

Harnessing Rf3/Rf4-Mediated WA-CMS Fertility Restoration for Hybrid Rice Breeding in India's North Eastern Hilly Region: Integrating Fertility Restoration, Heterosis, and Parental Genetic Divergence

Amit Kumar

amit4118@gmail.com

ICAR Research Complex for NEH Region

Avinash Pandey

ICAR-Indian Institute of Agricultural Biotechnology

Jayanta Bora

ICAR Research Complex for NEH Region

Letngam Touthang

ICAR Research Complex for NEH Region

Vikram Jeet Singh

Acharya Narendra Deva University of Agriculture and Technology

Simardeep Kaur

ICAR Research Complex for NEH Region

Veerendra Kumar Verma

ICAR Research Complex for NEH Region

Shridhar Ragi

Indian Agricultural Research Institute

S. P Das

National Research Centre for Orchids

Research Article

Keywords: Fertility restoration, genetic diversity, hybrid rice, combining ability

Posted Date: October 5th, 2026

DOI: <https://doi.org/10.21203/rs.3.rs-10626208/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

Hybrid rice has driven substantial yield gains across India, yet no hybrid built on north-eastern Indian landraces has reached large-scale cultivation, despite the region's distinct agro-ecology and growing exposure to climate variability. Closing this gap requires restorer lines that combine reliable fertility restoration, adequate genetic diversity from the cytoplasmic male sterile (CMS) parents they will be paired with, and strong combining ability, three requirements rarely evaluated together. Here, 33 candidate restorer lines and 5 CMS lines (IR 68888A, IR 58025A, RTN 10A, RTN 12A, and APMS 6A) were screened for the fertility-restoration genes *Rf3* and *Rf4* using gene-based/linked markers, profiled for genetic diversity with SSR markers, and evaluated as 165 test crosses in a modified line x tester design. *Rf4* was more prevalent than *Rf3*, and gene-based markers predicted realised spikelet fertility more reliably than gene-linked markers; applying a standard four-class fertility-restoration scheme classified 26 of the 33 lines (78.8%) as effective restorers (Fully or Partial), though most lines shifted class depending on the specific CMS tester used, showing that restoration is CMS-line-dependent rather than fixed. SSR-based clustering (STRUCTURE, UPGMA, PCoA) consistently resolved two major genetic groups, clearly separating Ratnagiri-derived CMS lines from north-eastern restorer germplasm, but this genetic distance only weakly predicted hybrid yield performance. Line x tester analysis showed non-additive gene action dominated yield and its component traits ($\sigma^2_{sca} > \sigma^2_{gca}$; dominance variance > additive variance), and combining this with commercial-heterosis screening identified IR 68888A/Sakur Dhan (156.67%), IR 68888A/MTU 1010 (148.57%), and AMPS-6A/DRR Dhan 44 (141.43%) as the top hybrids over the check Shahsarang. Classifying all test crosses by maturity duration showed that six of these eight top-heterosis hybrids were Mid-early (121–130 days) and none were Late, meaning the best-performing hybrids are also best positioned to escape terminal-season stress and enable double-cropping, linking this breeding result directly to climate-resilient food security in the north-eastern hill region

Introduction

Rice underpins food security for more than half the world's population, supplying a larger share of global dietary energy than any other cereal [1, 2, 3]. In India, the Green Revolution delivered a first major leap in production through high-yielding varieties, but that yield curve has since flattened even as demand keeps climbing: India's rice requirement is projected to reach around 146 million tonnes by 2030 [4], and because cultivable land is not expanding, that demand can only be met by raising yield per hectare rather than area sown.

Hybrid rice is the most proven route to this kind of yield step-change, offering a 20–25% advantage over the best conventional semi-dwarf varieties [5, 6, 7, 8], and India has released 127 rice hybrids to date [9, 10]. Yet adoption remains concentrated in a handful of states and covers barely 3 million hectares nationally, and, notably, not one of these hybrids has been built on the landraces of North-Eastern India, a region with a distinct agro-ecology, its own considerable varietal wealth, and growing exposure to the erratic, narrowing monsoon window that climate change is bringing to hill agriculture. This gap, between

what hybrid rice has demonstrated elsewhere and its near-total absence in the north-eastern hills, is the starting point for the present study.

Indian hybrid rice breeding is built almost entirely on the wild-abortive cytoplasmic male sterility (WA-CMS) system, which depends on three lines, cytoplasmic male sterile (A), maintainer (B), and restorer (R), whose interaction determines whether the F_1 regains pollen fertility [11]. Because fertility restoration is governed by nuclear-cytoplasmic gene interactions, correctly identifying true restorer lines is the first and most basic requirement of any hybrid programme [12]. This identification was historically done by hybridizing candidate lines and scoring pollen fertility by eye, a slow and land-intensive process; molecular markers for the two major restorer loci, *Rf3* (chromosome 1; [13]) and *Rf4* (chromosome 10; [14]), have since made this screening far faster [15, 16]. But marker-based screening only answers half the question, whether a line restores fertility, not whether it will make a genetically valuable, agronomically deployable parent, which is where diversity, combining ability, and duration come in.

Fertility restoration alone does not guarantee a good hybrid: heterosis depends on genetic complementation between parents, so restorer lines also need to be genetically diverse relative to the CMS lines they will be crossed onto [17, 18, 19] and, because CMS lines are few in number, finding restorers that combine well across multiple WA-CMS backgrounds is itself a bottleneck. Diversity and per-se performance are, however, imperfect predictors of hybrid value on their own; a phenotypically superior restorer can still combine poorly [20]. Combining ability analysis fills this gap, and among the available biometrical designs, line x tester analysis [21] is the most practical for screening many parental combinations at once, partitioning general combining ability (GCA, additive gene action) from specific combining ability (SCA, non-additive gene action) [22, 23, 24]. Molecular diversity data can, in principle, be linked forward to this combining-ability picture: [25] showed genetic distance could indirectly predict heterosis in maize without exhaustive testcrossing, a shortcut that would be valuable here if it holds.

Even a hybrid that restores fertility well, combines well, and is genetically well-matched to its CMS parent is only useful if its maturity duration fits the region's cropping calendar and climate risk. In the rainfed, terraced agriculture typical of the North-Eastern hills, a late-maturing hybrid is exposed to a longer window of unpredictable late-monsoon rainfall and early-winter cold at grain-filling, either of which can erase a yield advantage established earlier in the season; an early- or mid-early-duration hybrid, by contrast, escapes much of that risk and can also vacate the field in time for a second, rabi-season crop, directly raising cropping intensity, smallholder income, and household food security, outcomes that map directly onto the Sustainable Development Goals for zero hunger, climate action, and poverty reduction [26]. Despite this, duration is rarely classified and reported alongside heterosis and combining-ability results in hybrid rice studies, even though it determines whether a high-yielding hybrid is actually deployable in a given agro-ecology rather than merely superior on paper.

These four requirements, correct *Rf*-gene-based restorer identification, adequate parental genetic diversity, validated combining ability, and a maturity duration suited to the regional cropping system and climate risk, are rarely evaluated together in a single study, and have not been brought together at all for

restorer material from North-Eastern India. The present study was designed to close that gap directly: to screen candidate lines for *Rf3* and *Rf4* and classify their fertility-restoration ability; to profile the genetic diversity of the restorer lines and their WA-CMS testers using SSR markers; to develop and evaluate the resulting hybrids through line x tester analysis and test whether SSR-based genetic distance can indirectly predict the heterosis realised in them; and to classify the resulting hybrids by maturity duration, so that the parent combinations identified as superior are also assessed for their fit with climate-resilient, food-security-relevant cropping systems in the north-eastern hill region.

Materials and Methods

Experimental site and genetic materials

The present investigation was executed at the Plant Breeding farm of the ICAR Research Complex for NEH Region, Umiam, Meghalaya, India (latitude 25°39' N, longitude 91°54' E, elevation 966 m above MSL). A set of 33 restorer lines and 5 cytoplasmic genetic male sterile (CMS) lines were used for screening of fertility restorer genes (*Rf3* and *Rf4*) using gene-based/linked markers. Genotypes carrying *Rf3* and *Rf4* genes, either singly or in combination, were carried forward for hybridization, described next.

Development of experimental material

The potential restorers were used for hybridization with the five CMS lines, namely IR 68888A, IR 58025A, RTN 10A, RTN 12A, and APMS 6A, in a modified line x tester design where the testers (CMS lines) were used as females and the lines (potential restorer lines) were used as males, since CMS lines cannot themselves serve as males in a line x tester design (Kempthorne 1957). Spikelets of the five CMS parents were cut in the evening, followed by pollination from the 33 potential restorer parents (testers) the next morning (8:00-10:00 AM), giving a total of 165 test crosses that were then carried into field evaluation.

Evaluation of test crosses

The test crosses, along with their parents and standard checks (Shahsarang and Megha SA 2), were planted in the field during the 2022 *Kharif* season with three replications at two locations, namely Umiam (960 m above msl) and Upper Shillong (1500 m above msl). A 5-metre row length with 3 rows was used for each test cross, with a spacing of 20 x 10 cm. Standard crop-production practices, including recommended plant and row spacing, fertilizer application, and disease, insect, and pest management, were followed throughout to raise healthy crops. Five randomly chosen plants from the middle rows were used to record yield-related traits, namely plant height (cm), panicle length (cm), number of effective tillers per plant, number of filled grains per panicle, spikelet fertility percentage, test weight, and grain yield per plant (g). The days to 50% flowering was noted when the half the plants shown the flowering in entire plot. The average of the five plants constituted one replication value used for data analysis.

