Assessment of diversity of Indian aromatic rice germplasm collections for morphological, agronomical, quality traits and molecular characters to identify a core set for crop improvement

G. S. V. Prasad, G. Padmavathi, K. Suneetha, M. S. Madhav and K. Muralidharan*

Abstract

Background: Besides the Basmati, the aromatic rice germplasm (ARG) accessions are treasured for quality, medicinal value and aroma. The demand for aromatic rice is ever increasing. Genetic diversity is the source of variability to identify superior alleles controlling morphological, agronomic and quality traits, and molecular attributes. This study reports on the characterization of traits in ARG to identify a core set for breeding high-yielding varieties.

Methods: The genetic diversity was measured on the distinctness, uniformity and stability (DUS) of 46 traits in 208 Indian ARG in field, greenhouse and laboratory tests. We performed individual and combined analysis of DUS traits and molecular data generated using 55 SSR markers. The genetic distances between genotypes were estimated using Mahalanobis D^2 analysis and clustering by standardized Euclidean^2 distances, Ward Minimum variance, Gowers' similarity index and PowerMarker. The aim was to derive a core set of non-Basmati ARG using PowerCore to deploy in crop improvement.

Results: Eighty-two alleles were detected. Alleles per marker ranged from 2 (RM505) to 5 (RM276) with an average of 3.04 alleles. The markers are informative in analyzing the diversity as the PIC values estimated varied from 0.17 (RM577 on chromosome 1) to 0.72 (RM276 on chromosome 6) with an average of 0.54 per locus. RM276 with repeat motif of (AG)8A3(GA) 33 on chromosome 6 was the most informative (amplified 5 alleles). The combined analysis had shown genotypes in a few clusters to be more diverse than others. SSR markers RM289, RM505, RM577 and RM22866 were identified as genotype specific markers. With PowerCore, 46 genotypes (22%) were identified as a core set of ARG that represent all the alleles detected in the entire set investigated. 2-Acetyl-1-pyrroline is considered to impart aroma; it was not detected by GC–MS tests in many ARG.

Conclusions: Forty-six genotypes in the core set have different maturity periods, plant statures, grain types and grain quality traits. A parent can be selected from the core set to improve aromatic rice depending on the breeding objective. The olfactory sensing of strong aroma emitted by cooked kernels of all ARG was found more decisive than the costly GC–MS tests.

*Correspondence: muralidharan_km@yahoo.com
ICAR-Indian Institute of Rice Research (ICAR-IIRR), Hyderabad 500030, India

Prasad *et al. CABI Agric Biosci*   (2020) 1:13

Page 2 of 24

Background

Rice (*Oryza sativa* L. 2n = 24) belongs to the family *Graminae* and sub family *Oryzoidea.* Rice is the food for more than half the world's population of 7.8 billion (Worldmeters 2020; UN 2020). Besides dominating as an indispensable food component in Asia, rice is rapidly emerging as the chief food in Latin America and Africa. Arunachal Pradesh, an Indian state along the Indo-China border area is considered as the centre of origin and domestication of the cultivated rice species *Oryza sativa (*Muralidharan and Siddiq 2000). Distinct locally adapted aromatic germplasm are found in the northeast Indian states of Arunachal Pradesh, Assam, Manipur, Meghalaya, Mizoram, Nagaland, Sikkim and Tripura. Landraces of rice with aroma fetch higher price in the global markets including India, countries in the middle-east, Nepal, Bangladesh, Benin, Senegal and Guinea due to globalization, health consciousness, and culinary changes. Knowledge of the non-Basmati aromatic rice germplasm (ARG) gene pool of India will facilitate proper maintenance, conservation and utilization of this valuable resource for crop improvement. Rice has over 100,000 landraces and improved cultivars. This rich genetic diversity is crucial to improve rice crop performance and farmers' economy and to provide adequate food to a large population, especially in Asia. Characterization and quantification of genetic diversity has long been a major goal in evolutionary biology (Losos et al. 2013). Thousands of valuable allelic variations for traits of economic significance remain unutilized in nearly all crop plants. Less than 15% of the potential diversity in rice has been utilized (FAO 2008). Rice collections exhibit varied maturity periods, grain types, sizes and colours, and resistance to biotic and abiotic stresses (Muralidharan et al. 1996; Prasad et al. 2001; Rani et al. 2008). Landraces are Basmati and non-Basmati rice genotypes with many discrete traits in each one of them. Basmati and its derivatives are distinct from other rice germplasm accessions and varieties (Glaszmann 1987; Nagaraju et al. 2002; Narshimulu et al. 2011; Choudhary et al. 2013). There is a significant sub-group in non-Basmati rice possessing exquisite aroma, superior quality grains and medicinal properties. India had exported 4.415 million tonnes (mt) of Basmati to the global market valued at US$ 4722 million, and 7.6 mt of non-Basmati rice valued at US$ 3,048 million (APEDA 2020) between April 2018 and March 2019. The treatises or compendia on ancient Ayurveda, *Susruta* (~ 400 BC), *Charaka* (~ 700 BC), and *Vagbhata* (~ 700 AD) in India, refer to the medicinal value of many landraces both aromatic and non-aromatic rice (Bhishagratna 1907; Valiathan 2003; Wujastyk 2003). Attempts have been made in India to systematically study, elucidate properties and document some of these landraces (Leenakumary 2004).

Aromatic rice germplasm (ARG) constitute a small but an important sub-group of landraces. These are rated as the best in quality and fetch much higher price than high quality non-aromatic rice in the international market. Thus aromatic rice has become an important commercial commodity. Almost every state in India has its own collection of non-Basmati aromatic rice genotypes with mostly short and medium sized grains. Aromatic genotypes predominantly have long growth duration, photoperiod sensitivity, low yields and susceptibility to lodging. They are cultivated annually in ~ 600,000 hectares for domestic consumption in the states of Assam, Bihar, Chhattisgarh, Madhya Pradesh, Maharashtra, Odisha, Uttar Pradesh and West Bengal (Singh et al. 2003). A few studies have been made to analyse the genetic structure of Indian aromatic and quality rice (Nagaraju et al. 2002; Jain et al. 2004; Roy et al. 2015). It is necessary to conserve these landrace genotypes and understand available gene-pools to breed high-yielding aromatic rice varieties in the country.

Germplasm accessions are generally cataloged on the descriptors of DUS traits (Rani et al. 2006; IRRI 2002) and characters revealed by molecular markers. Among all different classes of molecular markers available for evaluating genetic diversity, microsatellites or simple sequence repeats (SSRs) are well known for their potentially high information content and versatility (Tautz 1989; Ishii et al. 2001). They help to establish the relationship among the individuals even with a lesser number of markers (McCouch et al. 1997). These molecular tools are used in genetic diversity studies due to their multi-allelic nature, high reproducibility, co-dominant inheritance, abundance and extensive genome coverage (McCouch et al. 2002; Sow et al. 2014). Our objective was to evaluate a collection of 208 ARG and document the genetic variations in 46 morphological, agronomical and grain quality traits, reaction to pests, and molecular features as revealed by SSR markers. We further examined by a combined analysis of all data matrices generated to select a representative core subset of ARG genotypes for use in breeding programs.

Prasad *et al. CABI Agric Biosci*   (2020) 1:13

Page 3 of 24

Methods

A total of 208 indigenous ARG were collected from four different rice germplasm conservation centres in India: they are from Indira Gandhi Agricultural University, Raipur, Chhattisgarh 492,012; Dr. Rajendra Prasad Central Agricultural University, Pusa, Samastipur 848,125, Bihar; Narendra Deva University of Agriculture and Technology, Faizabad 224,229 Uttar Pradesh; and Odisha University of Agriculture and Technology, Bhubaneswar 751,003 Odisha. Details on the list of ARG accessions used in this investigation are given in Additional file 1: Table S1.

Plant husbandry

All 208 ARG accessions were raised during the kharif or wet season in greenhouse and in irrigated fields to record morphological and agronomic characters and reaction to pests. Soils were primarily black vertisols consisting of montmorillonite clay (40%) with a pH of 8.2 to 8.5. Seeds of each of the germplasm accessions were sown in 1 m² plots in wet nursery beds fertilized with 10 t/ha of farmyard manure. Seedlings (21-days old) of each genotype were transplanted in four rows, each of 5 m in length and 25 cm apart in two replications in ICAR-IIRR field fertilized with 60 kg N/ha nitrogen, 60 kg P_2O_5/ha phosphorus and 60 kg K_2O/ha potash. P and K were applied as basal dressing and N was applied in three equal splits as basal and top dressings in 25 and 45 days after planting. Traits considered important in DUS tests were included for evaluation in this study and details are indicated in Additional file 1: Table S1. The guidelines for DUS test of rice (Rani et al. 2006) and the standard evaluation system (SES) (IRRI 2002) were adopted in evaluating traits. The grain classification of Ramaiah (Ramaiah 1985) was used. The standard protocol was used to analyse grain quality as defined by the ICAR-IIRR (ICAR-IIRR 2009). In preliminary examinations, to avoid confusion and errors in recording morphological, agronomic and quality descriptors, a standard methodology was established. The details of various traits measured and scores used to categorize traits in the present study are presented in Additional file 1: Table S1-Sheet 2. Reaction to certain devastating pests of rice believed as important traits are included. The SES (IRRI 2002) was used to estimate the reaction to diseases and insect pests. Data on 46 traits (15 morphological, 9 agronomic, 16 grain quality parameters and reactions to important 2 diseases and 4 insect pests) were carefully assessed and data saved for analysis. All the observations in field were recorded on representative five random plants and mean value in each replication was computed to record data on quantitative parameters.

Evaluation of reaction to pests

Seedlings were raised in a greenhouse. ARG genotypes were line-sown with a spacing of 5 cm between individual plants and 20 cm between lines in several sets of trays (60 × 40 cm) that were fertilized with farmyard manure. Twenty-one-day-old seedlings from a few sets of trays were transplanted in new sets of trays (60 × 40 cm) at a spacing of 5 cm between individual plants and 20 cm between lines. Trays were fertilized at the rate of 100 kg P_2O_5/ha applied as basal, and 50 kg N/ha applied in three equal splits as basal and topdressings in 25 and 45 days after planting. Sets of trays with 208 ARG were used for evaluation of reactions either on seedlings or on transplanted crop depending on the pest evaluated. In each of these tests two replications were maintained.

For identifying the reaction to blast (Bl), 15-day-old seedlings of ARG grown in trays were exposed to an inoculum of *Magnaporthe grisea* from heavily sporulating lesions in leaves of a highly susceptible local rice variety, HR12. During the infection period, nursery seedlings were subjected in the day to 28 ± 2 °C and 85% RH, and in the night to 25 ± 2 °C and ∼90% RH. Bl was recorded 15 days after inoculation using the SES (IRRI 2002). For reaction to bacterial leaf blight (BLB), transplanted plants grown in trays along with susceptible check variety TN1 were clip inoculated at the maximum tillering stage; in each ARG genotype, 2–5 leaves were clipped using inoculum of highly virulent IIRR-DX020 strain (10^9 cfu/ml). The lesion lengths were measured 15 days after inoculation.

Brown planhopper (BPH) or white-backed planthopper (WBPH) insects were maintained in isolation on the susceptible plants of the rice variety TN1 in Mylar cages. Twelve-day-old seedlings of ARG in trays were infested by releasing 3–5 first and second instar nymphs per seedling for 24 h. The trays with seedlings were covered with wire-mesh-lined cages with a wooden frame. Separate sets of trays in different greenhouse chambers were used for the two insect pests. The reaction of seedlings was scored using SES, 15 days after releasing the insects (IRRI 2002). For reaction to the virulent biotype 4 of gall midge (GM), 25 female and 10 male adult gall flies were released on to 15-day-old seedlings in trays covered with a nylon-mesh-lined wooden cage on two consecutive days. Trays were transferred to a high humidity chamber (> 90%RH) for the next two days to enable eggs to hatch and maggots to grow. Later, the trays were moved to ambient conditions and 20 days after release of nymphs, the genotypes were scored for reaction as percent plant damage (IRRI 2002). The natural incidence of yellow stem borer (YSB)

Prasad *et al. CABI Agric Biosci*   (2020) 1:13

Page 4 of 24

was estimated in field-grown ARG plants at harvest as per cent white earheads (IRRI 2002).

Analysis of quality parameters

Standard protocol as defined by the ICAR-IIRR (ICAR-IIRR 2009) was used to analyse grain quality. From each of the ARG plants, paddy grains were harvested individually and dried to 12–14% moisture content. The samples were stored after cleaning at room temperature. To determine all the properties other than grain dimensions, paddy was allowed to age for at least three months at an ambient temperature. Before analysis, sample material was exposed to the normal laboratory conditions for 1–2 days. A portion of the paddy (with husk) from each ARG was weighed and then decorticated (SATAKE Model THU 35A). After cleaning, the decorticated kernels were weighed again. Hulling percentage was derived as the ratio of the weight of the decorticated kernel to weight of paddy. Decorticated kernels were then polished by milling (SATAKE Model TM 05). The timing of the polisher was adjusted to obtain a 5% polishing of kernels. Milling percentage was derived as the ratio of the weight of the polished kernel to weight of unpolished kernels. Depending on the size of the grains, the polished kernels were passed through an appropriate rice grader containing grooves. The whole kernels were separated from the broken to quantify the head rice recovery. Kernels that were whole and three-quarters in size were weighed. Head rice recovery percentage was derived as the ratio of the weight of whole polished kernels to the weight of paddy.

After milling the polished rice, the length and breadth of more than 10 full grains possessing intact tips were measured. Based on the average length, breadth and ratio of length to breadth (L/B) of grains, the individual ARG genotype was classified (Ramaiah 1985) into five major grain types: long slender (LS, kernel length > 6 mm, and L/B of > 3.0); short slender (SS, kernel length < 6 mm, and L/B of > 3.0); medium slender (MS, kernel length < 6 mm, and L/B of 2.5–3.0); 4) long bold (LB, kernel length > 6 mm, and L/B of < 3.0); and short bold (SB, kernel length < 6 mm, and L/B of < 2.5). In each genotype, 1000 fully formed grains were weighed to record the grain weight (g). The degree of chalkiness (appearance of white belly or spots) on the endosperm of grains was graded as absent (A, none), very occasionally present (VOC, < 10%), occasionally present (OC, 11–20%) or present (> 20%).

