HORIZON
e-Publishing Group
HePG

RESEARCH ARTICLE

Identification of false smut - resistant donors in Rice (*Oryza sativa* L.) and analysis of their morpho-molecular diversity for resistance breeding

Mangudi Sowmiya[1], Swaminathan Manonmani[1*], Nallamuthu Ramya Selvi[1], Ramasamy Saraswathi[2], Ramalingam Suresh[1], Chellappan Gopalakrishnan[3], Muthurajan Raveendran[4] & Palanisamy Dhamotharan[5]

[1]Department of Rice, Centre for Plant Breeding and Genetics (CPBG), Tamil Nadu Agricultural University, Coimbatore , Tamil Nadu - 641 003, India

[2]Department of Plant Genetic Resources, Centre for Plant Breeding and Genetics (CPBG), Tamil Nadu Agricultural University, Coimbatore, Tamil Nadu - 641 003, India

[3]Department of Rice, Centre for Plant Protection studies (CPPS), Tamil Nadu Agricultural University, Coimbatore, Tamil Nadu -641 003, India

[4]Director of Research, Tamil Nadu Agricultural University, Coimbatore, Tamil Nadu -641 003, India

[5]Department of Genetics and Plant Breeding, Tamil Nadu Agricultural University, Coimbatore - 641 003, Tamil Nadu, India

*Email: manonmanis@tnau.ac.in

OPEN ACCESS

Check for updates

Abstract

False smut disease of rice caused by the pathogen *Ustilaginoidea virens* is a growing threat to the rice farmers as it affects both quality and quantity. Development of resistant variety becomes difficult, since very few resistant donors were available for false smut resistant breeding programme. Therefore, to identify potential donors for resistance breeding a total of 60 genotypes were screened at hotspot location (Gudalur) during *kharif* 2023 which led to identification of 12 highly resistant genotypes and the notable ones are Koolavalai, Periya chandikar, Kapikar selection and Earapalli. Genetic variability studies indicated the presence of additive gene action for all the agronomic and disease related traits. Principal component analysis revealed the first 5 principal components collectively contributing 78.79 % of the total variance with disease-related traits contributing significantly to divergence. Ten clusters were delineated using Mahalanobis D^2 statistics with clusters IX and V showing higher inter cluster distance (3453.64). Forty-one polymorphic markers were used to analyse the genotypes and The Unweighted Pair Group method with Arithmetic Mean (UPGMA) clustering by Jaccard distance formed 6 clusters. The Bayesian clustering classified the entire population into 2 subpopulations. False smut linked marker RM336 and RM218 were found to be the most informative marker with high Polymorphism Information Content (0.71, 0.69) and Heterozygosity Index (0.76, 0.73). The resistant genotypes such as IG71, Thulasi vasanai samba, Arupatham vellai, Kaltikar and Chinna aduku nel can be used in the future breeding programmes to develop the resistant cultivar and to identify the candidate genes governing resistance.

Keywords

False smut; phenotyping; genetic diversity; marker profiling; SSR markers

Introduction

Rice (*Oryza sativa* L.) is one of the most important staple crops consumed in the world and provides nearly half of the world's calories. The expanding global population is leading to an ever-increasing need for rice production, which can be done by breaking the yield barrier. However, rice production is

constantly threatened by various biotic and abiotic stresses. According to survey reported, the annual losses due to rice diseases are approximately 10-15 % worldwide (1). Among the diseases, Rice floral diseases are the most devastating diseases and false smut, a floral infecting fungal disease caused by the pathogen *Ustilaginoidea virens*, has emerged as a notable concern due to its emergence as epidemic disease in India, China and Japan (2, 3). The reason for the epidemic condition may be due to increased cultivation of high yielding varieties, hybrids, surplus usage of nitrogenous fertilizers, climatic variations (4). The disease incidence is favoured by high humidity (>90 %) and rainfall conditions with lower temperatures of about 25-30 °C during flowering time (5). The date of sowing also has a great impact on the disease severity (6). Depending on the rice cultivar and environmental conditions, a yield penalty of 3-70 % was observed globally (7). A yield loss of 0.5 - 75 % was reported in India (8). The typical disease symptom is the transformation of individual grains into yellowish orange smut balls, which eventually turns to a blackish green smut ball (2). This disease not only reduces the yield but also deteriorates the quality of rice by producing mycotoxins like *Ustiloxins* and *Ustilaginoidins*, posing substantial health hazards to animals and human beings. The most widely used approach to control the plant diseases is the use of chemical fungicide. Use of fungicides after the appearance of the false smut symptoms is little effective due to the fact that the actual infection process happens much earlier during booting stage of the crop. And also, it is very difficult to predict the disease epidemics as interaction between host, pathogen and environment is dynamic and highly complex (9, 10) as there is not much work done on the epidemiology of rice false smut. The only approach to tackle the disease relies heavily in the development and deployment of resistant rice varieties. But the process of developing resistance against false smut coupled with high yield through breeding depends on the availability of resistant sources, variability of the traits, heritability and genetic advance over percentage of mean (GAM) and divergence of the genotypes. Unfortunately, very limited information is available on these aspects. Germplasms have been recognised as the reservoir of various biotic and abiotic stresses resistance. Cultivated rice has restricted genetic base and limited degree of resistance. So, screening of germplasms for false smut resistance could help us to identify the best promising donors. Genetic diversity within the genotypes provides a vast reservoir of traits that can be harnessed to develop new, disease-resistant varieties. Enhanced understanding of the genetic relatedness of the breeding materials may also contribute for better germplasm conservation. Hence, the identification of diverse sources of resistance to false smut is crucial as it aids in parental selection and helpful in broadening the genetic base of cultivated rice. Mahalanobis D^2 statistics and Principal Component Analysis (PCA) are the multivariate statistical analysis tools used to study the genetic difference among genotypes based on the quantitative traits (11, 12). Diversity analysis based on morphological traits alone cannot give reliable results as these traits are highly influenced by environment (13) and controlled by additive gene effect. Therefore, molecular markers can be employed to find the distinctiveness of the genotypes as it is more reliable and time saving. A range of molecular markers like Restriction Fragment Length Polymorphism, Random Amplified Polymorphic DNA, Amplified Fragment Length Polymorphism, Simple Sequence Repeats (SSR), Inter Simple Sequence Repeats and Single Nucleotide Polymorphism were used for diversity analysis. Among all these molecular markers, the SSR marker was found to be capable of detecting the high polymorphism and is extensively used in rice (14). Although few studies have been conducted to identify false smut resistant donors, not much work has been carried out on analysing the variability and diversity of the genotypes. Therefore, this study aims to evaluate 60 rice accessions to identify the potential donors for false smut resistance and assess their genetic variability and genetic diversity at both morphological and molecular level to provide valuable insights.

Materials and Methods

Morphological screening of rice germplasms:

A total of 118 germplasms were initially screened in the *kharif* 2022 which resulted in 60 resistant lines. These 60 rice accessions were forwarded for further screening along with 2 resistant (RPHP42, IG25) and 2 susceptible checks (CO43, CO(R)50) for false smut resistance based on inoculation method (15). The seeds were collected from the Paddy Breeding Station, Coimbatore. The experiment was conducted in Hybrid Rice Evaluation Centre, Gudalur, Tamil Nadu, India during *kharif* 2023 which provides a very favourable weather conditions for disease development. The hotspot is located at 11.5 °N latitude and 76.5 °E longitude and an altitude of 1072 m above mean sea level. A minimum temperature of 16 °C to a maximum of 26 °C, with a relative humidity of 75-81 % was observed during the booting stage. Twenty-one-day old seedlings were transplanted to the main field in randomized block design with 60 test entries, 4 checks and 3 replications spaced at 20 x 20 cm apart. All the recommended standards of agronomic practices were executed. The list of genotypes and checks used are listed in Table 1.

Inoculum preparation:

The false smut pathogen *U.virens* was isolated from smut balls and surface sterilized with 1 % sodium hypochlorite or 70 % ethanol for 30 sec. The sterilized smut balls were washed with sterile water for further 30 sec and left to dry. Then the spore was scraped or cut out from the smut ball and plated in potato sucrose agar which was incubated at 25 ± 2 °C for 10-12 days. Then the fungal patches were taken and proliferated in milky grain broth for 7 days, with agitation of 120 rpm at 25 °C. Freshly prepared 100 mL potato broth was introduced to the culture and left in shaker for 1 week. Final spore suspension was diluted to a concentration of 2 x 10^5 conidia/mL using sterile water.

Table 1. List of genotypes studied.

Sl. No.	Genotypes	Sl. No.	Genotypes	Sl. No.	Genotypes
1	Panamara samba	23	Burma block	44	RP-Bio-226
2	Arupatham samba	24	Godavari samba	45	RH2-SM-1-2-2
3	Channagi	25	Vadivel	46	Red sirumani
4	Poongar selection	26	Karungam	47	katta samba
5	Kapikar selection	27	Pamani samba	48	ponmani samba
6	Varigarudan samba	28	Arasamba	49	IG75
7	Varigarundam samba selection	29	Koolavalai	50	Chinna aduku nel selection
8	Kama samba	30	Chinna aduku nel	51	IG12
9	Kama samba selection	31	Periya chandikar	52	Thillainayagan
10	Karthigai samba	32	RPHP125	53	Vellai kudaivelan
11	Purple puttu	33	kaltikar	54	RPHP163
12	Chithan samba	34	Earapalli	55	IG18
13	Ghandhasala	35	kalarkar	56	IG28
14	Norungan	36	sornavari	57	Nootripatum
15	Kallukar	37	RPHP103	58	Muthuvellai
16	Kallundikar	38	Ramakuruvaikar	59	Uppumolagai
17	Vadakathi samba	39	Murungankar	60	Rangoon samba
18	Sadai samba	40	IG71	61	CO 43 - susceptible check
19	Thulasi vasanai samba	41	Jeevan samba	62	RPHP42- resistant check
20	Rajamudi	42	RPHP104	63	CO(R)50- susceptible check
21	Arupatham vellai	43	IG49	64	IG25- resistant check
22	Mangam samba				

Artificial inoculation of false smut pathogen:

To screen the entries, an artificial injection method was done for successful infection to the panicles at the booting stage (Fig. 1). A 2 mL of spore suspension was injected in 5 tillers per plant across 5 plants per entry. All the morphological and false smut disease resistance related traits were recorded from the tagged plants. Disease related traits *viz.,* number of infected panicles/plant (NIPP), number of infected grains/panicle (NIGPa), number of infected grains /plant (NIGP), percentage of infected panicles/plant (PIPP) and percentage of infected grains / panicle (PIGPa), morphological traits *viz.,* days to 50 % flowering (DFF), plant height (PH), number of productive tillers /plant (NPT), panicle length (PL), number of florets / panicle (NFPP), number of chaffy grains /panicle (NCGP), percentage of spikelet sterility (PSS), hundred grain weight (HSW) and single plant yield (SPY) were recorded.

Phenotypic scoring:

The number of infected grains/panicles in each inoculated plant was assessed using a scale of 0-9 under 6 category and was evaluated to determine the resistance level of rice plants against false smut, based on the scoring system outlined which is cited in (16) as provided in Table 2.