Screening for fertility restoration genes (*Rf3*, *Rf4*) and association between genetic distance and heterotic potential

All 38 (33+5) parents were grown in soil-filled 96-hole plastic seedling trays. Fresh young leaves of five healthy plants, 10 days after sowing, were used for DNA extraction following the cetyl-tetramethyl ammonium bromide (CTAB) method [27]. DNA was diluted to a working standard after agarose gel-based quantitation using uncut λ genomic DNA and absorbance-based quantification with a Thermo Scientific NanoDrop 2000 (Thermo Fisher Scientific, USA), using the ratio of absorbance at 260 nm and 280 nm (A_{260}/A_{280} of 1.8-2.0). PCR was performed in a 20- μ l reaction mixture containing 1 μ l of genomic DNA, 0.8 μ l each of forward and reverse primers, 0.4 μ l of dNTPs, 0.125 μ l of Taq polymerase, 4 μ l of PCR buffer, and 12.875 μ l of nuclease-free water, using a 35-cycle profile of 94°C for 30 s (denaturation), 55°C for 30 s (annealing), and 72°C for 1 min (extension), flanked by an initial denaturation at 94°C for 5 min and a final extension at 72°C for 10 min. PCR products were separated on a 3.5% agarose gel and visualized under UV light using a Bio-Rad gel documentation unit, with a 100-bp DNA ladder used to size alleles. For the *Rf3* locus, both candidate gene-based markers, DRRM-*Rf3*-5 and DRRM-*Rf3*-10, were used, with DRRM-*Rf3*-10 being the more efficient of the two [28]; for the *Rf4* locus, the gene-linked marker RM6100 and the gene-based marker DRCG-RF4-14 were used (primer sequences, chromosomal location, and expected restorer/non-restorer amplicon sizes for all four markers are given in Table S1). A set of 33 standard SSR markers spanning the twelve rice chromosomes, drawn from the GCP panel of IRRI, was then applied to the same 38 genotypes to assess genetic distance and relationships among the parental lines, so that fertility-restoration genotype and genome-wide diversity could be examined side by side in the same material (Table 5).

Statistical analysis

Data recorded for the 38 potential restorer lines and checks for the various agro-morphological traits were used to compute descriptive statistics and genetic variability using R 4.2.1. Clustering (Stats and Factoextra packages) and principal component analysis (Factoshiny package) were performed in R 4.2.1, and combining-ability analysis was computed using the line x tester model of Kempthorne (1957) in R 4.2.1 with the agricolae package (R Core Team 2022). Genetic diversity parameters, namely polymorphic information content (PIC), number of alleles (N_a), number of effective alleles (N_e), gene diversity (GD), heterozygosity (H), and Shannon's information index (I), were estimated using PowerMarker 3.25 [29] and Popgene 1.32 [30]. Principal coordinate analysis was performed using GenAlEx v6.5 [31], and AMOVA and pairwise fixation index (F_{ST}) values were analyzed with Arlequin 3.5.2.2 [32]. Bayesian cluster estimation of population structure (global ancestry) was carried out with STRUCTURE 2.3.4, using ten replicates for each value of K (1 to 10), each with a burn-in of 100,000 iterations followed by 100,000 post-iterations; the replicate with the highest likelihood was retained for each K. Nei's (1983) genetic-distance-based clustering using the unweighted pair-group method with arithmetic mean (UPGMA) was performed with Power Marker 3.25 and visualized in Mega 11.0, and the correlation between Nei's genetic distance and standard heterosis was calculated in Microsoft Excel. In addition, since days to maturity is agronomically decisive for how a hybrid fits into regional cropping

systems, every test cross was classified by duration into four standard classes, Early (<120 days), Mid-early (121-130 days), Medium (131-140 days), and Late (>140 days), and this classification was cross-tabulated against CMS line and against the combining-ability/heterosis results in R 4.2.1. Fig. 1a summarizes how these parallel streams of analysis, molecular screening, phenotypic/duration evaluation, and combining-ability analysis, converge into the heterosis and sustainability assessment reported in the Results and Discussion.

Results

3.1 Fertility-restoration gene screening

Single-locus *Rf3* and *Rf4* allelic distribution

The presence of fertility restorer gene(s) was confirmed using fertility restorer gene-linked/based markers, with restorer and non-restorer alleles identified from amplicon size. For the *Rf3* locus, marker DRRM-*Rf3*-5 produced 160 bp (restorer allele) and 130 bp (WA-CMSline alleles) products, whereas DRRM-*Rf3*-10 produced 210 bp and 190 bp products in restorer and CMS lines, respectively (Supplementary Table S2). For the *Rf4* locus, DRCG-*Rf4*-14 produced 845 bp in restorer lines and 885 bp in CMS lines, while RM6100 produced 150 bp (restorer) and 140 bp (WA-CMSline) products. Figure 2 summarizes the resulting allelic-class distribution: with DRRM-*Rf3*-5, 7 of the 33 genotypes were homozygous restorer, 16 (48.48%) were homozygous non-restorer, and 10 were heterozygous for *Rf3*; with DRRM-*Rf3*-10, the corresponding counts were 3 homozygous restorer, 26 homozygous non-restorer, and 4 heterozygous, consistent with DRRM-*Rf3*-10 being the more chromosome-1-proximal and therefore more efficient marker. For *Rf4*, 6 genotypes (18.18%) and 14 genotypes (42.42%) were homozygous restorer using the gene-linked marker (RM6100) and the gene-based marker (DRCG-*Rf4*-14), respectively, while 11 and 18 genotypes were homozygous non-restorer using DRCG-*Rf4*-14 and RM6100, respectively, and 16 and 1 genotypes were heterozygous using DRCG-*Rf4*-14 and RM6100, respectively (Fig. 1b; full counts in Supplementary Table S2).

Combined *Rf3/Rf4* allelic combination classes

Combining these single-locus results, the genotypes were further classified into nine allelic combination classes of *Rf3* and *Rf4* (*Rf3Rf3Rf4Rf4*, *Rf3Rf3rf4rf4*, *rf3rf3Rf4Rf4*, *rf3rf3rf4rf4*, *Rf3rf3Rf4Rf4*, *Rf3Rf3Rf4rf4*, *Rf3rf3Rf4rf4*, *Rf3rf3rf4rf4*, and *rf3rf3Rf4rf4*; Fig. 1c; full class-wise counts in Supplementary Table S3), scored using four marker combinations: DRRM-*Rf3*-5 + DRCG-*Rf4*-14, DRRM-*Rf3*-10 + DRCG-*Rf4*-14, DRRM-*Rf3*-5 + RM6100, and DRRM-*Rf3*-10 + RM6100. With DRRM-*Rf3*-5 + DRCG-*Rf4*-14, 8 genotypes carried the restorer allele, 10 were heterozygous, and 15 carried the non-restorer allele at *Rf3*, while for *Rf4* the corresponding counts were 6, 15, and 12. With DRRM-*Rf3*-10 + DRCG-*Rf4*-14, the *Rf3* counts were 7 restorer, 10 heterozygous, and 16 non-restorer, and the *Rf4* counts were 6, 16, and 11. With DRRM-*Rf3*-5 + RM6100, the *Rf3* and *Rf4* counts were 7 and 14 restorer, 10 and 1 heterozygous, and 16 and 18 non-restorer, respectively, and with DRRM-*Rf3*-10 + RM6100 they were 3 and 14 restorer, 4 and 1 heterozygous, and 26 and 18 non-restorer, respectively. Across all four combinations, the frequency of

lines carrying *Rf4* was consistently higher than the frequency of lines carrying *Rf3*, a pattern that becomes important below when the same marker classes are related to the spikelet fertility of the derived test crosses.

Marker efficacy in identifying potential restorers

Having established which genotypes carried which *Rf* alleles, the next question was whether that molecular information actually predicted restoration in the field. Marker efficacy was calculated by comparing the presence of fertility restorer genes, singly or in combination, with the spikelet fertility percentage of the derived test crosses, using a cut-off of $\geq 60\%$ spikelet fertility to call a cross restored (Fig. 1d). The highest number of test crosses at or above this threshold came from WA-CMS lines APMS 6A and IR 68888A across all four marker combinations (DRRM-*Rf3*-5 + DRCG-*Rf4*-14: 19, 15; DRRM-*Rf3*-10 + DRCG-*Rf4*-14: 18, 14; DRRM-*Rf3*-10 + RM6100: 15, 9; DRRM-*Rf3*-5 + RM6100: 18, 13). With DRRM-*Rf3*-5 + DRCG-*Rf4*-14, average spikelet fertility across all crosses was 67.35%, with IR 58025A (73.68%) and APMS 6A (73.08%) highest; with DRRM-*Rf3*-5 + RM6100, the average was 61.29%, with APMS 6A (75.00%) and IR 58025A (66.67%) highest. With DRRM-*Rf3*-10 + DRCG-*Rf4*-14, the average was 72.29%, with RTN 12A (90.00%) and RTN 10A (77.78%) highest, and with DRRM-*Rf3*-10 + RM6100 the average was 72.41%, with RTN 12A (83.33%) and RTN 10A (83.33%) highest (full efficiency table in Supplementary Table S4). Gene-based markers were consistently more efficient than gene-linked markers at identifying restorer lines, and spikelet fertility was comparably high ($\geq 60\%$) whether *Rf3* and *Rf4* were present singly or together. Notably, a few crosses reached $\geq 60\%$ spikelet fertility even though the male parent carried no restorer allele at either locus, hinting that additional, as-yet-unmapped fertility-restoration loci may be segregating in this material and warrant dedicated follow-up mapping.

Classification of restorer lines by spikelet-fertility restoration

Beyond the single 60% cut-off used above to gauge marker efficacy, the 33 restorer lines were classified more finely using a standard four-class spikelet-fertility restoration scheme widely applied in WA-CMS characterization studies: Fully maintainer (FM, 0% spikelet fertility), Partial maintainer (PM, 0.10-50.00%), Partial restorer (PR, 50.10-75.00%), and Fully restorer (FR, $>75\%$). Applied to all 127 individual test crosses with recorded spikelet fertility, this classification placed 48 crosses (37.8%) as Fully restorer, 54 crosses (42.5%) as Partial restorer, 25 crosses (19.7%) as Partial maintainer, and no cross as a complete (0%) Fully maintainer (Fig. 5c); the 37 crosses with no recorded spikelet fertility value most likely correspond to complete or near-complete sterility that could not be reliably scored in the field, and are therefore functionally consistent with the fully maintainer response even though they could not be formally assigned a numeric class.

Aggregating each restorer lines mean spikelet fertility across the CMS testers it was crossed with (Fig. 5d) classified 7 lines as fully restorer (CR Dhan 801, Swarna, IURON 8/19, IRBN 43/19, Pusa 44, CR Dhan 802, and MTU 1010; mean SFP 76.7-85.2%), 19 lines as partial restorer (mean SFP 53.6-73.3%), and 7 lines as Partial maintainer (RCPL 1-431, RCPL 1-437, RCMR 44, RCMR 40, Shahsarang, Bhalum 4, and Kba Long; mean SFP 30.9-49.3%), with no line falling in the Fully maintainer class on average. Combining

the Fully restorer and Partial restorer classes, which both restore fertility to an agronomically usable extent (>50% spikelet fertility), 26 of the 33 lines (78.8%) qualify as effective restorers under this scheme, while the remaining 7 lines (21.2%) behave predominantly as maintainers and are not recommended for direct hybrid seed production, though they may still be useful as recurrent-selection donors for restoration-enhancing alleles. It is also worth noting that most lines, including several classed as Fully restorer on their mean, actually spanned two or even three fertility classes depending on which CMS tester they were crossed with (Table 1); this cross-specific variability confirms that fertility restoration in this material is not a fixed property of the restorer line alone but depends on the specific restorer-CMS combination, reinforcing why line x tester evaluation, rather than restorer screening in isolation, is necessary before a restorer can be certified for a specific CMS line.