Duplicate sets of six whole milled kernels of each ARG without cracks were placed in individual plastic boxes to find their alkali spreading value (ASV). Potassium hydroxide (10 ml of 1.7% KOH) was added to each sample, which was then left undisturbed in an incubator for 23 h. ASV was estimated in each of the ARG based on the level of alkali digestion and the gelatinization temperature. Amylose content was estimated per Juliano et al (1981) assay with the preparation of a standard curve. In a long test tube, rice flour (100 mg) of one genotype was added with rectified spirit (1 ml) and sodium hydroxide (NaOH, 9 ml). This solution was shaken in a hot water bath at 60 °C for 15 min. The sample after digestion was transferred to a volumetric flask, rinsed twice with hot distilled water and then made up to 100 ml. To a portion of this solution (5 ml), acetic acid (1 ml) and I_2-KI reagent were added, made up to 100 ml in the flasks, and the flasks were covered with a cloth for 20 min. Optical density of this solution was recorded at 620 nm using a spectrophotometer and amylose content was calculated on standard curve as waxy (1–2%), very low (2–9%), low (9–20%), intermediate (20–25%) and high (26–33%). To estimate gel consistency, rice flour (100 mg) of each genotype was added with ethanol (0.2 ml) containing thymol blue (0.25%) and potassium hydroxide (KOH, 2 ml) (Cagampang et al. 1973). Four replications were maintained for each accession. Each of the test tubes was covered with a glass marble to prevent heat loss and was kept in a water bath at 100 °C for 8 min. Then the samples were cooled for 5 min and placed in a low temperature water bath with ice. The tubes were laid horizontally on the table over a millimeter graph paper for one hour. Gel consistency was assessed as very hard (< 25 mm), hard (26–40 mm), medium (41–60 mm) or soft (61–100 mm) based on the length of the gel. To assess the aroma in sensory odor test, water (15 ml) was added to each rice sample (5 g) in a test tube and allowed to soak for 10 min after closing the tube with a marble. The samples were cooked in a hot water bath (90 °C) for 15 min, transferred to individual Petri dishes, covered with lids and kept in a refrigerator to cool. The fragrance that emanated when the Petri dishes were opened was scored as aroma.

Volume expansion ratio (VER) in raw milled to cooked kernels was determined by measuring the water displaced using a graduated cylinder. After measuring the length of kernels, a sample of kernels (5 g) of each genotype was dropped into a beaker containing water (15 ml) and the total volume was measured. A similar sample was cooked in a water bath at 90 °C for 20 min. The lengths were measured in more than 10 cooked kernels with intact tips at both ends. They were then dropped into water (50 ml) and expanded volume was measured. VER was derived by dividing the volume of cooked rice (ml) by the volume of raw rice (ml). Similarly the elongation ratio was derived by dividing the kernel length after cooking by the kernel length before cooking. Water uptake (WUP) was measured in a sample (2 g) of each genotype soaked in water (10 ml) for 30 min, and then boiled for 45 min at

Prasad *et al. CABI Agric Biosci*   (2020) 1:13

Page 5 of 24

80 °C in a water bath. After quickly cooling it using an ice cold water bath, the supernatant water was measured by decanting into a graduated cylinder. WUP was calculated as millilitre per gram of sample.

The genetic divergence was estimated for all the 46 traits following the Mahalanobis D^2 analysis and clustering was prepared using standardized Euclidean^2 distances and the Ward Minimum variance method. All other statistical analyses were done using Windostat software (https://www.directoryofscience.com/view/windostat_free_version).

Molecular analysis with SSR markers

The total genomic DNA was extracted per the protocol of Zheng et al (Zheng et al. 1995) from leaves collected from 21-day-old plants of each ARG grown in trays. In each genotype, leaves from 25 plants were bulked for DNA isolation. Limited sets of SSR markers have been demonstrated as adequate to discriminate even the most closely related genotypes in cereals; they varied from 11 (Plaschke et al. 1995) to 13 (Pathaichindachote et al. 2019), 15 (Struss and Plieske 1998), 23 (Russell et al. 1997), 32 (Aljumaili et al. 2018) or 63 (Nachimuthu et al. 2015). We used 55 SSR markers that were dispersed across the 12 rice chromosomes and were hyper polymorphic to estimate the diversity in ARG (Table 1). The PCR mixture contained 50 ng template DNA, 5 pmol of each primer (Integrated DNA Technologies, USA), 0.05 mM dNTPs (MBI Fermentas, Lithuania), $1 \times$ PCR buffer (10 mM Tris, pH 8.4, 50 mM KCl, 1.8 mM MgCl$_2$; Sanmar Fine Chemicals, India) and 1 U of *Taq* DNA polymerase (Sanmar Fine Chemicals, India) in a total reaction volume of 10 μl. The concentration of the DNA template, primers, dNTPs etc. mentioned will remain as a constant even if volumes vary. Template DNA was initially denatured at 94 °C for 5 min followed by 35 cycles of PCR amplification with the following parameters: a 30 s denaturation at 94 °C, a 30 s annealing at 55 °C and 1–2 min of primer extension at 72 °C. A final extension was done at 72 °C for 7 min. Genotyping by use of agarose gel electrophoresis is easily accepted by breeders due to its simple requirements and easy operation in the lab. For most breeders, agarose gel electrophoresis equipment is easy to obtain (Wei et al. 2020). Therefore, PCR amplified products were electrophoretically resolved in 3% agarose gels (Lonza, USA) in $0.5 \times$ TBE buffer at 100 V for 3.5 h in Hoeffer Super Submarine Electrophoresis unit (GE Biosciences, USA). SSRs are co-dominant and each amplicon is considered as separate. Amplicons were scored for their product size for each primer genotype combination. Even in the case of double bands, the presence or absence of amplicons can be clearly distinguished (Yadav et al. 2013). The gels were stained in ethidium bromide

(10 mg/ml) and the banding patterns were documented in an Alpha Imager gel documentation system (Alpha Innotech, USA). The sizes of the STMS PCR amplicons in all the genotypes were determined by comparison with 50 bp DNA ladders used as size standards. Markers were scored for the presence and absence of the corresponding band among the genotypes as 1 and 0, respectively. A data matrix comprising 1 and 0 was formed depending upon the visualized bands and analysed further.

Combined diversity analysis of data matrices on traits and SSR characterization

Polymorphic information content (PIC) represents the amount of polymorphism within a population. For a set of ARG accessions, genetic diversity parameters such as number of alleles per locus, allele frequency, heterozygosity and PIC values were estimated (https://www.agri.huji.ac) using the program PowerMarker Ver 3.25 (Liu and Muse 2005). Genetic similarity was also estimated between all genotype pairs using the similarity index proposed by Gower (Gower 1971). It uses both binary and quantitative morphological data to estimate a unique similarity index ranging from 0 to 1. The genetic similarity was estimated using the PowerMarker. It was converted into genetic dissimilarity according to the equation $D_{ab} = 1 - S_{ab}$. D_{ab} is the genetic dissimilarity between each pair of a and b genotypes and S_{ab} is the genetic similarity between each pair of a and b genotypes. The cluster analysis among the genotypes was performed based on Jaccard's dissimilarity matrix using DARwin-5 software. The dissimilarity matrix generated was used to construct an UPGMA dendrogram with the adjustment between the dissimilarity matrix and the dendrogram derived from the co-phenetic correlation coefficient (r) using the NTSYS pc 2.1 software.

Development of a core set of ARG

PowerCore is a program that applies the advanced M-strategy using a heuristic search (Kim et al. 2007). It was used to constitute a core set from the entire collection of ARG investigated (https://genebank.rda.go.kr/powercore/). Using PowerCore, four statistical parameters were analysed to compare the mean and variance ratio between core and entire collections of ARG. The significant differences were calculated between the core set and the entire set of 208 ARG as the mean (MD, %) and variance differences (VD, %) of traits. Coincidence rate (CR, %) and variable range (VR, %) were estimated (Hu et al. 2000) to evaluate the properties of the core set in comparison with the entire collection. The Shannon-Weiner (Shannon et al. 1949) diversity index (H^i) for allelic richness and evenness in the core set and in

Prasad *et al. CABI Agric Biosci*   (2020) 1:13

Page 6 of 24

Table 1 List of markers used to study molecular attributes of ARG genotypes

S. No.	Chr No.	Primer	Forward sequence	Reverse sequence
1	1	RM577	GCTTTCCCTCTAACCCCTCT	GGATGTACCGCTGACATGAA
2	1	RM8068	GGATGTACCGCTGACATGAA	TGAACATTTATTGATATGGTAAA
3	1	RM10843	CACCTCTTCTGCCTCCTATCATGC	GTTTCTTCGCGAAATCGTGTGG
4	2	RM485	CACACTTTCCAGTCCTCTCC	CATCTTCCTCTCTTCGGCAC
5	2	RM3688	GTTGAATCAAGCTGTGCAGC	AGCTAGGCAAAGCATGCATG
6	2	RM6318	TGCTGCTTCTGTCCAGTGAG	GGATCATAACAAGTGCCTCG
7	2	RM14140	CCTCCCTCTCCAAACACATTGC	TCATCAGCAGACAGAATGTTGACC
8	3	RM85	CCAAAGATGAAACCTGGATTG	GCACAAGGTGAGCAGTCC
9	3	RM251	GAATGGCAATGGCGCTAG	ATGCGGTTCAAGATTCGATC
10	3	RM6759	AAACGGAGTGGTCCATTAGC	GAAAGGTAGATGGGTTTGTGG
11	4	RM252	TTCGCTGACGTGATAGGTTG	ATGACTTGATCCCGAGAACG
12	4	RM273	GAAGCCGTCGTGAAGTTACC	GTTTCCTACCTGATCGCGAC
13	4	RM16701	TTCGCTGACGTGATAGGTTG	ATGACTTGATCCCGAGAACG
14	4	RM16738	AAACGGAGTGGTCCATTAGC	GAAAGGTAGATGGGTTTGTGG
15	4	RM17263	GAGTCGATGTAATGTCATCAGTGC	AGCAACAGCACAACTTGATG
16	4	RM3042	TATCGACGACGGCTGTTGAGAGG	GTTGGTTGGACCTTGGGTCTGG
17	4	RM17115	ATCGGCTTCCATCTGTATTCTCC	CACAGACATCTCGCTACTGACAGC
18	4	HRM17127	CTCAATGTTTCCCACAGTTACCG	TGTGTTATGTGTGCGTGATGAGC
19	4	HRM16592	CTTAGCACGGACACTCATATTTGG	CACAATACGTTTGATGGCTTGC
20	4	HRM16913	GTGTACGTGTTGGCTCTCTGTACG	GATGTTGCTTGTGCTGCAACC
21	5	RM122	GAGTCGATGTAATGTCATCAGTGC	GAAGGAGGTATCGCTTTGTTGGAC
22	5	RM289	TTCCATGGCACACAAGCC	CTGTGCACGAACTTCCAAAG
23	5	RM267	TGCAGACATAGAGAAGGAAGTG	AGCAACAGCACAACTTGATG
24	5	RM459	CTGCAATGCTGCATGACC	CACTTTCTCTGCAGCACCAG
25	6	RM276	CTCAACGTTGACACCTCGTG	TCCTCCATCGAGCAGTATCA
26	6	RM3463	TGCCTCTCTACCGCAACACAGC	CCCTTCCCTTACCTCTCCATTCC
27	6	RM19670	CAAGCAGAAGTGGAGGCAGTCG	GAGAAGGAGGTGGGAGGCATCG
28	6	RM19649	GATTAATGTGTGGGAGCTCATGG	GGATTTCACCCTTCTTGTTGTCG
29	6	RM19642	TCTCCGAGCACTTAAGAAAGTTGG	GAAGCGGTGATAGTGACACAAGG
30	6	RM2523	TTTGTGGGCTAGGATTACCTCACG	CACGAATCCTATGACTGATCCAACG
31	6	HRM20060	CACACATGAGTGGTTAGGTAAGATGC	CAGTGACAAGAGCGAAATGATCC
32	6	HRM20583	GAAGCTACTCCAAGTTCAGTCATTGC	TTAGTACGTATGCACATCCCATCC
33	7	RM11	TCTCCTCTTCCCCCGATC	ATAGCGGGCGAGGCTTAG
34	7	RM229	CACTCACACGAACGACTGAC	CGCAGGTTCTTGTGAAATGT
35	7	RM505	AGAGTTATGAGCCGGGTGTG	GATTTGGCGATCTTAGCAGC
36	7	RM21260	CTGCACAACCAGGAGAAATTAAGC	CTGACCACTCTAGCTTGCCTACC
37	8	RM23595	GCTCTGAGTTGGGATCATCACG	TCTAGGTGACGCTTGTAGGATGG
38	8	RM22837	ACCTGGGTCAGATGTCTGTTTGG	GGTAGAGCTCCATCCATCTTAGTGC
39	8	RM22866	TGCTTGTTCTTCCTCCCAAGC	CTTGCCCGTATTGTGGTCTCG
40	9	RM7424	AGAAGCCCATCTAGCAGCAG	TCAAGCTAGCCACACAGCTG
41	9	RM23899	GGAGGATCGAGGCGAAACTACC	AGCTGTTCGTGACCAGAATCTCC
42	9	RM24383	CACTCACTCACTCAGCTGTTTCTTGG	CTGATCAAGAACCGCAGGAACG
43	10	HRM25262	CAATGCAAAGTCTTGTACGG	GCTACATTGCATAGATCACTCG
44	10	HRM25754	TTTCATTCTACCACCCAGGTTGC	GGACCTCTTCTTTCTCATGTTCTGC
45	10	HRM25796	GAAGCTTCCTCCTACGCTTTCC	TCAAGACTCAAGAGCCACAGTGC
46	10	HRM24954	TGGACAGGTAATCATGTGTCTGC	TTTAGTTGGTGTGCACGTGAGG
47	10	HRM25310	CACGTAATTGCCCGCGTTGC	GATACGGTGATCTGGGAGGAATCG
48	11	RM26329	TAACCGGGACTAAAGATAGAGC	CTACGTCGAAATCGTAACTAGC
49	11	RM26829	GAGAAGGCCTGATGAGTACAAGG	GATTATTGTGCAGGTGAGAAGTGG

Prasad *et al. CABI Agric Biosci*   (2020) 1:13

Page 7 of 24

Table 1 (continued)

S. No.	Chr No.	Primer	Forward sequence	Reverse sequence
50	11	RM27311	TTACCAACCGGGACTAAAGATCG	CAATTCATAACGTCGGTCCTTCC
51	12	RM2851	CTAATATTAGCCATGAAACA	CTTATCAGTAGAACTGCAGA
52	12	RM27406	TGGTAGGTGTGCAATAGAAGTAGG	AATGCATGCAAACACAGTGG
53	12	HRM28616	CACCGGAGTTCCCTCAACTTACC	TACGTATGGCCAATTCAGACTGG
54	12	HRM28110	GATCGATCTTTAGTCCCGGTTGG	AACACAATTCACCGGTCCTTAGC
55	12	HRM27840	TTTGCGTGCTAGGGAGATTAGC	CATTATGTACTTACTCCCTCCCTTCC

the entire 208 ARG set was also computed and a core set of genotypes was identified.