Molecular assay:

Fresh and young green leaves were collected from 15 days old seedlings. Genomic DNA was extracted according to the CTAB method (17). A total of 72 markers covering 12 chromosomes were used for genotyping the 60 accessions. For polymerase chain reaction, a total reaction mixture of 10 µL : 1 µL of DNA template , 0.5 µL of forward primer, 0.5 µL of reverse primer, 3.5 µL of master mix and 4.5 µL of sterile water was prepared. The PCR profile involved an initial denaturation at 95 °C (5 min), followed by 35 cycles
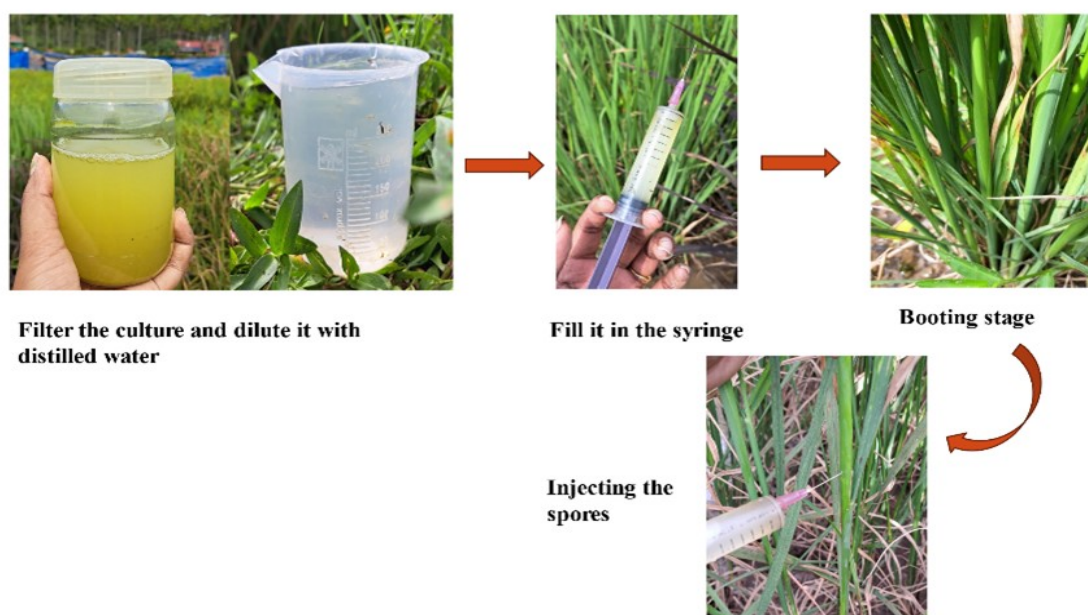

Fig. 1. Artificial inoculation of false smut pathogen through injection at booting stage.

Table 2. Disease scale for scoring.

Score	Number of infected florets/panicles	Category
0	0	Highly resistant
1	1	Resistant
3	2	Moderately resistant
5	3-6	Moderately susceptible
7	7-10	Susceptible
9	>11	Highly susceptible

of 94 ˚C (1 min), 55-65 ˚C (45 sec) and 72 ˚C (45 sec), with a final extension at 72 ˚C (10 min). The final PCR product was stored at 4 ˚C. The amplified DNA was separated on a 3 % polyacrylamide gel stained with ethidium bromide, run in 1X TBE (Tris-Boric-EDTA) buffer at 100V for 90 min, alongside a 100 bp ladder and detected under UV light using a Bio-Rad documentation unit.

Genetic diversity analysis:

For the morphological diversity all the 60 genotypes and 4 checks were included and for the molecular diversity analysis, 60 test genotypes along with one resistant (RPHP42) and one susceptible check (CO43) were analysed. Mahalanobis D² statistics was employed to discern distinct genotypes based on their morphological data, effectively capturing the multidimensional relationships and correlations among traits, which supports the breeding program by selecting desirable genotypes (11). PCA was used to transform a set of possibly correlated variables into a lower dimension uncorrelated variables yet retaining the maximum variance (12). To analyse molecular diversity, UPGMA clustering and Bayesian clustering was used in the study since it offers a perspective on the hybridization of genotypes from different clusters depending on their levels of resistance and susceptibility. The list of 41 polymorphic markers were listed in Table 3. Scoring of bands for the 41 polymorphic markers was performed based on the binary format '1' and '0' for the presence and absence of alleles. This binary format data was utilized to calculate the Jaccard dissimilarity coefficient. The Jaccard dissimilarity coefficient was used to form the clusters based on the Unweighted Pair Group Method with Arithmetic mean (UPGMA). Population structure analysis was done by using the base pair scoring data. The marker information parameters like polymorphism information content (PIC), observed heterozygosity, gene diversity, resolving power were calculated to identify the best informative marker. Resolving power (RP) of marker were calculated by using the formula (18),

$$RP = \Sigma \; Ib$$ where Ib (Band Informativeness) = (1-2|0.5-pI)

p = proportion of individuals containing the band.

Statistical Analysis:

Analysis of variance (ANOVA) for randomized block design was carried out with the replicated data. Variability study was conducted to analyse the genetic parameters for false smut traits and agronomic characters. Principal component analysis, a multivariate statistical technique was executed to identify the variation patterns. All these analyses were done using 'Agricolae', 'Variability', 'factomineR' and 'factoextra' packages of R studio. Mahalanobis D^2 Cluster analysis using tochers method was done by using the statistical package TNAUSTAT (19). The jaccard distance were calculated by 'PBPERFECT', an R-shiny based software (20). The marker information parameters like polymorphism information content, observed heterozygosity, gene diversity was calculated using powermarker v.3.25 software. Bayesian clustering was performed using the software STRUCTURE 2.3.4 and the results were obtained from STRUCTURE SELECTOR. Analysis of molecular variance (AMOVA), fixation index (Fst), rate of gene flow (Nm), observed (H_o) and expected (H_e) heterozygosity of the population were calculated using GenAlex 6.5 software (21).

Results

Morphological screening

Morphological screening of 60 accessions of rice germplasms for false smut resistance was conducted in the hotspot region (Gudalur), which revealed a diverse range of resistance levels when compared with the susceptible checks. The lines were divided into 6 categories namely highly resistant, resistant, moderately resistant, moderately susceptible, susceptible and highly susceptible based on the number of infected grains/panicle (Table 4). A total of 12 lines *viz.,* Channagi, Kapikar selection, Thulasi vasanai samba, Koolavalai, Periya chandikar, Earapalli, Sornavari, EC 728651, Thillainayagan, IG 18, Muthuvellai, RPHP163 showed zero symptoms making them highly resistant (Score 0) and can be considered as potential donors. Varigarudan samba, Sadai samba, Kallundikar, Rajamudi, Red sirumani, Ponmani samba, EC 738587, Nootripatum, Uppumolagai, Rangoon samba were found to be moderately susceptible (score 5). The accessions like Kama samba, Karthigai samba, Kodavari samba, Pamani samba, RPHP125, Mangam samba, Rama kuruvaikar, RH2-SM-1-2-1, Jeevan samba, Katta samba and Vellai kudaivelan were identified as moderately resistant lines (Score 3). A maximum of 8-9 smut balls per panicle were observed in the susceptible checks and 0 in the resistant check. All the disease related traits were significantly correlated (Table 5) with each other. Therefore, selection against any one of the traits is useful for further breeding program.

Morphological data

The mean data of the traits is provided in the supplementary file Table 1. The days to 50 % flowering ranged from 64.67 days (IG75) to 125.67 days (Vadakathi samba). The maximum plant height was observed for Chinna aduku nel (158.65 cm) and minimum for RP-Bio-226 (75.12 cm). An average of 19.40 productive tillers were observed. IG71 has the shortest panicle of 12.33 cm, lowest number of florets/ panicle (36.00) and a smaller number of chaffy grains (1.63). IG25 had the longest panicle of 32.67cm. More number of florets/ panicles was noted for Katta Samba (175.00). Vadivel recorded the highest percentage of spikelet sterility of 49.51 % and number of chaffy grains per panicle (61.67). The hundred seed weight

Table 3. List of polymorphic markers (chromosome number, forward and reverse primer sequences, annealing temperature, amplified product size range).

Marker	Chromosome number	Sequences	Annealing temperature ˚C	Amplified product size
RM3694	1	F: AAGGGGAAATCAACCTGTCC R:CAGAAGAGGCGAAGAAGACG	55	100-210
RM336	7	F: CTTACAGAGAAACGGCATCG R:GCTGGTTTGTTTCAGGTTCG	55	110-170
RM218	3	F: TGGTCAAACCAAGGTCCTTC R: GACATACATTCTACCCCCGG	55	110-170
RM242	9	F: GGCCAACGTGTGTATGTCTC TATATGCCAAGACGGATGGG	55	180-240
RM5638	1	F: GGCTTCCTCATCGCCATC R: CTGAGCAGCATTCCAGTCTG	55	200-260
RM144	11	F: TGCCCTGGCGCAAATTTGATCC R:GCTAGAGGAGATCAGATGGTAGTGCATG	55	190-280
RM235	12	F: AGAAGCTAGGGCTAACGAAC R: TCACCTGGTCAGCCTCTTTC	55	100-150
RM101	12	F: GTGAATGGTCAAGTGACTTAGGTGGC R: ACACAACATGTTCCCTCCCATGC	55	280-310
RM9	1	F: GGTGCCATTGTCGTCCTC R: ACGGCCCTCATCACCTTC	55	130-200
RM202	11	F: CAGATTGGAGATGAAGTCCTCC R: CCAGCAAGCATGTCAATGTA	55	160-200
RM264	8	F: GTTGCGTCCTACTGCTACTTC R: GATCCGTGTCGATGATTAGC	55	140-180
RM6085	11	F: GGTGAGAGATGGCTAAAGCG R: CATCGCCTCTAGCACCTCC	55	140-190
RM152	8	F: GAAACCACCACACCTCACCG R: CCGTAGACCTTCTTGAAGTAG	55	140-160
RM6925	8	F: TGAGAGGACGCTTGAAGAGG R: GCACCTAGTGACTGAAGGTTG	55	190-210
RM6208	8	F: TCGAGCAGTACGTGGATCTG R: CACACGTACATCTGCAAGGG	55	130-250
RM5609	12	F: CGCCAGTGTCGAATATGATG R: TCTTGGTGCAGTAGGTGCAC	55	140-180
RM339	8	F: GTAATCGATGCTGTGGGAAG R: GAGTCATGTGATAGCCGATATG	55	150-290
RM4589	12	F: GTTTAAACATGGGAGGTGTC R: CGAAATTTCTGAAATTTGGA	55	190-200
RM5926	11	F: ATATACTGTAGGTCCATCCA R: AGATAGTATAGCGTAGCAGC	55	100-160
RM211	2	F: CCGATCTCATCAACCAACTG R: CTTCACGAGGATCTCAAAGG	55	150-170
RM464	9	F: AACGGGCACATTCTGTCTTC R: TGGAAGACCTGATCGTTTCC	55	120-260
RM234	7	F: ACAGTATCCAAGGCCCTGG R: CACGTGAGACAAAGACGGAG	55	120-170
RM13679	2	F: AGATGACAAGGTGAGAGCACTGG R: TGGAGCCCAGAATTTCTAGATCG	55	270-290
RM229	11	F: CACTCACACGAACGACTGAC R: CGCAGGTTCTTGTGAAATGT	55	110-130
RM263	2	F: CCCAGGCTAGCTCATGAACC R: GCTACGTTTGAGCTACCACG	55	160-190
RM432	7	F: TTCTGTCTCACGCTGGATTG R: AGCTGCGTACGTGATGAATG	55	190-320
RM18451	5	F: ATATACAGCGCGGACATTGTGG R: CATGTCATCTTCACGCGAATCC	55	190-200
RM581	1	F: ACATGCGTGATCAACAATCG R: AATTGGATGTGGATGCACG	55	180-190
RM110	2	F: TCGAAGCCATCCACCAACGAAG R: TCCGTACGCCGACGAGGTCGAG	55	130-170
RM5341	12	F: TGCATTTTCCATACAATACG R: ATTTGATACATGGACGATGC	55	120-140
RM11	7	F: TCTCCTCTTCCCCCGATC R: ATAGCGGGCGAGGCTTAG	55	140-170
RM3773	9	F: CTGGATGAAAGGATACAACA R: CACATTATCTGTCAAGGTCC	55	180-190

