KASP™ based markers reveal a population sub-structure in temperate rice (Oryza sativa L.) germplasm and local landraces grown in the Kashmir valley, north-western Himalayas

Shikari, Asif Bashir ; Najeeb, Sofi; Khan, Gazala; Mohidin, Fyaz A.; Shah, Ashaq H.; Nehvi, Firdous A.; Wani, Shafiq A.; Bhat, Nazir A.; Waza, Showkat A.; Subba Roa, L.V. ; Steele, Katherine; Witcombe, John

Genetic Resources and Crop Evolution

DOI:
10.1007/s10722-020-01025-z

Published: 01/03/2021

Peer reviewed version

KASP^{TM} based markers reveal a population sub-structure in temperate rice germplasm and local landraces grown in the Kashmir valley, north-western Himalayas

[1]Asif Bashir Shikari, [1]Sofi Najeeb, [1]Gazala Khan, [1]Fayaz A. Mohidin, [1]Ashaq H. Shah, [2]Firdous A. Nehvi, [2]Shafiq A. Wani, [1]Nazir A. Bhat, [1]Showkat A. Waza, [3]L.V. Subba Rao, [4]Katherine A. Steele and [4]John R. Witcombe

[1]Mountain Research Centre for Field Crops, [2]Division of Plant Biotechnology,
Sher-e-Kashmir University of Agricultural Sciences & Technology of Kashmir, J&K, India, 192 102
[3]ICAR-Indian Institute of Rice Research, Hyderabad, India, 500 030
[4]Bangor University, U.K.

e-mail: asifshikari@skuastkashmir.ac.in

ORCID: 0000-0002-2911-5536

Name	Affiliation	e-mail
Asif Bashir Shikari (ABS)	*Mountain Research Centre for Field Crops, Sher-e-Kashmir University of Agricultural Sciences & Technology of Kashmir, J&K, India, 192 102*	asifshikari@skuastkashmir.ac.in
Sofi Najeeb (SN)	*Mountain Research Centre for Field Crops, Sher-e-Kashmir University of Agricultural Sciences & Technology of Kashmir, J&K, India, 192 102*	najeeb_sofi@rediffmail.com
Gazala Khan (GK)	*Mountain Research Centre for Field Crops, Sher-e-Kashmir University of Agricultural Sciences & Technology of Kashmir, J&K, India, 192 102*	moazin_khan@yahoo.co.in
Fayaz A. Mohidin (FAM)	*Mountain Research Centre for Field Crops, Sher-e-Kashmir University of Agricultural Sciences & Technology of Kashmir, J&K, India, 192 102*	famohiddin@rediffmail.com
Ashaq H. Shah (AS)	*Mountain Research Centre for Field Crops, Sher-e-Kashmir University of Agricultural Sciences & Technology of Kashmir, J&K, India, 192 102*	ahshah71@gmail.com
Firdous A. Nehvi (FAN)	*Division of Plant Biotechnology, Sher-e-Kashmir University of Agricultural Sciences & Technology of Kashmir, J&K, India, 192 102*	f.nehvi@rediffmail.com
Shafiq A. Wani (SAW)	*Division of Plant Biotechnology, Sher-e-Kashmir University of Agricultural Sciences & Technology of Kashmir, J&K, India, 192 102*	shafiqawani@gmail.com
Nazir A. Bhat (NAB)	*Mountain Research Centre for Field Crops, Sher-e-Kashmir University of Agricultural Sciences & Technology of Kashmir, J&K, India, 192 102*	nazirpathology@gmail.com
Showkat A. Waza (SW)	*Mountain Research Centre for Field Crops, Sher-e-Kashmir University of Agricultural Sciences & Technology of Kashmir, J&K, India, 192 102*	sahmad777@gmail.com
L. V. Subba Rao (LV)	*ICAR-Indian Institute of Rice Research, Hyderabad, India, 500 030*	lvsubbarao1990@gmail.com
Katherine A. Steele (KAS)	*Bangor University, U.K.*	k.a.steele@bangor.ac.uk
John R. Witcombe (JRW)	*Bangor University, U.K.*	j.r.witcombe@bangor.ac.uk

Abstract

The conservation and utilization of germplasm is contingent on its proper characterization at morphological or molecular levels. The present study aimed to elucidate the population sub-structure of 470 temperate rice germplasm collections of the Kashmir Valley. Analysis was carried out using KASP (Kompetitive Allele Specific PCR) assay on 213 genomic loci. Of these, a restricted set of 114 KASP loci were selected by the elimination of redundant, i.e. tightly linked markers based on map positions. STRUCTURE grouping was carried out to reveal three distinct sub-populations, K1, K2 and K3 comprising of 209, 156 and 105 germplasm accessions, respectively. Population F_{ST} values for K1, K2 and K3 were at 0.60, 0.24, 0.69, respectively, with highest pair-wise F_{ST} obtained between K2-K3 (0.53). Analysis using the restricted set of 114 markers gave a better inferred membership with a low average admixture of 15.1% compared with 22.6% based on the whole marker set. An improved agreement between STRUCTURE grouping and principal coordinate analysis was reached using the restricted marker set. Φ_{ST} values calculated based on nucleotide diversity also suggested three sub-populations: K2, mostly indica germplasm; K1 mostly exotic temperate japonica; and K3, local japonica varieties and landraces. Polymorphic SNPs and haplotypes were discovered which discriminated the three sub-populations. Fifteen KASP markers were most important in discriminating K2 from K1 and K3 and included SNPs associated with domestication within the *Wx*, *Ghd7* and *Ghd8* genes. KASP markers are cheaper than SSR markers. Some of the KASP markers were highly discriminatory, using both model and distance based approaches, and so can be used as a cost-effective tool for efficient maintenance and use of rice genetic resources.