3.2 Genetic diversity of parental lines

Marker polymorphism and gene diversity

All rice restorer and CMS lines were genotyped with 30 SSR (microsatellite) markers and 4 fertility-restorer markers, producing 51 alleles in total (Fig. 3b; full marker-wise statistics in Supplementary Table S6). The number of alleles per locus ranged from 1 to 3 (mean 1.50), with RM413 showing the highest count (3) and a group of markers (RM507, RM284, RM316, RM271, RM484, RM536, RM277, and others) showing the lowest (1); one marker had three alleles and 14 markers had two. Major allele frequency per locus ranged from 0.355 (RM413) to 0.947 (RM215), averaging 0.863, while the highest numbers of effective alleles per locus were seen at RM237 (2.00), RM495 (2.00), RM152 (1.99), and RM447 (1.82), averaging 1.33 overall. Nei's expected heterozygosity (H_e) ranged from 0.00 to 0.50 (RM495, RM237; mean 0.17); of the 34 markers tested, 15 showed H_e greater than zero and 18 showed zero. PIC values, which capture how informative each marker is, ranged from 0.095 to 0.591, with RM413 (0.59), RM237 (0.37), RM495 (0.37), RM151 (0.37), and RM447 among the most informative (Fig. 3c; genetic distance heatmap in Fig. 3a).

Population structure (STRUCTURE analysis)

Building on this marker-level picture, model-based clustering using a Bayesian approach was used to infer population structure across values of K from 1 to 10 (Fig. 2a, Fig. 2b). The ad hoc ΔK statistic identified $K = 2$ as optimal, splitting the population into two major groups: Group 1 with 18 lines, and Group 2 with 20 lines that included four CMS lines (IR 68888A, IR 58025A, RTN 10A, and RTN 12A). Lines and CMS lines with membership probability $\geq 80\%$ were classed as pure and the remainder as admixed; ten restorer lines (Bhalum 5, Swarna, DRR Dhan 44, IRBN 43/19, IURON 8/19, IRBN 15/19, RCMR 40, RCPL 1-438, UD 178, and RCMR 44) and two CMS lines (RTN 12A and APMS 6A) fell into this admixed category.

Distance-based clustering (UPGMA)

As an independent check on this structure, UPGMA clustering based on genetic distance grouped the restorer and CMS lines into the same two major clusters, plus an outgroup of two genotypes (UD 178 and RCPL 1-438) (Fig. 2c). Group 1 split into two subgroups of 10 lines each (26.317% of the panel each): Subgroup I comprised the four CMS lines (IR 68888A, IR 58025A, RTN 10A, and RTN 12A), while Subgroup II comprised restorer lines mostly from north-eastern India. Group II likewise split into two subgroups, with Subgroup I holding 5 restorer lines and Subgroup II, the largest cluster overall, holding 11 genotypes (28.947% of the panel); this Subgroup II was dominated by released varieties and advanced lines from Andhra Pradesh, Odisha, Himachal Pradesh, and Uttarakhand.

Principal coordinate analysis (PCoA)

Principal coordinate analysis (PCoA) of the full 33 restorer + 5 CMS panel again recovered the same two-group pattern in a two-dimensional scatterplot (Fig. 2d), with the first three coordinates jointly explaining 46.00% of the variability (PCoA1 = 18.60%, PCoA2 = 14.04%, PCoA3 = 13.37%; Supplementary Table S7), giving three independent lines of clustering evidence, STRUCTURE, UPGMA, and PCoA, that converge on the same two-group population structure.

Analysis of molecular variance (AMOVA) and genetic differentiation

To partition this diversity quantitatively, analysis of molecular variance (AMOVA) showed that 8.849% of total variation occurred among populations, 64.746% occurred among individuals within populations, and the remaining 26.405% occurred within individuals (Fig. 6a; Supplementary Table S8). Moderate F_{ST} (0.088) and high F_{IT} (0.736) values corroborated the presence of ample variability among the rice lines. Geographically, CMS lines from Ratnagiri were distantly related to genotypes from Andhra Pradesh, Odisha, Meghalaya, Manipur, and Sikkim, and genotypes from Odisha were distantly related to those from Manipur and Sikkim, whereas within the north-east, Meghalaya genotypes were relatively close to those from Sikkim and, to a lesser extent, Manipur (Fig. 3d). This geographic pattern is exactly the kind of complementary, distantly related parental combination (Ratnagiri CMS x north-eastern restorer) that the Discussion returns to when identifying promising cross combinations.

Association between genetic distance and hybrid performance

Having profiled both the fertility-restoration genotype and, separately, the SSR-based genetic distance of the same 38 parents, we next asked whether that genetic distance predicted hybrid performance directly. Based on genetic distance, RTN 12A (0.21) was the most diverse CMS line relative to the restorer panel, followed by APMS 6A (0.19) and IR 58025A (0.18). Restorers RCMR 40 (0.32), RCMR 48 (0.30), Bhalum 6 (0.26), MTU 1010 (0.26), and Kba-long (0.25) were most diverse from RTN 12A; RCMR 40 (0.29), UD 183 (0.27), MTU 1010 (0.26), Sambha Mahsuri (0.22), and Kba-long (0.22) from RTN 10A; UD 183 (0.25), RCPL 1-437 (0.25), and RCMR 40 (0.28) from APMS 6A; Kba-long (0.25), Sambha Mahsuri (0.25), UD 183 (0.24), and Pusa 44 (0.24) from IR 58025A; and RCMR 40 (0.28), Pusa 44 (0.24), UD 183 (0.22), Sambha Mahsuri (0.21), and Bhalum 3 (0.21) from IR 68888A (full matrix in Supplementary Table S5).

Associations between this genetic distance and the yield and component-trait performance of the

corresponding test crosses were computed using Pearson's correlation coefficients (Fig. 4a). Moderate positive correlations were found for the number of effective tillers per plant (APMS 6A, 0.22), panicle weight (RTN 12A, 0.35), and number of filled grains per panicle (RTN 12A, 0.16), while significant negative correlations were found for the number of effective tillers per plant (RTN 10A, -0.45), panicle weight (IR 68888A, -0.40), number of filled grains per panicle (IR 68888A, -0.48), spikelet fertility percentage (IR 68888A, -0.42), and yield per plant (IR 68888A, -0.28). Overall, no consistent, strong relationship was found between SSR-based parental genetic distance and the yield or component-trait performance of the derived crosses, a point returned to in the discussion alongside the marker-density limitation of the SSR panel used.

To test this relationship with the full performance dataset rather than only the CMS-line-wise summary above, genetic distance was matched directly against realised yield per plant (YPP) for all 123 individually genotyped and phenotyped test crosses for which both values were available (five crosses involving Vivek Dhan-44 could not be unambiguously matched between the molecular and phenotypic records and were excluded). This full-dataset check confirmed the pattern already indicated above: the overall correlation between genetic distance and YPP was negligible and non-significant ($r = -0.022$, $p = 0.81$, $n = 123$), and none of the five CMS-line-wise correlations reached significance (IR 68888A: $r = -0.28$, $p = 0.12$, $n = 32$; IR 58025A: $r = -0.20$, $p = 0.34$, $n = 24$; RTN 10A: $r = -0.02$, $p = 0.94$, $n = 17$; RTN 12A: $r = -0.15$, $p = 0.56$, $n = 18$; APMS 6A: $r = 0.14$, $p = 0.46$, $n = 32$). Notably, the eight test crosses with the highest commercial heterosis (see below) were not drawn from the most genetically distant parent pairs: their genetic distances (0.123-0.223) spanned essentially the full observed range and clustered close to the panel mean (0.184 ± 0.046) rather than at its upper extreme, so maximising parental genetic distance alone would not have been a reliable rule for shortlisting these hybrids in advance.

3.3 Field performance of test crosses

Per se performance of test crosses

Turning from parental diversity to the hybrids themselves, test crosses derived from RTN 12A had the greatest panicle length (26.68 cm), panicle weight (5.91 g), and number of filled grains per panicle (207.48), followed by those from APMS 6A. For yield per plant, APMS 6A-derived crosses were highest (16.27 g), followed by RTN 12A (15.67 g). Among the AMPS-6A-derived F1 hybrids, yield per plant ranged from 1 g (Kba long) to 33.8 g (DRR Dhan 44), with CR Dhan 801 (32.8 g), IRBN 43/19 (32.6 g), and IRBN 15/19 (31.67 g) also performing well; for IR 68888A, the best hybrids were with Sakur Dhan (35.93 g), MTU 1010 (34.8 g), IRBN 43/19 (33.4 g), and RCPL 1-426 (27.73 g); for IR 58025A, with IRBN 43/19 (30.2 g), Sakur Dhan (29.53 g), IRG 222 (27.87 g), Swarna (23.8 g), and Pusa 44 (21.87 g); for RTN 10A, with Sambha Mahsuri (30.73 g), Bhalum 6 (22.40 g), IRBN 15/19 (20.2 g), and Bhalum 3 (19.8 g); and for RTN 12A, with Sakur Dhan (28.80 g), IRBN 43/19 (25.73 g), IRBN 15/19 (25.00 g), and IURON 8/19 (19.93 g). Average CMS-line performance ranged from 10.57 g (IR 68888A) to 16.27 g (AMPS-6A), with RTN 10A and RTN 12A averaging 14.34 g and 15.67 g, respectively (full genetic-parameter summary in Supplementary Table S9).