RM6947	12	F: ATTAAACGTCCACTGCTGGC R: GCTAGGTTAGTGGTGCAGGG	55	150-170
RM216	10	F: GCATGGCCGATGGTAAAG R: TGTATAAAACCACACGGCCA	55	130-150
RM7056	1	F: GAAACGTGTAGCAGTACGCC R: ACCAAGCTCTTCATCAACGG	55	290-300
RM429	7	F: TCCCTCCAGCAATGTCTTTC R: CCTTCATCTTGCTTTCCACC	55	170-200
RM190	6	F: TTTGTCTATCTCAAGACAC R: TTGCAGATGTTCTTCCTGATG	55	100-120
RM244	9	F: CCGACTGTTCGTCCTTATCA R: CTGCTCTCGGGTGAACGT	55	170-180
RM26627	11	F: AGGTAGTACTTTGGCGCTAGTTACCC R: CCATAGTTGGGCACTGCTTGC	55	450-500
RM25292	10	F: TGTCCCTTTCTCCAATTCTCTCG R: TGTCTTCATCTTTGGCTCGATGG	55	100-140
RM104	1	F: GGAAGAGGAGAGAAAGATGTGTGTCG R: TCAACAGACACACCGCCACCGC	55	220-240

Table 4. Grouping of 60 genotypes based on the disease score.

Score	Number of infected florets/ panicle	Category	Number of genotypes	Genotypes
0	0	Highly resistant	12	IG71, Channagi, RPHP163, Earapalli, Sornavari, Thillainayagam, Muthuvellai, Thulasi vasanai samba, IG18, Kapikar selection, Koolavalai, Periya chandikar
1	1	Resistant	25	Panamara samba, Arupatham samba, Poongar selection, Kama samba selection, Purple puttu, Chithan samba, Ghandhasala, Norungam, Kallukar, Vadakathi samba, Arupatham vellai, Mangam samba, Burma block, Vadivel, Karungam, Arasamba, Chinna aduku nel, kaltikar, RPHP103, RPHP104, IG49, RP-Bio-226, Chinna aduku nel selection, IG12, IG28
3	2	Moderately resistant	11	Katta samba, RPHP125, RH2-SM-1-2-1, Ramakuruvaikar, Vellai kudaivelan, Kama samba, Godavari samba, Jeevan samba, Pamani samba, Karthigai samba, Kalarkar
5	3-6	Moderately susceptible	12	Kallundikar, Murungankar, Sadai samba, Varigarundam samba, Ponmani samba, IG75, Red sirumani, Nootripatum, Varigarundam samba selection, Rangoon samba, Rajamudi, Uppumolagai
7	7-10	Susceptible	-	-
9	>11	Highly susceptible	-	-

Table 5. Correlation between disease related variables.

	NIPP	PIPP	NIGPa	NIGP	PIGPa	SPY
NIPP	1.00	0.92	0.69	0.84	0.69	-0.10
PIPP	0.92	1.00	0.68	0.77	0.73	-0.13
NIGPa	0.69	0.68	1.00	0.90	0.93	-0.18
NIGP	0.84	0.77	0.90	1.00	0.82	-0.19
PIGPa	0.69	0.73	0.93	0.82	1.00	-0.22

Number of infected panicles/plant (NIPP), number of infected grains / panicle (NIGPa), number of infected grains /plant (NIGP), percentage of infected panicles/plant (PIPP) and percentage of infected grains /panicle (PIGPa), Single plant yield (SPY)

was maximum for Norungan (2.84 g) and minimum for Kaltikar (0.91 g). Highest single plant yield was observed for resistant genotype Ghandhasala (45.03 g) and lowest for the moderately susceptible genotype Rajamudi (10.51 g).

Genetic variability analysis

The presence of higher phenotypic coefficient of variation (PCV) than genotypic coefficient of variation (GCV) for all the traits indicates the influence of environment on its expression (Fig. 2). However, slight difference between the PCV and GCV reflects minimum environmental influence. All the traits except days to 50 % flowering, plant height and panicle length, exhibits high PCV and GCV. The moderate range of PCV and GCV was noted for days to 50 % flowering, panicle length and plant height. Broad sense heritability ranged from 61.45 to 98.63 %. High heritability (broad sense) coupled with high genetic advance over mean was observed for all the 5 disease related traits *viz.,* number of infected panicles/plants, number of infected grains/panicles, number of infected grains/plants, percentage of infected panicle/plant, percentage of infected grains/panicle and agronomic traits revealing the presence of additive gene action for its inheritance. The highest PCV, GCV, heritability and genetic advance over mean was observed for all the traits except for days to 50 % flowering, plant height and panicle length, so direct selection can be employed to improve these traits. Even though, days to 50 % flowering, panicle length and plant height have moderate PCV and GCV, the presence of high heritability and GAM favours its selection.
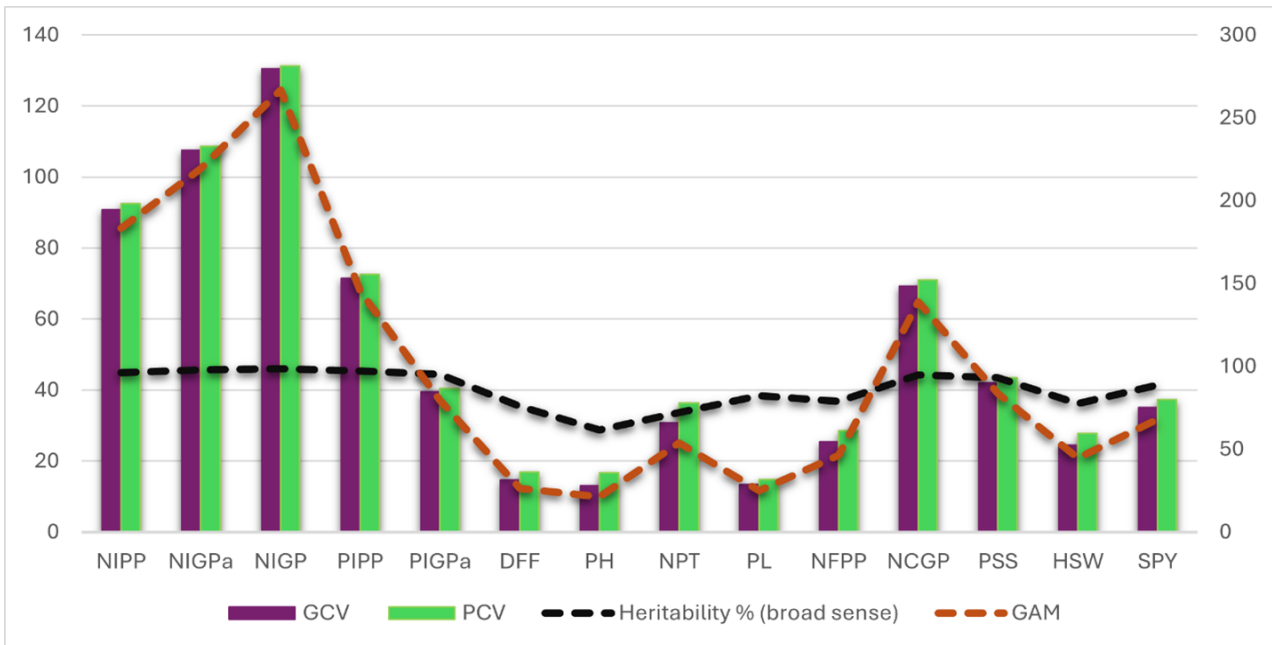
Fig. 2. Variability parameters of all the traits.

GCV- genotypic coefficient of variation; PCV- phenotypic coefficient of variation; GAM- Genetic Advance as percentage of mean.

NIPP- number of infected panicles/plant; NIGPa-number of infected grains/panicle; NIGP- number of infected grains/plant; PIPP- percentage of infected panicles/plant; PIGPa - percentage of infected grains/panicle; DFF- days to fifty per cent flowering; PH-plant height; PL-panicle length; NPT-number of productive tillers; NFPP -Number of florets/panicle; NCGP- number of chaffy grains/panicle; PSS- percentage of spikelet sterility; HGW- Hundred grain weight, SPY- single plant yield.

Morphological diversity

Principal component analysis

Analysis of variance (ANOVA) for the disease related traits and agronomic traits were presented in Table 6. The significant variation for the observed traits suggests the presence of sufficient diversity in the genotypes studied, which could be further used for developing false smut resistant cultivars. To analyse the structure of variance of the genetic diversity present in the germplasms, a principal component analysis (PCA) which is a multivariate dimensionality reduction tool was employed. PCA reduces the dimension by identifying the most significant traits that contribute for genetic variation. The eigen values, percent of variance, cumulative percent of variance and factor loadings for all the 14 traits given (Table 7). The scree plot (Fig. 3) depicts the relationship between each of

Table 6. Analysis of variance for disease and agronomic traits (ANOVA).

	Genotype	Replication	Residuals
Df	63	2	126
NIPP	21.352***	0.188	0.272
NIGPa	10.896***	0.016	0.079
NIGP	192.09***	0.47	0.89
PIPP	659.9***	4.3	6.1
PIGPa	0.8032***	0.0015	0.014
DFF	733.8***	37.5	70.6
PH	863.5***	3	149.3
NPT	117.62***	15.99	13.44
PL	29.108***	1.877	1.952
NFPP	2714.6***	4.4	223.9
NCGP	666.2***	20.4	12
PSS	335.5***	16.8	8
HSW	0.5593***	0.0778	0.0492
SPY	145.86***	0.8	5.93

***-Significant at 0.1 % probability level

Df- degrees of freedom, NIPP- number of infected panicles/plant NIGPa-number of infected grains/panicle, NIGP- number of infected grains/plant, PIPP- percentage of infected panicles/plant, PIGPa - percentage of infected grains/panicle, DFF- days to fifty per cent flowering, PH-plant height, PL-panicle length, NPT-number of productive tillers, NFPP -Number of florets/panicle, NCGP- number of chaffy grains/panicle, PSS- percentage of spikelet sterility, HGW- Hundred grain weight, SPY- single plant yield.

Table 7. Eigen value, proportionate variance, cumulative variance and factor loadings of first five principal components.