Key words: Rice, temperate, Structure, Diversity, SNPs, KASP

Introduction

The enormous range of diversity in cultivated rice (*Oryza sativa* L.) is represented by more than 120,000 varieties worldwide (Khush 1997, Vasudevan et al. 2014) including cultivars and large number of landraces, with around 50,000 of them present alone in India. Globally, the crop gene banks preserve and maintain a round 250,000 rice germplasm accessions which include cultivated types and their wild relatives (Jacob, et al., 2015). However, 95% of these valuable gene repositories have rarely been used in any breeding programme (Peng et al. 2009). Initially, the mapping of large number of microsatellites (SSR) markers (McCouch et al. 2002; Temnykh et al. 2001) and more recent genome saturation through discovery of SNPs (McNally et al. 2009; Singh et al. 2015; Trinh et al. 2017; Trung et al. 2017; Zhou et al. 2011) has helped to delineate genetic diversity and enabled a much better coverage of rice genome and the underlying trait association. Germplasm utilization activity itself depends upon the preliminary characterization and description of germplasm at population and individual level. The natural genetic variation in landraces preserved in gene banks (Diez et al. 2018) or in *ex situ* germplasm repositories (Vanniarajan et al. 2012) has been recently evaluated through the use of molecular markers.

The Himalayan tract represents a diversity hot spot of rice and is a home to tens of thousands of landraces. In India, temperate rice is grown in the North-Western Himalayan region (comprising Jammu and Kashmir, Himachal, and Uttrakhand) and North-Eastern hill states. The natural diversity within some areas of this broad region has been recently studied with the help of morphological and molecular markers (Choudhury et al. 2013; Roy et al. 2015; Salgotra et al. 2015; Umakanth et al. 2017). These studies have contributed to the evolutionary classification of rice in mountainous areas and has also helped to quantify the allelic diversity in populations. On the other hand, information on the detailed population structure and classification of high altitude rice from Kashmir is lacking. The valley of Kashmir located at 34^{o}N and 73^{o}E form the northern-most part where rice cultivation extends from an altitude 1500 to over 2200 m and is characterised by landraces and varieties having excellent resilience to cold stress (Parray and Shikari 2008). From a breeding point of view, introduced materials with early maturity and a certain degree of cold tolerance have been successfully utilized for trait improvement (Shikari et al. 2018). The present study was designed to assess the population structure of germplasm adapted to high altitude temperate ecology of Kashmir valley with the help of SNP genotyping using KASP assay. KASP is a homogeneous, uniplex, fluorescence-based genotyping technology based on allele-specific oligo extension and fluorescence resonance energy transfer (FRET) for signal generation. KASP has been reported to show improved cost-effectiveness and reliability as compared to some of the contemporary sequence-based markers (Semagn et al. 2014; Steele et al. 2018).

Materials and methods

Plant materials

We studied 470 rice (*Oryza sativa* L.) germplasm entries of diverse origin which are being maintained at the Mountain Research Centre for Field Crops (MRCFC), Khudwani, SKUAST-Kashmir. The repository comprised of 39 local landraces from Kashmir valley, 4 obsolete cultivars, 27 released varieties, 117 indigenous types (from parts of India other than Kashmir) and 63 exotic collections, as well as 220 advance breeding lines/ derivatives. The landraces and varieties included here mostly were of short round to medium slender grain type and belonged to japonica and indica ecotypes. The accessions designated as 'exotic' were those which have been procured / collected from sources other than belonging to Indian sub-continent. Over the years lines have undergone a process of purification, adaptive selection and acclimatization through generations of maintenance and evaluation (Supplementary Table S1).

SNP assay and Genotyping

Leaf samples were collected using a 96 well format 'Plant Sample Collection Kit' (Biosearch Technologies, Hoddesdon, Herts., U.K.) and subjected to genotyping. Each germplasm line was genotyped with KASP markers at 217 well-distributed genomic loci. Four of the loci with poor genotyping calls were dropped and thus 213 SNPs were pursued for analysis. The KASP markers selected here were designed previously using *indica* varieties and the *indica* reference genome and a large proportion of the loci are located within genes (Steele et al. 2018).

Population structure analysis

A Bayesian model-based clustering approach was implemented using the STRUCTURE v2.3.4 software (Pritchard et al. 2000) in order to define population sub-clustering across 470 germplasm accessions. STRUCTURE performs Bayesian assignment of individuals to a predefined number of K assumed sub-populations. An optimum number of sub-populations were inferred from the software, pre-set at admixture ancestry model with correlated allelic frequencies. SNP data was analysed at three replicate runs per K value, a burn-in period of 50000 and Markov Chain Monte Carlo (MCMC) simulations of 100000. MCMC process randomly assigns individuals to a pre-determined number of K groups followed by estimation of variant frequencies and re-designation of groups. The ideal K value was determined by using adhoc ΔK based on the rate of change in the log probability of data between successive K values (Evanno et al. 2005).

