenotypic diversity and aggressive competitive ability, it has become a persistent weed in diverse rice ecosystems (Liang et al., 2025). The problem has become more serious with adoption of mechanized direct seeding and harvesting, reduced availability of water for crop management, and cultivation of varieties with narrow genetic base (Singh et al. 2013; Cong et al., 2025). Management options like manual weeding and conventional herbicide application are ineffective due to phenotypic resemblance and close phylogenetic relationship. Although herbicide tolerant (HT) rice can address the problem, the high cross-compatibility of weedy rice with both genetically modified (GM) and non-GM HT-rice cultivars may sometimes lead to escape of the trait to weedy rice due to natural outcrossing. Before the extensive cultivation of semi-dwarf high yielding rice

varieties, farmers practicing direct seeding in India managed the weedy rice load below economic threshold levels through rotational cultivation of tall landraces having purple-pigmented and green leaf sheath, thereby selectively eliminating green- and purple leaf-sheathed weedy rice, respectively (Richharia 1964; Tewari 2008). However, almost all the present-day high yielding rice varieties lack leaf-sheath pigmentation and the practice of such rotational cultivation involving genotypes showing contrasting visual markers is almost forgotten. Although large scale adoption of transplanting effectively mitigated the threat over a long period of time, gradual build-up of weedy rice population has supported their resurgence even in transplanted fields of India, Japan and Taiwan in recent years (Singh et al. 2013; Imaizumi 2018; Hsu et al. 2025). In the Mekong delta of Vietnam, where yield loss up to 46% due to weedy rice is reported under the predominant direct seeding practice, farmers found it difficult to distinguish weedy rice and manage the problem through their conventional practice of using high seed rates (up to 300 kg ha^{-1}) (Chauhan et al., 2015). Phenotypic resemblance to cultivated varieties, varying degree of seed dormancy, early maturity and shattering habit make weedy rice a formidable and persistent competitor in rice fields. Therefore, understanding their population dynamics in cultivated environments is critical to decipher the adaptive strategies and design effective control measures (Nadir et al. 2017).

Botanically classified as *Oryza sativa* f. *spontanea*, weedy rice is considered to be originated through natural hybridization between cultivated rice (*Oryza sativa*) and its wild progenitors (*Oryza rufipogon* and *Oryza nivara*) through de-domestication and feralization (Vigueira et al., 2020; Cong et al., 2025). Among the two species, *O. rufipogon* is considered as the most common progenitor across geographical regions (Ma et al. 2008; Roma-Burgos et al. 2021) Such reverse-domestication process can be monitored by change in the key domestication-syndrome characters, such as dormancy, seed shattering, awn, plant height, and photoperiod sensitivity. These traits define the continuum linking wild, cultivated, and weedy forms under different selection regimes and shape their adaptive trajectories (Delouche et al. 2007; Huang et al. 2012). In contrast with the rice domestication, which was principally guided by positive artificial selection for few major traits, the de-domestication and feralization of weedy rice is principally affected by independent natural selection in different regions and time periods under cultivated or semi-cultivated environments. The effect of natural selection under such complex situations cannot be interpolated from the traditional gene flow models developed by assuming unrestricted natural environment (Migdałek and Żelawski 2022). Hence developing an effective model to understand weedy rice dynamics requires information about basal population structure and the associated changes occurring in both short and long term.

Gene flow between two species can be monitored by identification of interspecific recombinants containing fragments of genomes from either of the species (Migdałek and Zelawski 2022). Since weedy rice originate from natural hybridization of *O. sativa* and *O. rufipogon*, their interspecific derivatives containing fragments of *O. rufipogon* genome in the background of *O. sativa* and harbouring alleles governing weedy forms of domestication-syndrome traits can be effectively utilized to understand the population dynamics of weedy rice. Although *O. rufipogon*–derived populations have been widely exploited for yield enhancement and stress resilience (Xiao et al. 1998; Xie et al. 2007; Ge et al. 2024),

their potential to resolve the genetic architecture of crop-mimicry driven through domestication syndrome traits remains underexplored. Chromosome segment substitution lines (CSSLs) and backcross-derived inbred lines (BILs) with crop-mimicking de-domestication traits can therefore serve as a proxy for weedy rice. Syntenic, cross-transferable markers enabling systematic tracking of introgressed segments (Dalai et al. 2026) can further support the genetic dissection of those traits in uniform genetic backgrounds. In this study, we developed CSSLs and BILs of *O. rufipogon* in the genetic background of cultivated rice using a novel bulked backcrossing strategy guided by syntenic markers, and subsequently identified lines mimicking weedy rice. Selected BILs and NILs were further utilized to interpret phenotype-based selection as an evolutionary process influencing allele frequencies and population structure along the crop–weed continuum and to design future breeding strategies for sustainable management of this recalcitrant invasive weed.

Materials and methods

The study simultaneously involved development of introgression lines from wild progenitor, computational simulations for predicting population dynamics of weedy rice, and further utilization of both the resources to genetically dissect the population dynamics of domestication syndrome traits associated with crop mimicry. A schematic summary of the overall workflow has been presented in Fig. 1.

Development of introgression lines from wild progenitor

Plant materials

Four *O. rufipogon* accessions from India (IRGC 81885, AC 100444, AC 100015) and Bangladesh (IRGC 103404), along with 14 diverse *O. sativa* cultivars were utilized in the study (Supplementary Table 1). A set of CSSLs, SSSLs and BILs developed in an earlier study (Dalai et al. 2026) were included for validation purposes.

Identification of syntenic markers

A core set of 1K cross-amplifiable STMS markers for the *O. sativa* complex (1K-CoreCaOA) was previously developed as a resource for CSSL development across A-genome species (Dalai et al. 2026). From that basic dataset, common markers among *O. sativa* ssp. *japonica*, *O. sativa* ssp. *indica*, and *O. rufipogon* were selected. Markers showing intra- or inter-chromosomal translocations in one or more species or subspecies were excluded to select only a syntenic set.

Molecular marker profiling

Genomic DNA isolation and molecular marker profiling were carried out following standard protocols (Dalai et al. 2026). For pooled analysis, equal concentrations of DNA from three plants were combined after quantification. SNP genotyping was outsourced (Intertek India, Hyderabad) using the 1K rice custom amplicon (1k-RiCA) panel (Arbelaez et al. 2019).

Selection of cross combination for higher seed-setting

Approximately 1000 spikelets of the 14 *O. sativa* cultivars were emasculated for each cross combination, and separately pollinated with the four wild accessions to generate 56 interspecific hybrids. After hybridity confirmation through polymorphic markers, ~ 1000 emasculated spikelets of each recurrent parent were pollinated with bulked pollens of F_1 plants to generate BC_1F_1 progenies. Progenies from all cross combinations were characterized for the qualitative characters from a DUS descriptor (Shobha Rani et al. 2004) and compared with respective recurrent parents. In the crosses showing higher seed set with all the *O. rufipogon* accessions, BC_1F_1 plants were further characterized using one polymorphic marker per chromosome. In other crosses, marker analysis was restricted to phenotypically similar plants. Progenies morphologically resembling the recurrent parent, and not displaying heterozygosity for any of the 12 markers, were considered as selfing derived plants.

Modified breeding scheme for development of syntenic marker–anchored BILs and CSSLs

A population-based backcrossing strategy was adopted for BIL and CSSL development, replacing conventional methods that require extensive pedigree maintenance and genotyping in every generation (Fig. 2). The recurrent parent from the selected cross combination, chosen for higher seed set, was repeatedly backcrossed with bulked pollen collected from BC_1F_1 and BC_2F_1 progenies flowering on the same day.

In BC_3F_1, a total of 282 plants were randomly selected, and ninety-four pooled DNA samples (three plants per pool) were genotyped using 120 syntenic polymorphic STMS markers. The three plants from each pool were subsequently genotyped using only the heterozygous markers identified in the respective pools. A subset of plants representing all markers across the 12 chromosomes through multiple overlapping segments was selected. One panicle from each selected BC_3F_1 plant was sown, and panicle-to-progeny rows were raised in BC_3F_2. Eight plants per row were randomly chosen and advanced through the single-panicle descent method up to BC_3F_6, and one plant from each line was selected as a BIL.

Following SNP genotyping using the 1K-RiCA platform, a genetic map was constructed using IciMapping v4.1 (Meng et al. 2015). Markers with > 5% missing data or showing significant segregation distortion were excluded from linkage map construction. Genetic distances were calculated using the Kosambi mapping function (Kosambi 1943). Collinearity between the genetic and physical maps was assessed for the SNP markers and further validated through targeted profiling of the BILs with syntenic STMS

markers. Overlapping CSSLs anchored by both SNP and STMS markers were identified using the CSSL Finder tool (Lorieux 2012) and visualized through RStudio version 4.5.2.

Dissection and mapping of key domestication syndrome traits

Leaf sheath pigmentation and seed dormancy were dissected using BILs, CSSLs, single segment substitution line (SSSLs), and advanced substitution lines derived from the same cross combination generated in the present and a previous study (Dalai et al. 2026). Phenotypic variation was assessed, overlapping introgressed segments were compared, and common genomic intervals were identified. Recombinants within the target regions were further used for substitution mapping (Nguyen et al. 2021) to delimit the minimal marker intervals.

Mapping of leaf sheath pigmentation

All BILs and a previously developed CSSL set (Dalai et al. 2026) were phenotyped for purple versus green leaf sheath. Shared introgressed segments among lines with similar phenotypes were identified, and recombinants were used for fine mapping. An SSSL with a pigmented leaf sheath was crossed with the recurrent parent to generate ~ 2000 IC_1F_2 plants. A random subset was genotyped using two flanking STMS markers, and the genotype of IC_1F_2 plants for the gene controlling the trait was determined based on the segregation pattern of their $IC_1F_{2:3}$ progenies. Lines segregating for leaf sheath pigmentation were classified as heterozygous, whereas non-segregating lines were classified as homozygous. The allelic data of flanking markers and trait was utilized to calculate genetic distance. A non-segregating, purple-sheathed recombinant with upright tillers was selected as a NIL.

Mapping of seed dormancy

All BILs, the viviparous recurrent parent, and two non-viviparous checks (Swarna and CR 1014) were grown in an augmented field design to screen for seed dormancy. Germination tests were conducted with three biological replicates. Primary panicles from 20 plants flowering on the same day were tagged, and seeds were harvested 30 days later. Germination was assessed on the day of harvest and at 7, 14, and 21 days thereafter. The remaining seeds were dried, stored for three months, and retested with and without heat treatment (48°C for seven days) to distinguish dormancy from non-viability.

Lines showing no germination up to seven days after harvest but exceeding 50% germination within 21 days were validated across seasons. Overlapping introgressed segments were compared to identify associated marker intervals. Selected BILs carrying the target interval were crossed with the recurrent parent, reciprocal homozygous recombinants were identified and evaluated further. Previously developed CSSLs and BILs (Dalai et al. 2026) were further evaluated to reconfirm marker–trait associations. An SSSL carrying the target locus was crossed with the recurrent parent; IC_1F_2 progenies were space-planted and phenotyped for germination. Samples representing extreme and intermediate classes were

genotyped, and homozygous $IC_1F_{2:3}$ progenies for the candidate marker were comprehensively characterized for germination percentage over time.

Prediction of population dynamics of weedy rice

Baseline data collection

A total of 36 paddy fields (≥ 4000 m²) with varying levels of weedy rice infestation were sampled (Supplementary Table 2). Additionally, four fields with complete invasion by weedy rice swarms were selected. Observations were recorded at the booting to flowering stage of the crop. From each field, 100 randomly selected weedy rice plants were scored for the presence and length of awns [short (≤ 20 mm), medium (> 20–40 mm), or long (> 40 mm)], shattering habit (present/absent), and leaf sheath pigmentation (purple or green). Leaf sheath pigmentation of the rice varieties grown continuously over the years or under varietal rotation in the respective fields during the preceding three years was also recorded.

Computational simulations

Temporal changes in purple- and green–leaf sheath phenotypes in weedy rice were simulated using a probability-based, discrete-generation model implemented in R (R Core Team 2021). Model parameters were defined based on published reports and field surveys (Supplementary Table 3). Inheritance of the trait was assumed to follow two alternative genetic models: (i) single-gene control with complete dominance (F_2 segregation 3:1, purple: green), and (ii) complementary gene action (9:7), each present in equal proportion in the base population. The mating structure considered the predominantly autogamous nature of rice, assuming 90–95% selfing and 5–10% outcrossing per generation. The base population composition was set at a 70:30 ratio of crop to weedy rice for each season. Initial dimorphism in weedy rice was fixed at 30:70 for purple and green leaf sheaths, respectively.