Estimation of 2-acetyl-1-pyrroline (2AP) in some ARG genotypes

Quantitative assessment of 2AP was made on payment for GC–MS use at the CSIR-Indian Institute of Chemical Technology, Hyderabad, India. It was limited to only 90 ARG samples due to high cost. Therefore, this 2AP data was analysed separately. Unpolished brown kernels (300 mg) were powdered and loaded in an auto-sampler vial and 500 μl of acetonitrile containing 250 ppb of TMP (tri-methyl pyridine) was added. The vial was sealed, kept in an ultrasonic bath at 60 °C for 30 min and transferred to a refrigerator for 15 min. After centrifuging it at 6000 rpm for 10 min, 1 μl of supernatant solution was injected into the GC–MS system Agilent 6890 N GC equipped with a 5973 N inert mass selective detector (MSD) and 7683 series auto-sampler (Agilent Technologies, USA). A SPB-624 (Supelco, USA) fused silica capillary column of 30 m length, 250 μm internal diameter and 1.4 μm film thickness was used. The column oven was programmed from the initial 40 °C with 2 min hold-up time to the final temperature of 240 °C at a heating rate of 10 °C/min ramp. Helium was used as a carrier gas in constant flow mode at a flow rate of 1 ml/min. The inlet and GC–MS interface temperatures were maintained at 200 °C and 240 °C, respectively. The EI source and quadrupole were maintained at 230 and 150 °C, respectively. Using an auto-sampler, 1 μl of sample was injected into the GC–MS in a split less mode with 0.2 min split less time. MSD was operated in the SIM (selected ion monitoring) mode to scan the ion from m/z 30 to 750 in the EI mode with electron energy of 70ev. The retention times for 2AP and TMP were 12.13 ± 0.10 min and 13.21 ± 0.10 min, respectively. The peak areas obtained under these ions were considered for the calculation of 2AP content (ppm) with respect to TMP.

Results

Diversity in traits

ARG genotypes showed a remarkable diversity and a wide range of variability in all the studied traits (Table 2). A few characteristic variations are given for illustration. All the genotypes had thin stem (up to 0.40 cm) except for the Chini Kapoor genotype, which had a thick stem (> 0.55 cm). Specific leaf weight varied from 12.8 g in Lokti muchhi to 36.9 g in Lokti Musi. The plant habit was tall in 162, intermediate in 30 and semidwarf in 13 ARG genotypes. Panicles per plant varied from 4.9 in Ganjeikalli to 16.3 in RAU 3041; panicle length varied from 20 cm in Chhabiswa to 37 cm in Chini Kapoor; single plant yield varied from 5.1 g in Shyam Jira to 23.5 g in Heera kani; and 1000-grain weight varied from 8 g in Kon Joha 2 to 28 g in IGSR-3-1-40. The ratooning ability of ARG was high (> 75%) in 38 genotypes. The reaction of ARG genotypes to pests also varied: 96 were resistant (R) to blast, 8 were resistant to bacterial leaf blight; 1 was resistant (IGSR-3-1-40) to the brown planthopper; 3 were resistant to white backed planthopper; and 2 were moderately resistant (MR) to yellow stem borer. Head rice recovery was > 70% in 18 genotypes, 66–70% in 33 genotypes and 58–66% in 82 others. The highest head rice recovery of 73% was in Lajkuli badan and the lowest of 12.8% was in Barang. There was a considerable variation in 208 ARG for water uptake from 50 ml/g in ASGPC-39 to 285 ml/g in IGSR-3-1-5. Kernel length on cooking varied from 5.4 mm in Ambemohar to 13.2 mm in Kesr; it was 7–10 mm in 111 genotypes and ≥ 10 mm in 91 genotypes. Only in 16 genotypes did the alkali spread value indicate high alkali digestion and low gelatinization temperature. Gel consistency of cooked kernels was mostly intermediate or soft, and was hard only in 75 ARG genotypes. Aroma was detected in the cooked kernels of all 208 ARG genotypes in the sensory odor test.

ARG genotypes were grouped into 20 clusters in Ward's minimum variance dendrogram and Euclidean² distance (see Additional file 2: Figure S1) generated using the data on 46 traits examined. Among them, cluster 4 was the largest with 42 genotypes while cluster 10 comprised

Table 2 Variability in 208 ARG genotypes in morphological, agronomical and quality traits

No	Trait	Characteristics	Genotypes (No.)
	Morphological traits		
1	Leaf shape of ligule	Truncate	15
		Acute	53
		Split	140
2	Culm attitude	Erect	198
		Semi-erect	10
3	Stem thickness	Medium (0.40–0.55 cm)	207
		Thick (> 0.55 cm)	1
4	Leaf intensity of green colour	Light	117
		Medium	83
		Dark	8
5	Leaf pubescence of blade surface	Absent	77
		Weak	11
		Medium	101
		Strong	7
		Very strong	2
6	Panicle exsertion	Partly exserted	58
		Mostly exserted	74
		Well exserted	76
7	Flag leaf attitude of blade	Erect	58
		Semi-erect	91
		Horizontal	51
8	Colour of stigma in spikelet	White	193
		Light green	1
		Yellow	3
		Light purple	6
		Purple	5
9	Panicle—awns	Absent	160
		Present	46
10	Lemma and palea colour	Straw	142
		Gold and gold furrows on straw background	3
		Brown (tawny)	2
		Reddish to light purple	57
		Purple spots/furrows on straw	2
		Purple	1
		Black	1
11	Basal leaf sheath colour	Green	199
		Light purple	5
		Purple lines	2
		Uniform purple	2
12	Leaf colouration	Light green	8
		Green	195
		Dark green	2
		Purple blotch	2
		Purple	1
13	Leaf colouration of auricles	Colourless	204
		Light purple	4
14	Leaf colouration of collar	Absent	206
		Present	2

Prasad *et al. CABI Agric Biosci*   (2020) 1:13

Page 9 of 24

Table 2 (continued)

No	Trait	Characteristics	Genotypes (No.)
15	Leaf colour of ligule	White	198
		Light purple	5
		Purple	5
	Agronomical traits		
16	Plant height (cm)	Semidwarf (< 110 cm)	13
		Intermediate (110–130 cm)	29
		Tall (> 130 cms)	163
17	Panicles numbers per plant	A few (< 11)	201
		Medium (11–20)	4
18	Single plant yield (g)	Minimum	5.1
		Maximum	23.5
		Average	11.9
19	Grain weight of 1000 grains (g)	Very low (< 15 g)	85
		Low (15–20 g)	103
		Medium (21–25 g)	12
		High (26–30)	4
20	Leaf area (cm^2)	Minimum	5.1
		Maximum	23.5
		Average	11.9
21	Specific leaf weight	Minimum	12.8
		Maximum	36.9
		Average	19.2
22	Ratooning ability	> 75%	38
		50 to 75%	60
		25 to 50%	39
		15 to 25%	31
		< 15%	37
23	Time of heading (50% flowering)	Medium (91–110 days)	42
		Late (111–130 days)	153
		Very late (> 131 days)	10
24	Panicle length of main axis (cm)	Short (16–20 cm)	2
		Medium (21–25 cm)	85
		Long (26–30 cm)	81
		Very long (> 30 cm)	37
25	Leaf blast (Bl)	Resistant (R)	96
		Moderately resistant (MR)	67
		Susceptible (S)	40
26	Bacterial leaf blight (BLB)	Resistant (R)	8
		Moderately resistant (MR)	0
		Susceptible (S)	200
27	Brown planthopper (BPH)	Resistant (R)	1
		Moderately resistant (MR)	9
		Susceptible (S)	197
28	White backed planthopper (WBPH)	Resistant (R)	3
		Moderately resistant (MR)	10
		Susceptible (S)	195

Table 2 (continued)

No	Trait	Characteristics	Genotypes (No.)
29	Gall midge (GM)	Resistant (R)	6
		Moderately resistant (MR)	0
		Susceptible (S)	194
30	Yellow stem borer	Resistant (R)	0
		Moderately resistant (MR)	2
		Susceptible (S)	204
	Quality traits		
31	Hulling (%)	Minimum	66.5
		Maximum	82.8
		Average	78.5
32	Milling (%)	Minimum	53.7
		Maximum	74
		Average	68
33	Head rice recovery (HRR %)	Minimum	12.8
		Maximum	72.9
		Average	60.9
34	Decorticated grain length (mm)	Minimum	3.4
		Maximum	7.2
		Average	4.7
35	Decorticated grain width (mm)	Minimum	1.6
		Maximum	2.6
		Average	2.3
36	Decorticated grain length/breadth ratio	Minimum	1.3
		Maximum	4.4
		Average	2.3
37	Decorticated grain type or shape	Long slender	2
		Short slender	6
		Medium slender	79
		Long bold	4
		Short bold	117
38	Endosperm presence of chalk	Absent	78
		Very occasionally present	119
		Occasionally present	8
39	Volume expansion ratio (VER)	Minimum	4
		Maximum	5.6
		Average	4.8
40	Water uptake (WU ml)	Minimum	50
		Maximum	285
		Average	124
41	Kernel length after cooking (mm)	Minimum	5.4
		Maximum	13.2
		Average	9.3
42	Elongation ratio (ER)	Minimum	1.4
		Maximum	2.7
		Average	2
43	Alkali spreading value (ASV)	Low to intermediate	27
		Intermediate	163
		High	16

Prasad *et al. CABI Agric Biosci*   (2020) 1:13

Page 11 of 24

Table 2 (continued)

No	Trait	Characteristics	Genotypes (No.)
44	Endosperm amylose content (%)	Low (10–19%)	13
		Medium (20–25%)	179
		High (26–30%)	15
45	Gel Consistency (GC)	Hard gel (26–40 mm)	75
		Medium gel (41–60 mm)	107
		Soft gel (61–100 mm)	24
46	Aroma of cooked decorticated kernels	Absent	0
		Present	208

of 30 genotypes. Two clusters have only two genotypes, Kalakanhu and Amritbhog in cluster 3 and Kamini Bhog and Kapoor Chini in cluster 20. We also calculated the intra- and inter-cluster diversity based on Euclidean distance values; it ranged from 22.33 for cluster 16 to 150.23 for cluster 3 indicating that the genotypes within these two clusters were divergent (see for Ward's minimum variance cluster means data Additional file 3: Table S2). Euclidean distance values were higher between clusters 1 and 3, 3 and 18, and 1 and 18. ARG genotypes grouped in these three clusters were Shuklaphool, Nagri and Kanak Jeer (cluster 1), Kalakanhu and Amritbhog (cluster 3), and Bhogajoha, Jeerakasala, Chinoor B, Thurur Bhog, Tulsi Amrit and Rajabhog (cluster 18). ARG genotypes in clusters 4 and 10 were relatively closer to each other in comparison to other genotypes.

Molecular diversity

Of the 55 SSR markers tested only 27 (~ 2.3 markers per chromosome) were polymorphic. Details of markers along with the number of alleles detected, PIC values, and repeat motif with map position are given in Table 3.

A total of 82 alleles were detected and alleles per marker ranged from 2 (HRM16592, RM229, RM267, RM505, RM577 and RM23899) to 5 (RM276) with an average of 3.04 alleles per marker. The PIC values estimated for markers also varied from 0.171 (RM577 on chromosome 1) to 0.721 (RM276 on chromosome 6), with an average of 0.54 per locus which indicated the effectiveness of these markers in analyzing the diversity in ARG genotypes. SSR marker RM276 with the repeat motif of (AG)8A3(GA)33 on chromosome 5 was highly informative as it amplified five alleles with the highest PIC value (0.721). Cluster analysis based on Jaccard's similarity coefficient of polymorphic markers classified the 208 ARG genotypes into three main clusters as depicted in Fig. 1. Cluster I consisted of only six genotypes (Jaiphulla, Juhibengal-A, RAU 3056, NDR IRRI

3131, Kalanamak (Birdpur) and Sonth). Cluster II with 81 genotypes was further grouped into nine sub-clusters. Cluster III was the largest with 121 genotypes grouped in eight sub-clusters (Additional file 4: Figure S2).

All-inclusive data analysis

All the ARG genotypes were grouped into 20 clusters in the analysis of all data on morphological, agronomical and quality traits, and on molecular characterization using Gowers' similarity index-based average linkage dendrogram on the degree of divergence (see Additional file 5: Figure S3). The distribution pattern has revealed a single ARG genotype segregating into cluster 4 (Kota Basmati), cluster 9 (Thurur Bhog), cluster 10 (Narendra lalmati), cluster 13 (Raja Bhog), cluster 15 (Deulabhog) and cluster 20 (IGSR-3-1-40). Two genotypes were grouped into each of the cluster 12 (Nagri Dubraj and Ram Bhog B) and cluster 18 (Gatia and Champaran Basmati 2). Cluster 7 consisted of 46 ARG genotypes while cluster 11 had the highest of 57 genotypes. The clustering pattern in our study did not follow the known geographical distribution of the ARG genotypes. The inter-cluster distances in all the cases were higher than that of intra-cluster distances suggesting wider genetic diversity among the genotypes of different clusters (Table 4).

The intra-cluster D^2 values in all the clusters were low, which indicated the close relationship of the genotypes within a cluster. The D^2 values were higher between clusters 20 and 10 (0.340), between 18 and 15 (0.334), and between 17 and 10 (0.323). All ARG in a single-genotype clusters show similarity in 14 traits; they are culm attitude, medium stem thickness, colour of basal leaf sheath, colour of auricles, colour of collar, colour of ligule, alkali spreading value, endosperm amylose content and reactions to six pests. These genotypes however, show striking differences in other 32 traits.