Principal component analysis and GT biplot

To condense this multi-trait picture, principal component analysis (PCA) was carried out; three components with eigenvalue ≥ 1 jointly explained 67.48% of the variability (Fig. 6c; full eigenvalues and loadings in Supplementary Table S10). PC1 (40.60%) was driven mainly by panicle length (0.37), number of effective tillers per plant (0.56), panicle weight (0.81), number of filled grains per panicle (0.88), spikelet fertility percentage (0.90), and yield per plant (0.88); PC2 (15.72%) mainly by days to fifty percent flowering (0.72), plant height (0.75), panicle weight (0.37), and number of chaffy grains per panicle (0.43); and PC3 (11.16%) mainly by plant height (0.26), panicle length (0.64), number of effective tillers per plant (0.49), and test weight (0.31). A genotype-by-trait (GT) biplot combining genotype and trait vectors was then constructed to visually identify superior test crosses (Fig. 4b); the angle between trait vectors reflects their relationship, and in this biplot yield per plant grouped closely with spikelet fertility percentage, number of filled grains per panicle, number of effective tillers per plant, panicle weight, and panicle length, while it was negatively associated with days to fifty percent flowering and number of chaffy grains per panicle. Test crosses APMS 6A/IURON 8/19, RTN 12A/IURON 8/19, APMS 6A/DRR Dhan 44, APMS 6A/CR Dhan 801, and RTN 10A/Sambha Mahsuri sat closest to the favourable yield and component-trait vectors and are therefore flagged here as promising, a shortlist that is cross-checked against the combining-ability and heterosis results below.

Duration classification and its bearing on cropping-system fit

Because duration determines how a hybrid fits into the wider cropping calendar, not just how it performs in isolation, the 160 test crosses with recorded maturity duration were classified into four standard duration classes: Early (<120 days), Mid-early (121-130 days), Medium (131-140 days), and Late (>140 days). Mid-early duration crosses were the single largest class (79 crosses, 49.4%), followed by Medium duration (61 crosses, 38.1%), with Early (16 crosses, 10.0%) and Late (4 crosses, 2.5%) duration classes comparatively rare (Fig. 5a). This distribution was not uniform across CMS lines (Fig. 5b): IR 68888A produced the largest share of Early-duration crosses (8 of 33, 24.2%), while RTN 10A and RTN 12A produced most of the Medium- and all four of the Late-duration crosses, indicating that the female (CMS) parent, and not only the restorer, systematically shifts the maturity window of the resulting hybrid. Cross-referencing this classification with the eight highest-heterosis hybrids identified below showed that six of the eight (IR 68888A/Sakur Dhan, IR 68888A/MTU 1010, AMPS-6A/DRR Dhan 44, AMPS-6A/CR Dhan 801, AMPS-6A/IRBN 15/19, and AMPS-6A/Bhalum 4) fell in the Mid-early class, one (IR 68888A/IRBN 43/19) was Early, and one (RTN 10A/Sambha Mahsuri) was Medium duration; none of the top-heterosis hybrids fell in the Late-duration class.

3.4 Combining ability and heterosis

Line x tester ANOVA and variance partitioning

Analysis of variance (ANOVA) showed that the mean sum of squares for lines, testers, and F1 hybrids was significant at $p \leq 0.001$ for yield and all component traits (full ANOVA in Table 2), confirming

considerable genetic diversity among the genotypes for these traits. Lines (pollen parents) contributed more to this variation than testers (CMS females) for every trait, but for all traits the line x tester interaction term exceeded the contribution of lines alone (Fig. 4c; exact percentages in Supplementary Table S11), indicating that both parental lines and their specific interaction with testers shape trait expression. Consistent with this, variance due to specific combining ability (σ^2_{sca}) exceeded σ^2_{gca} , and dominance genetic variance exceeded additive genetic variance, for all seven traits, as reflected by the ratio of general to specific combining ability ($\sigma^2_{gca}/\sigma^2_{sca}$) and the degree of dominance (σ^2_D/σ^2_A)¹⁰², both smaller than unity (Fig. 6b; full genetic-component estimates in Table 3), pointing collectively to a preponderance of non-additive gene action in this material.

General combining ability (GCA) effects

General combining ability (GCA) for the parents ranged from -11.04 (Bhalum 4) to 13.85 (IRBN-15/19) for yield per plant, -86.84 (Bhalum 4) to 47.96 (CR Dhan 801) for number of filled grains per panicle, -1.52 (UD 178) to 1.32 (RCPL 1-426) for panicle weight, -1.86 (UD 216) to 4.64 (RCMR 46) for panicle length, -6.86 (Shahsarang) to 39.18 (Bhalum 3) for test weight, -2.74 (Bhalum 4) to 3.06 (MTU 1010) for number of effective tillers per plant, and -33.16 (Kba Long) to 20.41 (CR Dhan 801) for spikelet fertility percentage. Among the lines, IRBN-43/19 (13.85), Sakur Dhan (8.57), IRBN-15/19 (8.56), MTU 1010 (8.19), IURON-8/19 (4.79), Pusa 44 (4.57), RCPL 1-426 (4.37), and CR Dhan 801 (4.09) showed the highest, significant GCA for yield per plant and were also significant for several component traits, marking them as the strongest general combiners in this panel (Fig. 6d shows the full ranked distribution for yield per plant; full GCA effects for all traits in Table 4).

Specific combining ability (SCA) effects

Turning from general to specific combining ability, crosses of DRR Dhan 44 (15.56), CR Dhan 801 (12.08), Bhalum 5 (10.27), Vivek Dhan 44 (9.83), and UD 183 (8.32) with AMPS-6A showed the highest SCA; with IR-68888A, MTU 1010 (16.56), Sakur Dhan (16.36), RCPL 1-426 (12.36), RCPL 1-438 (9.54), and IRBN 43/19 (8.55) were the best combiners; with IR 58025A, IRG 222 (12.1) and Sakur Dhan (7.74); and with RTN 10A, Sambha Mahsuri (17.51), Bhalum 6 (10.84), and Shahsarang (7.98) (Fig. 6e shows the full SCA matrix for yield per plant; complete SCA effects and commercial heterosis for all crosses in Table 5).

Commercial heterosis over the check variety

Finally, commercial heterosis was computed relative to the mega variety Shahsarang, a high-yielding variety already tolerant to the abiotic stresses prevalent in Meghalaya. Heterosis across test crosses ranged from -100% (RTN 12A-based crosses at the low end) to 156.67% (IR 68888A/Sakur Dhan). The eight crosses with the highest commercial heterosis over Shahsarang were IR 68888A/Sakur Dhan (156.67%), IR 68888A/MTU 1010 (148.57%), AMPS-6A/DRR Dhan 44 (141.43%), IR 68888A/IRBN 43/19 (138.57%), AMPS-6A/CR Dhan 801 (134.76%), AMPS-6A/IRBN 15/19 (126.19%), RTN 10A/Sambha Mahsuri (119.52%), and AMPS-6A/Bhalum 4 (101.43%) (Fig. 4d); reassuringly, these are largely the same

crosses already flagged by the SCA effects above and by the GT biplot shortlist, giving three converging lines of evidence for the same set of superior hybrids.

Discussion

This study combined four analyses that are rarely reported together for the same restorer material: molecular screening for the fertility-restoration genes *Rf3* and *Rf4*, SSR-based profiling of parental genetic diversity, line x tester estimation of combining ability, and classification of hybrid maturity duration. Screening identified *Rf4* as the more prevalent gene and classified 26 of the 33 restorer lines (78.8%) as full or partial restorers; diversity profiling resolved a consistent two-group population structure that only weakly predicted hybrid performance; combining-ability analysis showed non-additive gene action dominating yield and its component traits; and duration classification showed that the highest-heterosis hybrids were predominantly Mid-early in maturity. The sections below interpret each of these results in turn, situating them against prior work and drawing out what each means for a regional hybrid rice programme.

Fertility-restoration screening and its bearing on restorer selection

Rf4 being consistently more frequent than *Rf3* across this panel agrees with [33], who found *Rf4* predominant among iso-cytoplasmic WA-CMS restorers with *Rf3* acting mainly in a synergistic role, and is broadly consistent with [34], who reported that roughly half of 570 Indian rice accessions carried an *Rf4Rf4/rf3rf3* genotype. That gene-based markers (DRCG-*Rf4*-14, and to a lesser extent DRRM-*Rf3*-10) tracked realised spikelet fertility more closely than the gene-linked marker RM6100 agrees with both [16] and [28], who separately argued that candidate-gene markers, being physically within the gene rather than merely linked to it, are inherently less prone to recombination-driven false positives. The small number of crosses reaching $\geq 60\%$ spikelet fertility without a scored restorer allele is not itself troubling: [35] and [36] both reported that fertility restoration in indica WA-CMS backgrounds can involve additional, still-unmapped minor *Rf* loci beyond *Rf3* and *Rf4*, consistent with that possibility rather than a marker failure. Operationally, this is why the four-class spikelet-fertility scheme used here (Fully maintainer, Partial maintainer, Partial restorer, Fully restorer) is more useful to a breeding programme than a binary restored/not-restored cut-off: it tells a breeder not just whether to keep a line, but what to do with it, advance it directly to hybrid seed production, pair it selectively with a compatible CMS tester, or redirect it to maintainer-line development, and the finding that most lines shifted between classes depending on the CMS tester used shows that restoration in this material has to be certified per cross combination rather than as a fixed varietal label.

Genetic diversity, population structure, and heterotic-pool design

The two-group population structure recovered independently by STRUCTURE, UPGMA, and PCoA in this study mirrors the two-subpopulation split reported by [20] for a hundred iso-cytoplasmic restorer lines genotyped with fifty IRRI SSR markers, and is consistent with the pattern documented for North-Eastern Indian rice germplasm by [37] and [38], both found that most usable genetic variation sat within, rather than between, geographic subpopulations, exactly as the AMOVA partitioning here showed (64.7% among individuals within populations versus 8.8% among populations). The allelic diversity recovered, 1.50 alleles per locus on average and PIC from 0.09 to 0.591, is lower than the 3.88 alleles per locus reported by [39] across a larger 150-variety panel, but comparable to the more modestly sized panels of [40] and [41], suggesting the lower diversity reflects sample composition rather than a shortcoming of the SSR panel itself. Practically, this structure functions as a heterotic-pool map: it identifies which restorers and CMS lines already sit in complementary pools, most notably the clear separation of the Ratnagiri-derived CMS lines from north-eastern restorer germplasm, a distinction a breeding programme can exploit directly when choosing parents, in much the same way heterotic-group classification has long guided parent choice in hybrid maize breeding.