Estimation of diversity statistics

Fixation index is a measure of the reduction in heterozygosity or allele sharing at any one level of a population hierarchy relative to another more inclusive level (Weir and Cockerham 1984; Weir and Hill 2002). F-statistics such as inbreeding coefficient (F_{IS}), Fixation index (F_{ST}), and the pairwise F_{ST}, were computed using GenAlEx 6.5. F_{IS} measures the extent of genetic inbreeding within subpopulations and is defined as the mean reduction in heterozygosity of an individual due to non-random mating within a sub-population. F_{IS} can range from –1.0 (all individuals heterozygous) to +1.0 (no observed heterozygotes). F_{ST} measures the extent of genetic differentiation among subpopulations and is defined as the mean reduction in heterozygosity of a subpopulation (relative to the total population) due to genetic drift among subpopulations. F_{ST} can range from 0.0 (no differentiation) to 1.0 (complete differentiation where subpopulations happen to be fixed for different alleles). Further, the parameter $\Phi_{ST} = (\pi_T - \pi_S)/\pi_T$), was calculated and provides an estimate of population differentiation based on nucleotide diversity (Excoffier et al. 1992). Here, π_T and π_S are analogous to H_T and $H_{S,}$ described above, and reflect nucleotide diversity. The Simple matching coefficients (Sokal and Michener 1958) based distance matrix was generated that was utilized for neighbour joining method of clustering (Saitou and Nei 1987) with the help of MEGA X (Kumar et al. 2018). An AMOVA (Analysis of molecular variance) (Peakall et al. 2003) was carried out using the GenAlEx 6.5 software (Peakall and Smouse 2012). It was done with 9999 permutations. The same program was used to carry out principal coordinate analysis across genotypic data. Mean genetic diversity (h) was calculated for each sub-population and was expressed as: $h = [1/m(1 - \sum_{i=1}^{n} pi^2)$, where, m is the number of marker loci, n is the number of individuals in a population, pi is the allelic frequency. Both h and number of effective alleles ($Ne = 1/ \sum_{i=1}^{n} pi^2$; with p_i as the allelic frequency) were worked out using Power Marker V3.0 software.

Restricted marker analysis

After the STRUCTURE analysis was drawn with the help of 213 KASP markers, a sub-set of 114 markers was chosen to repeat the estimation of population parameters. The marker sub-set was chosen after elimination of redundant markers occupying same loci. The purpose was to reverse the overrepresentation of certain chromosomal segments.

Secondly, those eliminated were mostly linked to functional genes related to biotic and abiotic stress tolerance and it avoided the clustering arising mainly from variability in such genes.

Results

Assessment of population sub-structure

A set of 470 rice germplasm accessions were investigated for various population parameters using 213 genome wide SNP markers spotted through KASP technology. STRUCTURE, a program based on Bayesian model was used to define population structure and yielded highest log likelihood estimate and peak ΔK value of 162.79 at K = 4, which suggested classification into four sub-populations. The four sub-populations were named as K1, K2, K3 and K4 (after *Khudwani*; location of our Research Centre) and turned out with an allocation of 84 (17.87%), 128 (27.23%), 76 (16.17%) and 182 (38.72%) genotypes, respectively (Supplementary Table S2). AMOVA (analysis of molecular variance) revealed that Φ_{PT}, an estimate of population genetic differentiation, was equal to 53% of the total molecular variance confirming a significant population structure. K2 versus K3 and K1 against K2, recorded highest pairwise Φ_{PT} of 0.647 and 0.616, respectively. The variability feature was further explained using three-tiered diversity parameters: H_I (mean observed heterozygosity per individual within subpopulations), H_S (mean expected heterozygosity within subpopulations) and H_T (expected heterozygosity in total population), which were subsequently used in the determination of population F-statistics. An important diversity parameter, F_{ST} was calculated for individual populations and appeared in an order: K3 (0.5947) > K1 (0.4875) > K4 (0.2943) > K2 (0.2213), thereby suggesting strong genetic sub-structure. In line with Φ_{PT} values mentioned above, highest pair-wise F_{ST} values were recorded for K2 – K3 (0.4651), K1 – K2 (0.4184) and K2 – K4 (0.3263) comparisons, therefore, explained discernible population differentiation. Principal Coordinates (PCs) were drawn on the data matrix with PC1 and PC2 explaining 53.15 and 5.57% of total variance with corresponding eigen values of 396.1 and 41.5, respectively. In addition to strong signal for admixture as was revealed by STRUCTURE based grouping, further it did not correlate with the pattern depicted by PCoA. Although, K2 plotted separately on negative PC1 axis against K1 and K3 which clustered together within a narrow factor range on positive axis of PC1 with a limited spread of -0.276 to 0.227 on PC2.

Restricted marker analysis: In an attempt to refine our population estimates, only a sub-set of KASP markers was chosen from a whole set of 213 (see Material and Methods). Analysis using limited marker set lead us to harvest only three sub-populations (instead of four) named K1, K2 and K3 (Table 1, Supplementary Fig. S1) and accommodated

209, 156 and 105 germplasm accessions, respectively. F_{ST} values recorded for the three populations stood at 0.60, 0.24, 0.69, respectively (Table 2). Highest pair-wise F_{ST} was obtained between K2-K3 (0.531) followed by K1-K2 (0.467) and lowest for K1-K3 (0.127) (Table 3). Ancestral relations were deepened through restrictive marker analysis with low average admixture levels of 15.1% compared with 22.6% on whole marker set. Individual sub-populations K1, K2 and K3 had 15.3%, 7.1% and 22.9% individuals with overlapping ancestry (Supplementary Table S3, S4, Fig. 1).

Principal Coordinate Analysis

The first two principal coordinates marked eigen values of 207.9 and 21.3 and explained cumulative variance of 58.7% (Supplementary Table S5). The grouping based on STRUCTURE and Principal Coordinate Analysis (PCoA) was observed to follow a similar pattern under restricted marker analysis. Individuals in K2 clustered on negative PC-1 in contrast to K1 and K3 those appeared in proximity along PC-1 with positive loadings (Fig. 2). The PCoA grouping corresponded well with pair-wise F_{ST} values among the three sub-populations.