Variables	PC$_1$	PC$_2$	PC$_3$	PC$_4$	PC$_5$
Eigenvalue	4.680	2.090	1.770	1.340	1.150
Variance percent	33.460	14.900	12.630	9.590	8.210
Cumulative variance %	33.460	48.360	60.990	70.570	78.790
NIPP	0.412	0.127	0.035	-0.088	0.010
NIGPa	0.410	0.156	-0.087	-0.009	-0.117
NIGP	0.430	0.141	-0.035	-0.007	-0.054
PIPP	0.404	0.131	-0.064	-0.040	-0.013
PIGPa	0.405	0.100	-0.177	0.024	-0.201
DFF	0.044	-0.063	0.157	0.646	-0.294
PH	-0.004	0.098	0.334	0.567	0.002
NPT	-0.081	0.024	0.352	-0.338	-0.306
PL	0.152	0.172	0.450	0.118	0.446
NFPP	0.027	0.311	0.479	-0.198	0.245
NCGP	0.193	-0.475	0.356	-0.112	-0.140
PSS	0.183	-0.546	0.222	-0.056	-0.252
HSW	-0.169	0.392	-0.043	0.105	-0.474
SPY	-0.138	0.307	0.291	-0.239	-0.442

NIPP- number of infected panicles/plant NIGPa-number of infected grains/panicle, NIGP- number of infected grains/plant, PIPP- percentage of infected panicles/plant, PIGPa - percentage of infected grains/panicle, DFF- days to fifty per cent flowering, PH-plant height, PL-panicle length, NPT-number of productive tillers, NFPP -Number of florets/panicle, NCGP- number of chaffy grains/panicle, PSS- percentage of spikelet sterility, HGW- Hundred grain weight, SPY- single plant yield.
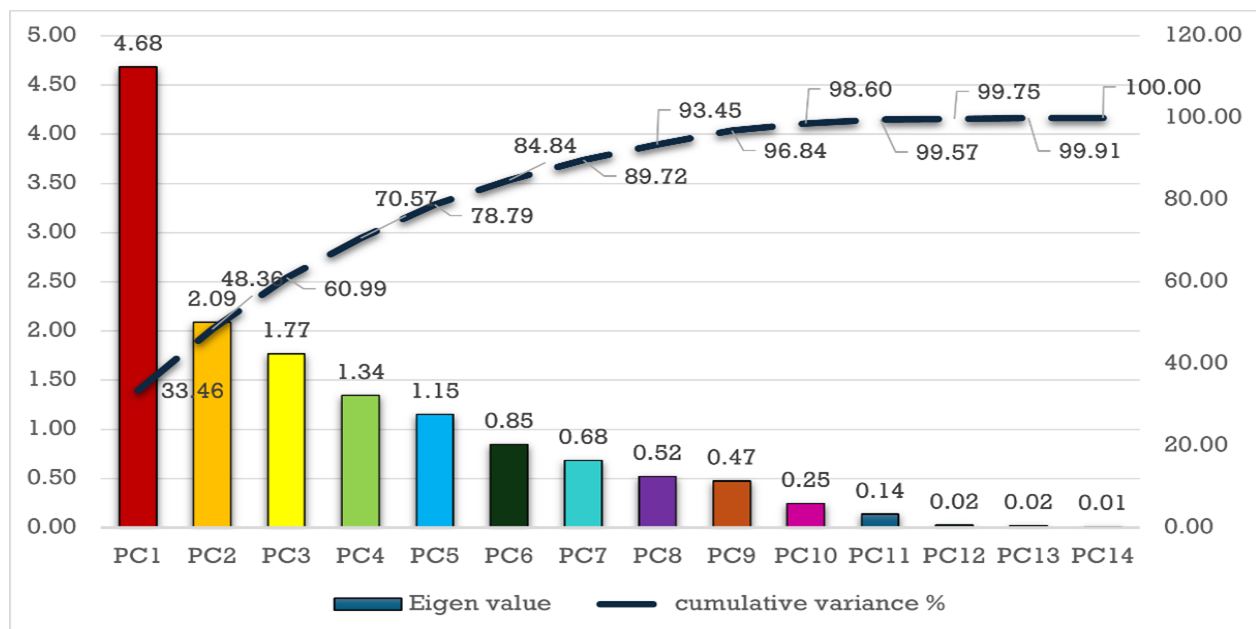
Fig. 3. Scree plot for eigen value and cumulative variance percent.

the principal components and the variance percent contribution. The first 5 principal components have eigen value of more than one accounting for 78.64 % of cumulative variance. PC_1 had eigen value of 4.68 with 33.45 % of total variability contributed by all the disease resistance traits *viz.,* NIPP, NIGPa, NIGP, PIGPa, PIPP followed by PC_2 with eigen value of 2.09 showing 14.90 % of total variance contributed by NCGP, PSS, NFPP and HSW in which NCGP and PSS exhibited negative contribution. PC_3 with 12.63 % of total variation is contributed by NFPP, PL, PH, NPT, NCGP, DFF, PH, NPT and SPY contributed 9.10 % of total variation to PC_4 with NPT and SPY contributing negatively. HSW, PL, SPY contributed to PC_5 with 7.79 % of total variance. The combined contribution of the other PCs to the overall divergence of the genotypes was 21.21 %.

From the factor loadings value, traits like NIPP (0.412), NIGPa (0.41), NIGP (0.43), PIPP (0.404) and PIGPa (0.405) have high positive significance for PC_1. All the traits except NCGP (-0.475), DFF (-0.063) and PSS (-0.546), contributes positive weightage for PC_2. For PC_3, DFF (0.157), PH (0.334), NPT (0.352), PL (0.45), NFPP (0.479), NCGP (0.356), PSS (0.222) and SPY (0.291) shows positive loadings and for PC_4, DFF (0.646), PH (0.566), PL (0.118) and HSW (0.105) gives positive high weightage. For PC_5, traits like PL (0.446) and NFPP (0.245) shows positive weightage.

The length of the loading vector for a principal component reflects the total contribution of all original variables to that component. The interaction between the traits contributing to variance was shown by the PCA biplot between PC_1 and PC_2 (Fig. 4 (A)). The longest vector is observed for NIGP followed by NIGP, NICGPa, NIPP, PIGPa, NCGP and PSS exhibiting its contribution to total divergence. The angle between the vector depicts the relation between them. If the angle is <90˚, it is positively related, >90˚ means negatively related and at 90˚ means uncorrelated (22). A negative correlation between single plant yield and all the disease related traits like NIPP, NIGPa, NIGP, PIGPa, PIPP was represented by an obtuse

angle between their corresponding vectors. Disease related traits namely NIPP, NIGP, PIGPa, NIGPa, PIPP produced an acute angle with NCGP, PSS and DFF indicating a positive correlation.

PCA biplot for genotypes (Fig. 4 (B)) depicts the relationship between the evaluated genotypes and variables. The most optimal genotypes pertaining to that specific characteristic are situated within the quadrant encompassing the variable vectors. The genotypes categorized as moderately susceptible for false smut disease like Kama samba, Varigarudan samba selection, Red sirumani, Ponmani samba, IG75, Uppumolagai, Rangoon samba and susceptible checks CO43, CO(R)50 were grouped in the first quadrant with the scoring vector NIGPa. The resistant genotypes *viz.,* Panamara samba, Kallukar, Poongar selection, Ghandhasala, Purple puttu, Mangam samba and Karungam performs better for single plant yield, 100 seed weight and number of productive tillers were grouped in the second quadrant. The highly resistant genotypes like Channagi, Kapikar selection, Thulasi vasanai samba, Koolavalai, Periya chandikar, RPHP140, IG25 congregated in the third quadrant.

Mahalanobis D² analysis

Diversity analysis is a very prominent tool for selecting the best and distant parents for hybridization programme. The diversity analysis was conducted using Mahalanobis D^2 statistics, while the clustering was performed utilizing Tocher's method. All 60 genotypes along with 2 resistant and 2 susceptible checks were categorized into 10 clusters according to their relative distances (D^2 values) (Table 8). The largest cluster is cluster I that includes 28 genotypes with only highly resistant and resistant genotypes, cluster II with 9 genotypes and cluster III with 11 genotypes. The clusters IV and VI comprised of 3 genotypes and clusters V, VII, IX, X contains 2 genotypes respectively. Cluster VIII is the solitary cluster. The distances within and between the clusters reflect the variation present within each cluster and across different clusters respectively. The maximum intra cluster distance was observed for cluster VII (273.74)
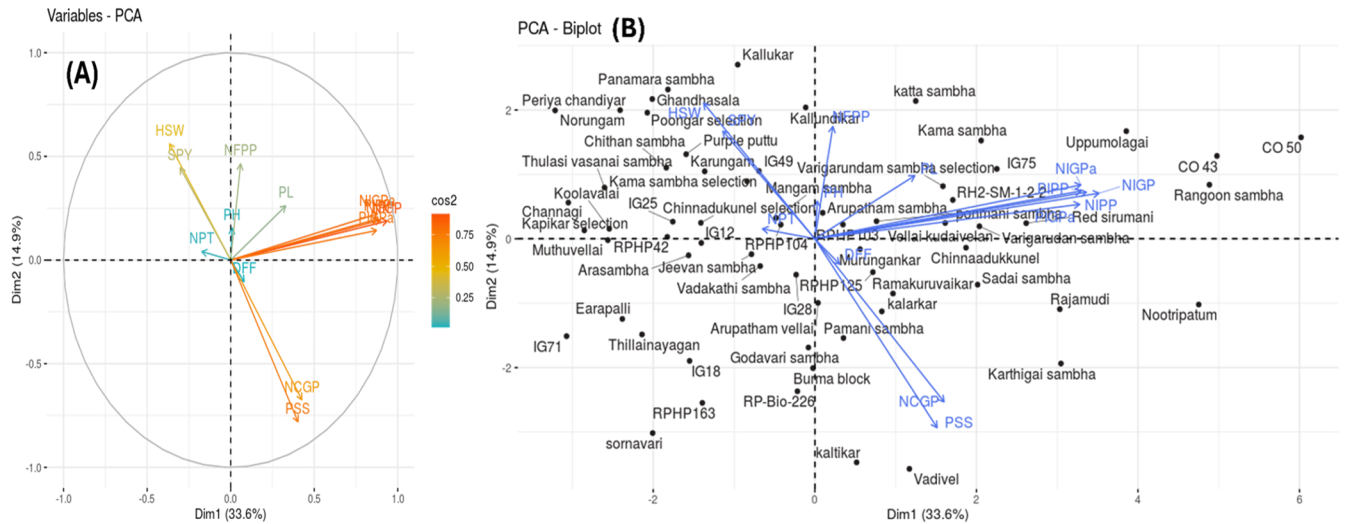
Fig. 4. (A) PCA biplot for dim1 and dim2 , (B) PCA Biplot for individual and variables

NIPP- number of infected panicles/plant; NIGPa-number of infected grains/panicle; NIGP-number of infected grains/plant; PIPP- percentage of infected panicles/plant; PIGPa - percentage of infected grains/panicle; DFF- days to fifty per cent flowering; PH-plant height; PL-panicle length; NPT-number of productive tillers; NFPP -Number of florets/panicle; NCGP- number of chaffy grains/panicle; PSS- percentage of spikelet sterility; HGW- Hundred grain weight, SPY- single plant yield.

Table 8. Grouping of clusters into 10 different clusters based on D^2 value.