The limited power of genetic distance to predict hybrid performance

Genetic distance correlated only weakly and inconsistently with yield and component-trait performance across the five CMS lines, and this held even when recomputed on the complete 123-cross performance dataset ($r = -0.022$, $p = 0.81$). This result diverges from [25], who found SSR-based genetic distance a workable indirect predictor of heterosis in tropical maize, but agrees with the more cautious position of [42], who, working with iso-cytoplasmic rice restorers, concluded that SSR distance alone was insufficient to substitute for phenotypic test-crossing and recommended denser, trait-linked marker systems such as genome-wide SNPs. The likely explanation is that SSR markers predominantly sample non-coding, presumably neutral regions of the genome with no direct mechanistic link to the loci controlling yield, so genome-wide distance and trait-specific heterosis need not track one another closely. This does not make the diversity data useless, it still disciplines heterotic-pool construction, as argued above, but it does mean yield itself cannot yet be predicted from genotyping data alone in this material.

Combining ability and the case for a hybrid, not pedigree, breeding strategy

Specific combining ability variance and dominance variance exceeded general combining ability and additive variance for every trait examined, a pattern also reported for rice yield traits by [43, 44] and earlier by [45] and consistent with the epistatic dominance effects implicated in heterotic combinations more broadly. This preponderance of non-additive variance is itself the justification for pursuing a hybrid rather than a pedigree-based breeding strategy on this germplasm, since selection on additive variance alone would leave most of the exploitable variation untouched. The specific combiners identified here, DRR Dhan 44, CR Dhan 801, and Bhalum 5 with AMPS-6A; MTU 1010, Sakur Dhan, and RCPL 1-426 with

IR-68888A, mirror the pattern seen in other Indian heterosis studies (46, 47, 48), where the best specific combinations were rarely predictable from general combining ability alone, reinforcing why GCA and SCA were treated here as complementary rather than interchangeable evidence.

Commercial heterosis, duration, and relevance to food security

The eight crosses with the highest commercial heterosis over Shahsarang, led by IR 68888A/Sakur Dhan (156.67%) and AMPS-6A/DRR Dhan 44 (141.43%), converge with the SCA rankings and the GT-biplot shortlist on a consistent set of candidate hybrids, giving three independent lines of evidence for the same material rather than a result specific to one analysis. That six of these eight hybrids were also Mid-early in duration, and none Late, connects this combining-ability result directly to deployment: in the rainfed, terraced agro-ecology of the north-eastern hills, an early-to-mid-early duration allows a hybrid to escape late-season drought or cold stress and to vacate the field in time for a second, rabi-season crop, a form of cropping-system intensification directly aligned with the Sustainable Development Goals framework for zero hunger, climate action, and poverty reduction [26]. Read together with the diversity and combining-ability results above, this duration profile is best understood not as an incidental agronomic detail but as the final filter that converts a statistically superior hybrid into one that is actually deployable in this specific region.

Conclusion

India's rice yield growth has plateaued even as demand keeps rising, and the one technology proven to deliver a further step-change, hybrid rice, has never been built on north-eastern Indian landraces, leaving farmers in one of the country's most climate-exposed hill ecosystems dependent on lower-yielding varieties poorly matched to an increasingly erratic monsoon. That mismatch, between a climate-vulnerable region and a yield technology it has never had access to, was the problem this study set out to address.

The solution attempted here was to treat restorer screening, parental diversity, combining ability, and hybrid duration as one connected pipeline rather than four separate exercises. Doing so identified 26 of 33 candidate lines as usable restorers, mapped which restorers and CMS lines are genetically complementary enough to cross productively, and converged, through combining-ability, biplot, and heterosis evidence together, on a short list of hybrids, IR 68888A/Sakur Dhan, IR 68888A/MTU 1010, and AMPS-6A/DRR Dhan 44 foremost, that are both high-yielding and, critically, Mid-early in duration. That duration match matters because it is what lets these hybrids escape the terminal-season drought and cold stress that climate change is making more frequent and less predictable in this hill agro-ecology, while also freeing the field early enough for a second crop and a second income.

These hybrids are therefore ready to move from this study into multi-location trials, not as a purely academic result but as the region's first concrete, climate-ready candidates for its own hybrid rice programme. The clearest limitation this work exposed, that SSR-based genetic distance only weakly

predicts which crosses will actually perform, points directly to what should come next: adopting denser, trait-linked marker systems so that future candidate hybrids can be screened faster and at lower cost, a priority that will only grow more urgent as a shifting climate narrows the window for testing each cross the slow way, in the field.

Declarations

CRedit Author contribution statement

Amit Kumar: Conceptualization, Investigation, Resources and software, Writing - preparation of the original draft, Writing - review and editing, Project administration. Avinash Pandey: Conceptualization, Investigation and Methodology. Jayanta Bora: Investigation and Methodology. Letngam Touthang: Resources and software. Vikram Jeet Singh: Resources and software, Writing - review and editing. Simardeep Kaur: Data curation, Writing - review and editing. Veerendra Kumar Verma: Resources and software, Validation. Shridhar Ragi: software, review and editing; S. P. Das: Data curation, Writing - review and editing.

Declarations

(a) Ethics approval and consent to participate: Not Applicable

(b) Consent for publication: Not Applicable

Funding

This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

Author Contribution

A. K: Conceptualization, Investigation, Resources and software, Writing - preparation of the original draft, Writing - review and editing, Project administration. A. P.: Conceptualization, Investigation and Methodology. Jayanta Bora: Investigation and Methodology. L.T.: Resources and software. V.J.S: Resources and software, Writing - review and editing. S.K.: Data curation, Writing - review and editing. V.K.V: Resources and software, Validation. S.R: software, review and editing; S.P.D: Data curation, Writing - review and editing.

Acknowledgement

The author would like to thanks Director, ICAR-RC for NEG region, Umiam, Meghalaya

Data availability

The data has been submitted in the Mendeley Data repository with DOI: 10.17632/9n864gnc99.1.

Tables

Tables 1-5 are available in the Supplementary Files section.

References

1. Singh VJ, Vinod KK, Krishnan SG, Singh AK. Rice adaptation to climate change: opportunities and priorities in molecular breeding. In: *Molecular Breeding for Rice Abiotic Stress Tolerance and Nutritional Quality*. 2021. pp. 1–25. 10.1002/9781119633174.ch1
2. Bin Rahman AN, Zhang J. Trends in rice research: 2030 and beyond. *Food Energy Secur.* 2022;12. 10.1002/fes3.390.
3. Yadav HK, Singh VJ, Bisen P, Sonu, Ragi S, Nandakumar S, et al. Multi-trait selection of drought tolerant rice genotypes based on seedling and root architecture responses to PEG-induced osmotic stress. *Front Sustainable Food Syst.* 2026;10:1865363. <https://doi.org/10.3389/fsufs.2026.1865363>.
4. Goyal SK, Singh JP. Demand versus supply of foodgrains in India: Implications to food security. Paper presented at: 13th International Farm Management Congress; 2002. Wageningen, The Netherlands. p. 20.
5. Yuan LP, Virmani SS, Mao CX. Hybrid rice - achievement and outlook. *Progress in Irrigated Rice Research*. Manila, The Philippines: International Rice Research Institute; 1989. pp. 219–35.
6. Sheeba NK, Viraktamath BC, Sivaramakrishnan S, Gangashetti MG, Khera P, Sundaram RM. Validation of molecular markers linked to fertility restorer gene(s) for WA-CMS lines of rice. *Euphytica.* 2009;167:217–27. 10.1007/s10681-008-9865-4.
7. Suresh Kumar D. Rice production in India: analysis of trend, constraints and technologies. *India Studies in Business and Economics*. Singapore: Springer; 2019. pp. 7–33. 10.1007/978-981-13-8363-2_2.
8. Shidenur S, Singh VJ, Vinod KK, Gopala Krishnan S, Ghritlahre SK, Bollinedi H, et al. Enhanced grain yield in rice hybrids through complementation of fertility restoration by rf3 and rf4 genes as revealed by multilocation evaluation of tropical japonica derived rice (*Oryza sativa*) hybrids. *Plant Breed.* 2020;139:743–53. 10.1111/pbr.12818.
9. Shidenur S, Singh VJ, Vinod KK, Krishnan SG, Ghritlahre SK, Bollinedi H, et al. Molecular detection of WA-CMS restorers from tropical japonica-derived lines, their evaluation for fertility restoration and adaptation. *Plant Breed.* 2019;138:553–67. 10.1111/pbr.12701.