Simulation methods were initially developed based on a single-locus model and later extended to incorporate two-gene interactions. Simulations were conducted with 100 replications. Model validation was performed by comparing the simulated decline in heterozygosity under partial selfing with theoretical expectations and by reproducing Hardy–Weinberg equilibrium under complete outcrossing maintained at constant allele frequencies across generations. This two-pronged strategy confirmed the correctness of the algorithms implemented in the simulation based on mating and segregation. The decline in heterozygosity was evaluated using the equation $H_t=H_0(1 - s/2)^t$, where 'H_t' represents heterozygosity at the t^{th} generation, 'H_0' is the heterozygosity in the base population, and 's' denotes the rate of selfing.

Simulations were conducted for both pure weedy rice swarms and a crop–weed continuum, assuming that the cultivated variety was uniformly purple or green and that each year, the crop was grown in the same field using the seed harvested from previous season without selection bias. Selection against plants phenotypically contrasting with the crop was modelled under two extreme scenarios: no selection

(0%) or complete unidirectional selection (100%). Phenotypic and allelic frequencies were updated in each generation based on genotypic frequencies through iterative computation (100 nos.) under different selection regimes. Mutation, migration, random genetic drift, penetrance, expressivity and any other factors affecting allele frequencies were assumed to act uniformly on both phenotypic classes.

Field testing of weedy rice population dynamics

Selection of BILs

All the BILs were initially characterized for the presence of awns. Lines possessing medium to long awns were further evaluated for seed shattering. BILs exhibiting spontaneous seed shattering along with short-term seed dormancy induced by the mapped gene were shortlisted. The selected lines were then genotyped using sd1(h), considered as a perfect marker for the *sd1* gene (Ellis and Spielmeyer 2002), and further classified according to leaf sheath pigmentation.

Using combinations of *sd1* alleles and leaf sheath pigmentation, four BILs were selected to represent contrasting adaptive phenotypic configurations of weedy rice. To constitute the mixed populations, seeds from each of the four selected BILs (10% each) were thoroughly combined with seeds of CR Dhan 307 or its purple leaf sheath NIL (60%) on a seed-count basis. Prior to mixing, all seeds were sun-dried and stored for six months.

Simulated field trial

The selected experimental plot had no history of rice cultivation during the preceding two seasons. Pre-sowing irrigation was applied before final land preparation. To control non-*Oryza* weed species, bispyribac sodium and a commercial formulation of florpyrauxifen-benzyl (2.13%) + cyhalofop-butyl (10.64% EC) were applied as early- and late-post-emergence herbicides, respectively, at recommended doses in all plots during both seasons.

The entire field was divided into two main plots, one following manual weeding (MP1) and the other without manual weeding (MP2). During season-1 (wet season 2024), each main plot was randomly subdivided into four subplots (SP1–SP4) comprising: (i) CR Dhan 307, (ii) purple leaf sheath NIL, (iii) a seed mixture of 60% CR Dhan 307 and 40% awned BILs, and (iv) a seed mixture of 60% purple leaf sheath NIL and 40% awned BILs. Subplots were bunded on all sides, and sowing was performed at a 20 cm row distance after leaving a 60 cm buffer zone from the bunds. At maturity, plants possessing awns were allowed to shatter seeds naturally. Manual shaking of panicles was also carried out before harvest to ensure uniform dispersal of remaining seeds within a sub-plot. All sides of the respective subplots were guarded with shade nets during shaking. Subsequently, all crop residues, including panicles, were removed from the field. After four weeks, the field was manually prepared using spades. During this gap period, all subplots were covered with shade nets to minimize seed movement by birds, squirrels, or rainfall.

During season-2 (dry season 2025), each subplot with residual shattered seeds from the previous season was further divided into two sub-subplots (SSP1 and SSP2), where CR Dhan 307 or the purple leaf sheath NIL were randomly sown. Three rows were skipped between the two sub-subplots in every sub-plot. Within MP1, all seedlings emerging between rows were manually removed during weeding, irrespective of their leaf sheath pigmentation. Additionally, any plant showing a distinct leaf sheath colour from the respective variety within the sub-subplot rows was removed up to 45 days after sowing during both seasons. Proportions of green- and purple-leaf-sheathed plants in every subplot or sub-subplot were recorded at 60 and 75 days after sowing. The number of hills bearing awned panicles was counted in each subplot (season-1) and sub-subplot (season-2) after flowering. For each subplot or sub-subplot, three randomly selected grids of 6 m^2 were sampled and treated as replications.

Statistical analysis and data representation

Statistical analyses were carried out using R (R Core Team 2021). The split-plot experiment was analyzed by considering replication, main plots, subplots, and sub-subplots according to the design structure. Treatment effects and their interactions were tested using appropriate error terms. Mean comparisons among genotypes and treatment combinations were performed using Tukey's honestly significant difference (HSD) test at the 5% significance level with the agricolae package. CSSL representation with anchored-markers, boxplots and violin plots were prepared in RStudio (version 4.5.2). Graphical genotyping was performed using GGT 2.0 (Villanueva et al. 2023) and MG2C (Chao et al. 2021).

Results

Selection of cross combination for introgression line development

Among the 56 cross combinations and their backcross progenies tested in the study, CR Dhan 307 recorded substantially higher F_1 and BC_1F_1 seed set with all four accessions of *O. rufipogon*. Seed setting was lowest with Heera, followed by CR1014. Variation was also recorded among the different *O. rufipogon* accessions. Better seed setting was observed for both AC100444 and AC100015 in all the crosses (Supplementary Fig. 1).

Identification of syntenic STMS markers

A total of 8,975 common cross-amplifiable markers were identified among *O. sativa* ssp. *japonica*, *O. sativa* ssp. *indica*, and *O. rufipogon*, of which 265 were non-syntenic (Fig. 3A, 3B; Supplementary Datasheet 1). Only one marker, AUT23344, showed intra-chromosomal translocation in *O. rufipogon*, where it was positioned at 7.27 Mb compared with ~ 24.16 Mb in *japonica* and 25.71 Mb in *indica*. All other non-syntenic markers exhibited inter-chromosomal translocations in either *indica* or *O. rufipogon*. Chromosome 11 harboured the highest number of non-syntenic markers (76), followed by chromosome 12 (41), with most translocations in these chromosomes observed in *indica*. The lowest number of

translocations was recorded on chromosome 9. In general, non-syntenic markers showed translocations in either *indica* or *O. rufipogon*. A rare exception was RM23936, located on chromosome 9 of *japonica*, chromosome 3 in *indica*, and chromosome 4 of *O. rufipogon*. Three closely linked markers (RM12185, RM20789, and RM17271) near the short arm telomeric region of chromosome 7 were found on chromosome 4 of both *indica* and *O. rufipogon* at similar physical positions. The syntenic markers were only considered for further studies (Fig. 3C; Supplementary Datasheet 2).

Development of synteny-guided BILs and identification of CSSLs

The CR Dhan 307/AC100444 cross combination was selected for BIL and CSSL development. All 398 polymorphic SNP markers detected between the parents could be traced among the BILs. However, due to significant segregation distortion, 73 markers were not considered for linkage map construction (Fig. 3A). The remaining 325 markers were mapped to 13 linkage groups (Fig. 4A). Only chromosome 3 was represented by two linkage groups due to a lack of sufficient polymorphic SNP markers to provide comprehensive coverage in one region. The marker order in the genetic map matched the physical positions in Nipponbare reference genome. The 120 polymorphic STMS markers previously verified for synteny were combined with the linear SNP markers to generate a physical map of the 12 chromosomes with better genome coverage (Fig. 4B). Targeted STMS marker profiling of the BILs for the introgressed segments of *O. rufipogon* revalidated the linkage map and achieved higher genome coverage. A total of 54 CSSLs comprehensively covering all the polymorphic marker intervals across the 12 chromosomes were identified (Fig. 4C). While chromosomes 1, 4, and 5 were represented by six CSSLs each, only two CSSLs covered chromosomes 2 and 10. All the CSSLs had at least one overlapping marker with their preceding and succeeding CSSLs without any gap. The CSSLs were named COr-B-CSSL (**C**R Dhan 307 and **O***ryza* **r***ufipogon* **B**ulk **C**hromosome **S**egment **S**ubstitution **L**ines). Each CSSL was further hyphenated with the chromosome number followed by the CSSL number for that chromosome. COr-B-CSSL-1-1 represents the first CSSL for chromosome 1, whereas COr-B-CSSL-1-6 represents the sixth CSSL of the same chromosome. Overall, the CSSLs covered ~ 358.54 Mb genomic region detected through all the polymorphic syntenic markers identified between parents. All the CSSLs also retained a few additional segments from the *O. rufipogon* genome (Supplementary Fig. 2).

Mapping and validation of locus for leaf sheath pigmentation

Among the 54 CSSLs, the only common introgressed region shared by the pigmented lines was located on chromosome 6 (Fig. 5A, 5B). Further survey of introgressed segments of the remaining 354 BILs delimited the trait to the interval between Chr06_4757948 and RM253. Two BILs showing recombination within this interval displayed contrasting phenotypes, one purple and the other green. The segregating IC_1F_2 progenies derived from the cross of SSSL-U2-6 and CR Dhan 307 mapped the leaf sheath pigmentation locus at a distance of 0.42 cM from RM253. Colocalization with reported genes and QTLs identified *OsC1* (*LOC_Os06g10350*) as the most likely candidate gene homologue. Substantial variation in

the intensity and onset of pigmentation was recorded among the homozygous progenies for RM253 (Fig. 5A).

In the IC_1F_2 generation of SSSL-U2-6/CR Dhan 307, 1492 plants showed pigmented leaf sheath and 503 were green, fitting the 3:1 segregation ratio ($\chi^2 = 0.048$, $p > 0.05$) expected for a single dominant gene. While 499 green-sheathed progenies were erect like CR Dhan 307, 1478 plants with purple leaf sheath were semi-spreading like SSSL-U2-6. The 14 erect and pigmented plants were selected, and in the $IC_1F_{2:3}$ generation, one line showing robust seedling growth, pigmentation from the early developmental stage, and no segregation for pigmentation or upright tillering was selected as the NIL (equivalent to BC_6F_3). Based on STMS and SNP allele profiles, the NIL showed 98.40% background recovery.

During identification of syntenic marker among *O. sativa* ssp. *japonica*, *O. sativa* ssp. *indica*, and *O. rufipogon*, ~ 22.39 syntenic STMS markers per Mb were identified (Supplementary Datasheet 2). Using that dataset, 17 additional syntenic STMS markers were identified between Chr06_4757948 and RM253, spanning a ~ 0.6 Mb physical distance, which may be utilized for further fine mapping after study of polymorphism (Fig. 5C).

Characterization and mapping of locus for seed dormancy

Among the 408 BILs, 56 lines with viable seeds did not show any germination immediately after harvest. Three lines remained dormant even after 140 days of harvest. The remaining 53 lines did not germinate up to seven days. However, within 21 days, > 50% of the seeds germinated (Fig. 6A, 6B). Two closely linked markers, Ghd7-2-02 and MSU7_7_9152479_C-G, were found as common among 49 BILs (Fig. 6C). Reciprocal recombinants between Chr07_5442983 and Ghd7-2-02 or between Ghd7-2-02 and Chr07_14729119, identified from $IC_1BC_1F_2$ progenies derived from three selected BILs, further confirmed the association of Ghd7-2-02 and MSU7_7_9152479_C-G with short-term seed dormancy (STSD) (Fig. 6D). Notably, two STMS markers, RM21263 (~ 7.42 Mb) and RM21323 (~ 8.87 Mb), located within the interval between Chr07_5442983 and Ghd7-2-02/MSU7_7_9152479_C-G, and another polymorphic SNP marker, Chr07_11290725, located at ~ 11.29 Mb, were non-syntenic, and therefore not included in the study (Fig. 3A).

CSSL-L-7-7, SSSL-L-7-7, and 19 other BILs out of 21 with the STSD feature identified from 215 backup BILs developed earlier from the same cross using a different approach (Fig. 2) reconfirmed the linkage of Ghd7-2-02 or MSU7_7_9152479_C-G with a key locus for STSD induction. The contrasting set of 94 plants selected from the IC_1F_2 generation of the SSSL-L-7-7/CR Dhan 307 cross further validated the role of the genomic region associated with Ghd7-2-02 or MSU7_7_9152479_C-G from *O. rufipogon* in STSD induction in the viviparous recurrent parent. All homozygous plants carrying the Ghd7-2-02 allele of *O. rufipogon* showed < 10% germination on the day of harvest and reached > 50% after 21 days. The heterozygous plants for Ghd7-2-02 recorded > 30.0% germination on the day of harvest, whereas the homozygous progenies carrying the Ghd7-2-02 allele of CR Dhan 307 showed 53.33–90.00% germination. The 47 homozygous plants for Ghd7-2-02 were retested in $IC_1F_{2:3}$, and STSD induction was found to be consistent during validation (Fig. 6E).

Prediction of population dynamics of weedy rice for leaf sheath pigmentation

Field surveys detected weedy rice populations under both direct-seeded and transplanted conditions. Except in certain fields of Cuttack (Odisha), where the traditional landrace Kalachampa with purple leaf sheath has been cultivated in recent years, weedy rice populations at other locations were nearly monomorphic for green leaf sheath. At the same location, although weedy rice in cultivated fields predominantly had green leaf sheaths, the invasive swarms in non-cropped areas exhibited both types of leaf sheath pigmentations in significant proportions (Supplementary Table 3).