Clusters	Genotypes
Cluster 1	Panamara samba, Arupatham samba, Channagi, Kapikar selection, Kama samba selection, Purple puttu, Chithan samba, Norungam, Vadakathi samba, Thulasi vasanai samba, Mangam samba, Burma block, Karungam, Arasamba, Koolavalai, Periya chandikar, Earapalli, RPHP104, IG49, Chinna aduku nel selection, IG12, Thillaninayagam, RPHP163, IG18, Muthuvellai, RPHP42, IG25.
Cluster 2	Poongar selection, Varigarudan samba, Varigarudan samba selection, Sadai samba, rajamudi, Red sirumani, IG75, Vellai kudaivelan, Uppumolagai.
Cluster 3	Kama samba, Kallukar, kallundikar, Arupatham vellai, Mangam samba, RPHP103, Rama kuruvaikar, Murungankar, RH₂-SM-1-2-1, Katta samba, Ponmani samba.
Cluster 4	Ghandhasala, Pamani samba, Godavari samba
Cluster 5	CO 43, CO(R)50
Cluster 6	Kaltikar, RP-Bio-226, Vadivel
Cluster 7	RPHP125, Nootripatum
Cluster 8	Jeevan samba
Cluster 9	IG71, Sornavari
Cluster 10	Karthigai samba, Chinna aduku nel

followed by cluster IX (268.35) and X (237.57), indicating extensive genetic diversity among the genotypes within this cluster (Table 9). Cluster VIII had zero intra cluster distance as it is a solitary cluster. The minimum intra cluster distance was witnessed in cluster V (49.26) followed by cluster I (152.01) highlighting that genotypes

within each cluster were closely related.

The inter-cluster distance values indicated the highest divergence between cluster V and IX (3453.64), followed by cluster V and I (3229.50) and cluster V and VII (3221.32). This suggests a greater diversity among the genotypes in these groups. Cluster V contains 2 susceptible check varieties therefore it is showing maximum diversity with other clusters having resistant genotypes. The cluster IV and VI (219.65) followed by cluster VIII and I (245.45) have lowest inter cluster distance highlighting the close relation among them.

Considerable importance should be given to the characters that contribute the most to the divergence when selecting genotypes and clusters (Fig. 5, Supplementary file Table 2). Number of infected grains/plants contribute more towards divergence (38.79 %) followed by number of chaffy grain/panicle (16.67 %), percentage of infected panicles/plant (10.07 %) and number of infected panicles/plant (7.04 %). This demonstrates the wide variation in genotypes for the number of infected panicles per plant. Based on the cluster means (Table 10), cluster IX recorded lowest mean values for number of infected panicles (0), number of infected grains/panicle (0), number of infected grains/plant (0), percentage of infected panicles/plant (0), percentage of infected grains/panicle (0.71) and days to 50 % flowering (85.5). Therefore, genotypes from these

Table 9. Average inter and intra cluster distance values of 10 clusters.

|	1	2	3	4	5	6	7	8	9	10
1	152.02	738.08	327.88	224.67	3229.51	313.04	1562.17	245.45	324.34	705.57
2		191.43	372.73	610.43	1232.72	646.33	379.59	703.16	1039.44	502.80
3			193.60	292.04	2289.09	331.22	866.25	312.87	648.07	406.77
4				185.56	2903.54	219.65	1280.34	293.17	489.81	550.23
5					49.27	3072.76	728.89	3221.32	3453.64	2329.26
6						158.46	1238.24	348.22	565.10	456.89
7							273.75	1501.06	1995.34	747.08
8								0.00	530.32	698.85
9									268.36	1157.08
10										237.58
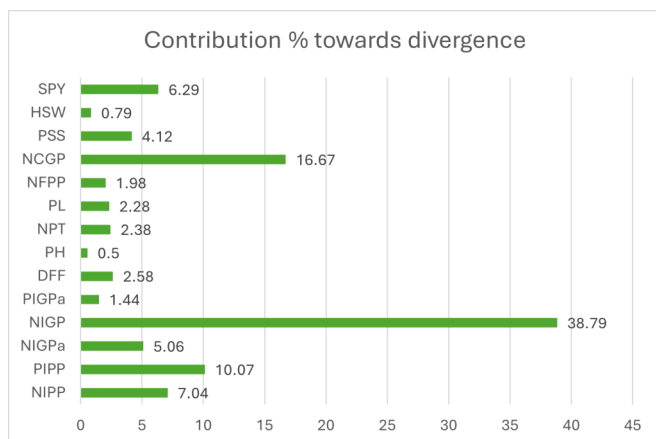
Fig. 5. Percent contribution of traits to total divergence.

NIPP- number of infected panicles/plant; NIGPa-number of infected grains/panicle; NIGP-number of infected grains/plant; PIPP- percentage of infected panicles/plant; PIGPa - percentage of infected grains/panicle; DFF- days to fifty per cent flowering; PH-plant height; PL-panicle length; NPT-number of productive tillers; NFPP -Number of florets/panicle; NCGP- number of chaffy grains/panicle; PSS- percentage of spikelet sterility; HGW- Hundred grain weight, SPY- single plant yield.

clusters can be selected as donors for developing false smut resistant cultivars. The highest mean values were recorded for single plant yield (31.61 g), 100 seed weight (1.84 g) and number of productive tillers (23.22) by cluster IV. Hence, choosing specific genotypes from these clusters would result in a wide range of variability in terms of yield and quality characteristics, thus allowing for additional selection for genetic enhancement.

Molecular diversity

Genetic profiling

Genotypic profiling of 60 genotypes along with one resistant (RPHP42) and one susceptible check (CO43) using SSR markers revealed 41 markers to be polymorphic. These 41 markers showed a total of 132 alleles with an average of 3.21 polymorphic bands per marker. The number of polymorphic bands ranged from 2 (RM58, RM7056, RM263, RM110, RM24, RM3773, RM6947, RM190, RM216, RM5341,

RM18451, RM25291) to 5 (RM9, RM3694, RM234, RM336, RM218, RM144) (Fig. 6). The attributes of individual SSR markers were evaluated by calculating their Polymorphism information content (PIC), gene diversity, resolving power (RP) (Table 11). PIC value ranged from 0.11-0.76 with an average of 0.43 per marker. High polymorphism information content (PIC) was observed for RM3694 (0.76) followed by RM336 (0.71) and RM218 (0.68). RM336 was found to be false smut linked marker which was reported (23) and RM218 (9). Lowest PIC value was observed by RM 5341 (0.11). RM3694 exhibited a high gene diversity value of 0.79, followed by RM336 at 0.75, with the average gene diversity across the samples being 0.48. The resolving power ranged from 1.77-2.94 with a mean of 2.09. The highest resolving power was identified in RM152 (2.94), RM9 (2.9) and RM6208 (2.87) following closely behind, while the lowest value was recorded for RM581 (1.77).

The Unweighted Pair Group method with Arithmetic Mean (UPGMA) clustering method grouped the genotypes into 6 clusters (Fig. 7, Supplementary file Table 3). The largest cluster, comprising 12 genotypes was cluster II, succeeded by cluster I and V, each containing 11 genotypes. Cluster III and IV are the smallest clusters with 10 genotypes each. Cluster VI contained 10 genotypes. Jaccard distance ranged from 0.28 to 0.84 with an average of 0.62 which indicated the presence of considerable genetic difference among the individuals (Supplementary file Table 4). High genetic dissimilarity (Jc=0.842) was observed between Norungan and RPHP125 followed by CO43 and IG49 (Jc=0.840). Minimum distance was observed between IG18 and RPHP163 (Jc=0.28) followed by Kallukar and Kallundikar (Jc=0.32).

Bayesian clustering

Forty-one markers were used to evaluate the population structure of 62 genotypes. The subgroups were examined across a range of K values from 1-10 and executed 3 times independently for each K value with a burn-in period of

Table 10. Cluster mean values of all the traits.

CLUSTER	NIPP	NIGPa	NIGP	PIPP	PIGPa	DFF	PH	NPT	PL	NFPP	NCGP	PSS	HSW	SPY
1	1.13	0.57	1.10	10.79	0.98	103.16	120.11	19.65	22.72	116.64	15.23	20.58	1.83	20.71
2	4.88	4.00	14.71	30.87	1.97	98.17	124.27	19.22	23.89	120.88	25.60	27.43	1.62	13.91
3	4.50	2.03	7.14	30.48	1.56	95.81	116.35	18.10	21.97	113.58	17.29	23.00	1.68	20.14
4	1.67	1.78	2.44	15.45	1.45	108.56	117.87	23.22	21.33	112.11	33.78	32.47	1.84	31.61
5	8.50	8.50	33.67	46.35	2.81	115.33	89.65	16.48	25.42	115.28	19.92	24.50	1.45	18.96
6	2.22	1.00	2.33	22.14	1.22	96.78	96.26	18.40	21.96	107.01	52.78	45.65	1.13	15.56
7	7.83	4.83	23.33	51.68	2.16	110.67	133.87	13.47	25.25	116.00	41.52	37.24	1.51	16.28
8	1.00	2.00	2.00	13.20	1.94	108.67	118.71	19.33	17.37	61.33	1.97	10.31	1.40	11.79
9	0.00	0.00	0.00	0.00	0.71	85.50	98.03	16.92	15.17	47.00	13.98	27.32	1.46	13.97
10	8.00	1.50	13.00	36.97	1.31	114.83	144.29	22.50	23.72	136.50	49.83	38.50	1.21	20.53

NIPP- number of infected panicles/plant NIGPa-number of infected grains/panicle, NIGP- number of infected grains/plant, PIPP- percentage of infected panicles/plant, PIGPa - percentage of infected grains/panicle, DFF- days to fifty per cent flowering, PH-plant height, PL-panicle length, NPT-number of productive tillers, NFPP -Number of florets/panicle, NCGP- number of chaffy grains/panicle, PSS- percentage of spikelet sterility, HGW- Hundred grain weight, SPY- single plant yield.
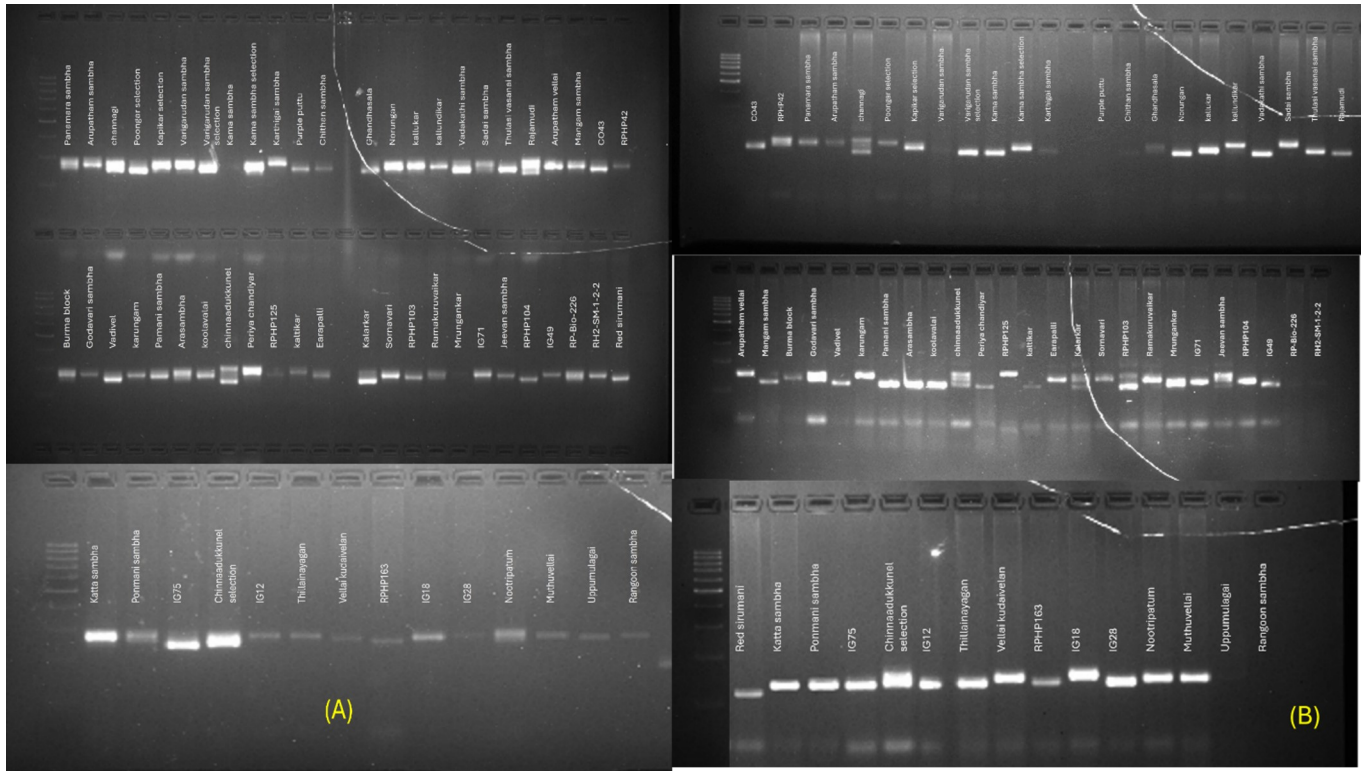
Fig. 6. Representation of banding pattern of (A) RM13679, (B)RM336.