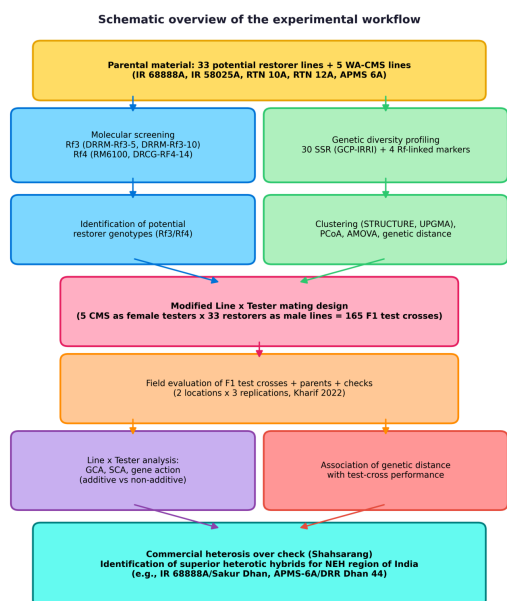
10. Kallugudi J, Singh VJ, Vinod KK, Krishnan SG, Nandakumar S, Dixit BK, et al. Population dynamics of wide compatibility system and evaluation of intersubspecific hybrids by indica-japonica hybridization in rice. *Plants*. 2022;11:1930. 10.3390/plants11151930.
11. Subbaiyan GK, Waters DL, Katiyar SK, Sadananda AR, Vaddadi S, Henry RJ. Genome-wide DNA polymorphisms in elite indica rice inbreds discovered by whole-genome sequencing. *Plant Biotechnol J*. 2012;10:623–34. 10.1111/j.1467-7652.2011.00676.x.
12. Praveen M, Anurag GU, Suneetha N, Umakanth AV, Patil JV, Madhusudhana R. Inheritance and molecular mapping of RF6 locus with pollen fertility restoration ability on A1 and A2 cytoplasm in sorghum. *Plant Sci*. 2015;238:73–80. 10.1016/j.plantsci.2015.05.020.
13. Zhang G, Lu Y, Bharaj TS, Virmani SS, Huang N. Mapping of the RF-3 nuclear fertility-restoring gene for WA cytoplasmic male sterility in rice using RAPD and RFLP markers. *Theor Appl Genet*. 1997;94:27–33. 10.1007/s001220050377.
14. Yao FY, Xu CG, Yu SB, Li JX, Gao YJ, Li XH, et al. Mapping and genetic analysis of two fertility restorer loci in the wild-abortive cytoplasmic male sterility system of rice (*Oryza sativa* L). *Euphytica*. 1997;98:183–7. 10.1023/a:1003165116059.
15. Alavi M, Ahmadikhah A, Kamkar B, Kalateh M. Mapping Rf3 locus in rice by SSR and CAPS markers. *Int J Genet Mol Biol*. 2009;7:121–6.
16. Revathi P, Medoju P, Singh AK, Sundaram RM, Raju S, Senguttuvel P, et al. Efficiency of molecular markers in identifying fertility restoration trait of WA-CMS system in rice. *Indian J Genet Plant Breed*. 2013;73:89. 10.5958/j.0019-5200.73.1.012.
17. Zeng L, Kwon TR, Liu X, Wilson C, Grieve CM, Gregorio GB. Genetic diversity analyzed by microsatellite markers among rice (*Oryza sativa* L.) genotypes with different adaptations to saline soils. *Plant Sci*. 2004;166:1275–85. 10.1016/j.plantsci.2004.01.005.
18. Sruthi K, Divya B, Senguttuvel P, Revathi P, Kemparaju KB, Koteswararao P et al. Evaluation of genetic diversity of parental lines for development of heterotic groups in hybrid rice (*Oryza sativa* L). *J Plant Biochem Biotechnol*. 2020. (2):236–52.
19. Singh VJ, Bhowmick PK, Vinod KK, Krishnan SG, Nandakumar S, Kumar A, et al. Population structure of a worldwide collection of tropical japonica rice indicates limited geographic differentiation and shows promising genetic variability associated with new plant type. *Genes*. 2022;13:484. 10.3390/genes13030484.
20. Kumar A, Singh VJ, Krishnan SG, Vinod KK, Bhowmick PK, Nagarajan M, et al. WA-CMS-based ISO-cytoplasmic restorers derived from commercial rice hybrids reveal distinct population structure and genetic divergence towards restorer diversification. *3 Biotech*. 2019;9. 10.1007/s13205-019-1824-3.
21. Kempthorne O. *An Introduction to Genetic Statistics*. New York: John Wiley & Sons Inc.; 1957. pp. 468–73.
22. Sprague GF, Tatum LA. General vs. specific combining ability in single crosses of corn. *J Am Soc Agron*. 1942;34:923–32.

23. Ahangar L, Ranjbar GA, Nouroozi M. Estimation of combining ability for yield and yield components in rice (*Oryza sativa* L.) cultivars using diallel cross. *Pakistan J Biol Sciences: PJBS*. 2008;11(9):1278–81. <https://doi.org/10.3923/pjbs.2008.1278.1281>.
24. Chen J, Zhou H, Xie W, Xia D, Gao G, Zhang Q, et al. Genome-wide association analyses reveal the genetic basis of combining ability in rice. *Plant Biotechnol J*. 2019;17:2211–22.
25. Reif JC, Melchinger AE, Xia XC, Warburton ML, Hoisington DA, Vasal SK, et al. Genetic distance based on simple sequence repeats and heterosis in tropical maize populations. *Crop Sci*. 2003;43:1275–82. [10.2135/cropsci2003.1275](https://doi.org/10.2135/cropsci2003.1275).
26. United Nations. *Transforming Our World: The 2030 Agenda for Sustainable Development*. New York: United Nations General Assembly; 2015.
27. Sarma B, Tanti B. Analysis of genetic diversity of certain species of *Aristolochia* using ISSR-based molecular markers. *Curr Life Sci*. 2017;3:47–53.
28. Balaji Suresh P, Srikanth B, Hemanth Kishore V, et al. Fine mapping of Rf3 and Rf4 fertility restorer loci of WA-CMS of rice (*Oryza sativa* L.) and validation of the developed marker system for identification of restorer lines. *Euphytica*. 2012;187:421–35. [10.1007/s10681-012-0737-6](https://doi.org/10.1007/s10681-012-0737-6).
29. Liu K, Muse SV. PowerMarker: an integrated analysis environment for genetic marker analysis. *Bioinformatics*. 2005;21:2128–9. [10.1093/bioinformatics/bti282](https://doi.org/10.1093/bioinformatics/bti282).
30. Yeh FC, Boyle TJB. Population genetic analysis of co-dominant and dominant markers and quantitative traits. *Belg J Bot*. 1997;129:157.
31. Peakall RO, Smouse PE. GENALEX 6: genetic analysis in Excel. Population genetic software for teaching and research. *Mol Ecol Notes*. 2006;6(1):288–95. <https://doi.org/10.1111/j.1471-8286.2005.01155.x>.
32. Excoffier L, Lischer HEL. Arlequin suite ver 3.5: a new series of programs to perform population genetics analyses under Linux and Windows. *Mol Ecol Resour*. 2010;10:564–7. [10.1111/j.1755-0998.2010.02847.x](https://doi.org/10.1111/j.1755-0998.2010.02847.x).
33. Kumar A, Bhowmick PK, Singh VJ, Malik M, Gupta AK, Seth R, et al. Marker-assisted identification of restorer gene(s) in ISO-cytoplasmic restorer lines of WA cytoplasm in rice and assessment of their fertility restoration potential across environments. *Physiol Mol Biol Plants*. 2017;23:891–909. [10.1007/s12298-017-0464-5](https://doi.org/10.1007/s12298-017-0464-5).
34. Katara JL, Verma RL, Nayak D, Ngangkham U, Ray S, Subudhi H, et al. Frequency and fertility restoration efficiency of rf3 and rf4 genes in Indian rice. *Plant Breed*. 2017;136:74–82. [10.1111/pbr.12401](https://doi.org/10.1111/pbr.12401).
35. Tan YP, Li SQ, Wang L, Liu G, Hu J, Zhu YG. Genetic analysis of fertility-restorer genes in rice. *Biol Plant*. 2008;52:469–74. [10.1007/s10535-008-0092-6](https://doi.org/10.1007/s10535-008-0092-6).
36. Shalini P, Manonmani S, Robin S. Genetic analysis of fertility restoration under CGMS system in rice (*Oryza sativa* L.) using three-way test-cross method. *J Genet*. 2015;94:9–16. [10.1007/s12041-015-0478-0](https://doi.org/10.1007/s12041-015-0478-0).

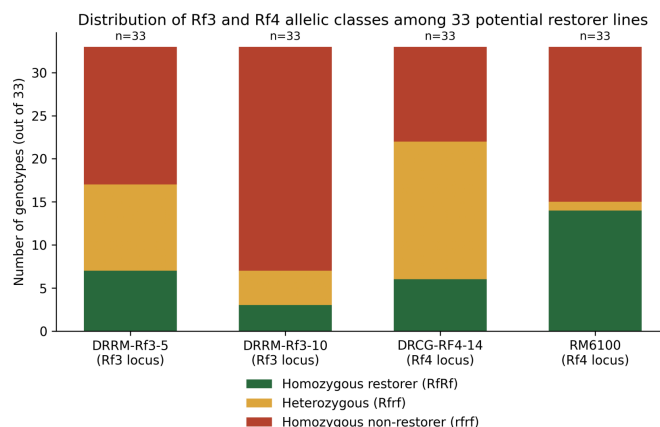
37. Balakrishnan C, Kumar A, Raj R, Verma VK, Touthang L, Kumar R, et al. Exploring genetic diversity, population structure and stability for yield related traits in rice germplasm of Northeastern India. *Genet Resour Crop Evol.* 2024;72:2631–51. 10.1007/s10722-024-02109-w.
38. Raj R, Kumar A, Balakrishnan C, et al. Unraveling the genetic diversity, population structure, and stability for yield-related traits of rice genotypes in mid-hills of northeastern India. *J Appl Genet.* 2024. 10.1007/s13353-024-00925-5.
39. Zhang P, Li J, Li X, Liu X, Zhao X, Lu Y. Population structure and genetic diversity in a rice core collection (*Oryza sativa* L.) investigated with SSR markers. *PLoS ONE.* 2011;6. 10.1371/journal.pone.0027565.
40. Hashimoto Z, Mori N, Kawamura M, Ishii T, Yoshida S, Ikegami M, et al. Genetic diversity and phylogeny of Japanese sake-brewing rice as revealed by AFLP and nuclear and chloroplast SSR markers. *Theor Appl Genet.* 2004;109:1586–96. 10.1007/s00122-004-1794-6.
41. Siwach P, Jain S, Saini N, Chowdhury VK, Jain RK. Allelic diversity among Basmati and non-Basmati long-grain indica rice varieties using microsatellite markers. *J Plant Biochem Biotechnol.* 2004;13:25–32. 10.1007/bf03263186.
42. Kumar A, Singh VJ, Gopala Krishnan S, Bhowmick PK, Vinod KK, Seth R, et al. Molecular marker-based estimates of genetic distance and prediction of heterosis in rice (*Oryza sativa*). *Indian J Agric Sci.* 2020;90:1439–44. 10.56093/ijas.v90i8.105939.
43. Azad AK, Sarker U, Ercisli S, Assouguem A, Ullah R, Almeer R, et al. Evaluation of combining ability and heterosis of popular restorer and male sterile lines for the development of superior rice hybrids. *Agronomy.* 2022;12(4):965. 10.3390/agronomy12040965.
44. Rahman MM, Sarker U, Swapan MA, Raihan MS, Oba S, Alamri S, et al. Combining ability analysis and marker-based prediction of heterosis in yield reveal prominent heterotic combinations from diallel population of rice. *Agronomy.* 2022;12:1797. 10.3390/agronomy12081797.
45. Sarker U, Biswas PS, Prasad B, Mian MAK. Heterosis and genetic analysis in rice hybrids. *Pak J Biol Sci.* 2002;5:1–5.
46. Banumathy S, Thiagarajan K, Vaidyanathan P. Study on magnitude of heterosis of rice hybrids for yield and its components. 2003; 287–93.
47. Singh NK, Singh AK, Sharma CL, Singh PK, Singh ON. Study of heterosis in rice using line x tester mating system. *Oryza.* 2007;44:260–3.
48. Roy SS, Senapati BK, Sinhamahapatra SP, Sarkar KK. Heterosis for yield and quality in rice. *Oryza.* 2009;46:87–93.