Table 3 Genetic diversity in ARG genotypes revealed by 27 SSR markers spread over 12 chromosomes of rice

S. No.	Chromosome (No.)	Primer name	Alleles (No.)	PIC values	Repeat motif
1	1	RM577	2	0.171	(TA)9(CA)8
2	1	RM10843	3	0.512	(TC)10
3	2	RM6318	3	0.662	(CTT)12
4	2	RM14140	3	0.514	(TTA)55
5	3	RM85	4	0.688	(TGG)5(TCT)12
6	3	RM6759	3	0.617	(TAT)53
7	4	HRM16592	2	0.300	(ATAC)29
8	4	HRM16913	4	0.711	(TAT)51
9	5	RM289	4	0.517	G11(GA)16
10	5	RM122	3	0.533	(GA)7A(GA)2A(GA)11
11	5	RM267	2	0.500	(GA)21
12	6	RM276	5	0.721	(AG)8A3(GA)33
13	6	HRM20060	3	0.585	(TTA)39
14	7	RM11	3	0.634	(GA)17
15	7	RM229	2	0.478	(TC)11(CT)5C3(CT)5
16	7	RM505	2	0.066	(CT)12
17	7	RM21260	4	0.686	(AG)18
18	8	RM23595	3	0.652	(GCAG)6
19	8	RM22866	3	0.367	(TA)19
20	9	RM7424	3	0.547	(GGAT)6
21	9	RM23899	2	0.382	(AT)32
22	9	RM24383	3	0.636	(CTC)8
23	10	HRM25754	3	0.634	(AT)47
24	11	RM26329	3	0.614	(AT)54
25	11	RM27311	3	0.653	(AT)72
26	12	RM27406	4	0.717	$(AT)_{72}$
27	12	HRM27840	3	0.603	(TAT)37

ARG genotype Kota Basmati in cluster 4 is a semidwarf from Rajasthan and has high single plant yield and leaf area, resistance to blast, short bold grains and soft gel consistency with aroma. Thurur Bhog in cluster 9 with aroma is tall with high ratooning ability, late maturing, moderately resistant to blast and a very high head rice recovery on milling. Narendra lalmati in cluster 10 is a semidwarf and mid-early maturing genotype with well-exserted panicle, high panicle numbers per plant and grain weight and resistance to blast disease. Raja Bhog from Madhya Pradesh in cluster 13 is tall, late-maturing, and has high kernel elongation after cooking with aroma, is high in amylose content, has hard gel consistency, and has moderate resistance to blast.

Deulabhog in cluster 15 has moderate resistance to blast and WBPH, and hard gel consistency. IGSR-3-1-40 in cluster 20 has tall plant stature, and high single plant grain yield, grain weight, and leaf area, and medium maturity duration; it has resistance to BPH, low water uptake on cooking kernels and medium to soft gel consistency with aroma.

Some ARG genotypes collected from different or same geographical region were grouped into same clusters, for instance Basamati and Basmati B from Uttar Pradesh in cluster 11, Champaran Basmati 3 and Champaran Basmati 4 from Bihar in cluster 7, and Kalanamak 1 and Kalanamak 2 from Uttar Pradesh in cluster 11 (Table 5). These genotypes did not show 100% similarity but exhibited variations in several traits and molecular attributes. In contrast, several other genotypes obtained from the same or different germplasm collection sites possessing similar or identical names were grouped in dissimilar clusters. They are represented by Bhanta Phool A (cluster 8) and Bhanta Phool B (cluster 3) from Uttar Pradesh; Champaran Basmati 1 (cluster 8) and Champaran Basmati 2 (cluster 18) from Bihar; Dhaniya B (cluster 3) and Dhaniya-B2 (cluster 7) from Uttar Pradesh; Juhibengal-A (cluster 19) and JuhiBengal-B (cluster 11) from West Bengal; Kanak jeer A (cluster 1) from Uttar Pradesh and Kanak jeer B (cluster 3) from Bihar; Kola Joha 1 (cluster 11), Kola Joha 2 (cluster 5) and

Fig. 1 Dendrogram showing genetic relationship among the 208 ARG genotypes based on Jaccard's similarity coefficient using 27 SSR markers (see Fig. S1, Additional file 4)
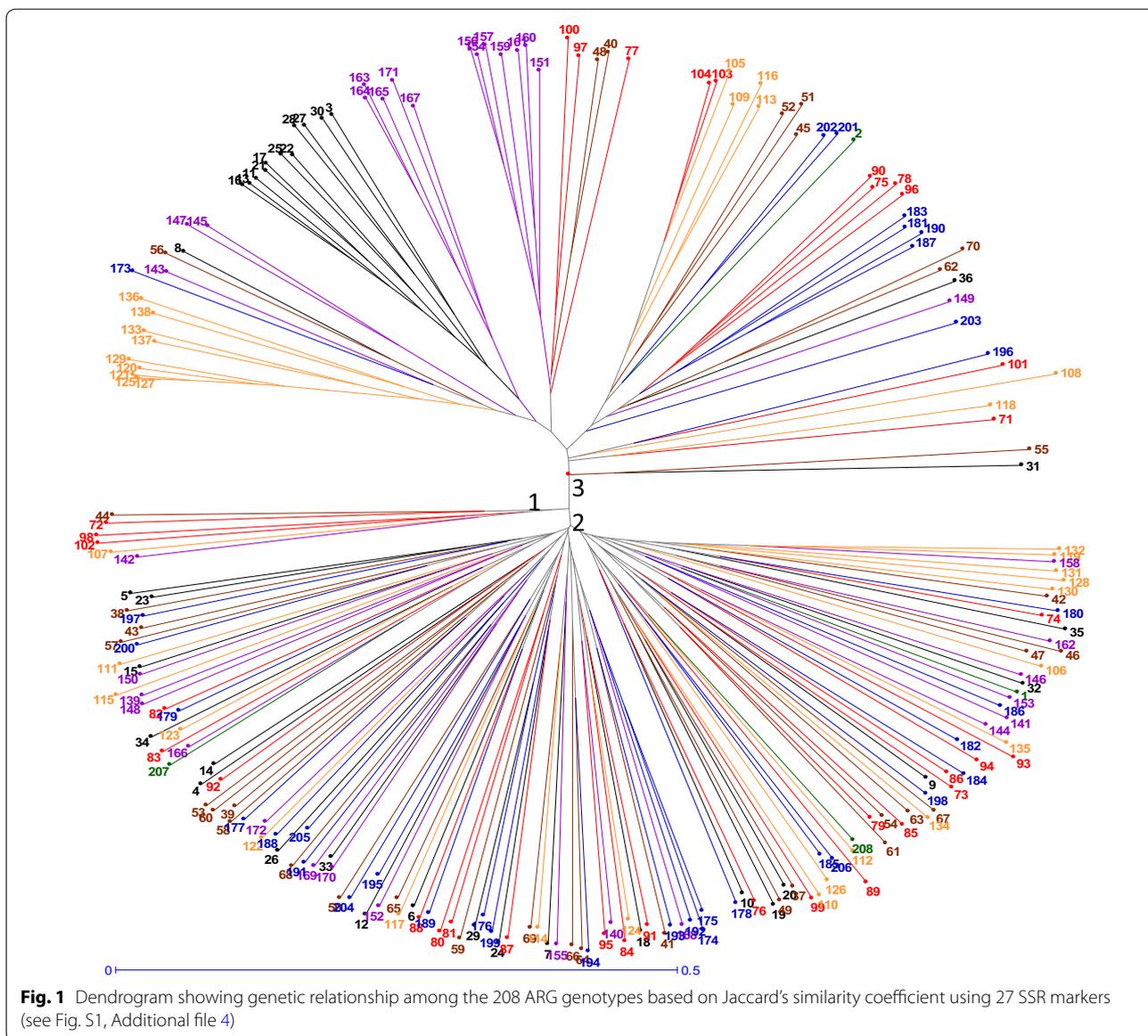
Kola Joha 3 (cluster 7) from Assam; Kon Joha 1 (cluster 5) and Kon Joha 2 (cluster 11) from Assam, Lectimachi (cluster 5) and Lectimachi B (cluster 16), Lokti muchhi (cluster 11) from Odisha; Lokti Musi (cluster 8), Loung choosi A (cluster 11) and Loung choosi B (cluster 3) from Chattisgarh; and Tulsiganthi (cluster 7) and Tulsi kanthi (cluster 5) from Odisha. Further, two ARG accessions with the same name were grouped in different clusters as with Atmashital (clusters 7 and 11), Bishnubhog (clusters 5 and 7), Ganjekalli (clusters 8 and 11), Kalanamak Birdpur (clusters 7 and 11), Kalajira (clusters 7 and 8), Neelabati (clusters 8 and 11), RAU 3043 (clusters 1 and 11), RAU 3048 (clusters 1 and 11) and RAU 3056 (clusters 1 and 11). Four ARG

accessions of Dubraj were grouped in different clusters apparently due to wider variations in all traits and molecular characters with RD 934 Dubraj in cluster 5, RD 1008 sel from Dubraj in cluster 7, RD 1214 Dubraj in cluster 11 and RD 1366 Dubraj in cluster 16. Some ARG genotypes in the same cluster exhibited similar fingerprint pattern with several SSR markers. Kamini Bhog and Kalijira 8−1 that are grouped together in a sub-cluster showed similar fingerprint patterns for 22 markers used. Another sub-cluster with three genotypes, Munibhog, Chitarsing and Dudaga also revealed similar fingerprints for 25 markers. In these two groups, we have detected similarity in the data for 40 of the 46 traits examined.

Prasad *et al. CABI Agric Biosci*   (2020) 1:13

Page 14 of 24

Table 4 Intra- and inter-cluster values among 20 clusters of 208 ARG genotypes

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
1	0.155	0.225	0.208	0.236	0.223	0.285	0.274	0.273	0.318	0.275	0.265	0.291	0.260	0.223	0.324	0.284	0.235	0.236	0.289	0.289
2		0.134	0.226	0.249	0.214	0.252	0.266	0.266	0.283	0.272	0.264	0.256	0.272	0.280	0.298	0.289	0.238	0.222	0.273	0.282
3			0.17	0.236	0.224	0.280	0.261	0.265	0.269	0.264	0.254	0.266	0.249	0.228	0.303	0.279	0.241	0.229	0.279	0.289
4				0	0.273	0.286	0.268	0.271	0.303	0.271	0.279	0.307	0.274	0.250	0.351	0.282	0.274	0.269	0.301	0.295
5					0.182	0.262	0.261	0.253	0.260	0.283	0.249	0.268	0.270	0.230	0.297	0.272	0.206	0.225	0.280	0.222
6						0.204	0.229	0.236	0.247	0.264	0.240	0.244	0.256	0.302	0.260	0.231	0.295	0.260	0.246	0.261
7							0.220	0.230	0.236	0.261	0.235	0.240	0.250	0.278	0.257	0.227	0.277	0.258	0.237	0.284
8								0.219	0.248	0.247	0.238	0.238	0.249	0.274	0.254	0.236	0.272	0.263	0.244	0.283
9									0.000	0.278	0.260	0.238	0.245	0.291	0.234	0.243	0.282	0.299	0.259	0.307
10										0.000	0.261	0.293	0.239	0.278	0.259	0.255	0.323	0.273	0.270	0.340
11											0.235	0.242	0.255	0.266	0.261	0.241	0.273	0.253	0.249	0.292
12												0.164	0.221	0.281	0.262	0.239	0.286	0.278	0.232	0.328
13													0.000	0.244	0.305	0.226	0.314	0.289	0.263	0.332
14														0.202	0.278	0.289	0.262	0.271	0.296	0.267
15															0.000	0.271	0.301	0.334	0.269	0.328
16																0.205	0.296	0.279	0.231	0.284
17																	0.115	0.222	0.297	0.236
18																		0.193	0.249	0.232
19																			0.207	0.297
20																				0.000

Based on combined data of morphological traits and 27 SSR molecular marker characterization

Genotype specific markers

A distinct band detected by a particular marker that is unique or less frequent at a particular base pair, is considered genotype specific. Of the 27 polymorphic markers used, four SSR markers (Fig. 2) were identified as genotype-specific markers: they are RM289 on chromosome 5, RM505 on chromosome 7, RM577 on chromosome 1 and RM22866 on chromosome 8. RM289 with the repeat motif G11(GA)16 showed a specific band at 150 bp in Magura, Bhulasapuri, Kota Basmati and Lectimachi. RM505 with the repeat motif (CT)12 revealed a specific band at 260 bp in RAU 3044 and Atmashital. RM577 with the repeat motif $(TA)_9(CA)_8$ showed a specific band at 250 bp in Gatia, RC 781 Chinor, IGSR-3-1-40, Kubri mohar and Kapoor Chini. RM22866 with repeat motif (TA)19 detected a specific band at 210 bp in IGSR-3-1-40 and Kubri mohar (Additional file 6: Figure S4).

Aroma and 2AP quantitative estimates

Strong aroma was detected in the sensory odor test with the cooked kernels of all the 208 ARG genotypes (see Additional file 1: Table S1). The quantitative estimation of 2AP with GC–MS was found to vary in powdered brown rice kernels in 90 ARG genotypes: it was absent (not detected) in 12 genotypes, 0.05 to 1 ppm in 25 genotypes, 1 to 2 ppm in 36 genotypes, 2 to 3 ppm in 12 genotypes, 3 to 4 ppm in 4 genotypes and 4 ppm in one genotype.

The highest level of 2AP in ARG genotypes tested was 4.49 ppm in RD 1214 Dubraj.

Development of a core set of ARG

A core set of 46 ARG genotypes were identified from the analysis of the data matrices on all traits and molecular characterization with 27 SSR markers (Table 6). The efficiency index was 0.84 and PIC value of the core set was 0.99 that established the presence of a maximum diversity of all the traits in the genotypes included. The coincidence rate retained by the core set was 96.93% that confirmed robustness and the retention of the allele richness of the entire ARG collection in the core set. A comparison of Shannon–Weaver diversity index (H′) among the core set and entire collection is presented in Table 7.