Gene diversity

A statistic, Φ_{ST} is a measure of population differentiation based on nucleotide diversity and was equal to 0.6795 (K1), 0.2915 (K2) and 0.7665 (K3) bearing a similar trend as that for Φ_{ST}. Sub-populations K1 to K3 recorded unbiased mean diversity (uh) estimates of 0.13 (K1), 0.25 (K2) and 0.10 (K3). The number of alleles per locus for a bi-allellic SNP marker has to be two in every case and as such N_e (number of effective alleles) were 1.14 (K1), 1.33 (K2) and 1.11 (K3) (Table 2). As regards the nature of marker polymorphism, the information on frequency of transversions was notably found to discriminate the sub-populations with respective values of 16.39% (K2), 7.58% (K3) and 7.04% (K1) across populations. Coefficients of Nei's Genetic identity among populations were highest (0.663) between K1 – K3 and lowest (0.000) between K2 – K3. These values corresponded to the relationship explained by pairwise Φ_{ST} coefficients and the results of the PCoA.

The Neighbour Joining method based on Simple Match Coefficients was applied to estimate the pattern of genetic divergence and clearly defined two major clusters at a molecular distance of around 0.50 (Fig. 3). Most of the japonica grouped into cluster-I and those of indica represented cluster-II. Out of a total of 470, Cluster-I and Cluster-II included 313 and 157 accessions, respectively. Cluster-I was further partitioned into two sub-clusters, Cluster-Ia and Cluster-Ib comprising of 166 and 147 accessions, respectively. Overall the individuals were categorized at an average divergence coefficient of 0.32. The highest distance coefficients were recorded between genotype *GS-592* against *Pusa Sugandh 3* (0.78) and HPR-2373 (0.77). Thirty nine local landraces originating from altitudinal range of 1500 to 2300 msl grouped closely within cluster-Ia (33) and cluster-Ib (4). Two other landraces *Yemberzul* and GS-23 appeared in cluster-II. The landraces with red pericarp namely, *Tangdhar Zag* and *Karnah Zag* and popular aromatic landraces, *Kamad* and *Mushk Budji* occupied similar clusters. Other temperate exotic and indigenous collections with japonica background occupied cluster-I. Cluster-II featured with almost all the indica and derivative lines. The 12 released and locally adapted varieties, spread across the tree circumference. Of these, high altitude japonica varieties *K-332* and *Shalimar Rice-5* grouped into Cluster-Ia and *Barkat* and *Kohsar* in Cluster-Ib. All the eight indica varieties (*China-988*, *China-1007*, *K-39*, *Chenab*, *Jhelum*, *Shalimar Rice-1*, *Shalimar Rice-2* and *Shalimar Rice-3*) grouped in cluster-I. Fine grained *Pusa Sugandh 3* and Basmati variety *Pusa Basmati 1509* clustered mid-way indica and japonica with proximity to the accessions adapted to North-western Himalayan region on one side and japonica group at the other. Germplasm accessions representing different clusters are given in Fig. 4.

Allelic polymorphism and distribution

STRUCTURE grouping into sub-populations K1, K2 and K3 revealed a pattern in terms with distribution of germplasm into indica and japonica. K2 mostly comprised of indica germplasm, K3 of local japonica (landraces) while K1 included other japonica collections. Graphical genotypes over 114 SNP loci depicted the discriminatory alleles (Fig. 5). At least fourteen SNPs discriminated K2 from K1 and K3 including two SNPs Waxy and Amy_W2_R_1, that are both associated with the Wx locus on chromosome 6. Likewise, Ghd7_05_SNP_ff_1 and Ghd8_SNP_ff_2 showed A/T and A/G polymorphism, respectively between K2 versus both K1-K3. SNP SSII_1_SNP_ff_1 on chromosome 10 is associated with Starch synthase II and produced A/G polymorphism between K2 / K1-K3 populations. In addition, 9 other SNP markers discriminated K2 from rest two populations, K1 and K3. A GAG haplotype on chromosome 7 differentiated K2 from other two sub-populations which carried CGA at corresponding sites. Likewise, K2 had the haplotype CC at two loci, on chromosome 9 against TT for K1 and K3. Markers RM171_SNP_nn_1, RM147_SNP_nn_3 and RM590_SNP_ff_1 on chromosome 10 amplified a haplotype GAG in K1 and K3 against ACT in K2. The unique locus OsR498G0713985600_SNP_ff_1 differentiated between K1 and K3 with C/T polymorphism. (Fig. 5; Supplementary Table S6; Supplementary Fig. S2).

Near about 90% of accessions in K3 were local landraces and 83% of K1 were exotic (japonica) germplasm. Of the 220 advanced breeding lines, more than 90% were grouped into K1 and K2 which indicates that varietal breeding programmes have largely been carried through utilization of exotic and indigenous germplasm while landraces have been promoted in their original form (Table 4). Pertinently, six KASP loci largely differentiated local landraces from temperate exotic germplasm and included RM9B_SNP_nn_2; OsR498G0510120000_SNP_ff_3; ALK_SNP_ff_1; RM51_SNP_nn_2; OsR498G0713985600_SNP_ff_1 and CRG4_SNP_nn_1 (Supplementary Fig. S3).