Figures

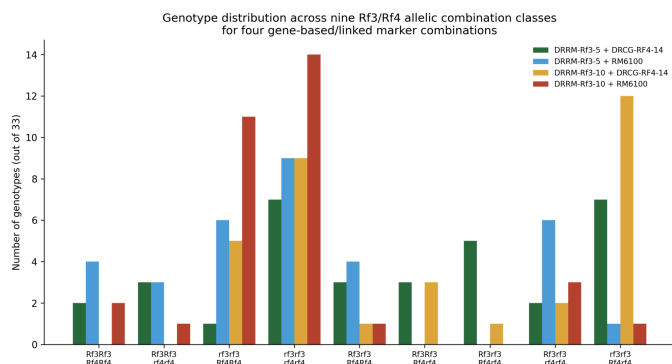
a



b



c



d

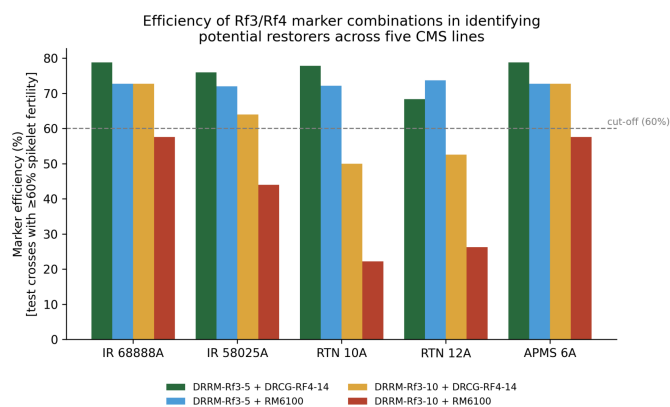


Figure 1

Experimental workflow and molecular screening for *Rf3/Rf4* fertility restorer genes.

(a) Schematic overview of the experimental workflow, showing how molecular screening for *Rf3/Rf4* restorer genes and SSR-based genetic-diversity profiling of the 38 parental lines feed into the line x tester test-cross evaluation, combining-ability analysis, and the identification of heterotic hybrids. (b) Distribution of *Rf3* and *Rf4* allelic classes (homozygous restorer, heterozygous, homozygous non-restorer) among the 33 potential restorer lines, for each of the four *Rf3/Rf4* markers used. (c) Number of genotypes assigned to each of the nine *Rf3/Rf4* allelic combination classes, for the four gene-based/gene-linked marker combinations used for screening. (d) Efficiency of the four *Rf3/Rf4* marker

combinations in identifying potential restorers, expressed as the percentage of test crosses reaching $\geq 60\%$ spikelet fertility, for each of the five CMS lines; the dashed line marks the 60% cut-off used to call a cross restored.

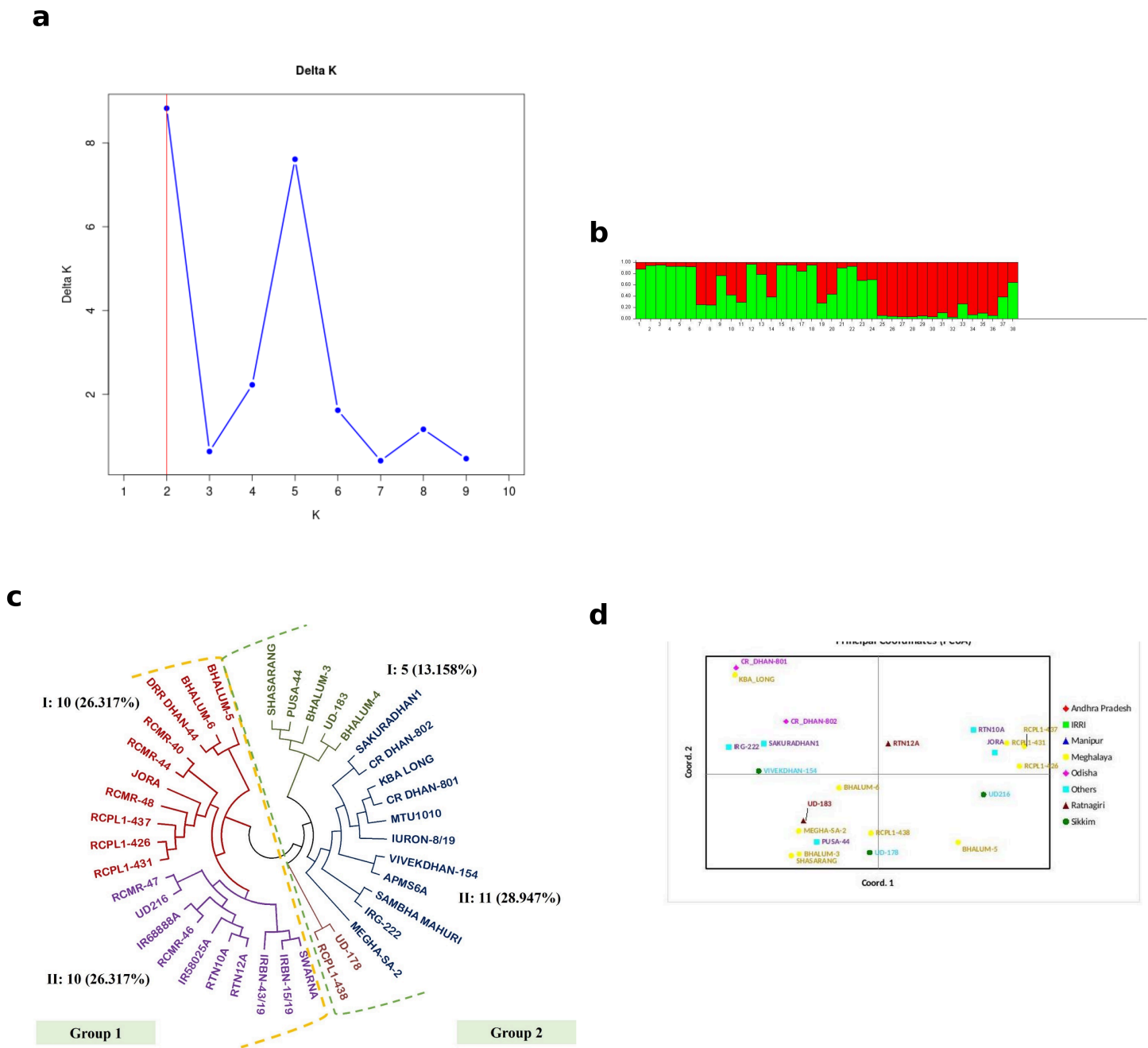


Figure 2

Population structure of the 38 parental lines based on SSR markers.

(a) Estimated ΔK across $K = 1-10$ from STRUCTURE analysis, identifying $K = 2$ as the optimal number of clusters. (b) STRUCTURE bar plot showing individual membership coefficients at $K = 2$. (c) UPGMA dendrogram of genetic distance among the 33 restorer and 5 CMS lines. (d) Two-dimensional scatterplot from principal coordinate analysis (PCoA).

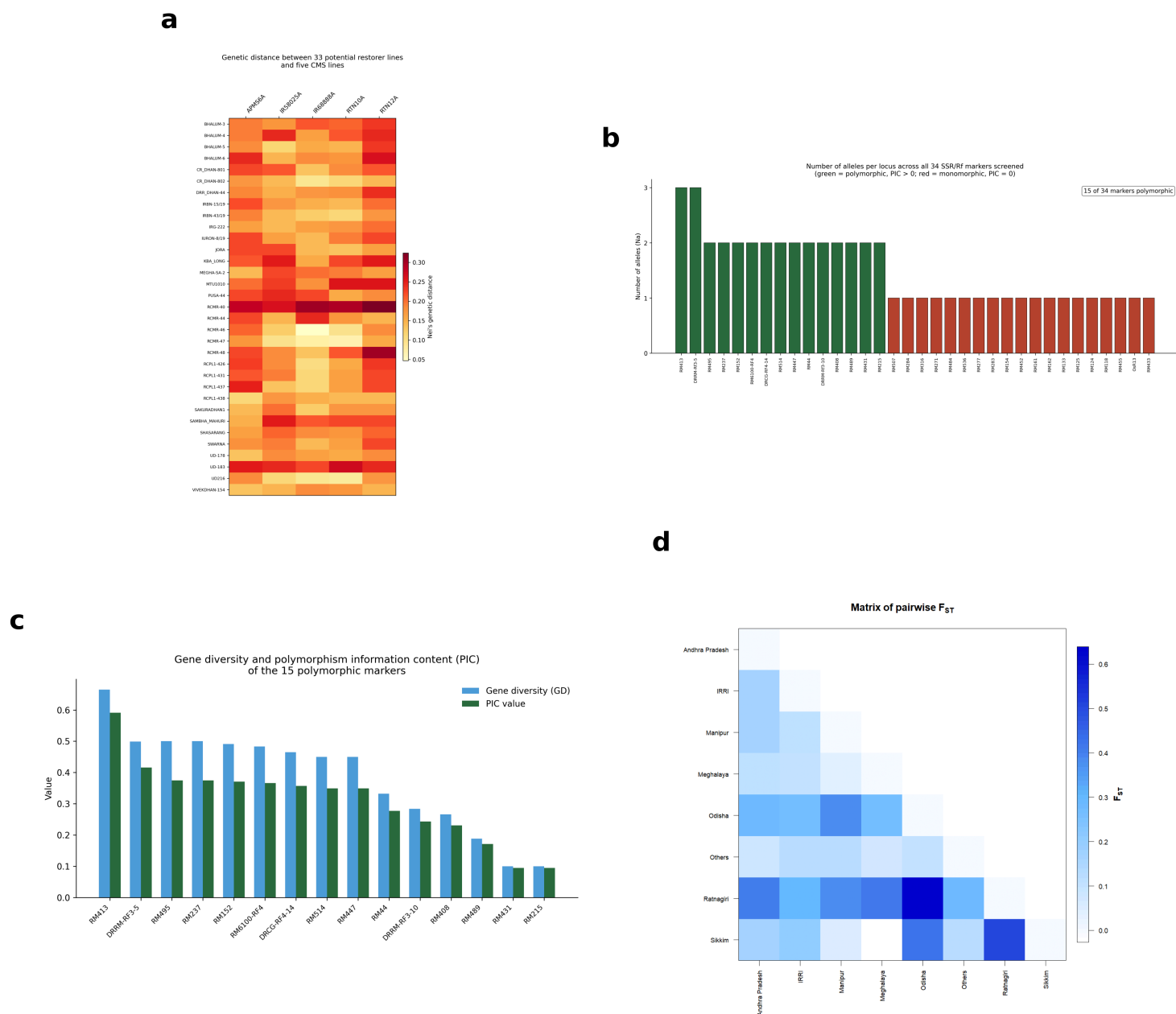


Figure 3

Marker diversity statistics and geographic differentiation.