Table 11. Number of alleles, polymorphism information content (PIC), gene diversity, observed heterozygosity, Resolving power (RP) of the polymorphic markers.

S.No	Marker	Number of alleles	Gene Diversity	Heterozygosity	PIC	RP
1	RM3694	5	0.8	0.03	0.77	1.81
2	RM336	5	0.76	0.03	0.71	2.03
3	RM218	5	0.73	0	0.69	1.94
4	RM242	4	0.72	0.05	0.66	2.03
5	RM5638	4	0.69	0.15	0.65	1.81
6	RM144	5	0.69	0.05	0.65	2.06
7	RM235	4	0.69	0	0.63	2
8	RM101	4	0.67	0	0.62	2
9	RM9	5	0.66	0.45	0.61	2.9
10	RM202	4	0.67	0.03	0.61	2.03
11	RM264	3	0.66	0.02	0.58	2.03
12	RM6085	4	0.64	0	0.58	1.97
13	RM152	3	0.64	0.48	0.56	2.94
14	RM6925	3	0.64	0.02	0.56	2
15	RM6208	3	0.62	0.45	0.55	2.87
16	RM5609	3	0.61	0	0.54	2
17	RM339	3	0.59	0.26	0.53	2.55
18	RM4589	3	0.59	0	0.52	2
19	RM5926	4	0.56	0	0.51	2
20	RM211	3	0.51	0.02	0.45	2.03
21	RM464	5	0.44	0.08	0.42	2.16
22	RM234	5	0.46	0.02	0.41	2.03
23	RM13679	3	0.45	0.02	0.4	2.03
24	RM229	3	0.45	0.02	0.38	2.03
25	RM263	2	0.5	0.02	0.37	1.97
26	RM432	3	0.18	0.05	0.17	2.1
27	RM18451	2	0.48	0	0.37	2
28	RM581	2	0.41	0	0.33	1.77
29	RM110	2	0.43	0	0.33	2
30	RM5341	2	0.12	0	0.11	2
31	RM11	3	0.35	0.02	0.31	2.03
32	RM3773	2	0.35	0	0.29	2
33	RM6947	2	0.33	0	0.28	2
34	RM216	2	0.32	0.18	0.27	2.35
35	RM7056	2	0.31	0	0.26	2.03
36	RM429	3	0.27	0.02	0.24	2.03
37	RM190	2	0.25	0	0.22	2
38	RM244	2	0.24	0.02	0.21	2.03
39	RM26627	3	0.23	0	0.21	2
40	RM25292	2	0.21	0	0.19	2
41	RM104	3	0.17	0.08	0.16	2.16

100000. A peak was observed at K= 2, Δk = 158.98 when plotted against ad hoc statistics (ΔK) (24), suggesting that the population could be divided into 2 distinct groups (Fig. 8 (A) and (B)). The subpopulation 1 (SP_1) consists of 39 genotypes with 69.23 % pure and 30.76 % admixtures. The subpopulation 2 (SP_2) consists of 23 genotypes with 69.56 % pure and 30.43 % admixtures. Both the populations had a mixture of all categories of genotypes from highly resistant genotype to susceptible genotype. Resistant and susceptible check was admixtured in nature, indicating that it can be grouped in either of the population. Analysis of Molecular Variance (AMOVA) depicts the total genetic variation in the population. Maximum variation of 84 % was observed among the individuals within the population and among the population, the divergence was only 4 %. (Table 12, Supplementary file Fig. 1). The 41 polymorphic markers exhibited an average fixation index of 0.033, suggesting a limited diversity within subpopulations (SP). The rate of Gene flow (Nm) was higher (7.263) which reflects a
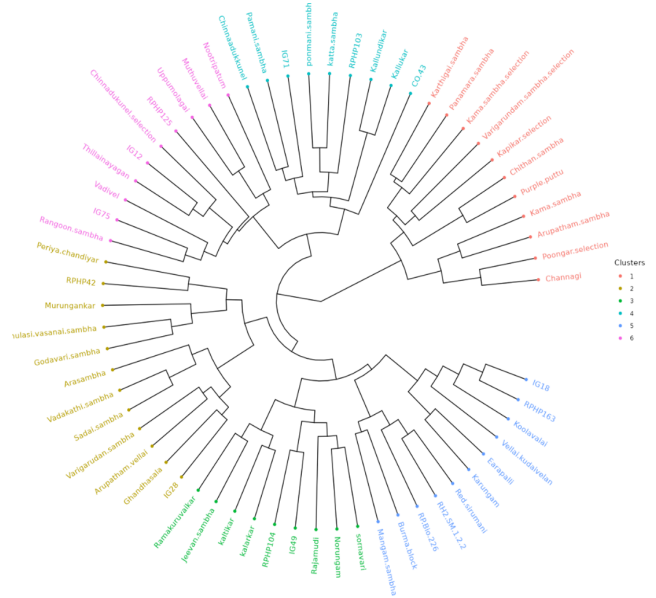

Fig. 7. Dendrogram based on Jaccard coefficients.

Table 12. Analysis of molecular variance among and between subpopulations.

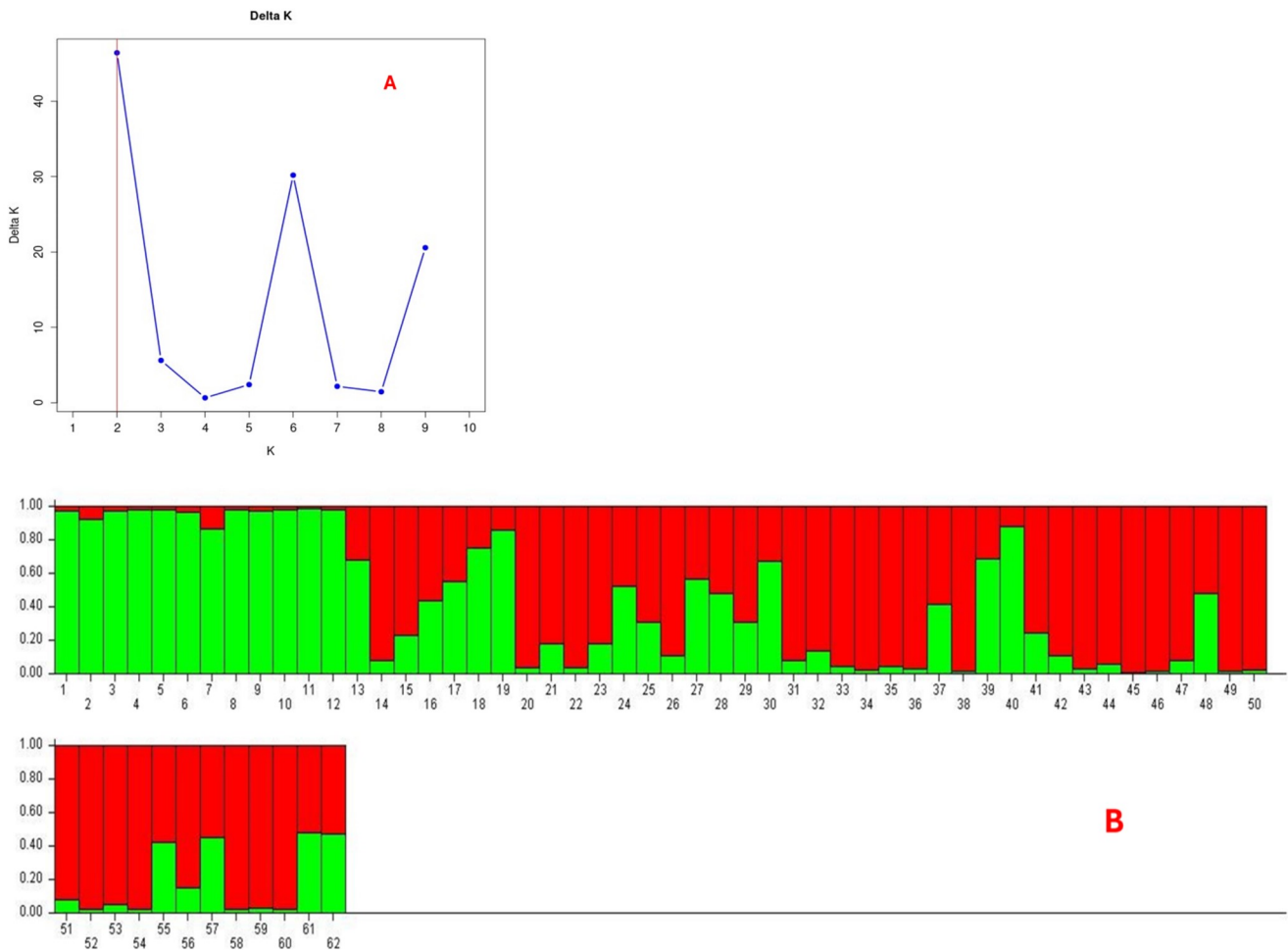| Source	Df	SS	MS	Estimated Variance	% of variation
Among Population	1	43.660	43.660	0.431	4 %
Among Individuals	60	1121.622	18.694	8.714	84 %
Within Individuals	62	78.500	1.266	1.266	12 %
Total	123	1243.782		10.411	100 %

Fig. 8. (A) Graphical representation of delta K value, (B) Bayesian clustering of rice genotypes into subpopulations (Red – SP_1, Green – SP_2). For genotypes name (Refer Table 1).

Table 13. Pairwise population matrix (Fst) and rate of gene flow (Nm) value of subpopulations along with heterozygosity index.

	pop1	pop2	Ho	He
pop1	0.000	7.263	0.061	0.486
pop2	0.033	0.000	0.063	0.448

Below diagonal- fst, above diagonal- Nm

substantial level of gene exchange among populations. The expected heterozygosity of the population was higher for SP_1 (0.486) (Table 13).