The predictions showed that, in the absence of any selection pressure against a particular leaf sheath pigmentation, weedy rice populations in natural swarms can maintain dimorphism even when only 30% of the initial population possesses purple leaf sheath. After a short-term decline in purple- and increase in green-sheathed plants, the trends are generally reversed. If the trait is governed by a single dominant gene, irrespective of the outcrossing percentage, the proportions of green- and purple-sheathed plants stabilize rapidly (Fig. 7A_1–A4). When complementary gene action governs the trait, after a steep decline during the first season, the proportion of purple leaf-sheathed plants gradually increases and, at a certain point, both leaf sheath colours reach equal frequencies. Higher outcrossing rates accelerate the process, and the populations gradually stabilize (Fig. 7B1–B4). When single dominant gene and complementary gene action are present in equal magnitudes, both phenotypes reach equal proportions at a higher speed (Fig. 7C_1–C_4).

In a crop–weed continuum without selection pressure, the proportion of weedy rice plants with contrasting leaf sheath pigmentation may change constantly at a very slow rate, provided that a single dominant gene and complementary gene action equally control the inheritance of pigmentation in the base population. A certain proportion of heterozygotes for purple leaf sheath continues to survive in every generation under such conditions (Fig. 8A_1–A_4; 4B_1–B_4). Even under 100% selection pressure to remove all weedy rice plants with contrasting leaf sheath pigmentation, purple-sheathed weedy rice can persist at very low frequency for 25–30 crop seasons/generations among green-sheathed plants comprising both weedy rice and the crop variety (Fig. 8C_1–C_4). Under the opposite scenario, green-sheathed weedy rice plants may be eliminated within 7–8 crop seasons/generations (Fig. 8D_1–D_4). The outcrossing rate significantly influences the population dynamics when selection pressure is not applied. However, under unidirectional selection, the role of the outcrossing rate is negligible.

Selection of weedy derivatives for field study

Among the BILs, substantial variation was recorded for awn morphology (Fig. 9A). The awns were purple or white during flowering. Notably, when the genotype was awned, all plants with purple leaf sheath also exhibited pigmented awns. The majority of the BILs with medium to large awns showed grain shattering habit. All medium- to large-awned BILs with grain shattering features were further screened for Ghd7-2-

02 marker and retested for STSD. We detected 17 such lines among the 623 BILs. Further classification based on *sd1* and RM253 identified four types of awned genotypes with STSD and shattering habit, viz., tall plants with purple leaf sheath; semidwarf plants with purple leaf sheath; tall plants with green leaf sheath; and semidwarf plants with green leaf sheath (Fig. 9B). Considering a 40% basal infestation level, 10% seeds of each of the four BILs were mixed with 60% seeds of CR Dhan 307 or its purple leaf sheath NIL on seed count basis.

Population dynamics of weedy rice under different selection regimes

During season-1 (wet season 2024), almost all the plants were present within their respective rows in different treatments. Within MP1, manual weeding reduced the population of weedy derivatives in both subplots (MP1SP3 and MP1SP4) where mixed populations were sown (Fig. 10A and 10B). However, weedy derivatives with similar leaf sheath colour were present in both subplots. Weedy populations were substantially higher in MP2SP3 and MP2SP4 compared to MP1SP3 and MP1SP4.

During season-2 (dry season 2025), *Oryza* seedlings were also found between the rows in both main plots. Manual weeding removed such plants from MP1. Within MP1SP3, where a mixed population of CR Dhan 307 and its weedy derivatives was grown in season-1, the population of weedy derivatives was substantially reduced in sub-subplot 1 (MP1SP3SSP1), which grew the purple leaf sheathed NIL. However, in MP1SP3SSP2, the population of weedy types was significantly higher. On the contrary, the weedy population in MP1SP4SSP1 was like that in MP1SP3SSP2. The population of awned weedy plants in MP1SP4SSP2 was at par with MP1SP3SSP1. The populations of weedy derivatives in both sub-subplots of MP2SP3 and MP2SP4 were nearly three times higher than those in MP1SP3SSP2 or MP1SP4SSP1 (Fig. 10A, 10B, 10C, and 10D).

Discussion

This study identified, validated, and systematically utilized syntenic STMS and SNP markers in a modified breeding scheme to develop overlapping introgression lines from *O. rufipogon* with comprehensive genome coverage at substantially reduced cost and efforts compared with conventional strategies. Beyond trait discovery and pre-breeding, the introgressed population and marker resources were deployed to dissect crop-weed evolutionary dynamics and to examine the effects of imposing directional selection pressure against adaptive traits of weedy rice.

Synteny-guided introgression as an efficient framework for wide hybridization

Modern pre-breeding programmes require structured populations and strategic deployment of molecular markers, especially during the utilization of wild species. Syntenic cross-transferable markers facilitate precise tracking of wild genomic segments during backcrossing (Ray et al. 2016). However, no marker

system is suitable at every stage of a breeding programme for all laboratories. For example, custom amplicon-based SNP arrays are excellent tools; however, they are suitable only when whole-genome information is required. Moreover, such genotyping services are not available in every country or laboratory at all times. On the contrary, PCR-based markers can be routinely used in laboratories with basic facilities but are not economical to use at a very large scale (Dalai et al. 2021).

The present study systematically combined a modified breeding strategy with both marker systems for CSSL development. Bulk backcrossing without any genotyping until BC_2F_1, genotyping of pooled subsets in BC_3F_1 with a limited set of anchored STMS markers, advancing the population until BC_3F_5 without genotyping, genetic map construction using custom SNP amplicons in BC_3F_6, and selective reuse of STMS markers to fill genomic gaps minimized the genotyping cost, eliminated the need for exhaustive pedigree maintenance, and substantially improved genome coverage compared with conventional strategies (Dalai et al. 2026).

Moreover, ~ 22.5 syntenic STMS markers per Mb was identified between *O. rufipogon* and the two subspecies of Asiatic rice, facilitating high-density coverage of delineated genomic intervals after a locus is mapped. The same population and marker framework were directly utilized to identify loci governing two domestication-related traits. These overlapping introgression lines enabled evaluation of domestication-related traits in a largely uniform genetic background. This is particularly valuable for traits placed at the interface of domestication and de-domestication, which often remain masked in the cultivated gene pool of rice (Li et al. 2017).

Domestication-related traits in the crop-weedy rice continuum

Leaf sheath pigmentation, plant height, seed dormancy, awn development, and grain shattering are critical for the persistence of weedy rice in cultivated fields. Mapping the leaf sheath pigmentation locus near RM253 and its colocalization with *OsC1* homologue align with earlier reports (Xiong et al. 1999; Lorieux et al. 2000; Fan et al. 2008; Chen et al. 2024; Jiang et al. 2024). However, it may be noted that the purple leaf sheath phenotype in rice may occur due to other genes, besides *OsC1*. We have been able to produce purple leaf sheath NILs through independent introgression of both *OsC1* homologue from *O. rufipogon* and *OsDFR* (*LOC_Os01g44260*) gene from a landrace in the same recurrent parent ‘Swarna’ (unpublished data). This also explains why different F_2-segregation ratios (3:1, 9:7 or even complex higher order) are observed in different populations for the trait (Chin et al. 2016). Among all the *OsC1* introgressed BILs of CR Dhan 307, whenever present, the awns were always purple. If the awns were absent, only the apiculus regions were pigmented. This also indicate the possible pleiotropic effect of *OsC1* or tight linkage among the genes controlling those tissue specific pigmentations. Although the *O. rufipogon* homologue ensured pigmentation, variation in the stage and intensity of expression suggests quantitative modulation by additional loci. Dominant inheritance and selection of NILs expressing high pigment intensity at the early growth stage further enhanced its value as a visual marker. The tight coupling-phase linkage between the genes for leaf sheath pigmentation and tiller angle could also be

circumvented through visual selection in large populations. In other cases, high-density syntenic STMS markers may help to reduce linkage drags, which often create a major hurdle in utilization of wild species in crop improvement.

Variable levels of seed dormancy also play a critical role in the persistence of weedy rice populations (Olajumoke et al. 2016). Short-term dormancy enables staggered germination in rice–rice cropping systems, whereas deep dormancy sustains populations under rice–fallow or crop rotations (Singh et al. 2013; Ajaykumar et al. 2025). Use of a semidwarf recurrent parent with viviparous germination, along with selection of grain-shattering BILs with short-term seed dormancy enabled completion of the simulated field experiment within two consecutive seasons. A genic SNP marker Ghd7-2-02 associated with short-term dormancy was colocalized with *Ghd7*, which is known for its role to induce dormancy and prevent vivipary by increasing the abscisic acid (ABA)/gibberellin (GA_3) ratio during rice seed germination (Hu et al. 2021). While *Ghd7-0* is a null (Xue et al. 2008) and *ghd7* is the recessive allele (Hu et al. 2021), *Ghd7-1* and *Ghd7-3* are the two fully functional dominant alleles of the gene (Saito et al. 2019). *Ghd7-2* is a comparatively weaker allele of *Ghd7* (Wang et al. 2021). Although the nearest polymorphic flanking markers on both sides were > 4.0 Mb apart, the trait could still be traced directly to the genic SNP marker Ghd7-2-02 through substitution mapping. This would otherwise have been difficult to precisely locate and quantify through conventional interval mapping approaches as the three other polymorphic markers predicted as closely linked to Ghd7-2-02 as per Nipponbare reference genome, were non-syntenic and showed significant segregation distortion in the population. Another study to map QTLs for seed dormancy from a weedy rice accession ‘Ludao’ identified chromosome 7 as the sole genomic hotspot for the trait (Nguyen et al. 2019).

Evolutionary basis of the traditional weedy rice management practice

Short- and long-term projections of weedy rice population dynamics in natural swarms and crop–weed continuums provided an evolution guided explanation for the traditional practice of rotating landraces with contrasting leaf sheath pigmentation. Simulations indicated that complementary gene action and natural outcrossing allow dimorphism for leaf sheath pigmentation to persist and stabilize, even when uniformly pigmented varieties are cultivated over extended periods. Under unidirectional selection, conspecific weeds phenotypically resembling crop varieties are favored, although the number of generations required for complete elimination of contrasting phenotypes differs significantly between green and purple sheath types. When phenotypically uniform varieties are cultivated in long term, such as modern semi-dwarf varieties with green leaf sheaths, selection promotes convergence toward crop mimicry. In contrast, bidirectional selection disrupts this convergence, cyclically discriminate between crop variety and weedy plants, and thereby enables elimination of weedy types at the vegetative stage. These findings have practical implications in determining the frequency of varietal rotation required in the continuum of a crop and its conspecific weeds.

Field observations were consistent with these projections. The substantial reduction in resurgence of weedy derivatives in the second season after the introduction of contrasting pigmentation indicates that phenotypic divergence may constrain their persistence across generations. Although complete elimination was expected under rigorous weeding up to 45 days, a few plants escaped selection and reached maturity, possibly due to scattered germination or variable expressivity of pigmentation trait during early vegetative stage in few weedy plants. Continuous monitoring and rouging of leftover weedy plants at early flowering stage, therefore, should be an essential component of any sustainable management strategy.

Implications for evolution-guided management under different production systems

Direct seeding creates ecological conditions that favor rapid invasion of weedy rice. Continuous monoculture and exclusive reliance on chemical control can progressively select for adaptive responses against such management strategies. Integrating evolutionary principles into varietal development, however, can offer a sustainable alternative. Breeding varieties or near-isogenic lines that enable creation of artificial selection pressure against weedy derivatives mimicking the crop without dependence on specific herbicides can broaden their applicability. Where herbicide tolerant varieties are possible to cultivate, development of near-isogenic lines of herbicide-tolerant varieties with phenotypic contrasts can help in developing stewardship strategies against gene flow.

Transplanted rice has historically suppressed early emerging weeds through seedling age advantage and puddling. However, our surveys in some parts of India, along with recent reports from Japan (Imaizumi 2018) and Taiwan (Hsu et al. 2025) indicate that weedy rice is steadily invading transplanted paddy fields. Introducing contrasting pigmentation through varietal rotation will also support the removal of weedy derivatives shortly after transplanting. Hence, irrespective of crop establishment method, adoption of evolution-guided management of weedy rice represents an example of ecological intensification. By harnessing the genetic diversity of crop wild relatives to alter population dynamics of conspecific weeds, genomic resources and introgression lines can be utilized beyond trait mapping and marker-assisted breeding.

Conclusion and future perspectives

The study demonstrates how synteny-guided introgression lines derived from a wild progenitor can be used as a platform to integrate genomics with the genetics of crop domestication and apply these insights to crop management and breeding. Strategic deployment of traits to overcome crop mimicry in breeding programmes can create directional selection pressure against a recalcitrant conspecific weed and reduce its fitness. A traditional agronomic practice involving low-yielding landraces was reintroduced as an evolution-guided management strategy suitable for modern direct-seeded systems with semidwarf high-yielding varieties. By integrating wild germplasm, genomic tools, and ecological understanding, the study opens new avenues for sustainable management of conspecific weeds while

expanding the scope for incorporating the diversity of crop wild relatives into modern breeding programs.