Discussion

The restricted markers analysis procedure helped us to estimate population sub-structure among a set of 470 germplasm lines. The markers were selected to give a more uniform genome representation by elimination of redundant, i.e., tightly linked loci that were mostly trait-based markers. An average of 50% (0.51) of variability was explained by population sub-structure across 470 germplasm lines. The strong pair-wise F_{ST} values obtained between K2 and other two populations were in line with the evolutionary expectations, since K2 was mostly comprised of indica accessions and the japonica accessions were concentrated in K1 and K3. Inbreeding coefficient (F_{IS}) for all the three populations was high (> 0.88) as expected for self-fertilizing species. The close proximity of K1-K3, as evident from low pair-wise F_{ST} estimates (0.127), suggests high allele sharing between these two sub-populations. The average F_{ST} of progenitor *Oryza rufipogan* measures 0.18 against 0.55 for domesticated *O. sativa* (Huang et al. 2010). Indica are believed to have descended from Or-I (*O. rufipogan*-I) group with preservation of 75% of total genetic diversity and F_{ST} = 0.17. On the other hand japonica and aromatic rice have descended from Or-III with strong bottleneck with representation of 33% divergence and high level of population differentiation (F_{ST} =0.36) (Huang et al. 2012). These theories are supportive of the population differentiation levels for japonica (K1 and K3) and indica (K2) in our materials. Between K1 and K3, the latter contained most of the landraces and recorded a higher F_{ST} than the former. Landraces symbolize an intermediate stage of evolution between wild and cultivated germplasm. 'Inferred ancestry' on individual basis was calculated from Q-Q plots based on percentage admixtured individuals in a sub-population, i.e., where an admixture population was defined as having a greater than 15% probability of belonging to another subpopulation. . The estimate for admixture was 15.1% for 114 SNPs, while for the whole marker analysis (213 SNPs) it was 22.6%, thereby validating the usefulness of elimination of redundant markers. The highest value for admixture was in K3 (22.9%) followed by K1 (15.3%) and K2 (7.1%). High admixture levels in K3 were because of considerable genome sharing with K1 as both populations mostly represent japonica. However, local landraces of Kashmir in sub-population K3 were highly differentiated from those in K1, probably reflecting their distinct ancestry

STRUCTURE operates on assigning membership coefficients of individual samples towards sub-populations (Pritchard et al. 2009), while PCoA aligns samples along meaningful coordinates (Mohammadi and Prasanna, 2003). In our case, the two approaches showed a similar clustering pattern although the fine difference between K1-K2 was dissipated in PoCA but was clearly resolved through STRUCTURE. The population defined through STRUCTURE analysis under whole marker set differed to that produced from more uniformly distributed markers that proved to reflect a more reliable grouping. The over-representation of certain parts of genome may lead to false conclusions when estimating genetic diversity or population sub-structure. This statement is supported by the close agreement of the results from STRUCTURE grouping and the PCoA on the restricted set of markers. Secondly, the restricted marker analysis gave lower admixture levels compared with the analysis using all of the markers.

The more useful nucleotide diversity coefficients, Φ_{IS}, Φ_{ST} and Φ_{IT} were computed from the SNP data. The coefficients are analogous to F-coefficients but are not dependent on heterozygosity. The Φ_{ST} coefficients are based on the nature of SNP polymorphisms and, thereby, substantiate the presence of the population structure determined by other methods. There was a varying proportion of transversions among the sub-populations, and this pattern explains the differing Φ_{ST} of the sub-populations. While F_{ST} has been regarded as the outcome of recent sharing of alleles, Φ_{ST} is an outcome of a long evolutionary history and, therefore, possess higher values (Excoffier et al. 1992).

Estimates of genetic diversity: The high genetic diversity of sub-population K2 is an outcome of greater allele sharing in indica as compared to japonica. The level of genetic diversity happens to be low in japonica compared to indica which is in agreement with the findings of Choudhary et al. (2013). Further, the individuals in sub-population K3 mostly belong to higher hills and have a lower diversity compared with K1 and K2 which originate from plains. The diversity gradient across altitude has been mentioned by Roy et al. (2016). Since the SNPs are bi-allellic markers, N_e (number of effective alleles) was less than two in all three sub-populations. Coefficients of Nei's Genetic identity among populations were highest (0.66) between K1 – K3 and lowest (0.00) between K2 – K3. These values corresponded to the relationship explained by pairwise F_{ST} coefficients. The genetic distance (GD) was computed based on Simple Match coefficients followed by grouping through Neighbour joining principle. Two major cluster were identified at inter-cluster distance of ~0.52. Cluster-I comprised of japonica and Cluster-II with most of indica germpalsm lines. Basmati and other derived accessions grouped mid-way. In spite of the correlations that were found of GD with geographical origin, the division into indica, japonica and derivative class dominated over the clustering pattern based on geographical area. For example, the varieties bred and released for same geographical area occupied separate clusters: Shalimar Rice-1, -2, -3 (all indica from SKUAST-Kashmir) fell in Cluster-II, whereas K-78, K-332, Shalimar Rice-5 grouped into Cluster Ia (all japonica from SKUAST-Kashmir). Earlier we performed the principal component analysis on a set of 150 germplasm lines using 31 agro-morphological traits (Shikari et al., 2009) The study delineated the accessions into two major clusters with some accessions falling mid-way. Our results are in close conformity with the classification based on morphological markers. Although morphological markers are governed by genic loci they may, in many cases, be different from molecular markers which also originate from non-genic regions. Recently, Gaur et al. (2019) performed analysis based on kernel dimensions and found that the local landraces plotted across two clusters. In the present study, we classified them in two sub-clusters, namely Ia and Ib. Some of the landraces which are grown under mid-mountains like *Mushk Budji* and *Kamad* and belong to japonica, clustered together. Similarly, the red pericarp landraces, *Tangdhar Zag* and *Karnah Zag*, which belong to same region were hardly differentiated. Recently we carried out the studies on expression of quality related genes where these two showed similar expression levels for quality (Hussain et al. 2020). On the other hand, even though some accessions were placed within the same cluster within low genetic distance, they belonged to different ecologies.