Discussion

The persistent occurrence of diverse diseases in rice cultivation has significantly impacted crop yields, raising concerns about food security (25). Among these, rice false smut has recently emerged as a notable epidemic disease leading to a decreased crop production and grain quality deterioration (3). Management of rice false smut through fungicides is effective only if they are applied prophylactically. Availability of false smut resistant donors for breeding programmes is limited. Under this background, identification of potential donors for disease-resistant breeding is crucial for sustainable disease management using host plant resistance. Therefore, conducting a thorough analysis of the genetic diversity and population structure within rice germplasm collections is vital for developing the effective breeding plans. By selecting superior and diverse resistant genotypes as parents, it is possible to enhance disease resistance and ensure long-term agricultural sustainability.

Phenotypic screening for false smut resistance was conducted in the hotspot region to ensure disease development. The planting was done in such a manner that flowering coincides with the monsoon rainfall. In addition, artificial inoculation was carried during late booting stage by syringe inoculation method to ensure that the screening was fool-proof. In the present study, the morphological screening of 60 accessions of rice germplasm at the hotspot region along with the resistant and susceptible check resulted in wide range of variation in resistance. The study identified 12 highly resistant, 25 resistant, 11 moderately resistant and 12 moderately susceptible lines. The 2 landraces, Purple puttu and Vadakathi samba, which were reported to be resistant (10), were also found to be resistant in our screening process. The highly resistant lines Koolavalai, Periya chandikar, Earapalli and Kapikar selection were identified to be high yielding. However, the yield of the remaining resistant lines was slightly lower compared to the high-yielding susceptible checks CO43 and CO(R)50 though they exhibited enhanced resistance. Similar findings of high resistance accompanied with lower yield were observed previously (10). Hence, these genotypes cannot be released directly because of their low yielding capacity. The decline in yield associated with false smut is not only due to the presence of smut balls but is also linked to an increase in chaffiness (8, 26). This study showed that the moderately susceptible lines Varigarudan samba,

Rajamudi, Sadai samba, Red sirumani, Nootripatum, Uppumulagai had higher spikelet sterility percentage and lower yield compared to highly resistant and resistant genotypes. Though the genotypes Kallundikar, IG75, Rangoon samba were moderately susceptible, they exhibited higher yield than the checks. So, these accessions can be improved by introgressing with the resistant lines for obtaining significant genetic gains.

Association studies for disease-related traits namely number of infected panicles, number of infected grains/panicles, number of infected grains/plants, percentage of infected grains/plant, percentage of infected panicles/plant revealed that all these traits were positively correlated with each other. Additionally, the negative correlation was observed between the single plant yield and disease related traits which indicate that yield loss occurs due to infection. Therefore, these traits should be considered while selecting the resistant genotypes. This is in accordance with the earlier studies (10, 27).

The genetic variability study showed that PCV was slightly higher than GCV which suggested less environmental influence on trait expression. High PCV, GCV, heritability and GAM were obtained for all the disease related traits *viz.,* number of infected panicles, number of infected grains/panicles, number of infected grains/plants, percentage of infected grains/plant, percentage of infected panicles/plant and yield attributes *viz.,* single plant yield, hundred grain weight, panicle length, number of productive tillers. The wide range of variability in these traits coupled with additive gene action, facilitates their importance in selection for developing resistant plant types. Similar findings of high variability accompanied with additive gene action for disease-related traits were reported for false smut traits (28) and for the number of blast spots per plant in rice (29). Comparable outcomes for yield attributing traits were also reported (30, 31).

Principal component analysis (PCA) delineates distinct axes of variation and quantifies the proportion of variance explained by each axis, highlighting the significance of traits in driving divergence along these axes (32). All the disease related traits contributed to the PC_1 with 33.41 % variability. Genotypes such as Ghandhasala, Kallukar, Poongar selection, Kapikar selection, Mangam samba, Purple puttu, IG49, Channagi and Thulasi vasanai samba which displayed high resistance and excellent performance for traits contributing to yield were clustered together near the relevant vectors. Although the genotypes like Thillainayagam, Sornavari, RPHP163 and IG18 showed high resistance they were positioned in the other quadrant due to low yield. The disease resistant genotypes coupled with good yield attributes can be selected for resistance breeding (33).

The analysis of Mahalanobis D^2 statistics recorded with 10 distinct clusters indicating the presence of adequate genetic divergence among the 64 genotypes studied. Most of the highly resistant and resistant genotypes were categorized collectively in cluster I with 29 genotypes. Within the cluster I, the intra cluster distance was low suggesting that the genotypes may share a common

ancestry. Cluster VII and Cluster IX showed a high intra-cluster distance, which could be attributed to the influence of past breeding practices on trait development (34). High inter cluster distance was also noticed for cluster V with clusters I, VIII and IX. The presence of 2 susceptible check varieties in Cluster V (CO43, CO(R)50) results in the highest level of diversity with other clusters (I, VIII, IX) which contain resistant genotypes. The genotypes from distinct clusters which showed high inter cluster distances can be crossed to obtain transgressive segregants in rice (35). The traits *viz.,* number of infected grains/panicle and number of chaffy grains/panicles contributed more for divergence in our study which could be considered as best discriminating factor and can be selected directly for crop improvement. But the disease-related trait (AUDPC) for bacterial leaf blight and number of chaffy grains/panicles were found to be least contributor in the findings of (36, 37). Low mean values for disease related traits were noticed in the clusters I, VIII and IX. The short duration genotypes have less chance of false smut disease occurrence in rice (38). The genotypes in cluster IX (IG71, Sornavari) were found to be highly resistant in addition to low mean value for days to 50 % flowering. But these two genotypes might have been escaped from the disease due to early flowering. The medium duration and highly resistant lines like Channagi, RPHP163, Earapalli, Thillainayagam, Muthuvellai, Thulasi vasanai samba, IG18, Kapikar selection, Koolavalai, Periya chandikar from cluster I may be considered as potential donors as they were exposed to conducive disease environment after confirming their resistance through repeated screening. The resistant genotype Ghandhasala from Cluster IV can also serve as a donor for developing high-yielding resistant cultivars as this cluster exhibits high mean values for yield-contributing traits.

Clustering the 62 genotypes with the use of SSR markers based on the jaccard distance formed 2 major clusters with 6 subclusters. These clusters were formed despite of their resistance levels. In the findings, 24 landraces evaluated by 12 SSR markers were clustered into 6 subclusters (39) whereas analysed 32 rice accessions with 34 SSR primers and reported the formation of 4 clusters (40). In the current study, the numbers of alleles and the average PIC was 3.21 and 0.43 respectively which were closely comparable with the findings of (41). The number of alleles is positively correlated with PIC and gene diversity which indicates significant allelic variations in the SSR markers (42). Our present study also showed that SSR primers RM3694, RM336 and RM218 showed higher values for PIC and gene diversity emphasizing it as the highly informative markers. Hence the false smut linked marker RM336 and RM 218 can be used in further research for marker assisted screening. From our findings, RPHP125 and Norungan exhibited a high dissimilarity coefficient (0.842) and were placed in different clusters confirming their genetic distinctiveness. High jaccard distance of 0.82 between Kapikar selection and Mangam samba, 0.78 between Kallukar and Kapikar selection, 0.77 between IG49 and Purple puttu were noted among the high performing genotypes identified through PCA. In the findings, good performing genotypes based on PCA had high genetic

distance and clustered separately (30). The susceptible check CO43 exhibited high genetic distance with resistant lines viz., IG49 (0.84), Mangam samba (0.83) and Karungam (0.82). These genotypes can be hybridised to develop the mapping population for unravelling the genomic regions responsible for false smut resistance. The resistant lines like IG71, Thulasi vasanai samba, Arupatham vellai, Poongar selection, Kaltikar and Chinna aduku nel were grouped in different clusters in both the clustering method (D^2 and UPGMA). Hence, the genes governing the resistance may be different in these genotypes. However, the highly resistant Channagi and Kapikar selection were clustered in the same group based on both morphological and molecular methods, which may be due to the involvement of similar kind of genes governing resistance in these genotypes.

In the Bayesian clustering, the population were divided into two sub-populations. Our findings were similar to the findings who conducted an evaluation on the genetic diversity and population structure of 81 genotypes using 30 SSR markers (43). AMOVA confirmed the presence of considerable amount of variation in the population. The presence of admixtures in some germplasms may be due to allelic reshuffling (44). The variation among the individuals within the population is 84 %, within the individuals is 12 % and a less variation of 4 % among the populations has been observed. The significant diversity within the population indicates a high degree of genetic differentiation. Similar variation patterns in AMOVA were observed (45). Fixation index (Fst) is used to quantify the genetic variation between the populations. In the present study, low Fst of 0.03 indicates the less amount of diversity between the subpopulations. An Fst value close to 0 suggests that the subpopulations are genetically similar, implying a high degree of gene flow. Similar report was also observed while investigating the diversity of rice germplasms for early seedling vigour (46). Spontaneous mutation and outcrossing may contribute for the high gene flow rate (47). In addition, our study also showed a high Nm value of 7.263 suggested high gene flow rate in the subpopulation and the observed heterozygosity of the population is lower than the expected heterozygosity. Thus, the low heterozygosity observed may be due to the autogamous nature of rice. These results indicate that gene flow within populations is significant with various allele combinations. Moreover, the frequency of allelic exchange within populations is significantly higher than that among the populations.

The grouping pattern is different in both UPGMA and D^2 clustering methods. This difference may also be due the attributes associated with environmental influence and genotype-environment interactions or may be due to different geographical origin. This discrepancy between morphological and molecular clusters was also reported by (30, 48). This study reveals significant variation among the genotypes for reaction against false smut disease. Under the hotspot conditions, 12 genotypes were highly resistant which may be used in the future breeding programmes aimed to map the Quantitative Trait Loci (QTL) for genes linked to the false smut resistance in rice. Moreover, these genotypes can also be effectively utilized for the

development of high yielding false smut resistant rice varieties.

Conclusion

Both D^2 statistics and UPGMA clustering analysis revealed a significant genetic divergence in the genotypes. The highly resistant and high yielding genotypes viz., koolavalai, Periya chandikar, Earapalli and Kapikar selection could be used as potential donors in resistance breeding. The resistant lines such as IG71, Thulasi vasanai samba, Arupatham vellai, Poongar selection, Kaltikar and Chinna aduku nel, which are grouped into different clusters in both the grouping methods, could be utilized to develop a mapping population for the identification of resistance genes. These genotypes can be used for developing resistant variety through introgression using marker assisted backcross breeding as well.

Acknowledgements

Authors' contributions

MS Carried out the research, performed all the statistical analysis and wrote the manuscript, SM participated in conceptualization, investigation and supervision of the research work, NR contributed in design of the study, RS, RS, CG and MR reviewed and corrected the manuscript, PD contributed in data analysis. All authors read and approved the final manuscript.

Compliance with ethical standards

Conflict of interest: Authors do not have any conflict of interests to declare.

Ethical issues: None.

Did you use generative AI to write this manuscript?

No, the AI tools are not used for writing the manuscript

References

1. Savary S, Willocquet L, Elazegui FA, Castilla NP, Teng PS. Rice pest constraints in tropical Asia: quantification of yield losses due to rice pests in a range of production situations. Plant Disease. 2000;84(3):357-69. https://doi.org/10.1094/PDIS.2000.84.3.357

2. Ladhalakshmi D, Laha GS, Singh R, Karthikeyan A, Mangrauthia SK, Sundaram RM, et al. Isolation and characterization of *Ustilaginoidea virens* and survey of false smut disease of rice in India. Phytoparasitica. 2012;40:171-76.https://doi.org/10.1007/s12600-011-0214-0

3. Singh AK, Pophaly DJ. An unusual rice false smut epidemic reported in Raigarh district, Chhattisgarh. International Rice Research Notes. 2010;35:1-3.