Declarations

Compliance with ethical standards

The authors of this publication declare that the experiments conducted for the study have followed the extant national and international laws.

Conflict of interest

It is hereby declared that there is no conflict of interest among the authors of this publication.

Funding

This study was financially supported by Indian Council of Agricultural Research through the “Consortium Research Project on Molecular Breeding (EAP-211)” and ICAR-CRRI in-house project no. 1.3.

Author Contribution

S.A.D., D.D., A.H., A.B.K..: Marker profiling, field works, phenotyping, data recording; A.C., K.A.M., P.S., D.B.: Genetic modelling, data analysis, writing– review & editing; VK, KKJ, NPM, CKP, TM: Guidance, resources, data visualization, writing– review & editing; S.R.: Bioinformatics; D.R.P., M.K.K.: Breeding works, field evaluations, data collection; resources; writing– review & editing; M.C.: Conceptualization, designing experiments, breeding works, bioinformatics, supervision, funding acquisition, formal analysis, data curation. data-visualization, validation, writing- original draft. This work is a part of Ph.D. thesis of the first author. All authors read, reviewed and approved the manuscript.

Acknowledgement

The team is grateful to Post Graduate Department of Botany, Utkal University, India for approving the first author to carry out her research work ICAR-CRRI, Cuttack. We acknowledge the support received from Dr. S.K. Das and Dr. S. Sarkar from ICAR-CRRI, Cuttack for providing resources to synthesize large number of STMS markers. International Rice Research Institute, Philippines provided two wild germplasm accessions used in the study. Administrative support was provided by Director, ICAR-CRRI. Large Language Model (LLM) ChatGPT was used for grammar and language check during initial stage of

manuscript preparation. However, authors take all responsibilities for correctness of sentences/statements.

Data Availability

All the relevant tables, additional data in the form of datasheets or required figures have been provided as electronic supplementary material. Specific additional information, if required, will also be shared on request.

References

1. Ajaykumar R, Sivasabari K, Vigneshwaran R, Kumaresan P (2025) Weedy rice (Oryza sativa f. spontanea) complexes on direct seeded rice (DSR) and its management strategy: A review. Agricultural Reviews 46: 62–69. https://doi.org/10.18805/ag.R-2538
2. Arbelaez JD, Dwiyanti MS, Tandayu E, Lantada K, Jarana A, Ignacio JC, Platten JD, Cobb J, Rutkoski JE, Thomson MJ, Kretzschmar T (2019) 1k-RiCA (1K-Rice Custom Amplicon) a novel genotyping amplicon-based SNP assay for genetics and breeding applications in rice. Rice 12(1):55. https://doi.org/10.1186/s12284-019-0311-0
3. Chao J, Li Z, Sun Y, Aluko OO, Wu X, Wang Q, Liu G (2021) MG2C: a user-friendly online tool for drawing genetic maps. Molecular Horticulture 1(1):16. https://doi.org/10.1186/s43897-021-00020-x
4. Chauhan BS, Namuco OS, Ocampo LAL, Son TTN, Thu TTA, Nam NN, Phuong LN, Bajwa AA (2015) Weedy rice (*Oryza sativa* f. *spontanea*) problems and management in wet direct-seeded rice (*O. sativa* L.) in the Mekong Delta of Vietnam. Crop Protection 78:40–47. https://doi.org/10.1016/j.cropro.2015.08.016
5. Chen HJ, Sawasdee A, Lin YL, Chiang MY, Chang HY, Li WH, Wang CS (2024) Reverse mutations in pigmentation induced by Sodium Azide in the IR64 rice variety. Current Issues in Molecular Biology 46(12):13328–13346. https://doi.org/10.3390/cimb46120795
6. Chin HS, Wu YP, Hour AL, Hong CY, Lin YR (2016) Genetic and evolutionary analysis of purple leaf sheath in rice. Rice 9(1):8. https://doi.org/10.1186/s12284-016-0080-y
7. Cong Y, Gui Y, Yong K, Zhu M, Liu K, Liu J, Qiu J (2025) Deciphering rice feralization: insights from genomics of weedy rice. Genomics Communications 2(1): e007. https://doi.org/10.48130/gcomm-0025-0007
8. Dalai D, Chakraborti M, Mondal TK, Ray S, Kar MK, Chakraborty K, Pani DR, Sarkar S, Bose LK, Behera M, Chattopadhyay K, Deepa, Vijayan J, Dash SK, Pradhan C, Patra BC, Marndi BC (2021) The core set of sequence-tagged microsatellite sites markers between halophytic wild rice *Oryza coarctata* and *Oryza sativa* complex. Euphytica 217(4):57. https://doi.org/10.1007/s10681-021-02790-3
9. Dalai D, Pani DR, Dhal SA, Behera M, Mondal TK, Azaharudheen MTP, Vijayan J, Sarkar D, Ghose P, Roy A, Molla KA, C A, Bose LK, Mohapatra T, Ray S, Kar MK, Chakraborti M (2026) A robust 1K-core marker set for wild germplasm management and targeted pre-breeding of rice: development and

applications. Theoretical and Applied Genetics 139(1):23. https://doi.org/10.1007/s00122-025-05133-1

10. Delouche JC, Burgos NR, Gealy DR, De San Martin GZ, Labrada R, Larinde M, Rosell C (2007) Weedy rices: Origin, biology, ecology, and control. In. FAO Plant Production and Protection Paper No.188. Food and Agriculture Organization (FAO) of the United Nations, Rome, Italy.
11. Ellis MH, Spielmeyer W (2002) “Perfect” Markers for the Semidwarfing Gene *sd1* in Rice. International Rice Research Notes. 27(2):1–1. https://doi.org/10.5281/zenodo.6823534
12. Fan FJ, Fan YY, Du JH, Zhuang JY (2008) Fine mapping of C (Chromogen for anthocyanin) gene in rice. Rice Science 15(1):1–6. https://doi.org/10.1016/s1672-6308(08)60012-8
13. Ge Y, Chen G, Cheng X, Li C, Tian Y, Chi W, Li J, Dai Z, Wang C, Wang C, Duan E, Liu Y, Sun Z, Li J, Wang B, Xu D, Sun X, Zhang H, Zhang W, Wang C, Wang C, Wan J (2024) The superior allele $LEA12^{OR}$ in wild rice enhances salt tolerance and yield. Plant Biotechnology Journal 22(11):2971–2984. https://doi.org/10.1111/pbi.14419
14. Hsu YT, Wang YC, Du PR, Li CP, Wu DH (2025) Morphological diversity and crop mimicry strategies of weedy rice under the transplanting cultivation system. Agronomy 15(4):984. https://doi.org/10.3390/agronomy15040984
15. Hu Y, Song S, Weng X, You A, Yongzhong X (2021) The heading-date gene *Ghd7* inhibits seed germination by modulating the balance between abscisic acid and gibberellins. The Crop Journal 9(2):297–304. https://doi.org/10.1016/j.cj.2020.09.004
16. Huang X, Kurata N, Wei X, Wang Z-X, Wang A, Zhao Q, Zhao Y, Liu K, Lu H, Li W, Guo Y, Lu Y, Zhou C, Fan D, Weng Q, Zhu C, Huang T, Zhang L, Wang Y, Feng L, Furuumi H, Kubo T, Miyabayashi T, Yuan X, Xu Q, Dong G, Zhan Q, Li C, Fujiyama A, Toyoda A, Lu T, Feng Q, Qian Q, Li J, Han B (2012) A map of rice genome variation reveals the origin of cultivated rice. Nature 490(7421):497–501. https://doi.org/10.1038/nature11532
17. Imaizumi T (2018) Weedy rice represents an emerging threat to transplanted rice production systems in Japan. Weed Biology and Management 18(2):99–102. https://doi.org/10.1111/wbm.12146
18. Jiang L, Lyu S, Yu H, Zhang J, Sun B, Liu Q, Mao X, Chen P, Pan D, Chen W, Fan Z, Li C (2024) Transcription factor encoding gene *OsC1* regulates leaf sheath color through anthocyanidin metabolism in *Oryza rufipogon* and *Oryza sativa*. BMC Plant Biology 24(1):147. https://doi.org/10.1186/s12870-024-04823-0
19. Kosambi DD (1943) The estimation of map distances from recombination values. Annals of Eugenics 12(1):172–175. https://doi.org/10.1111/j.1469-1809.1943.tb02321.x
20. Li LF, Li YL, Jia Y, Caicedo AL, Olsen KM (2017) Signatures of adaptation in the weedy rice genome. Nature Genetics 49(5):811–814. https://doi.org/10.1038/ng.3825
21. Liang X, Zhao C, Liu K, Wang W, Huo Z, Song X, Qiang S (2025) Advances in research on the biological characteristics of weedy rice. Plants 14(20):3188. https://doi.org/10.3390/plants14203188

22. Lorieux M (2012) MapDisto: fast and efficient computation of genetic linkage maps. Molecular Breeding 30(2):1231–1235. https://doi.org/10.1007/s11032-012-9706-y
23. Lorieux M, MN. Ndijiondjip MN and Gesquiere A (2000) A first interspecific *Oryza sativa* × *Oryza glaberrima* microsatellite-based genetic linkage map. Theoretical and Applied Genetics 100:593–601. https://doi.org/10.1007/s001229900061
24. Ma D, Li M, Sun J (2008) Studies on biological diversity and genetic differentiation of Liaoning weedy rice. Journal of Shenyang Agricultural University 39 265–269. DOI: 10.1007/s10499-007-9164-4
25. Meng L, Li H, Zhang L, Wang J (2015) QTL IciMapping: Integrated software for genetic linkage map construction and quantitative trait locus mapping in biparental populations. The Crop Journal 3(3):269–283. https://doi.org/10.1016/j.cj.2015.01.001
26. Migdałek G, Żelawski M (2022) Measuring population-level plant gene flow with topological data analysis. Ecological Informatics 70:101740. https://doi.org/10.1016/j.ecoinf.2022.101740
27. Nadir S, Xiong HB, Zhu Q, Zhang XL, Xu HY, Li J, Dongchen W, Henry D, Guo X-Q, Khan S, Suh H-S, Lee DS, Chen L-J (2017) Weedy rice in sustainable rice production. A review. Agronomy for Sustainable Development 37(5):46. https://doi.org/10.1007/s13593-017-0456-4
28. Nguyen CD, Zheng SH, Sanada-Morimura S, Matsumura M, Yasui H, Fujita D (2021) Substitution mapping and characterization of brown planthopper resistance genes from indica rice variety, 'PTB33' (*Oryza sativa* L.). Breeding Science 71(5):497–509. https://doi.org/10.1270/jsbbs.21034
29. Nguyen T, Zhou C, Zhang T, Yu J, Miao R, Huang Y, Zhu X, Song W, Liu X, Mou C, Lan J, Liu S, Tian Y, Zhao Z, Jiang L, Wan J (2019) Identification of QTL for seed dormancy from weedy rice and its application to elite rice cultivar 'Ninggeng 4'. Molecular Breeding 39:123. https://doi.org/10.1007/s11032-019-1031-2
30. Olajumoke B, Juraimi AS, Uddin MK, Husni MHA, Alam MA (2016) Competitive ability of cultivated rice against weedy rice biotypes: A review. Chilean Journal of Agricultural Research 76(2):242–251. https://doi.org/10.4067/s0718-58392016000200015
31. R Core Team (2021) R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. https://www.R-project.org
32. Ray S, Bose LK, Ray J, Ngangkham U, Katara JL, Samantaray S, Behera L, Anumalla M, Singh ON, Chen M, Wing RA, Mohapatra T (2016) Development and validation of cross-transferable and polymorphic DNA markers for detecting alien genome introgression in *Oryza sativa* from *Oryza brachyantha*. Molecular Genetics and Genomics 291(4):1783–1794. https://doi.org/10.1007/s00438-016-1214-z
33. Richharia RH. 1964. Wild rice menace and possible methods of its control. Central Rice Research Institute (CRRI) publication, India.
34. Roma-Burgos N, Sudo MPS, Olsen KM, Werle I, Song BK (2021) Weedy rice (*Oryza spp.*): what's in a name? Weed Science 69:505–513. doi: 10.1017/wsc.2021.22