Among the SNPs which differentiated the sub-population K2 from K1-K3, Waxy_SNP, Amy_W2_R_1 associated with Wx locus, SNP SSII_1_SNP_ff_1 related to endosperm starch synthesis, Ghd7_05_SNP_ff_1 and Ghd8_SNP_ff_2 linked to heading date were most prominent. Among the six KASP loci which differentiated local landraces from exotic temperate germplasm, the marker, ALK_SNP_ff_1 at the ALK locus determines kernel starch properties. The *ALK* gene is linked to amylopectin chain-length in rice endosperm, and it co-segregates with starch synthase II enzyme that determines gelatinization temperature in rice (Gao et al. 2011). Strong selection under domestication has been reported for several important genes and include Wx for amylose (Wang et al. 1995), qSH1 for seed shattering (Konishi et al. 2006), Rc for pericarp colour (Sweeney et al. 2007), and Ghd 7 related to heading date (Huang et al. 2012). InDel (Sahu et al. 2017) and SSR (Vanniarajan et al. 2012) markers have also been reported to differentiate indica and japonica populations. Such markers or genes reveal high degree of polymorphism between indica and japonica and have possibly evolved before the divergence of the two ecotypes from a common progenitor.

The present level of genetic divergence points towards possible useful variability for traits of economic importance and grain quality. Few of the germplasm accessions studied here were previously evaluated for cold tolerance (Sanghera et al. 2011), apart from the work on characterization of landraces for stress resistance (Umakanth et al. 2017). We recently identified certain specific alleles for blast resistance (Shikari et al. 2014) and also revealed differential expression for γ-ammino butyric acid among rice landraces (Hussain et al. 2020). Besides, the characterization and genetic improvement of landraces was carried out for resistance towards rice blast (Khan et al. 2018). The process of germplasm characterization helps in the documentation and the long-term conservation of germplasm which, in turn, may help in the better utilization of genetic resources for the development of improved rice varieties. Further, the genotyping process can help define a core germplasm set and may also help in the selection of a population for mapping useful alleles linked to traits of economic importance.

KASP markers are more cost effective than SSR markers (Steele et al., 2018) that have previously been commonly used to characterise germplasm in rice (Yang et al., 2019), wheat (Roncallo et al., 2019) and Brassica (Li et al. (2019). The markers effectively divided a population of germplasm of the temperate region of the Kashmir valley into sub-populations with the greatest distinction between indica and japonica groupings. A small number of KASP markers were highly discriminatory and were usually associated with domestication traits. KASP markers, and specifically highly discriminatory markers, can be used as a cost-effective tool for the more efficient maintenance and use of rice genetic resources.

Compliance with Ethical Standard

Conflict of interest: The authors declare that they have no conflict of interest.

Author Contributions: ABS carried out the field work for the study; ABS, KAS, and JRW were involved in the KASP genotyping process; KAS identified the KASP markers for the study; ABS, GK performed statistical analysis; SN, NAB, LV facilitated the access to and maintenance of germplasm; FAN, SAW supported for critical inputs; SW helped in editing the analyses; ABS and GK wrote the article with assistance from JRW

KASP^{TM} based markers reveal a population sub-structure in temperate rice germplasm and local landraces grown in the Kashmir valley, north-western Himalayas

[1]Asif Bashir Shikari, [1]S. Najeeb, [1]Gazala Khan, [1]F. A. Mohidin, [1]Ashaq H. Shah [1]Showkat A. Waza, [2]F. A. Nehvi, [2]Shafiq A. Wani, [1]N. A. Bhat, [3]L.V.S. Rao, [4]K. A. Steele and [4]J. R. Witcombe

[1]Mountain Research Centre for Field Crops, [2]Division of Plant Biotechnology,
Sher-e-Kashmir University of Agricultural Sciences & Technology of Kashmir, J&K, India, 192 102
3ICAR-Indian Institute of Rice Research, Hyderabad, India, 500 030
[4]Bangor University, U.K.
e-mail: asifshikari@skuastkashmir.ac.in

ORCID: 0000-0002-2911-5536

Acknowledgement:

Authors duly acknowledge the funding and technical support facilitated by Innovate U.K. and Biosearch Technologies, Hoddesdon, Herts, U. K. for the conduct of this study

References

Choudhary G et al. (2013) Molecular Genetic Diversity of Major Indian Rice Cultivars over Decadal Periods PloS one 8:e66197 doi:10.1371/journal.pone.0066197

Choudhury B, Khan ML, Dayanandan S (2013) Genetic structure and diversity of indigenous rice (Oryza sativa) varieties in the Eastern Himalayan region of Northeast India SpringerPlus 2:228 doi:10.1186/2193-1801-2-228

Diez MJ et al. (2018) Plant Genebanks: Present Situation and Proposals for Their Improvement. the Case of the Spanish Network Frontiers in plant science 9:1794 doi:10.3389/fpls.2018.01794

Evanno G, Regnaut S, Goudet J (2005) Detecting the number of clusters of individuals using the software STRUCTURE: a simulation study Molecular ecology 14:2611-2620 doi:10.1111/j.1365-294X.2005.02553.x