4. Andargie M, Li L, Feng A, Zhu X, Li J. Mapping of the quantitative trait locus (QTL) conferring resistance to rice false smut disease. Current Plant Biology. 2018;15:38-43. https://doi.org/10.1016/j.cpb.2018.11.003

5. Hedge YR, Anahosur KH. Effect of false smut of rice on yield components and growth parameters. Indian Phytopathology. 2000;53(2):181-84.

6. Ahonsi MO, Adeoti AA, Erinle ID, Alegbejo MD, et al. Effect of variety and sowing date on false smut incidence in upland rice in Edo state, Nigeria. International Rice Research Notes. 25(1):1-14.

7. Zhou L, Mubeen M, Iftikhar Y, Zheng H, et al. Rice false smut pathogen: implications for mycotoxin contamination, current status and future perspectives. Frontiers in Microbiology. 2024;15. https://doi.org/10.3389/fmicb.2024.1344831

8. Baite MS, Raghu S, Prabhukarthikeyan SR, Mukherjee AK, et al. Yield loss assessment in rice (*Oryza sativa*) due to false smut infection. The Indian Journal of Agricultural Sciences. 2020;90(2):361-64. https://doi.org/10.56093/ijas.v90i2.99023

9. Zhou YL, Xie XW, Zhang F, Wang S, Liu XZ, et al. Detection of quantitative resistance loci associated with resistance to rice false smut (*Ustilaginoidea virens*) using introgression lines. Plant Pathology. 2014;63(2):365-72. https://doi.org/10.1111/ppa.12091

10. Doddakoppalu Govindaiah M, Selvaraj R, Kadirimangalam SR, Sundararajan A, Nachimuthu VV, et al. Genetic dissection of false smut resistance in rice through genome wide association mapping. Journal of Phytopathology. 2022;170(5):282-99. https://doi.org/10.1111/jph.13069

11. Mahalanobis PC. On the generalized distance in statistics. Sankhya: The Indian Journal of Statistics, Series A (2008). 2018;80:S1-7. https://www.jstor.org/stable/48723335

12. Talekar SC, Praveena MV, Satish RG. Genetic diversity using principal component analysis and hierarchical cluster analysis in rice. Int J Plant Sci. 2022;17(2):191-96. https://doi.org/10.15740/HAS/IJPS/17.2/191-196

13. Jasim Aljumaili S, Rafii MY, Latif MA, Sakimin SZ, et al. Genetic diversity of aromatic rice germplasm revealed by SSR markers. BioMed Research International. 2018;2018(1). https://doi.org/10.1155/2018/7658032

14. Chukwu SC, Rafii MY, Ramlee SI, Ismail SI, Oladosu Y, et al. Genetic analysis of microsatellites associated with resistance against bacterial leaf blight and blast diseases of rice (*Oryza sativa* L.). Biotechnology and Biotechnological Equipment. 2020;34(1):898-904. https://doi.org/10.1080/13102818.2020.1809520

15. Ashizawa T, Takahashi M, Moriwaki J. et al. A refined inoculation method to evaluate false smut resistance in rice. J Gen Plant Pathol. 2011;77:10-16. https://doi.org/10.1007/s10327-010-0279-5

16. Huang S, Liu L, Wang L, Hou Y. Research on advance of rice false smut *Ustilaginoidea virens* (Cooke) Takah worldwide: Part I. Research status of rice false smut. Journal of Agricultural Science. 2019;11(15):240. https://doi.org/10.5539/jas.v11n15p240

17. Doyle JJ, Doyle JL. A rapid DNA isolation procedure for small quantities of fresh leaf tissue. Phytochemical Bulletin. 1987;19:11-15.

18. Prevost A, Wilkinson MJ. A new system of comparing PCR primers applied to ISSR fingerprinting of potato cultivars. Theoretical and Applied Genetics. 1999;98:107-12. https://doi.org/10.1007/s001220051046

19. Manivannan N. TNAUSTAT-Statistical package. 2014. https://sites.google.com/site/tnaustat

20. Allan V. PB-Perfect: a comprehensive R-based tool for plant

breeding data analysis. PB-Perfect. 2023. https://allanbiotools.shinyapps.io/pbperfect/

21. Peakall ROD, 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

22. Christina GR, Thirumurugan T, Jeyaprakash P, Rajanbabu V. Principal component analysis of yield and yield related traits in rice (*Oryza sativa* L.) landraces. Electronic Journal of Plant Breeding. 2021;12(3):907-11. https://doi.org/10.37992/2021.1203.125

23. Saraswathi R. Elucidation of gene action and identification of QTLs or genes for false smut resistance in rice *Oryza sativa* L. India: TamilNadu Agricultural University;2024. Available from: http://hdl.handle.net/10603/513362

24. Evanno G, Regnaut S, Goudet J. Detecting the number of clusters of individuals using the software STRUCTURE: a simulation study. Mol Ecol. 2005;14:2611-20. https://doi.org/10.1111/j.1365-294X.2005.02553

25. Li R, Chen S, Matsumoto H, Gouda M, et al. Predicting rice diseases using advanced technologies at different scales: present status and future perspectives. Abiotech. 2023;4(4):359-71. https://doi.org/10.1007/s42994-023-00126-4

26. Chib HS, Tikoo ML, Kalha CS, Gupta BR, et al. Effect of false smut on yield of rice. Indian Journal of Mycology and Plant Pathology. 1992;22(3):278-80.

27. Rani R, Sharma VK, Lore JS, Pannu PP, Kaur Y. Screening of rice germplasm against *Ustilaginoidea virens*, the incitant of false smut. Agricultural Research Journal. 2016;53(1):134-37. https://doi.org/10.5958/2395-146X.2016.00026.0

28. Selvi NR, Saraswathi R, Geetha S, Gopalakrishnan C. Genetic studies on false smut disease resistance and yield components in rice (*Oryza sativa.* L). Electronic Journal of Plant Breeding. 2023;14(1):96-106. https://doi.org/10.37992/2023.1401.013

29. Momotaj begum MD, Mamunur rashid MD, Hasan MS, Hasan MZ. Genetic variability and crossing among blast resistant and susceptible fine aromatic rice (*Oryza sativa*) land races. Res on Crops. 2000;19(3):450-57.

30. Nivedha R, Manonmani S, Kalaimagal T, Raveendran M, Kavitha S. Assessing the genetic diversity of parents for developing hybrids through morphological and molecular markers in rice (*Oryza sativa* L.). Rice. 2024;17(1):17. https://doi.org/10.1186/s12284-024-00691-2

31. Hosagoudar GN, Babu U, Kovi BS. Genetic variability studies of short duration fine grain paddy varieties having resistance to blast diseases. J Pharmacogn and Phytochem. 2018;7(4):935-39.

32. Nadarajan N, Manivannan N, Gunasekaran M. Quantitative genetics and biometrical techniques in plant breeding. Kalyani, New Delhi, India. 2016;227-32.

33. Mohapatra C, Chand R, Singh AK, Dixit GP. Principal component analysis for quantitative traits and powdery mildew resistance in pea (*Pisum sativum* L.). Journal of Food Legumes. 2017;30 (2):43-47.

34. Dinesh K, Devi MS, Sreelakshmi C, Paramasiva I. Exploring the genetic diversity for yield and quality traits in indigenous landraces of rice (*Oryza sativa* L.). Electronic Journal of Plant Breeding. 2023;14(2):502-10. https://

doi.org/10.37992/2023.1402.055

35. Bhati PK, Singh SK, Sharma SD. Genetic divergence for quantitative traits in rice germplasm. Electronic Journal of Plant Breeding. 2015;6(2):528-34.

36. Singh A, Singh PK, Singh SP. Estimation of genetic divergence analysis for yield and bacterial leaf blight (BLB) disease resistance in rice (*Oryza sativa* L.). Journal of Pharmacognosy and Phytochemistry. 2019;8(2):1124-28.

37. Sai BH, Reddy BR, Priya MS, Vemireddy LR. Genetic diversity analysis for yield and yield attributing characters in rice (*Oryza sativa* L.). Electronic Journal of Plant Breeding. 2022;13(4):1282-87. https://doi.org/10.37992/2022.1304.161

38. Fu RT, Chen C, Wang J, Zhao LY, Chen XJ, Lu DH. Evaluation and screening of rice germplasm resources resistant to rice false smut. Journal of Southern Agriculture. 2022;53(1):78-87.

39. Vabna FA, Islam MZ, Prince MF, Hoque ME. Molecular diversity analysis in Boro rice (*Oryza sativa* L.) landraces using SSR markers. Asian Journal of Biology. 2021;12(1):36-48. https://doi.org/10.9734/ajob/2021/v12i130156

40. Sarif HM, Rafii MY, Ramli A, Oladosu Y, Musa HM, et al. Genetic diversity and variability among pigmented rice germplasm using molecular marker and morphological traits. Biotechnology and Biotechnological Equipment. 2020;34(1):747-62. https://doi.org/10.1080/13102818.2020.1804451

41. Pradhan AK, LNR Vemireddy, B Tanti. Assessment of the genetic variability and population structure in boro rice cultivars of Assam, India using candidate gene based SSR markers. Genet Resour Crop Evol. 2023;70(6):1747-65. https://doi.org/10.1007/s10722-022-01533-0

42. Subramani P, Kalipatty Nalliappan G, Narayana M, Veerasamy R, Natesan S. Molecular diversity, population structure analysis and assessment of parent hybrid relationships in fodder maize. Crop Breed Appl Biotechnol. 2024;24(3). https://doi.org/10.1590/1984-70332024v24n3a28

43. Nath A, D Nath, T Dasgupta. Population structure and diversity analysis of Indica and Japonica rice genotypes. Bulgarian Journal of Agricultural Science. 2021;27(1):120-27.

44. Nithya N, Beena R, Abida PS, Sreekumar J, Stephen R, et al. Genetic diversity and population structure analysis of bold type rice collection from Southern India. Cereal Res Commun. 2021;49:311-28. https://doi.org/10.1007/s42976-020-00099-w

45. Pandita D, Mahajan R, Zargar SM, Nehvi FA, et al. Trait specific marker-based characterization and population structure analysis in rice (*Oryza sativa* L.) germplasm of Kashmir Himalayas. Mol Biol Rep. 2023;50(5):4155-63. https://doi.org/10.1007/s11033-023-08324-5

46. Anandan A, Anumalla M, Pradhan SK, Ali J. Population structure, diversity and trait association analysis in rice (*Oryza sativa* L.) germplasm for early seedling vigor (ESV) using trait linked SSR markers. PLoS One. 2016;11(3). https://doi.org/10.1371/journal.pone.0152406

47. Nuijten E, van Treuren R, Struik PC, Mokuwa A, Okry F, et al. Evidence for the emergence of new rice types of interspecific hybrid origin in West African farmers' fields. PLoS One. 2009;4 (10). https://doi.org/10.1371/journal.pone.0007335

48. Pathak P, Singh SK, Korada M, Habde S, Singh DK, et al. Genetic characterization of local rice (*Oryza sativa* L.) genotypes at