35. Saito H, Okumoto Y, Tsukiyama T, Xu C, Teraishi M, Tanisaka T (2019) Allelic differentiation at the *E1/GHD7* locus has allowed expansion of rice cultivation area. Plants 8(12):550. https://doi.org/10.3390/plants8120550
36. Shobha Rani N, Shobha Rao LV, Viraktamath BC, Mishra B (2004) National guidelines for the conduct of tests for distinctiveness, uniformity and stability. Directorate of Rice Research, Hyderabad, Telangana, India: 6–13.
37. Singh K, Kumar V, Saharawat YS, Gathala M, Ladha JK, Chauhan BS (2013) Weedy rice: An emerging threat for direct seeded rice production systems in India. Rice Research: Open Access 1:1000106. http://dx.doi.org/10.4172/jrr.1000106
38. Tewari AN (2008) Weedy rice and its management. In: Singh Y, Singh VP, Chauhan B, Orr A, Mortimer AM, Johnson DE, Hardy B (Eds). Direct seeding of rice and weed management in the irrigated rice–wheat cropping system of the Indo-Gangetic Plains. International Rice Research Institute, Los Baños, Philippines, pp: 205–212.
39. Van Berloo R (2008) GGT 2.0: Versatile software for visualization and analysis of genetic data. Journal of Heredity 99(2):232–236. https://doi.org/10.1093/jhered/esm109
40. Varshney JG, Tiwari JP (2008) Studies on weedy rice infestation and assessment of its impact on rice production. Indian Journal of Weed Science 40(3–4):115–123.
41. Vigueira PA, Olsen KM, Wagner CR, Chittick ZB, Vigueira CC (2020) Weedy rice from South Korea arose from two distinct de-domestication events. Frontiers in Agronomy 2:602612 https://doi.org/10.3389/fagro.2020.602612
42. Villanueva G. Rosa-Martínez E, Vilanova S, Gramazio P, Prohens, J, Plazas M (2023) Chromosome visualization with introgressions: a tutorial using tassel and GGT2 software. EDULEARN23 Proceedings. p. 8419–8422. https://doi.org/10.21125/edulearn.2023.2187
43. Wang Q, Su Q, Nian J, Zhang J, Guo M, Dong G, Hu J, Wang R, Wei C, Li G, Wang W, Guo H-S, Lin S, Qian W, Xie X, Qian Q, Chen F, Zuo J (2021) The *Ghd7* transcription factor represses *ARE1* expression to enhance nitrogen utilization and grain yield in rice. Molecular Plant 14(6):1012–1023. https://doi.org/10.1016/j.molp.2021.04.012
44. 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(1):e16189. https://doi.org/10.1371/journal.pone.0016189
45. Xiao J, Li J, Grandillo S, Ahn SN, Yuan L, Tanksley SD, McCouch SR (1998) Identification of trait-improving quantitative trait loci alleles from a wild rice relative, *Oryza rufipogon*. Genetics 150(2):899–909. DOI: 10.1093/genetics/150.2.899
46. Xie X, Jin F, Song MH, Suh JP, Hwang HG, Kim YG, McCouch SR, Ahn SN (2007) Fine mapping of a yield-enhancing QTL cluster associated with transgressive variation in an *Oryza sativa × O. rufipogon* cross. Theoretical and Applied Genetics 116(5):613–622. https://doi.org/10.1007/s00122-007-0695-x

47. Xiong LZ, Lin KD, Dai XK, Xu CG and Zhang Q (1999) Identification of genetic factors controlling domestication-related traits of rice using an F_2 population of a cross between *Oryza sativa* and *O. rufipogon*. Theoretical and Applied Genetics 98: 243–251. https://doi.org/10.1007/s001220051064
48. Xue W, Xing Y, Weng X, Zhao Y, Tang W, Wang L, Zhou H, Yu S, Xu C, Li X, Zhang Q (2008) Natural variation in *Ghd7* is an important regulator of heading date and yield potential in rice. Nature Genetics 40(6):761–767. https://doi.org/10.1038/ng.143

Figures

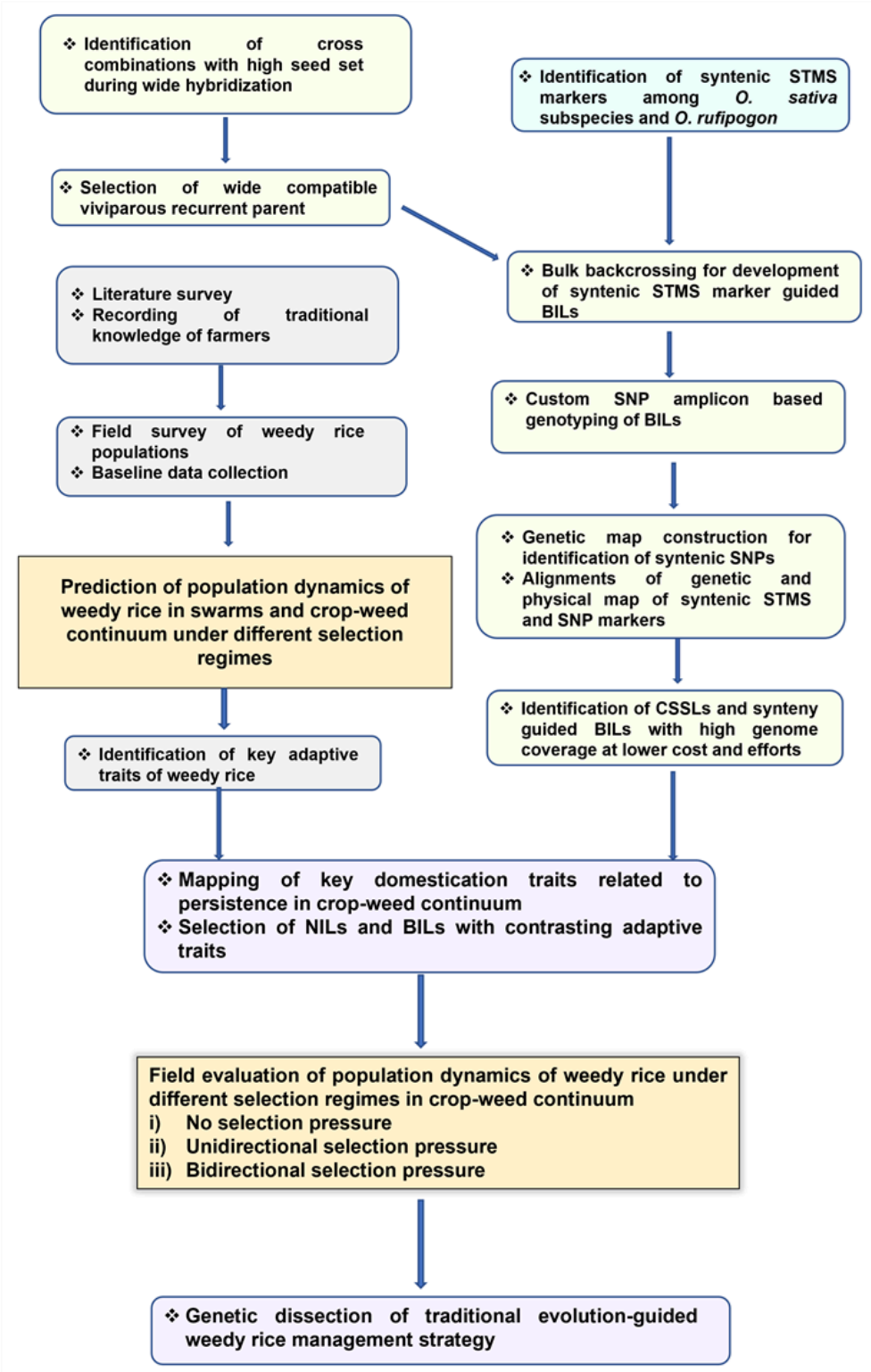


Figure 1

A schematic summary of the overall workflow followed for dissecting the population dynamics of domestication syndrome traits associated with crop mimicry in weedy rice

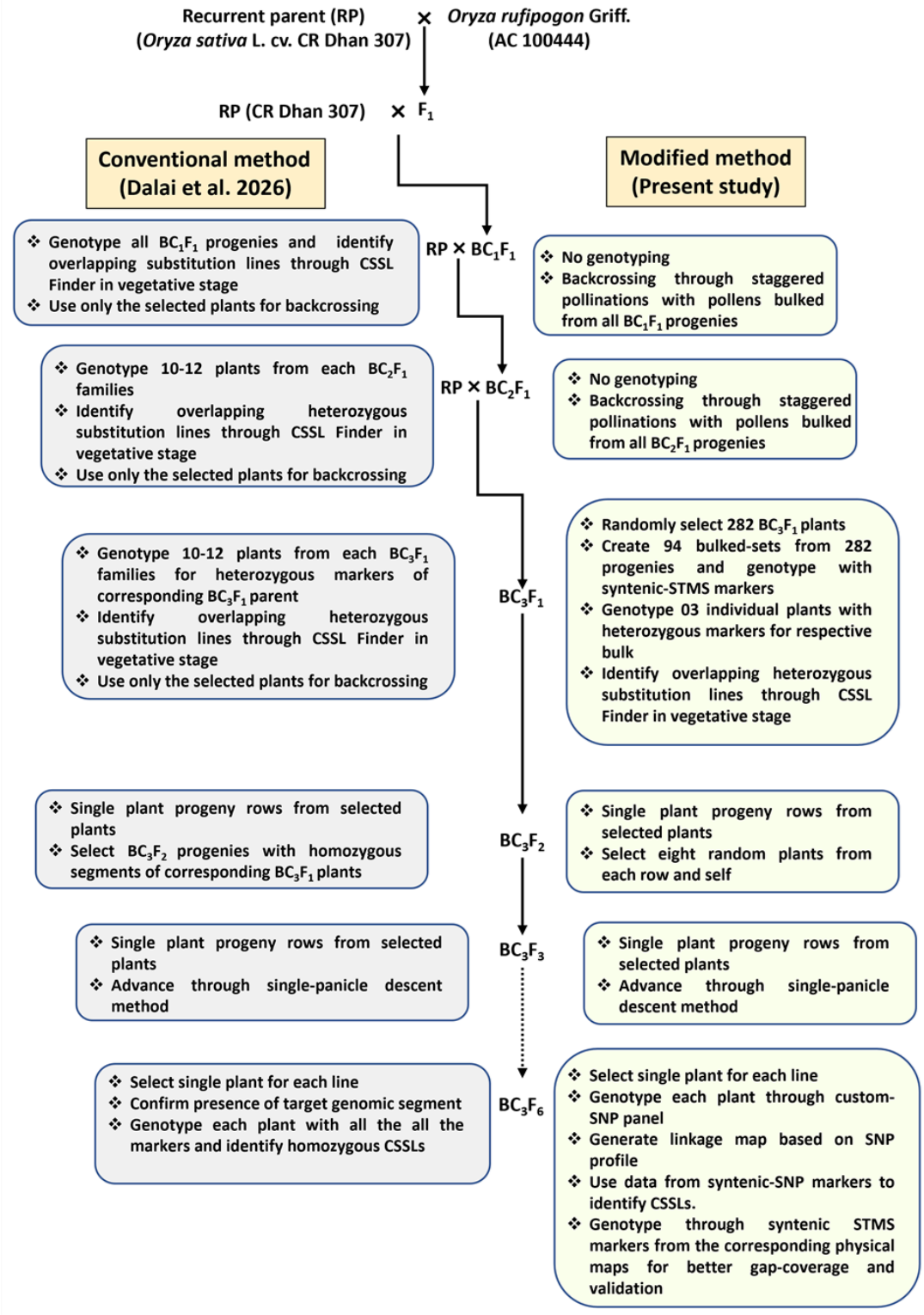


Figure 2

Schematic diagram illustrating the modified methodology followed for developing synteny-guided backcross-derived inbred lines (BILs) and chromosome segment substitution lines (CSSLs) from *O. rufipogon*.