The estimates on mean, standard deviation and coefficient of variation (CV) for 46 genotypes of the core set over all the traits, and for 46 traits over these genotypes, were derived (see Additional file 1: Table S1). ARG genotypes were ranked on the basis of coefficient of variation (CV). Estimates on ARG genotypes across all traits revealed that the CV ranged from 29–42 per cent (see Additional file 7: Table S3). This narrow CV range of genotypes for all traits studied in the field experiment indicated a robust and equally good expression of traits in them. In contrast, the CVs of 46 traits varied widely from 2 to 93 per cent when derived over all the ARG

Table 5 Clustering pattern of 208 ARG genotypes based on 46 traits and 27 SSR marker characterizations

Cluster	ARG genotypes	Total
1	Badsha, Barang, Dindli, Kankjeer A, RAU 3036, RAU 3043, RAU 3048 (ARG 26), RAU 3056 (ARG 88), Rukmini bhog, Shuklaphool, Shyam Jira and Sitabhog	12
2	Bans patri, Bayasa Bhog, Bhulasapuri, Chitar Sing, Dudaga, Dudhraj, Jhili Safri, Munibhog, Rajin 12, Ram kali and Tulsi Bhog	11
3	Adam Chini B, ASGPC 14, ASGPC 38, Bhanta Phool B, Dhaniya B, Kalakanhu, Kanak Jeer B, Loung Choosi B and NDR 625	9
4	Kota Basmati	1
5	Bishnubhog (ARG 143), Chhabiswa, Govind Bhog, Kamod, Kola Joha 2, Kon Joha 1, Lectimachi, RB 287 Dubraj Bandi, RC 781 Chinnor, RD 1205 Dudgi, RD 934 Dubraj, RK 2512, Kadam Phool, RR 399 sel from Raja bhog and Tulasikanthi	14
6	Boga Joha, Bor Joha 1, Kubrimohar, Mohun Bhog, Pimpudibasa, Tulsi Amrit, and Tulsi Prasad	7
7	Atma Shital (ARG 58), Badshapasand, Bansa Parijat, Basnsapuri, Basuabhava, Bhatta Phool, Bishnu bhog (ARG 10), Champaran Basmati 3, Champaran Basmati 4, Chebdra Chhal, Chhatri Bhog, Dhaniya B2, Dhoiabankoi, Dhuriawa, Dubrajsena, Durga Bhog, Ganga Baru, Gangaballi, Gangabarud, Heera kani, IGSR-3-1-5, Kala Joha, Kalajira (ARG 104), Kalanamak (Birdpur) (ARG 107), Kali Kamod, Karunji phool, KB 13, Khorika Joha, Kola Joha 3, Krishna Joha, Lajkuli badan, Mahulakuchi, Moongphali B, Parbatjira, Parijatak, Randhunipaugal, RAU 3030, RAU 3079, RD 1008 (sel from Dubraj), RK 1019 (sel from Krishnabhog), Shankar chini-B, Shrabanamasi, Sonachoor, Til Kasturi, Tulasiganthi and Tulasiphulla	46
8	Ambemohar, Bans Phool A, Bhanta Phool A, Bhatagundi, Bikoni, Champaran Basmati 1, Chinoor B, Ganjekalli (ARG 41), Hankesh, Jaiphulla, Kalajira (ARG 42), Kanak Jeer, Kopousali Joha, Lokti Musi, Maguraphulla, Malaysia, Modak B, Neelabati (ARG 36), Seetabhog and Seetakeshari	20
9	Thurur Bhog	1
10	Narendra lalmati	1
11	Amritbhog, ASGPC 12, ASGPC 19, Atmashital (ARG 76), Basmati, Basmati B, Basnadhan, Dhanaprasad, Dhanprasad, Ganjekalli (ARG 100), IGSR 2-1-6, Jala, Jeerakasala, Jhali, Jhilpanjari, Jiraphool, JuhiBengal-B, Kalajira 801, Kalanamak (Birdpur) (ARG 49), Kalanamak (Nichlaul), Kalanamak 1, Kalanamak 2, Kalikati, Kamini Bhog, Kanak Jeera, Kankjeer, Kapoor Chini, Karnal Local B, Kataribhog, Kesar, Kheerasai, Khosakani, Khudi Joha, Kola Joha 1, Koli Joha, Kon Joha 2, Lokti muchhi, Loung Choosi A, Nagri, NDR IRRI 3131, Neelabati (ARG 101), Rajin 7, Ramdhuri, RAU 3041 (ARG 27), RAU 3041 (ARG 87), RAU 3043, RAU 3044, RAU 3048 (ARG 85), RAU 3055, RAU 3056 (ARG 79), RAU 3061, RAU 3076, RAU 724-48-33, RD 1214 Dubraj, Sheetalkani, Sonth and Urai Butta	57
12	Nagri Dubraj and Ram Bhog B	2
13	Raja Bhog	1
14	Karpurkranti and Manasi	2
15	Deulabhog	1
16	ASGPG 39, Chini Kapoor, Duban Mua, Dudh Nag, Dudikanth, Ganjo, Garra kat, Jhingisiali, Joha Bora, Kamod 119, Laung Choor B, Lectimachi B and RD 1366 Dubraj	13
17	Magura and Sujata	2
18	Champaran Basmati 2 and Gatia	2
19	Bubraj (Raipur), Juhibengal-A, Kamini Joha, Krushnabhoga and Thakurabhog	5
20	IGSR-3-1-40	1

genotypes. Aroma in cooked decorticated kernels was present in all these ARG genotypes.

The core set comprised medium (13), late (30) and very late (3) maturing tall (35), intermediate (5) and semidwarf (6) ARG genotypes (Table 6). They represent collections from 11 states of India. These genotypes have short bold (24), medium slender (16), short slender (3) and medium bold grains (3 genotypes). Panicles of ARG genotypes including those in the core set showed a wide variation (see Additional file 8: Figure S5). Head rice recovery varied from 23% in RAU 3041 to 72% in Magura. Amylose content was low (18%) in Juhibengal-A and high (28%) in RC 781 Chinoor. The gel consistency of cooked kernels was soft in 8 genotypes, medium in 23 genotypes and hard in others. The cooked kernels of all genotypes in the core set were found to be aromatic.

Discussion

The genetic proximity of aromatic and *japonica* types are reported to be the reasons for less efforts to trace origin of aromatic rice (Zhao et al. 2011; Wang et al. 2014). The center of diversity of aromatic rice is along the Himalayan foothills where they are traditionally grown in India (Glaszmann 1987; Khush 2000). The group V designation attributed by Glaszmann (Glaszmann 1987) has not been universally accepted and conflicting reports describe this group as *aromatic* (Khush 2000; Civan et al. 2015; Civan et al. 2020). The genomic history of the aromatic population was unraveled by Civan et al. (2020) from the diversity found in the unique genome-wide *sativa–rufipogon* data set. They have identified *aus* as the original crop of the Indian

Fig. 2 Detection of specific bands in ARG genotypes with SSR markers, RM289, RM505, RM577 and RM22866 (see Fig. S2, Additional file 6)
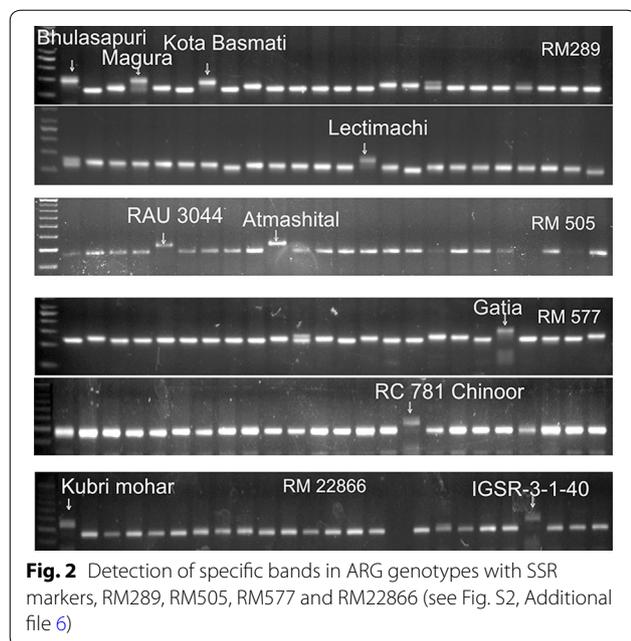
subcontinent, *indica* and *japonica* as later arrivals, and aromatic a specific product of local agriculture.

Basmati with its long grains has dominated the domestic and international market for aromatic rice. Indian farmers continue to grow indigenous non-Basmati landraces with medium and short grains that excel equally for aroma and cooking qualities besides proving medicinal benefits for domestic consumers. These fragrant landraces are locally considered as superior for other quality traits such as amylose content, grain length, grain elongation after cooking and texture of cooked rice and have high cultural significance. Bindli, a small-grained aromatic rice possesses aroma and its cooked kernels elongate by more than 3-folds surpassing that of Basmati. Cooked kernels of the landrace Radhunipagal is known to produce strong and intense aroma (Singh et al. 2003). Cooked kernels of Muhulakuchi and Kesr of the core set identified in our study also showed 3-folds elongation. Yet, scientists, merchants and exporters have not paid adequate attention in non-Basmati aromatic rice improvement and marketing them as a priced commodity.

Generally tall and late maturing aromatic rice genotypes produce poor grain yields. Therefore, commercial cultivation of aromatic rice remains unprofitable despite the premium price offered for its grains. It is essential to assess aromatic rice germplasm accessions for a selective breeding to improve their yield while retaining unique aroma and other quality features. Analysis of characterization data on morphological, agronomical and 16 physico-chemical traits showed considerable variation among the aromatic rice germplasm (Kapoor et al. 2019). Characterization of native aromatic rice landraces that are highly preferred by consumers can help in their identification and conservation (Choudhury et al. 2001). Morphological descriptors are reliable and easy to study but require skill in making assessments and the availability of field, greenhouse and laboratory facilities. In spite of these drawbacks, morphological descriptors continue to be important in characterization of germplasm lines due to low cost. ARG collections from various geographical regions of India are preserved at certain national research centres in India. We obtained 208 ARG genotypes from four such centres, purified and characterized for 46 morphological, agronomical and quality traits considered important for crop improvement. In addition, molecular characterization of genotypes was done with 27 polymorphic SSR markers. Sensory aroma assessment was made on the fragrance arising from cooked kernels of all the genotypes. 1-Acetyl-2-pyrroline (2AP), the compound considered responsible for aroma was quantitatively estimated in the kernels limited to only 90 genotypes due to high processing and detection costs.

We have shown ARG genotypes to exhibit a remarkable variability in traits of plant habit, stem thickness, specific leaf weight, panicle exsertion, panicle number per plant and panicle length, grain types and its weight, reaction to pests and grain quality. In Ward's minimum variance dendrogram generated using 46 traits, ARG genotypes were grouped into 20 clusters with Euclidean distance ranging from 22.33 to 150.23, and with only two genotypes Kalakanhu and Amritbhog in one cluster. Euclidean distance values were higher in clusters with Shuklaphool, Nagri and Kanak Jeer (cluster 1), Kalakanhu and Amritbhog (cluster 3), and Bhogajoha, Jeerakasala, Chinoor B, Thurur Bhog, Tulsi Amrit and Rajabhog (cluster 18). Some ARG collected from different or same region were grouped into same clusters despite differences in traits. Several other genotypes obtained from the same or different sources possessing same or identical names were grouped in dissimilar clusters. Use of morphological descriptors may have less significance due to low polymorphism and heritability, late expression and vulnerability to environmental influences. In contrast, DNA markers that are widely distributed in the genome remain unaffected across different stages of plant growth, seasons, locations and agronomic practices, and thus provide highly reproducible inferences (McCouch et al. 1997).

SSR markers have proved to be very effective tools in the study of genetic diversity and organism relationships due to their high polymorphic nature and transferability (Ishii et al. 2001). Genotyping of 192 *O. sativa* germplasm lines using 61 SSRs showed 205 alleles with the PIC value

Table 6 Core set of 46 genotypes identified representing all alleles detected 208 ARG genotypes

S. No.	ARG No.	Name	Source	Height	Grain type	Grain wt. (g)	Aroma	2AP#
Medium heading (91–110 days)[a]								
1	1	Shukla Phool	CH	Tall	MS	19	Present	-
2	3	Kubri Mohar	CH	Tall	SS	23	Present	0.66
3	4	Bhatta Phool	UP	Tall	SB	13	Present	-
4	6	Bans patri	CH,MH	INT	MS	12	Present	-
5	8	Chini Kapoor	CH	Tall	SB	15	Present	2.02
6	13	Moongphali—B	UT,UP	SD	SB	15	Present	-
7	14	Shyam Jira	BI	Tall	SB	13	Present	-
8	61	Muhulakuchi	OD	Tall	SB	15	Present	0.91
9	67	Kesr	UP	SD	SB	18	Present	-
10	72	Juhibengal-A	WB	Tall	MS	15	Present	3.64
11	166	Kanak Jeer	UP	Tall	MS	16	Present	0.17
12	201	Kapoor Chini	CH	SD	MS	17	Present	-
13	208	Narendra lalmati	UP	SD	SS	17	Present	1.51
Late heading (111–130 days)								
14	15	Dindli	UP	Tall	MS	16	Present	-
15	27	RAU 3041	BI	SD	SB	12	Present	0.91
16	29	RAU 724-48-33	BI	SD	MS	17	Present	-
17	31	Maguraphulla	OD	Tall	SB	21	Present	0
18	41	Ganjeikalli	OD	Tall	SB	17	Present	0
19	44	Jaiphulla	OD	Tall	SB	20	Present	0
20	48	Nagri Dubraj	OD	Tall	MS	16	Present	0.13
21	50	Basnadhan	OD	Tall	SB	13	Present	-
22	62	Thakurabhog	OD	Tall	SB	13	Present	-
23	64	Kalikati	OD	Tall	SB	12	Present	0.98
24	76	Atmashital	CH	Tall	SB	11	Present	-
25	77	Amritbhog	UP	Tall	SB	17	Present	-
26	83	RAU 3079	BI	Tall	SB	13	Present	2.51
27	84	RAU 3030	BI	Tall	MS	12	Present	0.20
28	86	RAU 3043	BI	INT	SB	13	Present	-
29	93	Pimpudibasa	OD	Tall	LB	11	Present	-
30	97	IGSR-3-1-5	CH	INT	SB	12	Present	-
31	98	IGSR-3-1-40	CH	Tall	LB	28	Present	-
32	101	Neelabati	OD	Tall	SB	11	Present	1.65
33	109	Malaysia	UP	Tall	MS	16	Present	-
34	116	ASGPC-19	UP	Tall	SB	12	Present	-
35	121	Bikoni	CH	Tall	SB	27	Present	-
36	126	Dudaga	MP	INT	MS	16	Present	-
37	151	RD 1008 Sel from Dubraj	CH	Tall	SS	15	Present	1.59
38	157	RC 781 Chinnor	CH	Tall	LB	26	Present	-
39	159	RB 287 Dubraj Bandi	CH	Tall	MS	16	Present	-
40	170	Champaran Basmati 1	BI	Tall	MS	18	Present	1.35
41	190	Khorika Joha	AS	Tall	MS	18	Present	0
42	199	Kamini Bhog	UP	Tall	SB	10	Present	2.01
43	205	Kala Namak (Nichlaul)	UP	Tall	MS	19	Present	3.39
Very late heading (> 131 days)								
44	46	Magura	OD	Tall	SB	15	Present	-
45	207	Jeerakasala	KE	INT	SB	21	Present	0
46	32	Deulabhog	OD	Tall	MS	-	Present	-