Excoffier L, Smouse PE, Quattro JM (1992) Analysis of molecular variance inferred from metric distances among DNA haplotypes: application to human mitochondrial DNA restriction data Genetics 131:479-491

Gao Z, Zeng D, Cheng F, Tian Z, Guo L, Su Y. Yan M, Jiang H, Dong G, Huang Y, Han B (2011) ALK, the Key Gene for Gelatinization Temperature, is a Modifier Gene for Gel Consistency in Rice Journal of integrative plant biology, 53(9), pp.756-765

Gaur A, Parray GA, Shikari AB and Najeeb S (2019). Capturing the Genetic Diversity for Grain Quality Attributes in a Set of Temperate Rice (Oryza sativa L.) Germplasm by Cluster Analysis and the Assessment of Wx gene Polymorphism Int. J. Pure Applied Biosciences, 7(3): 67-73

Huang X et al. (2010) Genome-wide association studies of 14 agronomic traits in rice landraces Nature genetics 42:961

Huang X et al. (2012) A map of rice genome variation reveals the origin of cultivated rice Nature 490:497-501

Hussain SZ, Jabeen R, Naseer B, Shikari AB (2020) Effect of soaking and germination conditions on γ-aminobutyric acid and gene expression in germinated brown rice Food Biotechnology 34:132-150 doi:10.1080/08905436.2020.1744448

Iqbal AM, Chrungoo N, Shikari A, Najeeb S, Gayle A, Mujtaba A (2017) Molecular Diversity of Rice Germplasm Grown Under High Altitude Conditions Plant Cell Biotechnology And Molecular Biology: 481-488

Jacob SR, Tyagi V, Agrawal A, Chakrabarty SK, Tyagi RK (2015) Indian plant germplasm on the global platter: an analysis. PloS one, 10(5), p.e0126634.

Khan GH et al. (2018) Marker-assisted introgression of three dominant blast resistance genes into an aromatic rice cultivar Mushk Budji Scientific reports 8:4091 doi:10.1038/s41598-018-22246-4

Khush GS (1997) Origin, dispersal, cultivation and variation of rice Plant molecular biology 35:25-34

Konishi S, Izawa T, Lin SY, Ebana K, Fukuta Y, Sasaki T, Yano M (2006) An SNP caused loss of seed shattering during rice domestication Science 312:1392-1396

Kumar S, Stecher G, Li M, Knyaz C, Tamura K (2018) MEGA X: Molecular Evolutionary Genetics Analysis across Computing Platforms Molecular biology and evolution 35:1547-1549 doi:10.1093/molbev/msy096

Li, P. et al. (2019) Development of a core set of KASP markers for assaying genetic diversity in Brassica rapa subsp. chinensis Makino Plant Breeding 138.3: 309-324 doi.org/10.1111/pbr.12686

McCouch SR et al. (2002) Development and mapping of 2240 new SSR markers for rice (Oryza sativa L.) DNA research 9:199-207

McNally KL et al. (2009) Genomewide SNP variation reveals relationships among landraces and modern varieties of rice Proceedings of the National Academy of Sciences of the United States of America 106:12273-12278 doi:10.1073/pnas.0900992106

Mohammadi SA and Prasanna BM (2003) Analysis of genetic diversity in crop plants—salient statistical tools and considerations. Crop science, 43(4), pp.1235-1248

Najeeb S, Parray GA, Shikari AB, Zaffar G, Kashyp SC, Ganie MA, Shah A (2017) Farmers participatory selection of new rice varieties to boost production under temperate agro-ecosystems Journal of Integrative Agriculture (17)61810 doi:10.1016/s2095-3119

Parray G, Shikari AB (2008) Conservation and characterization of indigenous rice germplasm adapted to temperate/cooler environments of Kashmir valley ORYZA-An International Journal on Rice 45:198-201

Peakall R, Ruibal M, Lindenmayer DB (2003) Spatial autocorrelation analysis offers new insights into gene flow in the Australian bush rat, Rattus fuscipes Evolution; international journal of organic evolution 57:1182-1195 doi:10.1111/j.0014-3820.2003.tb00327.x

Peakall R, Smouse PE (2012) GenAlEx 6.5: genetic analysis in Excel. Population genetic software for teaching and research--an update Bioinformatics 28:2537-2539 doi:10.1093/bioinformatics/bts460

Peng ZY et al. (2009) Characterization of the genome expression trends in the heading-stage panicle of six rice lineages Genomics 93:169-178 doi:10.1016/j.ygeno.2008.10.005

Porras-Hurtado L, Ruiz Y, Santos C, Phillips C, Carracedo Á, Lareu M (2013) An overview of STRUCTURE: applications, parameter settings, and supporting software Frontiers in genetics 4:98

Pritchard JK, Stephens M, Donnelly P (2000) Inference of population structure using multilocus genotype data Genetics 155:945-959

Pritchard V, Metcalf J, Jones K, Martin A, Cowley D (2009) Population structure and genetic management of Rio Grande cutthroat trout (Oncorhynchus clarkii virginalis) Conservation Genetics 10:1209

Roncallo PF et al. (2019) Genetic diversity and linkage disequilibrium using SNP (KASP) and AFLP markers in a worldwide durum wheat (Triticum turgidum L. var durum) collection PloS one 14.6 doi.10.1371/journal.pone.0218562

Roy S et al. (2015) Genetic Diversity and Population Structure in Aromatic and Quality Rice (Oryza sativa L.) Landraces from North-Eastern India PloS one 10:e0129607 doi:10.1371/journal.pone.0129607