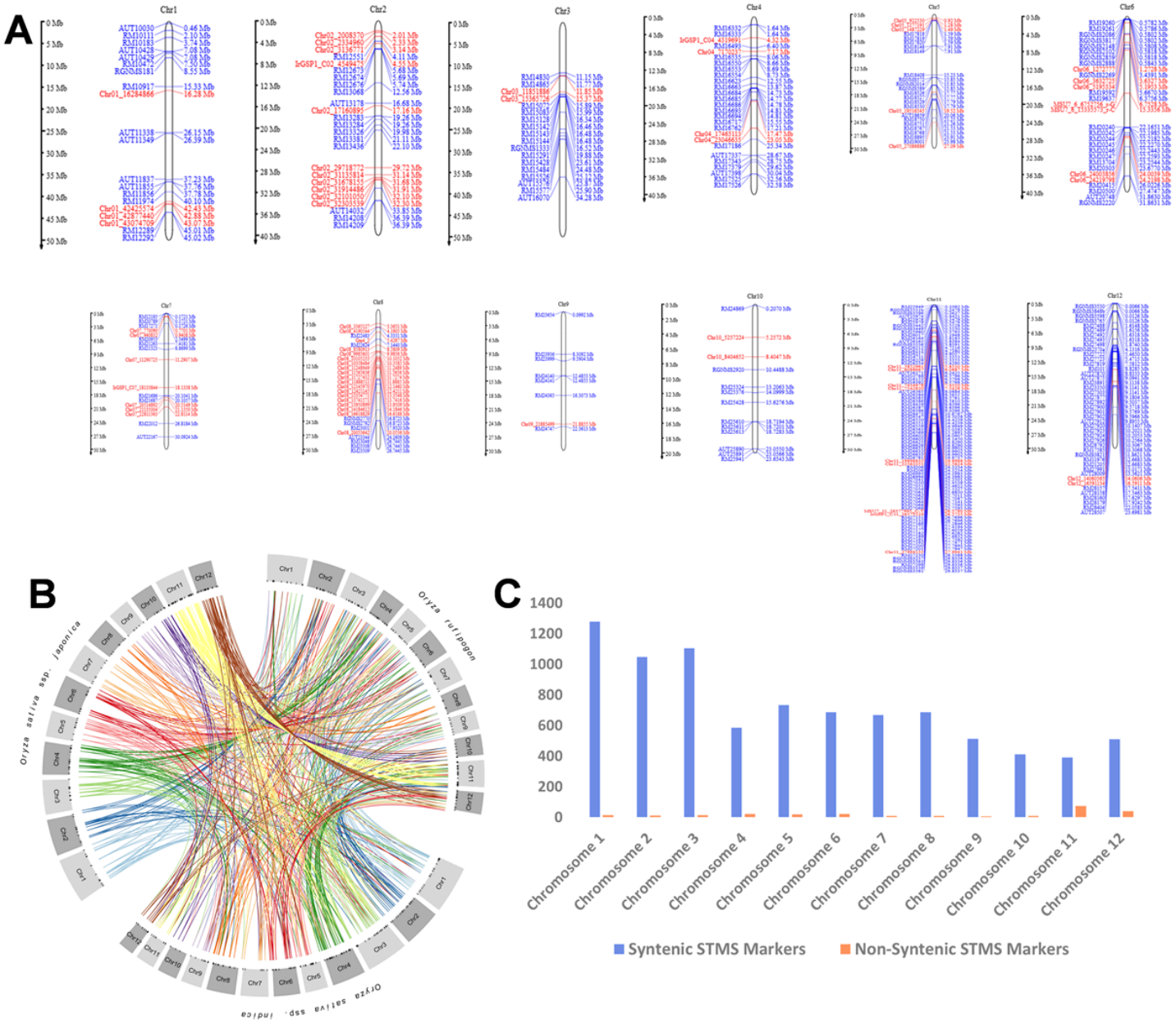


Figure 3

Identification and categorization of common cross-amplifiable STMS markers among *O. sativa* ssp. *japonica*, *O. sativa* ssp. *indica*, and *O. rufipogon into syntenic and non-syntenic classes.* ***A.*** *Non-syntenic STMS markers identified from whole-genome sequences (blue) and polymorphic SNPs from the 1k-RiCA panel showing segregation distortion (red) in the* BC_3F_6 *population of CR Dhan 307 × O. rufipogon accession AC100444.* ***B.*** *Graphical representation of intra- and inter-chromosomal translocations of the non-syntenic STMS markers among O. sativa* ssp. *japonica*, *O. sativa* ssp. *indica*, and *O. rufipogon. Marker names and predicted physical positions in different subspecies or species have been provided in Supplementary Datasheet 1.* ***C.*** *Relative proportions of syntenic and non-syntenic STMS markers identified*

across the 12 chromosomes. Marker names and predicted physical positions of syntenic STMS markers have been provided in Supplementary Datasheets 1 and 2.

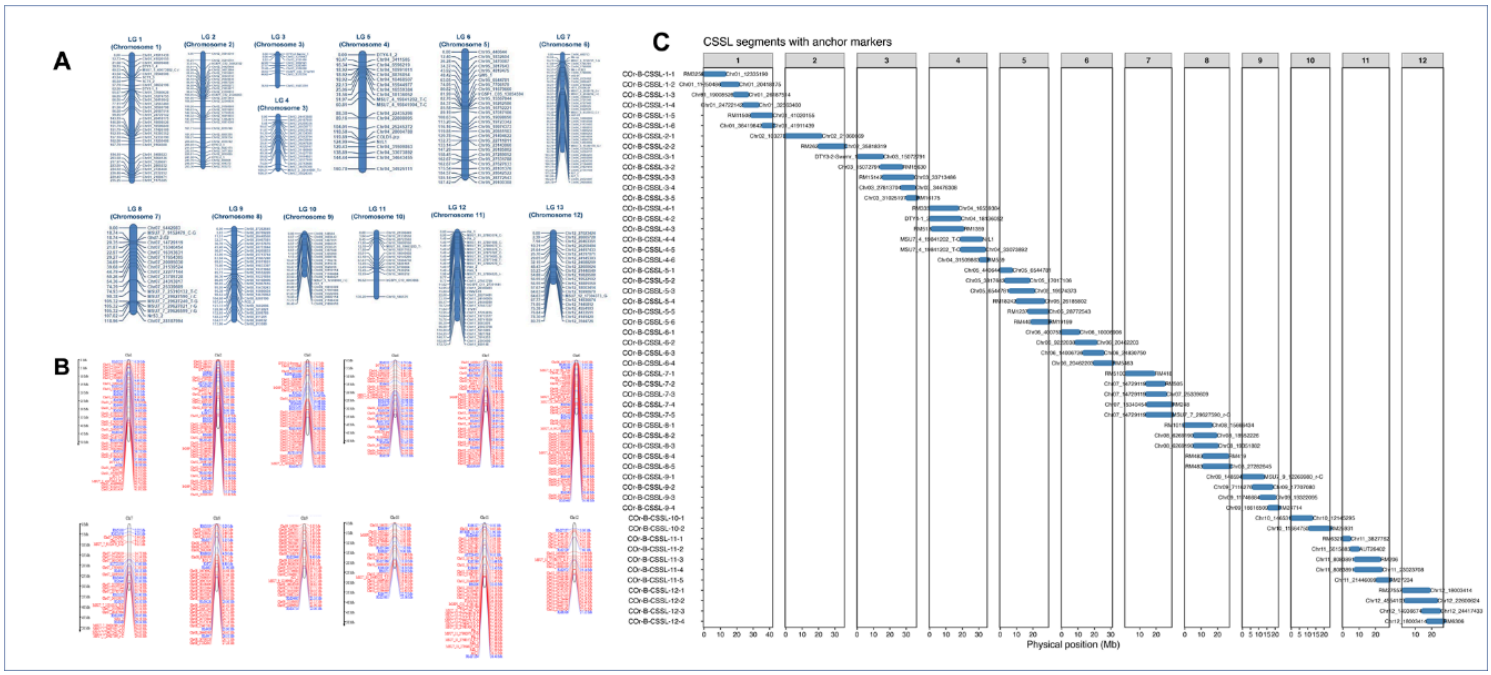


Figure 4

Genetic map of SNP markers and consensus physical map of syntenic STMS and SNP markers used for the development and characterization of backcross-derived inbred lines (BILs) and identification of chromosome segment substitution lines (CSSLs) from the BC_3F_6 population of *CR Dhan 307 × O. rufipogon accession AC100444*. ***A.*** *Genetic map of* 325 SNP markers without segregation distortion. **B.** Consensus physical map showing 120 syntenic and polymorphic STMS markers (blue) and 325 SNP markers (red) between *CR Dhan 307 and O. rufipogon accession AC100444*. ***C.*** *A total of 54 CSSLs were identified from the population, each represented with its left- and right-anchored markers (SNP/STMS).*

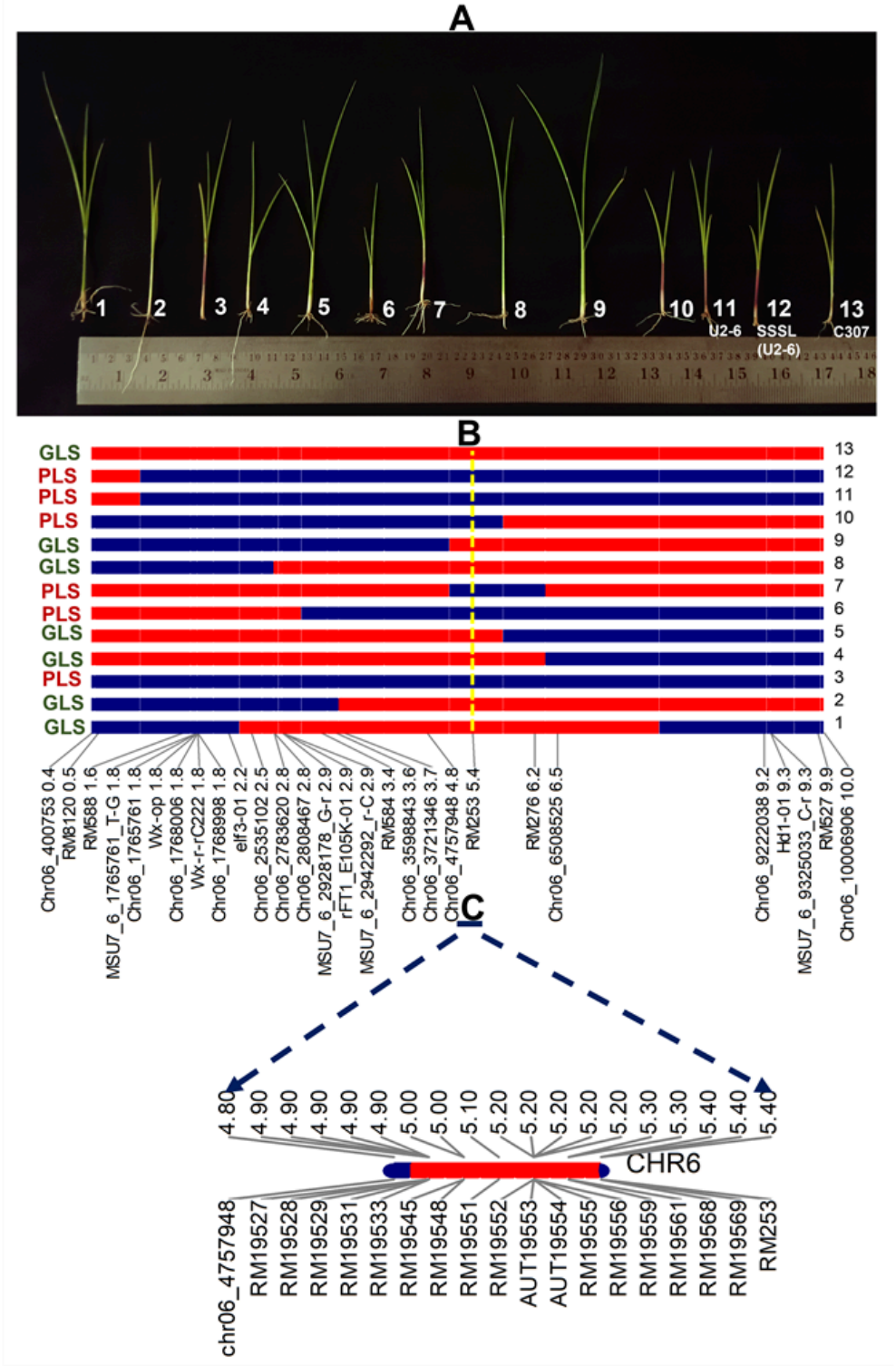


Figure 5

Mapping of the locus for leaf sheath pigmentation using introgression lines, including chromosome segment substitution lines (CSSLs), backcross-derived inbred lines (BILs), and single segment substitution lines (SSSLs) developed in the present study and a previous work (Dalai et al. 2026) from backcross progenies of *CR Dhan 307 × O. rufipogon accession AC100444.* ***A.****Phenotypic variation among selected introgression lines for leaf sheath pigmentation at the early seedling stage.* ***B.*** *Substitution*

mapping with introgression lines delimited the trait to the genomic interval between Chr06_4757948 and RM253 on chromosome 6. Multiple earlier studies with both cultivated and wild rice indicated that one of the gene (*OsC1*) associated with purple leaf sheath of rice is located on chromosome 6 and show a tight linkage with RM253 (Xiong et al. 1999; Lorieux et al. 2000; Fan et al. 2008; Chen et al. 2024). *GLS: Green Leaf Sheath; PLS: Purple Leaf Sheath. Genotype numbers (1–13) in the graphical genotyping diagram correspond to the plant numbers shown in 7A.* **C.** Seventeen additional syntenic STMS markers identified within the interval between Chr06_4757948 and RM253 from syntenic marker dataset (Supplementary Datasheet 2) as further fine mapping resource.