Based on morphological, agronomical and quality traits and molecular characterization with 27 SSR markers. [a] Days to 50% flowering. *AS* Assam, *BI* Bihar, *CH* Chhattisgarh, *HA* Haryana, *KE* Kerala, *MP* Madhya Pradhesh, *MH* Maharashtra, *OD* Odisha, *UP* Uttar Pradesh, *UT* Uttarakhand and *WB* West Bengal, *Int* Intermediate, *SD* semidwarf, *LS* long slender, *SS* short slender, *MS* medium slender, *LB* long bold; *SB* short bold; *2AP* 1-Aetyl-2-pyrroline; # Not tested

Source data are provided in Additional file 1: Table S1

Prasad *et al. CABI Agric Biosci*   (2020) 1:13

Page 18 of 24

of 0.756. Population structure analysis using model-based and distance-based approaches revealed two distinct subgroups (Nachimuthu et al. 2015). Analysis of 18 primers dependent genetic polymorphism revealed 89 shared and 91 unique allelic variants generated in the form of amplified products amongst 18 traditional and improved varieties. PIC values of the primers ranged from 0.60 to 0.88 with an average of 0.80. Ample genetic differentiation and divergence amongst entries were deduced by a comparison of similarity coefficients (Saheewala et al. 2019). In Thai and exotic rice germplasm, 110 alleles were detected with an average of 8.46 alleles per locus using 13 markers distributed over 12 rice chromosomes. The averages of gene diversity, heterozygosity and polymorphic information content were 0.59, 0.02 and 0.56, respectively (Pathaichindachote et al. 2019). The allelic diversity and relationship among 48 *aus* rice landraces determined through 11 SSR markers showed 3 alleles (RM234 and RM277) to 15 alleles (RM493) (Ahmed et al. 2019). The PIC values ranged from 0.19 (RM277) to 0.90 (RM493) with an average 0.70. In our study with 208 ARG genotypes, 82 alleles were detected and alleles per marker ranged from 2 (RM505) to 5 (RM276). Average number of 3.04 alleles per marker found in our study on indigenous group of ARG genotypes was comparable to the previous studies (Nagaraju et al. 2002; Cho et al. 2000). The PIC value in the present study is also similar to the previous studies (Sivaranjani et al. 2010; Wong et al. 2009), higher than the earlier studies (Singh et al. 2004; Akagi et al. 1997) and lower than those reported by Ni et al. (Ni et al. 2002). Further, directed selection during domestication might have led to preferential accumulation of favorable alleles. The high average number of alleles per locus reported in some previous studies might be due to global assessment of diversity using large and diverse collections rather than a group from specific geographic region; it may also be partly due to the higher resolution power of the amplicons in the sequencing gel detection system (Ni et al. 2002).

Among 55 SSR markers, 28 were not considered for analysis as they were either monomorphic or produced ambiguous amplification. Only 27 SSR markers and ~2.3 markers per chromosome were polymorphic, and they grouped 208 ARG genotypes into three clusters with PIC values varying from 0.171 to 0.721 indicating their effectiveness in analyzing the diversity. Cluster I consisted of only six genotypes viz., Jaiphulla, Juhibengal-A, RAU 3056, NDR IRRI 3131, Kalanamak (Birdpur) and Sonth. There was no distinct trend in the diversity of morphological traits of genotypes based on geographic origin. Genotypes that showed high divergence have been reported to produce wide variability with transgressive segregates in exotic rice (Qian and He 1991), semi-dry

rice (Rao and Gomathinayagam 1997) and lowland rice (Pankaj et al. 2006). Grouping of some ARG genotypes of the same origin into different clusters was also observed, which indicated their broad genetic base. The Gower similarity index had grouped ARG genotypes into 20 clusters on morphological, agronomical and quality traits and molecular markers characterization. The distribution pattern revealed clusters with one or two or more ARG genotypes with higher intra-cluster distances than inter-cluster distances suggesting wider genetic diversity between clusters. However, in clusters with only one ARG accession, genotypes showed striking differences in several traits. Some ARG genotypes in the same cluster exhibited a similar fingerprint pattern with several SSR markers, but showed difference with other markers.

Markers with PIC values of 0.5 or higher are highly informative for genetic studies and are extremely useful in distinguishing the polymorphism rate of a specific locus. In the present study, 14 SSR markers viz., HRM27840, RM26329, RM6759, RM11, HRM25754, RM24383, RM23595, RM27311, RM6318, RM21260, RM85, HRM16913, RM27406 and RM256 with average PIC value of > 0.6 are identified as informative markers and are useful for genetic diversity studies. Further, four SSR markers RM289, RM505, RM577 and RM22866 (Fig. 2) were identified as genotype specific markers. This DNA finger printing can help us identify germplasm duplicates, genotype diversity and novel alleles in the germplasm. It can also help us formulate a core collection for use, register plant variety, establish essentially derived status and maintain germplasm seed purity.

Often, qualitative and quantitative data are used separately to assess genetic diversity of crop genotypes. When varieties with gene pools were examined using molecular markers, extremely high similarity measures are produced that are linked to morphological similarities in the crop (Roldan-Ruiz et al. 2001). The significant correlation indicates that these two independent sets of data likely reflect the same pattern of genetic diversity and validate the use of these methods for diversity estimation and also in grouping of genotypes. The degree of correspondence between the clusters formed may not agree with each other in some cases. Such ambiguity is avoided in our study by assessing the diversity based on the combined data of qualitative and quantitative traits. Clustering based on the combined morphological parameters and molecular markers suggested that genotype IGSR-3-1-40 present in cluster 20 was more diverse amongst genotypes evaluated. This genotype is an evolved ARG possessing many superior attributes, when compared to others. Dissection of phenotypic and molecular variability in ARG genotypes has revealed a wide genetic variation for the rice grain quality, especially amylose content

Table 7 PowerCore PIC values for traits and SSR molecular marker characterization

Trait/Marker	Core collection		Entire collection	
	Shannon Weaver	Allele	Shannon Weaver	Allele
Leaf shape of ligule	0.908	3	0.800	3
Stem thickness	0.213	3	0.125	3
Leaf intensity of green colour	1.125	4	0.877	4
Leaf blade pubescence	1.282	5	1.028	5
Panicle exsertion	1.088	3	1.093	3
Stigma colour	0.631	5	0.348	5
Panicle awns	0.374	2	0.376	2
Lemma palea at flowering	1.072	6	0.815	6
Basal leaf sheath colour	0.539	4	0.221	4
Leaf blade colour	0.612	5	0.301	5
Leaf margin colour	0.432	2	0.221	2
Auricle colour	0.245	2	0.095	2
Collar colour	0.107	2	0.054	2
Ligule colour	0.424	3	0.226	3
Plant height	1.326	5	1.204	5
Panicle number	0.149	2	0.049	2
Single plant yield	1.101	4	1.010	4
1000-grain weight	1.077	4	0.955	4
Leaf area	1.594	6	1.521	6
Specific leaf weight	0.852	4	0.718	4
Days to flowering	0.822	3	0.704	3
Panicle length	1.226	5	1.120	5
Hulling	1.045	4	0.707	4
Milling	0.958	4	0.771	4
Head rice recovery	1.040	5	0.841	5
Kernel length	1.465	6	1.371	6
Kernel breadth	0.850	3	0.791	3
Length/breadth ratio	0.946	3	0.799	3
Grain type	1.056	5	0.861	5
Volume expansion ratio	1.298	4	1.239	4
Water uptake	1.212	5	1.192	5
Kernel elongation after cooking	1.744	7	1.679	7
Elongation ratio	0.850	3	0.743	3
Amylose content	0.718	5	0.618	5
Gel consistency	1.363	4	1.331	4
Aroma	0.674	2	0.619	2
Blast	1.474	5	1.404	5
Bacterial leaf blight	0.466	4	0.307	4
Brown planthopper	0.423	5	0.250	5
White backed planthopper	0.497	5	0.309	5
RM10843	1.010	3	1.031	3
RM85	1.173	4	1.180	4
RM6759	1.545	5	1.519	5
RM16592	1.037	3	1.031	3
RM505	0.451	2	0.391	2
RM11	0.883	3	0.866	3
RM021260	1.202	4	1.205	4

Table 7 (continued)

Trait/Marker	Core collection		Entire collection	
	Shannon Weaver	Allele	Shannon Weaver	Allele
RM289	0.706	4	0.616	4
RM22866	1.080	3	1.046	3
RM23595	1.082	3	1.086	3
RM23899	1.500	5	1.464	5
RM24383	0.749	3	0.726	3
RM7424	0.572	3	0.517	3
RM25754	1.462	5	1.476	5
RM27311	0.728	3	0.694	3
RM27406	1.750	6	1.716	6
RM27840	1.368	5	1.311	5

and alkali-spreading value as reported in the USDA rice mini-core collection that represents worldwide rice germplasm lines (Song et al. 2019).

A core collection of rice can serve to conserve and promote use of the germplasm, to understand the inherent diversity of grass genomes, and to exploit synteny among grass genomes. The choice of germplasm accessions for breeding or research is based on the existence of a particular useful trait that is expressed by an accession (Jackson 1997). A core collection should contain, with minimum repetitiveness, a good representation of the diversity of the germplasm. The use of core subsets of a size greater than or equal to 10% of the whole collection satisfied the requirement of representativeness with regard to diversity of maize landraces stored in the International Maize and Wheat Improvement Center (CIMMYT) germplasm bank (Franco-Duran et al. 2019). Sample sizes greater than 10% of the whole population size retained more than 75% of the polymorphic markers for all selection strategies and types of sample; lower sample sizes showed more instability among repetitions. A core collection of 150 rice germplasm lines (78% of the entire set) were assembled based on the analysis of population structure and genetic relatedness in 192 lines (Nachimuthu et al. 2015). This extremely high a representation in the core set was apparently due to inclusion of germplasm collected from nine different states of India as well as from Argentina, Bangladesh, Brazil, Bulgaria, China, Colombia, Indonesia, The Philippines, Taiwan, Uruguay, Venezuela and the United States. Still the perfect ratio and fixed size for all core set of genotypes does not exist, since different crops or different construction goals need different sampling percentages (Hu et al. 2000). There is no universally accepted method for constructing a core set as many factors affect representativeness of the core set, such as sampling percentage,

data type, number of traits observed, genetic diversity of germplasm, grouping method and sampling method (Rao et al. 2012). A core set of 160 short duration rice germplasm (87% of total diversity) from over 1000 accessions held by Kerala Agricultural University in India was formed based on 24 quantitative characters using the PowerCore (Saini et al. 2013). A mini-core of 21 submergence tolerant rice genotypes representing the genetic diversity derived from 5716 accessions was developed using PowerCore on the basis of one qualitative and five quantitative traits (Jambhulkar et al. 2018). Selection of germplasm accessions on the basis of passport data is usually the primary criterion for inclusion in any core collection. However, none of these core sets reported earlier had the full complement of alleles detected in the entire set of genotypes studied. The size of the core collection should also clearly reflect users' needs (Roy et al. 2015; Brown 1989). With PowerCore, we identified a core set of 46 ARG genotypes (22% of total collection) from the analysis of data matrices on 46 traits and molecular characterization with 27 SSR markers (Table 6). The efficiency index was 0.84 and PIC of the core set was 0.99, which established the presence of a maximum diversity. This core set contained all the 82 alleles detected in the entire 208 ARG collection investigated. Relatively more genotypes are in this core set than that reported in many of self-pollinated crops. Rice genotypes are preferentially grown based on maturity duration, plant stature and grain type. Hence, the presence of more genotypes with variations for these traits in the core set that represent fully the entire collection studied has a high value to breeders seeking to improve ARG.

The coefficient of variation (CV) is a standardized, dimensionless measure of dispersion relative to a data set's average. It enables the comparison of several datasets on genotypes with different units of measurement (Ospina and Marmolejo-Ramos 2019). It is also used as a measure to compare the robustness of different biological traits (Félix and Barkoulas 2015). The CV was below 20% in 21 traits indicating the robustness of these traits in all the genotypes of the core set; these traits are aroma, hulling, time of heading, milling, plant height, water uptake, endosperm amylose content, head rice recovery, panicle length, gel consistency, specific leaf weight, leaf area, 1000-grain weight, volume expansion ratio, stem thickness, kernel length after cooking, endosperm chalk, single plant yield, panicle number per plant, decorticated grain type or shape and decorticated grain length. All the remaining 25 traits are apparently more variable in expression amongst ARG genotypes. A further scrutiny of all data on traits of genotypes in the core set can be used to refine and select a particular ARG genotype from the core collection to suit the breeding objectives.

Soft and fluffy texture of cooked rice, with a pleasant fragrance and appealing taste are the most desired by consumers (Singh et al. 2018). Juhibengal-A or Jeerakasala with low amylose content (18–20%) in kernels can be used to make flat rice noodles, and those genotypes with hard gel consistency are suitable for extruded noodles. Further they may be used in making rice paper and rice wines. RC 781 Chinnor and Ganjeikalli are the best to produce parboiled rice as they contain high amylose (27–28%) in grain. Genotypes with low or high amylose content in grains and with hard, medium or soft gel consistency are available in the core set that can be used to develop exclusive value-added products like aromatic biscuits, fermented saki etc. for domestic consumption and export.

Sixty-five crosses have been developed utilizing part of this identified core germplasm. These elite cultures are in multi-location tests under the All India Coordinated Rice Improvement Programme. Thus the intensive efforts in this study have led to the identification of diverse genotypes for use in the hybridization to develop high-yielding semidwarf aromatic rice varieties suitable for commercial cultivation and quality non-Basmati aromatic grain export.