(a) Heatmap of Nei's genetic distance between each of the 33 potential restorer lines and the five CMS lines. (b) Number of alleles (N_a) per locus across all 34 SSR/Rf markers screened, distinguishing the 15 polymorphic markers from the 19 monomorphic markers (full statistics in Supplementary Table S6). (c) Gene diversity (GD) and polymorphism information content (PIC) of the 15 polymorphic SSR/Rf markers, ranked by PIC value. (d) Pairwise F_{ST} matrix among the potential restorer and CMS lines, grouped by state/region of origin.

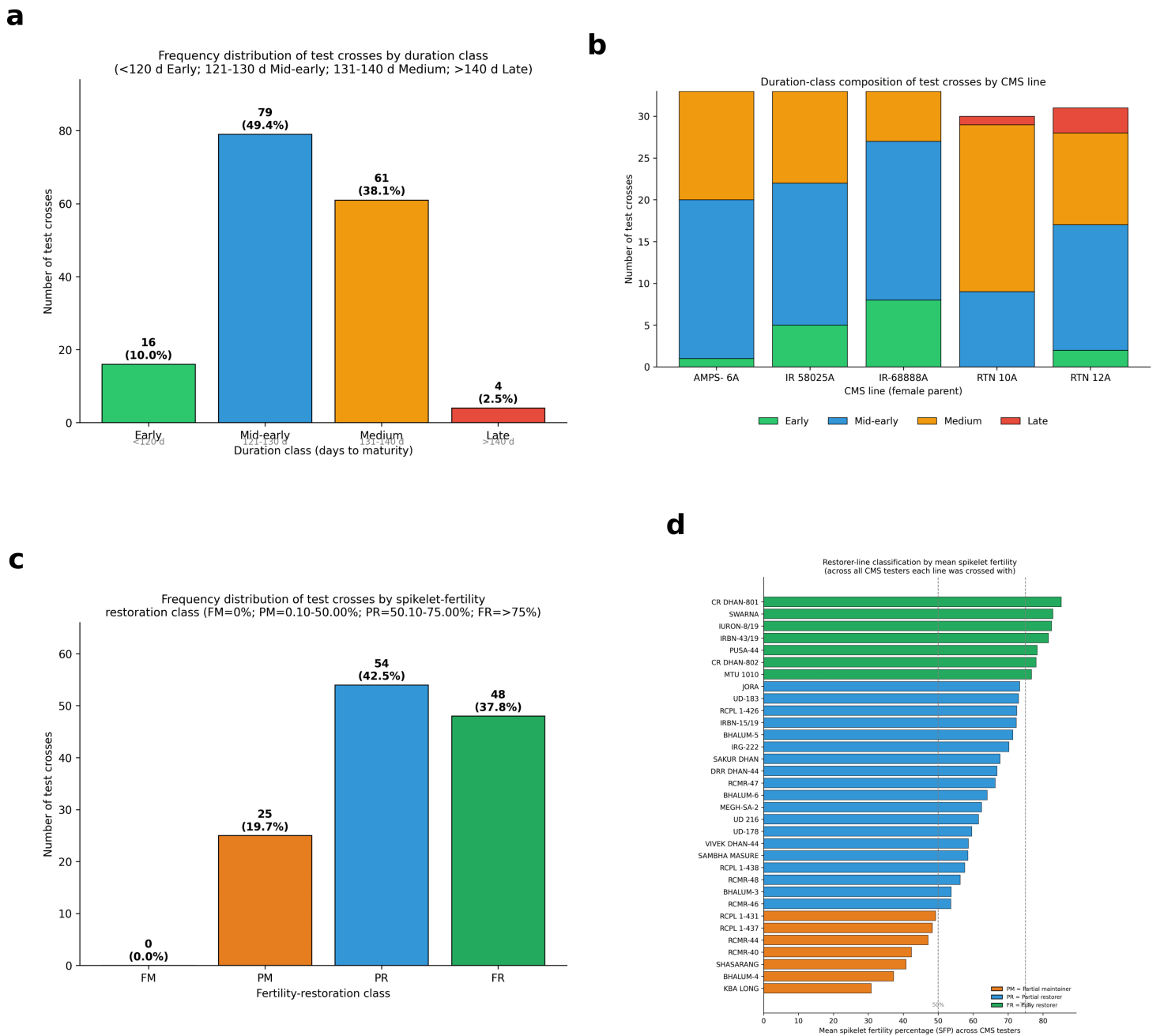


Figure 5

Duration and spikelet-fertility restoration classification of test crosses.

(a) Frequency distribution of the 160 test crosses with recorded maturity duration across four duration classes (Early <120 days; Mid-early 121-130 days; Medium 131-140 days; Late >140 days). (b) Duration-class composition of test crosses broken down by CMS line (female parent), showing that the female parent systematically shifts the maturity-duration profile of the resulting hybrids. (c) Frequency distribution of the 127 phenotyped test crosses across four spikelet-fertility restoration classes (FM = Fully maintainer, 0%; PM = Partial maintainer, 0.10-50.00%; PR = Partial restorer, 50.10-75.00%; FR = Fully restorer, >75%). (d) All 33 restorer lines ranked by mean spikelet fertility percentage across their CMS testers and colour-coded by classification (full underlying data in Table 1).

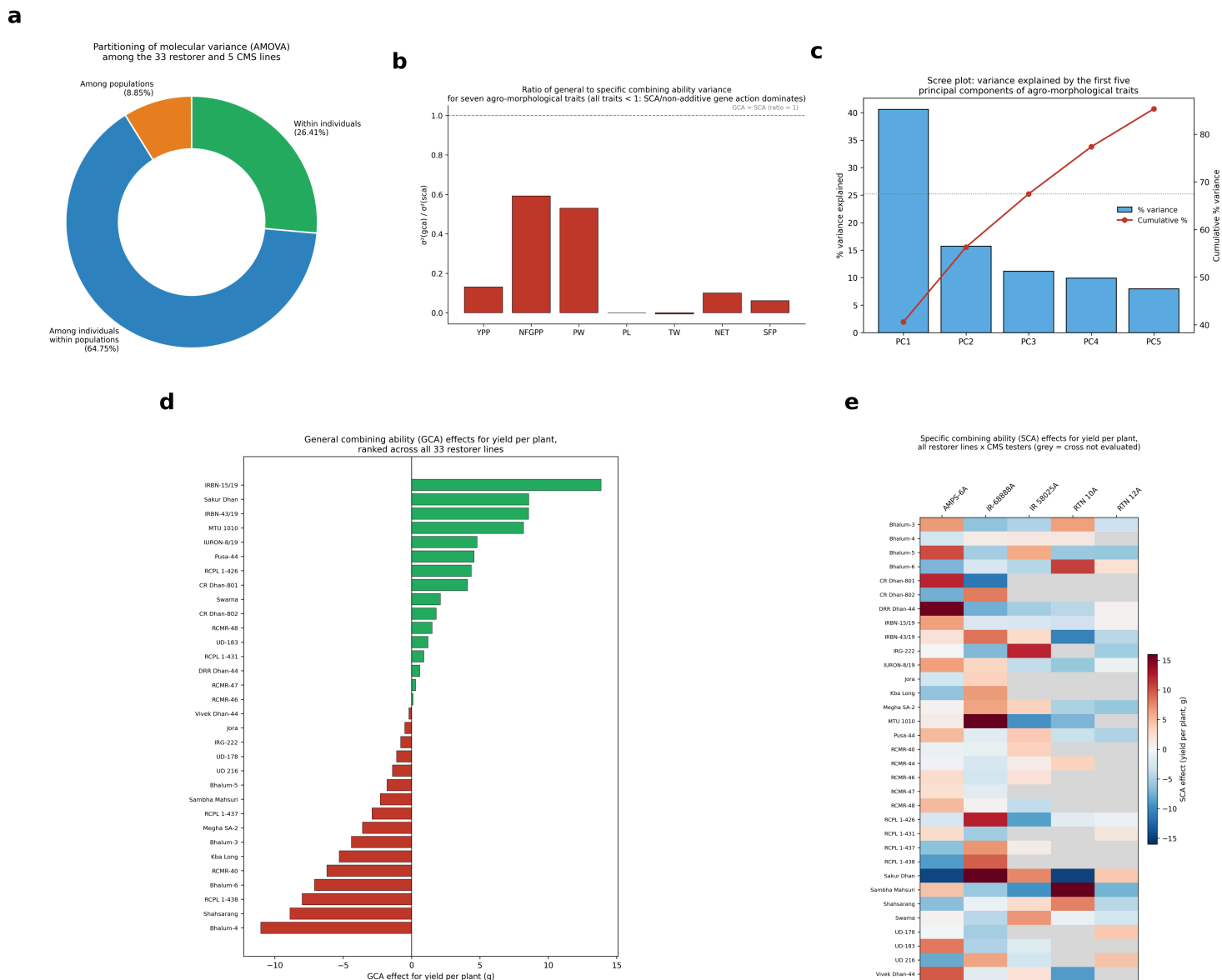


Figure 6

Genetic components, combining ability, and trait architecture underlying the heterosis results.

(a) Partitioning of molecular variance (AMOVA) among the 33 restorer and 5 CMS lines. (b) Ratio of general to specific combining ability variance ($\sigma^2_{gca}/\sigma^2_{sca}$) for seven agro-morphological traits; all values below the dashed reference line (ratio = 1) indicate that specific combining ability/non-additive gene action dominates. (c) Scree plot of variance explained by the first five principal components of agro-morphological traits. (d) General combining ability (GCA) effects for yield per plant, ranked across all 33 restorer lines. (e) Specific combining ability (SCA) effects for yield per plant across all evaluated restorer-line x CMS-tester combinations (grey cells = cross not evaluated).

Supplementary Files

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

- [Tables.docx](#)
- [Supplementarytable.docx](#)