Roy S, Marndi BC, Mawkhlieng B, Banerjee A, Yadav RM, Misra AK, Bansal KC (2016) Genetic diversity and structure in hill rice (Oryza sativa L.) landraces from the North-Eastern Himalayas of India BMC genetics 17:107 doi:10.1186/s12863-016-0414-1

Sahu PK, Mondal S, Sharma D, Vishwakarma G, Kumar V, Das BK (2017) InDel marker based genetic differentiation and genetic diversity in traditional rice (Oryza sativa L.) landraces of Chhattisgarh, India PloS one 12:e0188864 doi:10.1371/journal.pone.0188864

Saitou N, Nei M (1987) The neighbor-joining method: a new method for reconstructing phylogenetic trees Molecular biology and evolution 4:406-425

Salgotra RK, Gupta BB, Bhat JA, Sharma S (2015) Genetic Diversity and Population Structure of Basmati Rice (Oryza sativa L.) Germplasm Collected from North Western Himalayas Using Trait Linked SSR Markers PloS one 10:e0131858 doi:10.1371/journal.pone.0131858

Sanghera G, Hussaini A, Anwer A, Kashyap S (2011) Evaluation of some IRCTN rice genotypes for cold tolerance and leaf blast disease under temperate Kashmir conditions Journal of Hill Agriculture 2:28-32

Semagn K, Babu R, Hearne S, Olsen M (2014) Single nucleotide polymorphism genotyping using Kompetitive Allele Specific PCR (KASP): overview of the technology and its application in crop improvement Molecular breeding 33:1-14

Shikari AB, Parray GA, Rather AG and Sheikh, FA (2009). Principal component analysis for evaluation of rice (*Oryza sativa* L.) germplasm. Journal of Rice Research, 2(1): 16-22

Shikari AB, Najeeb S, Khan GH, Ali G, Parray G, Zargar S, Sheikh F (2018) DNA fingerprinting of rice (Oryza sativa L.) varieties cultivated in Kashmir SKUAST Journal of Research 20:32-36

Shikari AB et al. (2014) Identification and validation of rice blast resistance genes in Indian rice germplasm Indian Journal of Genetics and Plant Breeding (The) 74:286-299

Singh N et al. (2015) Single-copy gene based 50 K SNP chip for genetic studies and molecular breeding in rice Scientific reports 5:11600 doi:10.1038/srep11600

Sokal R, Michener C (1958) A statistical method for evaluating systematic relationships. iniv. kansas sci. bull., 38: 1409–1438 Prim Product Ecol Factors Lake Maggiore 127

Steele KA et al. (2018) Accelerating public sector rice breeding with high-density KASP markers derived from whole genome sequencing of indica rice Molecular breeding : new strategies in plant improvement 38:38 doi:10.1007/s11032-018-0777-2

Sweeney MT, Thomson MJ, Cho YG, Park YJ, Williamson SH, Bustamante CD, McCouch SR (2007) Global dissemination of a single mutation conferring white pericarp in rice PLoS genetics 3

Temnykh S, DeClerck G, Lukashova A, Lipovich L, Cartinhour S, McCouch S (2001) Computational and experimental analysis of microsatellites in rice (Oryza sativa L.): frequency, length variation, transposon associations, and genetic marker potential Genome research 11:1441-1452 doi:10.1101/gr.184001

Trinh H et al. (2017) Whole-Genome Characteristics and Polymorphic Analysis of Vietnamese Rice Landraces as a Comprehensive Information Resource for Marker-Assisted Selection International journal of genomics 2017:9272363 doi:10.1155/2017/9272363

Trung KH, Nguyen TK, Khuat HBT, Nguyen TD, Khanh TD, Xuan TD, Nguyen XH (2017) Whole Genome Sequencing Reveals the Islands of Novel Polymorphisms in Two Native Aromatic Japonica Rice Landraces from Vietnam Genome biology and evolution 9:1816-1820 doi:10.1093/gbe/evx135

Umakanth B et al. (2017) Diverse Rice Landraces of North-East India Enables the Identification of Novel Genetic Resources for Magnaporthe Resistance Frontiers in plant science 8:1500 doi:10.3389/fpls.2017.01500

Vanniarajan C, Vinod KK, Pereira A (2012) Molecular evaluation of genetic diversity and association studies in rice (Oryza sativa L.) Journal of genetics 91:9-19 doi:10.1007/s12041-012-0146-6

Vasudevan K. Vera Cruz CM, Gruissem W, Bhullar NK (2014) Large scale germplasm screening for identification of novel rice blast resistance sources. Frontiers in plant science, 5, p.505 doi: 10.3389/fpls.2014.00505

Wang ZY et al. (1995) The amylose content in rice endosperm is related to the post-transcriptional regulation of the waxy gene The Plant Journal 7:613-622

Weir BS, Cockerham CC (1984) Estimating F-Statistics for the Analysis of Population Structure Evolution; international journal of organic evolution 38:1358-1370 doi:10.1111/j.1558-5646.1984.tb05657.x

Weir BS, Hill WG (2002) Estimating F-statistics Annual review of genetics 36:721-750 doi:10.1146/annurev.genet.36.050802.093940

Yang G et al. (2019) Development of a core SNP arrays based on the KASP method for molecular breeding of rice Rice 12: 21 doi.org/10.1186/s12284-019-0272-3

Zhou M et al. (2011) Genome-wide analysis of clustering patterns and flanking characteristics for plant microRNA genes The FEBS journal 278:929-940 doi:10.1111/j.1742-4658.2011.08008.x