Humans sense aroma due to the production and release of more than 300 volatiles in rice grains. Rice aroma is estimated by using sensory evaluation, gas chromatography–mass spectrometry (GC–MS), gas chromatography–olfactometry (GC–O), and sensory evaluation or electronic nose (E-nose) (Hu et al. 2020). The presence of 2-acetyl-l-pyrroline (2AP) in grains is considered as the key aroma compound present in almost all the aromatic rice genotypes. Bradbury et al. (2005) identified a single recessive gene responsible for aroma. This gene is a defective allele of a gene encoding betaine aldehyde dehydrogenase BADH2. The deletion observed in exon 7 of this (*BADH2)* gene generates a premature stop codon and presumably results in loss of activity. It was hypothesized that loss of BADH2 activity causes 2AP accumulation (Bradbury et al. 2008). The 2AP content varied from 0 to a maximum level of 4.2 ppm in RD 1214 Dubraj (ARG 153). Although fragrance was found in the cooked kernels in sensory odor tests, 2AP was not detected with the costly GC–MS assay of brown rice samples of 12 ARG genotypes: they are Dhuriawa (ARG 16), Maguraphulla (ARG 31), Krushnabhoga (ARG 33), Ganjeikalli (ARG 41), Jaiphulla (ARG 44), Seetabhog (ARG 66), Lajkuli badan (ARG 89), Kataribhog (ARG 103), Badshapasand (ARG 110), Khorika Joha (ARG 190), Kali Kamod (ARG 195) and Jeerakasala (ARG 207). Applying an improved headspace solid-phase micro-extraction (HS-SPME) method combined with a GC–MS assay to a collection of 228 aromatic rice germplasm, the 2AP

Prasad *et al. CABI Agric Biosci*   (2020) 1:13

Page 21 of 24

content in 34 genotypes was reported to be below the limit of detection (Xie et al. 2019). In addition to 2AP, six volatile compounds like nonanal, hexanal, hydroxyl methyl furfural, indole, benzyl alcohol, and guaiacol were detected at similar concentration in the aromatic germplasm (Peddamma et al. 2018). This suggested that it is premature to identify one or more of these compounds as responsible for the desired aroma considering the specificity of olfactory receptors and their highly varying odor detection threshold for individual compounds. These volatiles are present in the rice grains or generated during cooking. Locating the key compounds that contribute to aromatic rice has been difficult. Un-milled black rice contained significantly larger amounts of total volatiles than milled black rice (Choi et al. 2019). Cultural practices and growth environments of aromatic rice and storage of grains after harvest also significantly affect fragrance. The human nose has a theoretical odor detection limit of about 10^{-19} mol (Wilkie et al. 2004), making sensory evaluation a valuable and sensitive method for rice aroma analysis. Therefore, an exclusive reliance on 2AP estimates obtained using GC–MS may lead us to erroneous conclusions on the presence or absence of aroma. Further, most assessments on 2AP were done using brown rice or processed rice products. But the trade and consumers rely only on the odor of cooked kernels. The sensory odor evaluation of fragrance in cooked kernels adopted in our investigations has detected the presence of aroma in all the 208 ARG genotypes. On the basis of the odor active values, 18 odor-active compounds were determined as key aroma compounds (Yu et al. 2019). For most volatiles with odor descriptions, odor thresholds and retention indices are available for *japanica* rice. The universal aromatic scale (UAS) and the rice aromatic scale (RAS) were used to evaluate flavor attributes of cooked rice samples (Arroyo and Seo 2017). Sensory methods are not quantitative and are affected by physical condition and environment. Large variations are experienced in repeated evaluation and in odor sensitivities between individual tests of a genotype. Sensory fatigue is likely to be a source of error. Different proportion of characteristic volatiles might result in different perceptual results on the intensity of perceived odor and its attributes. Clear understanding on the characteristic volatiles that release aroma and their relationship with sensory results are still needed (Hu et al. 2020).

Rice aroma is a highly heritable trait (Wakte et al. 2017). Genetic analysis showed that primary aroma traits were genetically controlled by recessive monogenes, independent of cytoplasmic genes. However, aroma was also studied as a quantitative trait, and many genes were included in the expression (Hashemi et al. 2013). Therefore, genes for rice aroma appear to be complex, with several major and minor genes controlling rice aroma. Further, three QTLs *aro3-1*, *aro4-1* and *aro8-1* were identified on Basmati chromosomes 3, 4 and 8, respectively (Singh et al. 2007) that were independent of any association with the *badh2* gene on chromosome 8. A major QTL *qGE6.1* on chromosome 6, and another *qGE6.2* on chromosome 4 were identified (Arikit et al. 2019). Transcript profiling enabled detection of close relations between aroma-related genes like *TPI*, *PDH* and *OAT*, and between *P5CS* and *P5CDH* in the grains of Gobindobhog, Kalonunia, Tulaipanji and Radhunipagal (Ghosh and Roychoudhury 2020). A simple, co-dominant functional marker has been developed that targets the InDel polymorphism in *badh2* gene. This marker was shown to co-segregate with the trait of fragrance in the mapping population (Sakthivel et al. 2009).

Grain aroma in Basmati rice is controlled by a single recessive gene located on chromosome 8 (Singh et al. 2007). By using the backcross inbred lines and chromosome segment substitution lines, major-effect QTLs can be detected for several agronomic traits, including heading date, eating quality, grain shape, grain chalkiness, culm width, 1000-grain weight, and resistance to pests. Using such methodologies genes have been identified for grain quality: *GS3* for grain length and ALK gene for alkali spreading value in Basmati (Singh et al. 2018). Background selection was performed using 65 well-distributed SSR markers to preserve most of the KDML105 genetic background for the adaptability and market quality of the Thai Jasmine rice (Vanavichit et al. 2018). The quantitative trait loci have been identified in a F_2 derived from Ranjit and Kola Joha (ARG) using SSR markers linked to aroma. In aromatic non-Basmati rice from Assam, of two QTLs for grain aroma each on chromosome 5 and chromosome 8, only one was in a position similar to the aroma gene of Basmati rice (Talukdar et al. 2017). The discovery of genes and QTLs responsible for grain aroma is the key to develop and use marker-aided selection to improve aromatic non-Basmati rice.

Conclusions

Forty-six morphological, anatomical and quality traits and 27 SSR markers reveled characteristics were studied in 208 genotypes of the non-Basmati ARG accessions maintained in India. A wide genetic variability was evident from the data on traits, molecular characterization, alleles identified and segregation in one-, two- or three-genotype clusters with large genetic distance in diversity analysis. The core set identified with 46 genotypes represented all the 82 alleles detected

Prasad *et al. CABI Agric Biosci*   (2020) 1:13

Page 22 of 24

in the entire collection of ARG. These genotypes have different plant statures, maturity periods and quality attributes to exploit in breeding high-yielding aromatic rice. The sensory odor test of fragrance from cooked kernels of ARG was more conclusive than the expensive 2AP estimation with the costly GC–MS that failed to detect fragrance in brown rice of many aromatic genotypes. It is essential that we develop simple and easy characterization of aroma with descriptors to use in crop improvement, and in domestic and international trade. Quantitative assessments are needed on aroma from non-Basmati dehusked brown kernels, milled kernels and parboiled rice before and after subjecting them to various storage periods and conditions. Accelerated research is essential to discover genes and QTLs responsible for grain aroma to develop and use marker-aided selection to improve aromatic non-Basmati rice.

Supplementary information

Supplementary information accompanies this paper at https://doi.org/10.1186/s43170-020-00013-8.

Additional file 1: Table S1. Sheet 1 data on Assessment of morphological, agronomical and quality traits of aromatic germplasm (ARG), 2-Acetyl-1-pyrroline content (ppm) and core set; & Sheet 2- Details on descriptors, characteristics, score assessment and decimal code for growth stages used for ARG.

Additional file 2: Figure S1. Ward's minimum variance dendrogram and Euclaidean^2 distance.

Additional file 3: Table S2. Ward's minimum variance, Cluster means data.

Additional file 4: Figure S2. Dendrogram showing genetic relationship among the 208 ARG genotypes based on Jaccard's similarity coefficient using 27 SSR markers.

Additional file 5: Figure S3. Average linkage dendrogram and Gowers' similarity index.

Additional file 6: Figure S4. Detection of specific bands in ARG genotypes with SSR markers, RM289, RM505, RM577 and RM22866.

Additional file 7: Table S3. Estimates on mean, standard deviation and coefficient of variation (CV) for 46 ARG genotypes over all the traits, and for 46 traits over all the ARG genotypes investigated.

Additional file 8: Figure S5. Variability in panicles of ARG genotypes including those in the core set.

Abbreviations
ARG: Aromatic Rice Germplasm; ICAR-IIRR: Indian Institute of Rice Research (Formerly DRR Directorate of Rice Research, and All-India Coordinated Rice Improvement Project), Hyderabad 500,030 India; DUS: Distinctness, uniformity and stability; SSR: Simple sequence repeats; RM: Rice marker; 2AP: 2-Acetyl-1-pyrroline; GC–MS: Gas chromatography–mass spectrometry; SES: Standard evaluation system; Bl: Blast; BLB: Bacterial leaf blight; BPH: Brown planhopper; WBPH: White-backed planthopper; GM: Gall midge.

References
APEDA. Agricultural & processed food products export development authority, Government of India, New Delhi. 2020. https://apeda.gov.in/apedawebsite/six_head_product/cereal.htm. Accessed 11 July 2020.

Ahmed MR, Islam MM, Emon RM, Rana MS, Haque MS, Nuruzzaman M. DNA Fingerprinting and diversity analysis of some *aus* rice landraces. Indian J Sci Technol. 2019;12:28.

Akagi H, Yokozeki Y, Inagaki A. Highly polymorphic microsatellites of rice consist of AT repeats, and a classification of closely related cultivars with these microsatellite loci. Theo Appl Genet. 1997;94:61–7.

Aljumaili et al. Genetic Diversity of Aromatic Rice Germplasm Revealed By SSR Markers. 2018; 7658032. https://doi.org/10.1155/2018/7658032

Arikit S, Wanchana S, Khanthong S, Saensuk C, Thianthavon T, Vanavichit A, Toojinda T. *QTL-Seq* identifies cooked grain elongation QTLs near soluble starch synthase and starch branching enzymes in rice (*Oryza sativa* L.). Sci Rept. 2019;9

Arroyo SEJ, Seo H-S. Effects of the type of reference scale on descriptive sensory analysis of cooked rice: Universal aromatic scale versus rice aromatic scale. J Sens Stud. 2017;32:e12295. https://doi.org/10.1111/joss.12295.

Bhishagratna K. An English translation of the *Sushruta Samhita*, based on original Sanskrit text. Calcutta 1907. Calcutta, India.

Bradbury LM, Fitgerald TL, Henry RJ, Jin Q, Waters DLE. The gene for fragrance in rice. Plant Biotech J. 2005;3:363–70.

Bradbury L, Gillies S, Brushett D, Waters D, Henry R. Inactivation of an amino-aldehyde dehydrogenase is responsible for fragrance in rice. Plant Mol Biol. 2008;68(4–5):439–49.

Brown AHD. Core collections: a practical approach to genetic resources management. Genome. 1989;1989(31):818–24.

Cagampang GB, Perez CM, Juliano BO. A gel consistency test for eating quality of rice. J Sci Food Agric. 1973;1973(24):1589–94. https://doi.org/10.1002/jsfa.2740241214.

Cho YG, Ishii T, Temnykh S, Chen X, Lipovich L, McCouch SR, Park WD, Ayers N, Cartinhour S. Diversity of microsatellites derived from genomic libraries and gene bank sequences in rice (Oryza sativa L). Theor App Genet. 2000;100:713–22.

Choi S, Seo HS, Lee KR, Lee S, Lee J, Lee J. Effect of milling and longterm storage on volatiles of black rice (*Oryza sativa* L.) determined by headspace solid-phase microextraction with gas chromatography-mass spectrometry. Food Chem. 2019;276:572–82.

Choudhary G, Ranjitkumar N, Surapaneni M, Deborah DA, Vipparla A, Anuradha G, Siddiq EA, Reddy VL. Molecular genetic diversity of major Indian

Prasad *et al. CABI Agric Biosci*   (2020) 1:13

Page 23 of 24

rice cultivars over decadal periods. PLoS ONE. 2013;8(6):e66197. https://doi.org/10.1371/journal.pone.0066197.

Choudhury PR, Kohli S, Srinivasan K, Mohapatra T, Sharma RP. Identification and classification of aromatic rices based on DNA fingerprinting. Euphytica. 2001;118(3):243–51.

Civan P, Ali S, Batista-Navarro R, Drosou K, Ihejieto C, Chakraborty D, Ray A, Gladieux P, Brown TA. Origin of the aromatic group of cultivated rice (Oryza sativa L) traced to the Indian subcontinent. Genome Biol Evol. 2020;11(3):832–43. https://doi.org/10.1093/gbe/evz039.

Civan P, Craig H, Cox CJ, Brown TA. Three geographically separate domestications of Asian rice. Nat Plants. 2015;1:15164.

FAO. Genetic diversity in rice. 2008. https://www.fao.org/3/y4751e/y4751e0b.htm. Accessed 11 July 2020.

Franco-Duran J, Crossa J, Chen J, Hearne SJ. The impact of sample selection strategies on genetic diversity and representativeness in germplasm bank collections. BMC Plant Biol. 2019;19:520. https://doi.org/10.1186/s12870-019-2142-y.

Félix MA, Barkoulas M. Pervasive robustness in biological systems. Nat Rev Genet. 2015;16:483–96. https://doi.org/10.1038/nrg3949.

Glaszmann JC. Isozymes and classification of Asian rice varieties. Theore Appl Genet. 1987;74:21–30.

Gower JC. A general coefficient of similarity and some of its properties. Biometrics. 1971;27:857–72.

Ghosh P, Roychoudhury A. Differential regulation of genes associated with aroma production in *indica* rice cultivars during grain developmental stages. Vegetos. 2020; 33: 2

Hashemi FSG, Rafii MY, Ismail MR, Mahmud TMM, Rahim HA, Asfaliza R, et al. Biochemical, genetic and molecular advances of fragrance characteristics in rice. Crit Revu Plant Sci. 2013;32(6):445–57.

Hu X, Lu L, Guo Z, Zhu Z. Volatile compounds, affecting factors and evaluation methods for rice aroma: a review. Trends Food Sci Technol. 2020;97:136–46.

Hu J, Zhu J, Xu XM. Methods of constructing core collections by stepwise clustering with three sampling strategies based on the genotypic values of crops. Theor Appl Genet. 2000;101:264–8.

ICAR-IIRR. Laboratory manual on rice grain quality: physico-chemical & molecular procedures. ICAR-Indian Institute of Rice Research (Directorate of Rice Research). India: Hyderabad; 2009. p. 29.