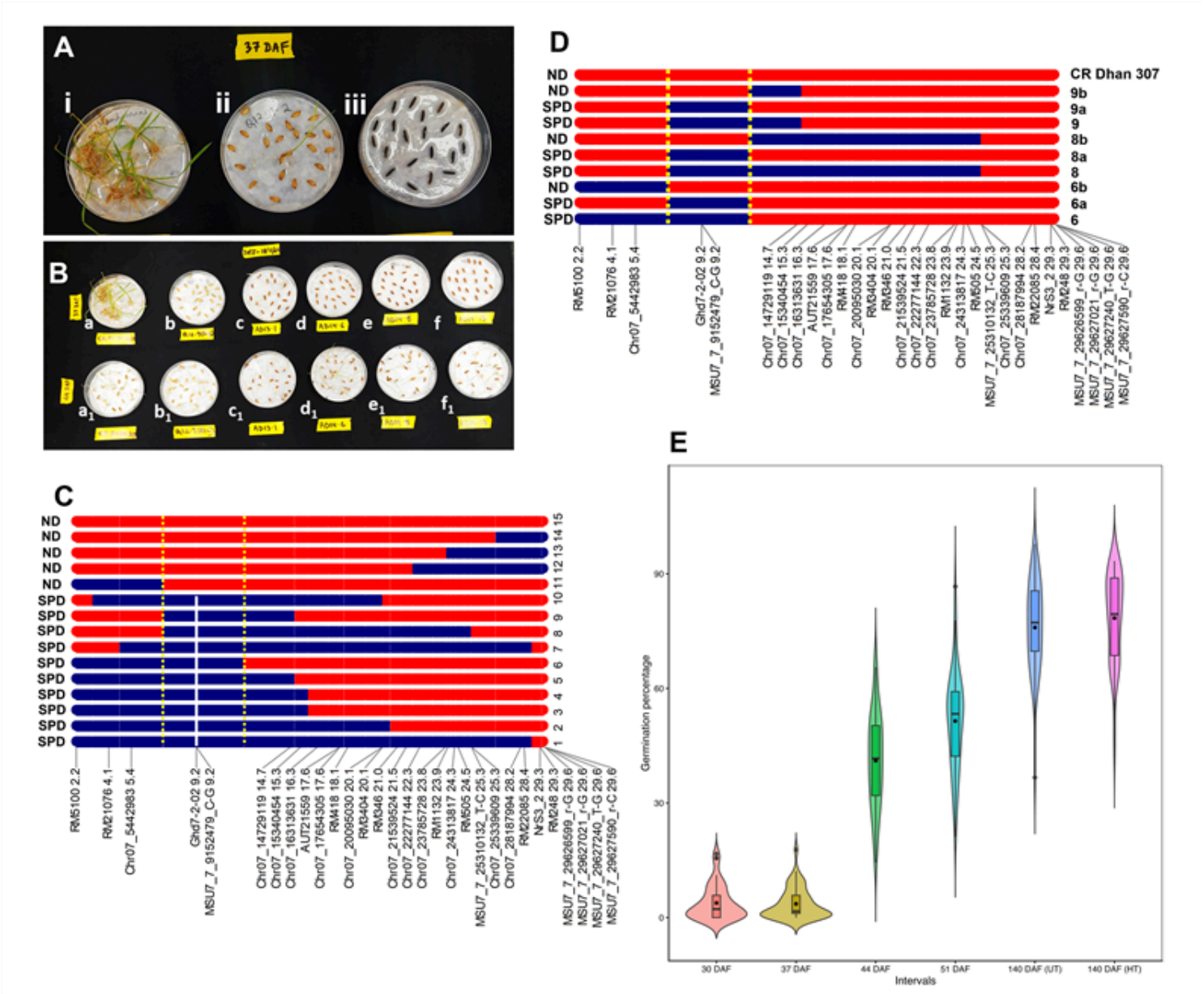


Figure 6

Mapping of the locus for seed dormancy using introgression lines, including chromosome segment substitution lines (CSSLs), backcross-derived inbred lines (BILs), and single segment substitution lines (SSSLs) identified from backcross progenies of *CR Dhan 307 × O. rufipogon accession AC100444* developed in the present study and a previous work (Dalai et al. 2026). ***A.*** *Representative photograph*

showing germination percentage of BIL Q12-9 (ii), compared with CR Dhan 307 (i) and O. rufipogon accession AC100444 (iii). These representative seeds were tested for germination seven days after harvest, and the photograph was captured 14 days thereafter. ***B.*** *Contrasting germination patterns between CR Dhan 307 and BILs showing short-term seed dormancy (STSD): a–f, seeds tested seven days after harvest (37 days after flowering) and photographed 14 days thereafter;* a_1*–*f_1*, seeds tested 14 days after harvest (44 days after flowering) and photographed seven days thereafter. While a &* a_1 *represent CR Dhan 307, b–f and* b_1*–*f_1 *represent selected BILs showing STSD.* **C.** Substitution mapping with BILs and CSSLs predicted a locus regulating STSD, located either between Chr07_5442983 and Ghd7-2-02 or between Ghd7-2-02 and Chr07_14729119. **D.** Reciprocal recombinants between Chr07_5442983 and Ghd7-2-02 or between Ghd7-2-02 and Chr07_14729119 (6a/6b; 8a/8b; 9a/9b), identified from $IC_1BC_1F_2$ progenies of three selected BILs (shown as 6, 8, and 9 in Fig. 8C), confirmed the association of Ghd7-2-02 and MSU7_7_9152479_C-G with STSD. **E.** Variation in germination percentage among 47 $IC_1F_{2:3}$ progenies homozygous for the Ghd7-2-02 allele of *O. rufipogon* identified from IC_1F_2 progenies of SSSL-L-7-7/CR Dhan 307 at different durations after flowering (30, 37, 44, 51, and 140 days after flowering). UT: Untreated; HT: Heat treated to break dormancy.

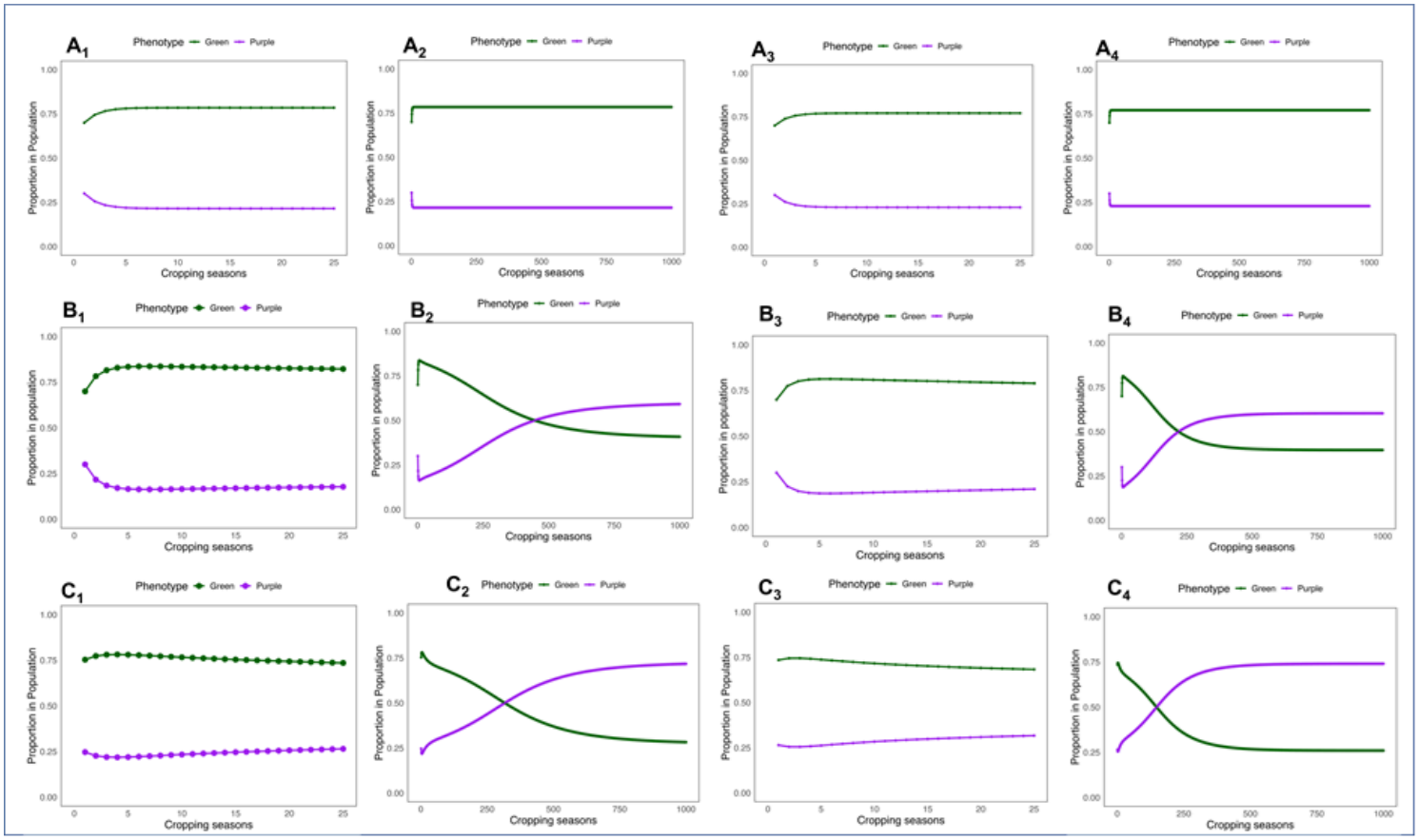


Figure 7

Predicted short- and long-term changes in weedy rice population dynamics for leaf sheath pigmentation (purple vs. green) in weedy rice swarms. Trait inheritance was modeled under single-gene dominance and/or complementary gene action, assuming 90–95% selfing and 5–10% outcrossing per generation.

Initial weedy rice populations were set at 30:70 for purple and green leaf sheaths. **A:** Only **s**ingle dominant gene controlling the trait in the base population**;** $\mathbf{A_1}$**,** $\mathbf{A_2}$**:** 5% outcrossing; $\mathbf{A_3}$**,** $\mathbf{A_4}$**:** 10% outcrossing. **B:** Only **c**omplementary gene action controlling the trait in the base population; $\mathbf{B_1}$**,** $\mathbf{B_2}$**:** 5% outcrossing; $\mathbf{B_3}$**,** $\mathbf{B_4}$**:** 10% outcrossing. **C:** Equal contribution of single dominant gene and complementary gene action in the base population; $\mathbf{C_1}$**,** $\mathbf{C_2}$**:** 5% outcrossing; $\mathbf{C_3}$**,** $\mathbf{C_4}$**:** 10% outcrossing.

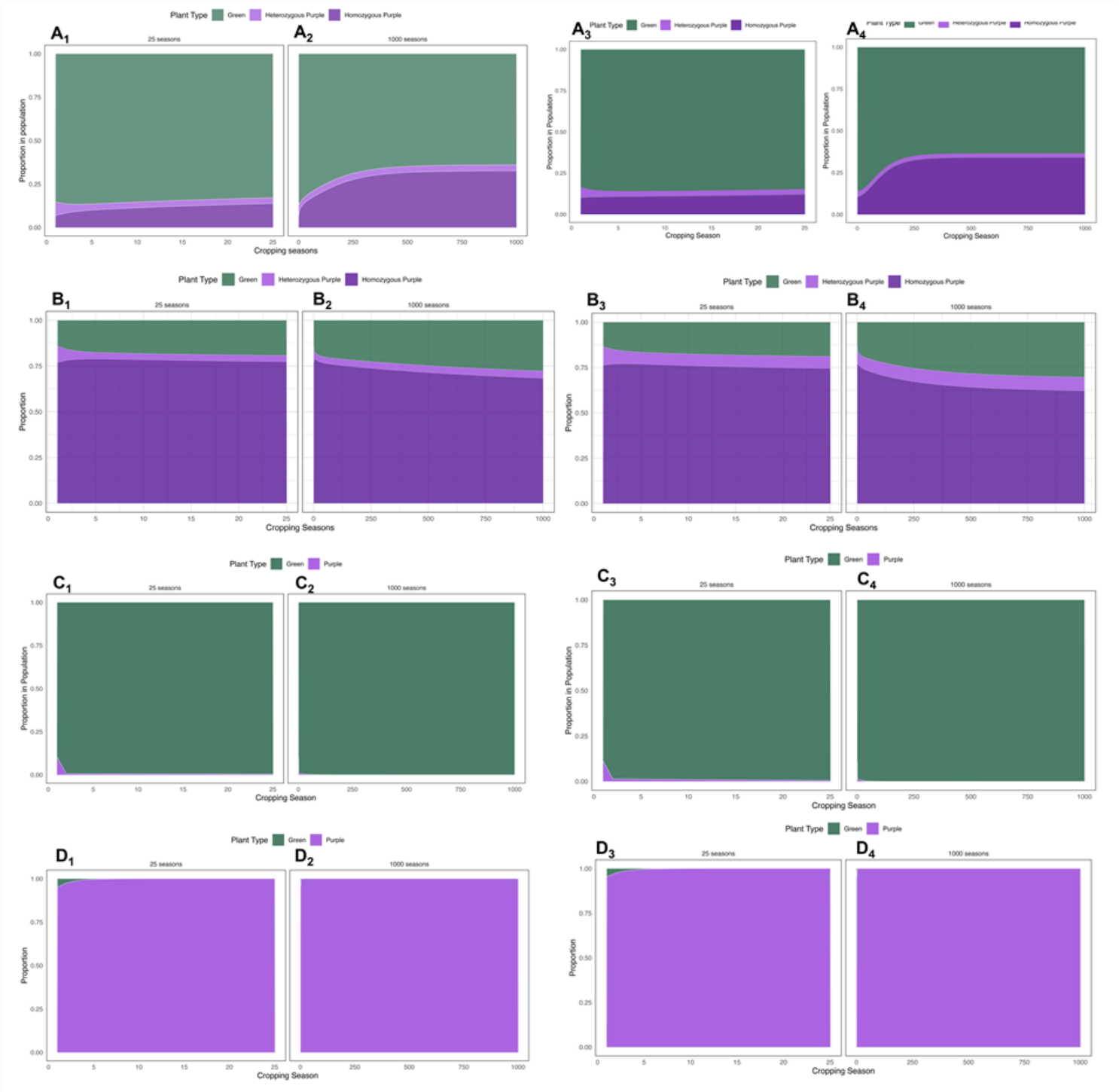


Figure 8

Prediction of temporal changes (short- and long-term) in weedy rice population dynamics for leaf sheath pigmentation (purple vs. green) in a crop–weed continuum under differential selection pressure. Trait inheritance was assumed to be controlled by both single gene with complete dominance and complementary gene action, present in equal proportion in the base population of weedy rice. The mating structure assumed 90–95% selfing and 5–10% outcrossing per generation. The base population was set at a 70:30 ratio of crop to weedy rice for each season, with initial polymorphism in weedy rice fixed at 30:70 for purple and green leaf sheaths, respectively. **A:** No selection pressure applied and the continuously grown cultivar has green leaf sheath. **A_1, A_2:** 5% outcrossing; **A_3, A_4:** 10% outcrossing. **B:** No selection pressure applied and the continuously grown cultivar has purple leaf sheath. **B_1, B_2:** 5% outcrossing; **B_3, B_4:** 10% outcrossing. **C:** Absolute selection pressure applied against purple leaf-sheathed plants and the continuously grown cultivar has green leaf sheath. **C_1, C_2:** 5% outcrossing; **C_3, C_4:** 10% outcrossing. **D:** Absolute selection pressure applied against green leaf-sheathed plants and the continuously grown cultivar has purple leaf sheath. **D_1, D_2:** 5% outcrossing; **D_3, D_4:** 10% outcrossing. All the models were validated with 100- replications and reiterations.

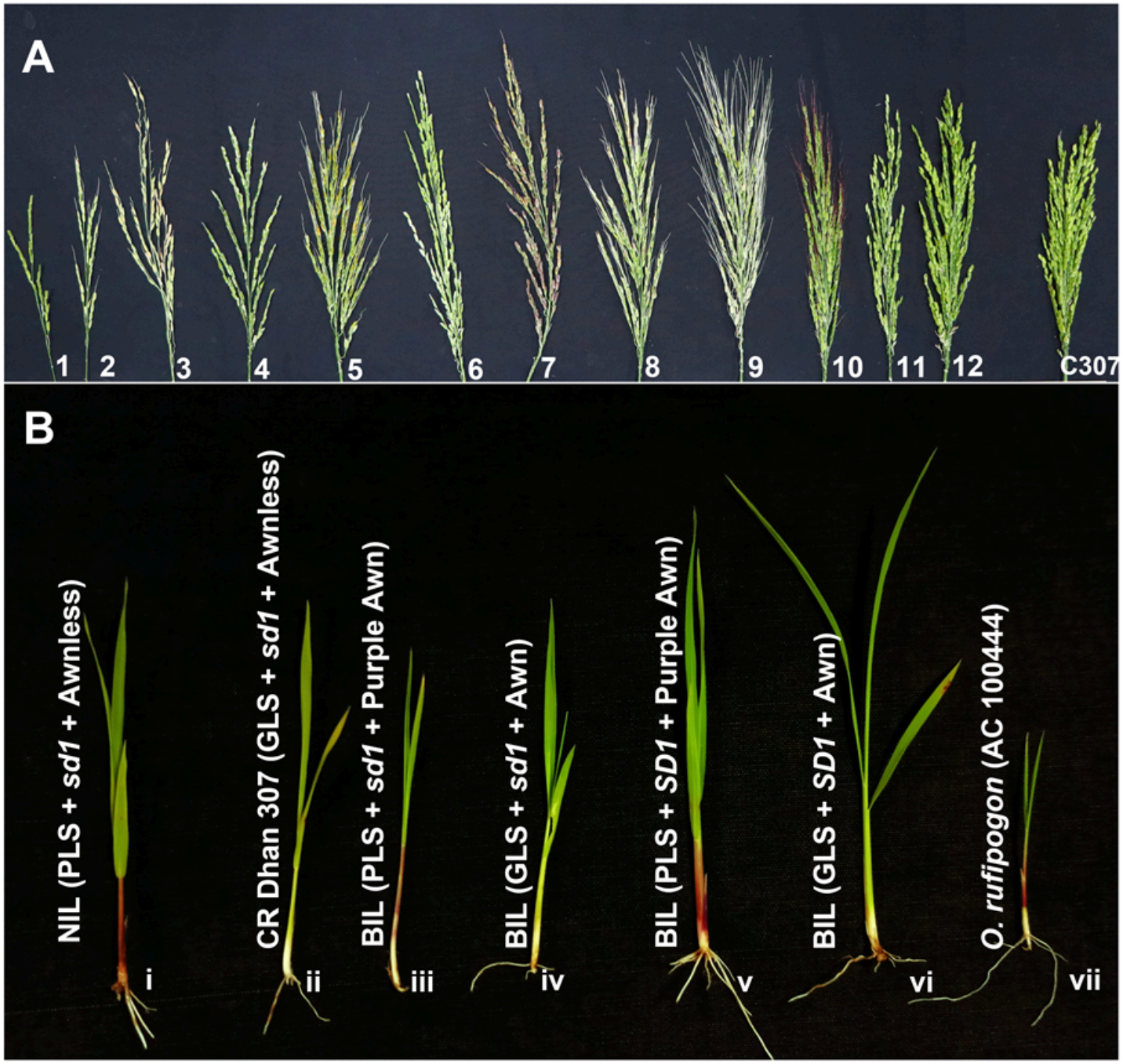


Figure 9

Representative variation for awn morphology and leaf sheath pigmentation at the seedling stage among introgression lines of *CR Dhan 307 × O. rufipogon accession AC100444*. ***A.*** *Variation in awn morphology among BILs (1–12) compared with the recurrent parent CR Dhan 307 (C307).* **B.** Variation in leaf sheath pigmentation among the BILs and NIL selected for the field simulation study: i. near isogenic line (NIL) derived from IC_1F_2 progenies of SSSL-U2-6/CR Dhan 307; ii. CR Dhan 307; iii. BIL with purple leaf sheath and *sd1* allele; iv. BIL with green leaf sheath and *sd1* allele; v. BIL with purple leaf sheath and *SD1* allele, vi. BIL with green leaf sheath and *SD1* allele; vii. *O. rufipogon accession AC100444. PLS: Purple leaf sheath; GLS: Green leaf sheath; sd1: semi-dwarfing allele of sd1 gene. SD1: wild type (tall phenotype) allele*

of sd1 gene. The seedlings shown here were nursery raised during January 2026 at ICAR-CRRI, Cuttack. The seedlings with purple basal pigmentation also showed healthier growth without any cold injury symptoms when compared to those with green leaf sheath.

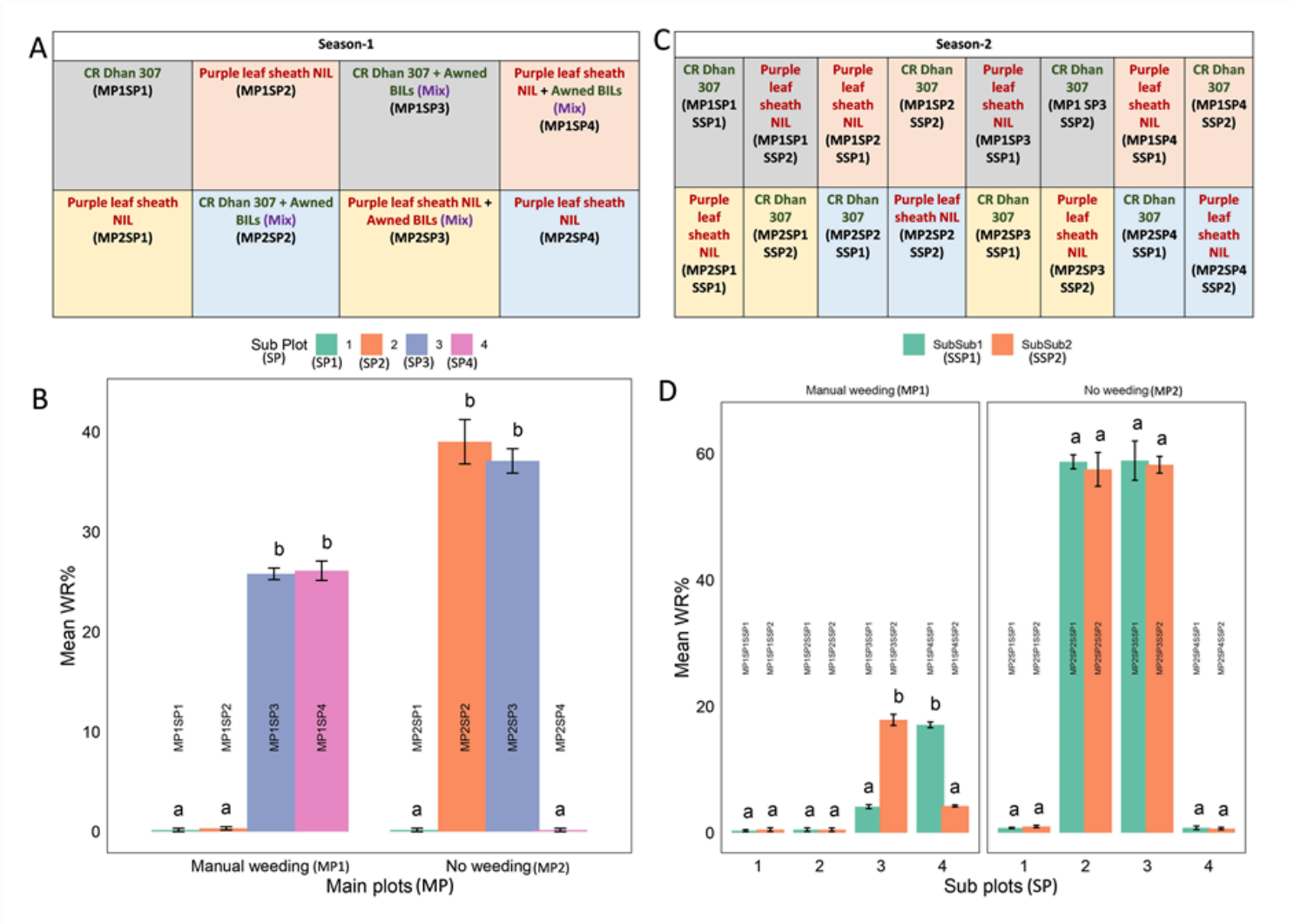


Figure 10

Simulated field experiment for dissection of population dynamics of weedy derivatives under different selection regimes. **A.**Experimental design for season-1. The field was divided into two main plots: one with manual weeding (MP1) and the other without manual weeding (MP2). Each main plot was randomly subdivided into four subplots (SP1–SP4) comprising: (i) CR Dhan 307, (ii) purple leaf sheath NIL, (iii) a seed mixture of 60% CR Dhan 307 and 40% awned BILs, and (iv) a seed mixture of 60% purple leaf sheath NIL and 40% awned BILs. Any variant of the main cultivar for leaf sheath pigmentation in respective subplots within MP1 were removed within 45 days. The experiment was laid out in a split-plot design. **B.** Relative frequencies of weedy derivatives observed after flowering in the four subplots (SP) of the two main plots (MP). **C.** Experimental design for season 2. Each subplot with residual shattered seeds from the previous season in both main plots was further divided into two sub-subplots (SSP1 and SSP2), where CR Dhan 307 or the purple leaf sheath NIL were randomly sown. Any variant for leaf sheath pigmentation of the main cultivar in the respective sub-subplots of all subplots within MP1 were

removed within 45 days. The experiment was laid out in a split-plot design. **D.** Relative frequencies of weedy derivatives observed after flowering in the two sub-subplots (SSP) of all the four subplots (SP) under the two main plots (MP). Alphabetical letters on the histograms indicate Tukey's HSD test-based comparisons. Different letters (a, b) indicate statistically significant differences at 5% level. In 10B, comparisons were made among subplots (SP) within a main plot (MP); and in 10D, comparisons were made between the two sub-subplots (SSP) within a subplot (SP).

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- SupplementaryDatasheet1.xlsx
- SupplementaryDatasheet2.xlsx
- SupplementaryFigure1and2.pdf
- SupplementaryTables.docx