IRRI. Standard evaluation system for rice. International Rice Research Institute. Manila: Philippines; 2002. p. 56.

Ishii T, Xu Y, McCouch SR. Nuclear- and chloroplast microsatellite variation in A-genome species of rice. Genome. 2001;44:658–66.

Jackson MT. Conservation of rice genetic resources - the role of the International Rice Genebank at IRRI. Plant Mol Biol. 1997;35:61–7.

Jain S, Jain KR, McCouch SR. Genetic analysis of Indian aromatic and quality rice (Oryza sativa L.) germplasm using panels of fluorescently-labeled microsatellite markers. Theor Appl Genet. 2004;109:965–77.

Jambhulkar NN, Patra BC, Reddy JN, Sarkar RK. Developing mini core of rice germplasm for submergence tolerance. Indian J Plant Genet Res. 2018;31:89–96.

Juliano BO, Perez CM, Blakeney AB, Castillo T, Kongseree N, Laignelet B, Lapis ET, Murthy VVS, Paule CM, Webb BD. International cooperative testing on the amylose content of milled rice. Starch. 1981;33:157–62. https://doi.org/10.1002/star.19810330504.

Kapoor RKA, Chettri PK, Gopi R, Kalita H, Rao DS, Babu VR. Morphometric, physico-chemical and micronutrient characterization of rice (Oryza sativa l) landraces of Sikkim Himalayas. Indian J Plant Genet Res. 2019;32:327–39.

Khush GS. Taxonomy and origin of rice. In: Singh RK, Singh US, Khush GS, editors. Aromatic rices. Oxford & IBH Publishing, New Delhi. 2000; p 5–13.

Kim KW, Chung HK, Cho GT, Ma KH, Chandrabalan D, Gwag JG, Kim TS, Cho EG, Park YJ. PowerCore: a program applying the advanced M strategy with a heuristic search for establishing core sets. Bioinformatics. 2007;23(16):2155–62.

Leenakumary S, Genetic diversity, characterization and evaluation of medicinal rice in Kerala. In: Compendium of papers of science society interface on medicinal and aromatic rice, 20–21 March 2004, (mimeographed). Pattambi, Palakkad, Kerala, India: Regional Agricultural Station; 2004.

Liu K, Muse SV. PowerMarker: an integrated analysis environment for genetic marker analysis. Bioinformatics. 2005;21:2128–9.

Losos JB, Arnold SJ, Bejerano G, Brodie ED III, Hibbett D, Hoekstra HE, et al. Evolutionary biology for the 21st Century. PLoS Biol. 2013;11(1):e1001466.

McCouch SR, Chen X, Panaud O, Temnykh S, Xu Y, Cho YG, Huang N, Ishii T, Blair M. Microsatellite marker development, mapping and applications in rice genetics and breeding. Plant Mol Biol. 1997;35:89–99.

McCouch SR, Teytelman L, Xu Y, Lobos KB, Clare K, Walton M, Fu B, Maghirang R, Li Z, Xing Y, Zhang Q, Kono I, Yano M, Fjellstrom R, DeClerck G, Schneider D, Cartinhour S, Ware D, Stein L. Development and mapping of 2240 new SSR markers for rice (Oryza sativa L.). DNA Res. 2002;9(6):199–207.

Muralidharan K, Prasad GSV, Rao CS. Breeding for rice improvement, where do we stand? Curr Sci. 1996;71:438–48.

Muralidharan K, Siddiq EA. Agriculture in India - From 10000 BC to 2000 AD. National Agricultural Sciences Museum. p 328. New Delhi: Indian Council of Agricultural Research; 2000. p. 110001.

Nachimuthu VV, Muthurajan R, Duraialaguraja S, Sivakami R, Pandian BA, Ponniah G, Gunasekaran K, Swaminathan M, Suji KK, Sabariappan R. Analysis of population structure and genetic diversity in rice germplasm using SSR markers: an initiative towards association mapping of agronomic traits in *Oryza sativa*. Rice. 2015;8:30. https://doi.org/10.1186/s12284-015-0062-5.

Nagaraju J, Kathirvel M, Rameshkumar R, Siddiq EA, Hasnain SE. Genetic analysis of traditional and evolved Basmati and non-Basmati rice varieties by using fluorescence-based ISSR-PCR and SSR markers. Proc Natl Acad Sci USA. 2002;99:5836–41.

Narshimulu G, Jamaloddin M, Vemireddy LR, Anuradha G, Siddiq EA. Potentiality of evenly distributed hypervariable microsatellite markers in marker assisted breeding of rice. Plant Breed. 2011;130:314–20.

Ni J, Colowit PM, Mackill DJ. Evaluation of genetic diversity in rice subspecies using microsatellite markers. Crop Sci. 2002;42:601–7.

Ospina R, Marmolejo-Ramos F. Performance of some estimators of relative variability. Front Appl Math Stat. 2019. https://doi.org/10.3389/fams.2019.00043.

Pankaj KS, Murali NM, Dipak KH, Med RV. Genetic divergence in lowland rice of north eastern region of India. Comm Bio Crop Sci. 2006;1:35–40.

Pathaichindachote W, Panyawut N, Sikaewtung K, Patarapuwadol S, Muangprom A. Genetic diversity and allelic frequency of selected Thai and exotic rice germplasm using SSR markers. Rice Sci. 2019;26:393–403.

Peddamma SK, Ragichedu PK, Maddala S, Rao DS, Lella VSR, Konne K, Sripada P, Krishnan SGM, Singh AK, Maganti SM. Insight of aroma in brown rice through chemical assessment of 2-acetyl-1-pyrroline (2AP) in aromatic germplasm of India. Cereal Chem. 2018;95:679–88.

Plaschke J, Ganal MW, Röder MS. Detection of genetic diversity in closely related bread wheat using microsatellite markers. Theor Appl Genet. 1995;91:1001–7. https://doi.org/10.1007/BF00223912.

Prasad GSV, Muralidharan K, Rao CS. Stability and yield performance of genotypes, proposal or re-grouping world rice area into mega environments. Curr Sci. 2001;81:1337–466.

Qian TW, He KM. Utilization of exotic rice germplasm resources in Guang-dong province. Crop Genet Resour. 1991;2:36–7.

Ramaiah K. Grain classification. In: Rice Research in India. 1985; p 629. Indian Council of Agricultural Research, New Delhi 110001.

Rani, NS, Prasad GSV, Rao LVS, Sudharshan I, Pandey M, Babu R, Ram T, Padmavati G, Muralidharan K, Pasalu IC, Viraktamath BC. High yielding rice varieties in India. Tech. Bull. 2008; p 188. ICAR-Indian Institute of Rice Research (Directorate of Rice Research), Hyderabad, India.

Rani NS, Rao LVS, Viraktamath BC. National guidelines for the conduct of distinctness, uniformity and stability: rice. 2006; p 1–39. ICAR-Indian Institute of Rice Research (Directorate of Rice Research), Hyderabad, India.

Rao TP, Gomathinayagam P. Genetic diversity in semidry rice under different environments. Madras Agr J. 1997;84:314–7.

Rao ES, Kadirvel P, Symonds RC, Geethanjali S, Ebert AW. Using SSR markers to map genetic diversity and population structure of *Solanum pimpinellifolium* for development of a core collection. Plant Genet Resour. 2012;10:38–48.

Roldan-Ruiz I, van Euwijk FA, Gilliland TJ, Dubreuil P, Dillmann C, Lallemand J, de Loose M, Baril CP. A comparative study of molecular and morphological methods of describing relationships between perennial ryegrass (*Lolium perenne* L.) varieties. Theor Appl Genet. 2001;103:1138–50.

Roy S, Banerjee A, Mawkhlieng B, Misra AK, Pattanayak A, Harish GD, Singh SK, Ngachan SV, Bansal KC. Genetic diversity and population structure in aromatic and quality rice (Oryza sativa L.) landraces from north-eastern India. PLoS ONE. 2015;10(6):e0129607.

Prasad *et al. CABI Agric Biosci*   (2020) 1:13

Page 24 of 24

Russell J, Fuller J, Young G, Thomas B, Taramino G, Macaulay M, Waugh R, Powell W. Discriminating between barley genotypes using microsatellite markers. Genome. 1997;40:442–50.

Saheewala H, Sharma VK, Shahi VK. Molecular characterization and genetic divergence analysis of traditional and improved aromatic rice varieties using microsatellite markers. Indian J Plant Genet Res. 2019;32:377–84.

Saini P, Francies RM, Bastian D, Vigneshwaran V. Genetic assessment of core set developed from short duration rice accessions held by Kerala Agricultural University. J Trop Agric. 2013;51:126–31.

Sakthivel K, Rani NS, Pandey MK, et al. Development of a simple functional marker for fragrance in rice and its validation in Indian Basmati and non-Basmati fragrant rice varieties. Mol Breed. 2009;24:185–90. https://doi.org/10.1007/s11032-009-9283-x.

Shannon CE, Weaver W, The mathematical theory of communication. University of Illinois Press. IL: Urbana; 1949.

Singh RK, Sharma RK, Singh AK, Singh VP, Singh NK, Tiwari SP, Mohapatra T. Suitability of mapped sequence tagged microsatellite site markers for establishing distinctness, uniformity and stability in aromatic rice. Euphytica. 2004;135:135–43.

Singh R, Singh A, Sharma T, Singh A, Singh N. Fine mapping of aroma QTLs in Basmati rice (*Oryza sativa* L.) on chromosomes 3, 4 and 8. J Plant Biochem Biotech. 2007. https://doi.org/10.1007/BF03321978.

Singh RK, Singh US, editors. A treatise on the scented rices of India. 2003. New Delhi: Kalyani Publishers, India.

Singh V, Singh AK, Mohapatra T, Krishnan G, Ellur RK. Pusa Basmati 1121 – a rice variety with exceptional kernel elongation and volume expansion after cooking, Rice. 2018;11–19.

Sivaranjani AKP, Pandey MK, Sudharshan I, Ramkumar G, Madhav MS, Sundaram RM, Varaprasad GS, Rani NS. Assessment of genetic diversity among basmati and non-basmati aromatic rices of India using SSR markers. Curr Sci. 2010;99(2):221–6.

Song J-M, Arif M, Zhang M, Sze S-H, Zhang H-B. Phenotypic and molecular dissection of grain quality using the USDA rice mini-core collection. Food Chem. 2019;284:312–22.

Sow M, Ndjiondjop M-N, Sido A, Mariac C, Laing M, Bezançon G. Genetic diversity, population structure and differentiation of rice species from Niger and their potential for rice genetic resources conservation and enhancement. Genet Resour Crop Evol. 2014;61(1):199–21313.

Struss D, Plieske J. The use of microsatellite markers for detection of genetic diversity in barley populations. Theor Appl Genet. 1998;97:308–15. https://doi.org/10.1007/s001220050900.

Talukdar PR, Rathi S, Pathak K, Chetia S, Baruah AR, Sarma RN. QTL analysis in aromatic rice of Assam. India Rice Res. 2017;5(4):1–7.

Tautz D. Hypervariability of simple sequences as a general source for polymorphic DNA markers. Nucleic Acids Res. 1989;7(16):6463–71.

UN. World population day 11 July 2020, United Nations. https://www.un.org/en/observances/world-population-day). Accessed 11 July 2020

Valiathan MS. The legacy of Charaka. 2003. Orient Longman. ISBN 81–250–2505–7

Vanavichit A, Kamolsukyeunyong W, Siangliw M, Siangliw JL, Traprab S, Ruengphayak S, Chaichoompu E, Saensuk C, Phuvanartnarubal E, Toojinda E, Tragoonrung S. Thai HomMali rice: Origin and breeding for subsistence rainfed lowland rice system. Rice. 2018;11:20. https://doi.org/10.1186/s12284-018-0212-7.

Wakte K, Zanan R, Hinge V, Khandagale K, Nadaf A, Henry R. Thirty-three years of 2-acetyl-1-pyrroline, a principal Basmati aroma compound in scented rice (*Oryza sativa* L.): A status review. J Sci Food Agric. 2017;97(2):384–95.

Wang J, Guan Y, Wang Y, Zhu L, Wang Q. A strategy for finding the optimal scale of plant core collection based on Monte Carlo simulation. Sci World J. 2014. https://doi.org/10.1155/2014/503473.

Wei H, Tianhao Z, Pengfei W, Bo W, Jiaming S, Zhongmin H, Lingling C, Kede L, Yongzhong X. Development of whole-genome agarose-resolvable lindel markers in rice. Rice. 2020;13:1. https://doi.org/10.1186/s12284-019-0361-3.

Wilkie K, Wootton M, Paton JE. Sensory testing of Australian fragrant, imported fragrant, and non-fragrant rice aroma. Int J Food Prop. 2004;7(1):27–36.

Wong SC, Yiu PH, Bong STW, Lee HH, Neoh PNP, Rajan A. Analysis of Sarawak barrio rice diversity using microsatellite markers. Am J Agri Biol Sci. 2009;4(4):298–304.

Worldmeters. Current world population, July 2020. https://www.worldometers.info/world-population/ Accessed 11 July 2020

Wujastyk D. The roots of Ayurveda. 2003. London etc.: Penguin. ISBN 0–14–044824–1.

Xie L-H, Tang SQ, Wei X-J, Jiao G-A, Sheng Z-H, Hu P-S. An optimized analysis of 2-acetyl-1-pyrroline content diversity in the core collection of aromatic rice germplasm (*Oryza sativa* L.). Cereal Chem. 2019;96:698–707.

Yadav S, Singh A, Singh MR, Goel N, Vinod KK, Mohapatra T, Singh AK. Assessment of genetic diversity in Indian rice germplasm (*Oryza sativa* L.): use of random versus trait-linked microsatellite markers. J Genet. 2013;92:545–57.

Yu H, Xie T, Xie J, Ai L, Tian H. Characterization of key aroma compounds in Chinese rice wine using gas chromatography-mass spectrometry and gas chromatography-olfactometry. Food Chem. 2019;293:8–14.

Zhao K, Tung C, Eizenga G, et al. Genome-wide association mapping reveals a rich genetic architecture of complex traits in *Oryza sativa*. Nat Commun. 2011;2:467. https://doi.org/10.1038/ncomms1467.

Zheng K, Subudhi PK, Domingo J, Magpantay G, Huang N. Rapid DNA isolation for marker assisted selection in rice breeding. Rice Genet Newsl. 1995;12:255–8.

Publisher's Note