Allele mining for the grain number gene An-1 in rice (Oryza sativa L.)

R. Shobica Priya¹, T. Kalaimagal*,¹ S. Rajeswari¹, R. Ajay Prasanth² and M. Raveendran²

¹ Centre for Plant Breeding and Genetics, TNAU, Coimbatore, Tamil Nadu, India
² Centre for Plant Molecular Biology and Biotechnology, TNAU, Coimbatore, Tamil Nadu, India
* E-Mail: kalaimagal.t@gmail.com

Abstract
Rice yield has attained a plateau and hence the enhancement of grain yield is indispensable to feed the growing population, which could be achieved by the identification of superior alleles in the existing germplasm. Any variation in the pleiotropic gene, An-1 (yield gene) leads to enhanced grain number and grain size in rice. Hence, the gene was chosen for analyzing the allelic diversity/haplotype variation with 150 lines of 3K RG panel which revealed that, the gene An-1 has 20 Single Nucleotide Polymorphisms and 10 INDELs encompassing both intronic and exonic regions. The genotypes were divided into four haplotypes in the combination of seven SNPs with the maximum number of genotypes in the first haplotype and the least number of genotypes in fourth haplotype. From the study, H1 was identified as a superior haplotype. The haplo-pheno analysis identified the superior donors viz., SIGARDIS, GENIT and DAMNOEUB KAUN KHMOM harbouring superior haplotype combinations, which may be further utilized in haplotype-based breeding for the development of high yielding rice varieties.

Key words: rice, An-1, grain number, haplotype, SNPs, allelic diversity.

INTRODUCTION
Rice (Oryza sativa L.) is life for almost half of the global human population. The world faces the challenge of feeding 9.7 billion people by around 2050 (Source: United Nations Department of Public Information). Owing to an increase in the global population and a decrease in arable land, upsurging grain yield has become a priority. Rice yield has attained a plateau for the past one decade. Hence, enhancement of grain yield is the foremost objective in rice-breeding programs (Umadevi et al., 2019, Singh et al., 2020). The yield potential of the crop can be increased by altering the photosynthetic rate (C4 rice), identification of novel genes from wild, distant relatives/germplasm (allele mining), or by the creation of novel alleles through targeted mutagenesis. Grain yield in rice, which is a quantitative trait is governed by the components viz., the number of panicles per plant, the number of grains per panicle, and grain weight (Xing and Zhang, 2010). Among these, grain number per panicle is a highly variable component that mainly depends on the length of the panicle, number of primary and secondary branches, and percentage of filled grains (Deshmukh et al., 2010). The untapped novel alleles in primitive cultivars/landraces are of immense value for developing superior cultivars (Fess et al., 2011). Exploring allelic variations/haplotype variations pave way for identifying suitable donors with the desired trait of interest that can be further employed in crop improvement programmes (Varshney et al., 2018).

Domestication of rice was achieved through selection for deletion mutant, but genetic improvement in rice depends...
Allele mining for the grain number gene

on the search of novel alleles and assembly of them in favourable genetic background. In recent days, several genes related to grain yield in rice have been identified. Some of the key notable genes were An1, Gn1a, DEP1, DEP3, APO1, GS3, GW5, PROG1, LRK1, OsSPL14, EP3, and SP1 which modulate grain number per panicle directly or indirectly (Gouda et al., 2019). Among them, An-1 is a pleiotropic gene that regulates grain size, grain number, and awn development in rice. Transgenic studies have confirmed that the An-1 gene positively regulates the extension of lemma (i.e. awn elongation) and negatively regulates the grain number per panicle in rice. Attempts have been made to increase the grain number per panicle by downregulating the An-1 gene through RNAi technology. There is an increase in grain number per panicle by around 13.60 to 38.4 per cent in RNAi lines. Hence, the downregulation of the An-1 gene leads to an upregulation of the LOG (LONELY GUY) gene, which in turn leads to an enhanced cytokinin level and meristematic activity, resulting in increased grain number per panicle and thus yield (Luo et al., 2013). Though numerous studies have been conducted for yield enhancement in rice, the impact of alleles related to yield traits are not very much clear and hence it is a pre-requisite to survey the allelic diversity in the germplasm of rice (Gouda et al., 2020). Allele mining of yield-related traits has been attempted by very few research groups on genes like Gn1a, GS312, DEP114, Ghd713, and sd115 (Vemireddy et al., 2019). This study was aimed at unravelling the allelic variations in An-1 encompassing 150 genotypes from 3K RG panel, a core collection of 3000 resequenced rice accessions and harnessing its potential opportunities in future plant breeding programmes.

MATERIALS AND METHODS
A total of 150 genotypes belonging to the 3K RG panel (Li et al., 2014) were raised in augmented design at Paddy Breeding Station, Tamil Nadu Agricultural University during Rabi, 2020 (Table 1). Twenty-five days old seedlings were transplanted in the main field at a spacing of 20 x 20 cm and the recommended package of practices were carried out with adequate fertilizer application of 150:50:50 N:P:K kg/ha, respectively. The number of grains per panicle was recorded on three randomly selected competitive plants in each genotype.

Descriptive statistics including mean, median, range (minimum, maximum) and coefficient of variation (CV) for the whole population were estimated using Minitab 19 Statistical Software. Frequency distribution and box plot graphs were also constructed using Minitab 19 Statistical Software (Allen, 2019).

The allelic diversity for the An-1 gene (LOC_Os04g28280) was carried out for the studied genotypes by downloading their sequence using the Rice-SNP seek database (http://snp-seek.irri.org/) encompassing the non-synonymous SNPs and INDELs. Nipponbare (Oryza sativa ssp. japonica) sequence of An-1 gene was taken as a reference one and aligned against the study sequences using BioEdit software (Hall, 2011).

The haplotype analysis was carried out by downloading the allelic variations of the An-1 gene for the respective genotypes from the Rice-SNP seek database in PLINK format and later converted into haplovieview format using PLINK software (Jonathan, 2010). Haplotype groups and Linkage Disequilibrium blocks of the allelic variation were constructed using Haplovieview 4.2 software with parameters including HW p-value (Hardy-Weinberg p value) of 0.001, minimum genotype per cent of 75 and minimum minor allele frequency of 0.001 (Barrett et al., 2005). The significant differences between the constructed haplotype groups were proved by Dunnett’s test using Minitab 19 Statistical Software (Allen, 2019).

RESULTS AND DISCUSSION
Any distinct variation in the gene, An-1 (yield gene) leads to increased grain number and grain size in rice and hence this gene was chosen for assessing its allelic diversity.

The phenotypic data of the 150 genotypes were subjected to descriptive statistics, which delineates the simple measures of variability. Estimates for mean, median, range (minimum, maximum) and coefficient of variation were (Table 2). It was observed that the coefficient of variation was high i.e. 27.35 (> 20%) indicating that there is an ample amount of variation in the studied population and hence, the selection of suitable donors for the respective trait can be carried out with the available genotypes. From the box plot curve, it is understood that the range varies from 47 to 223 number of grains per panicle with an average of 113.34 grains per panicle (Fig. 1). The histogram implies that the population follows a normal distribution (Fig. 2).

Multiple Sequence Alignment of 150 genotypes with reference nucleotide sequence for An-1 gene identified 20 SNPs and 10 INDELS in the both intronic and exonic region, which is given in Table 3 along with their minor allele frequency (MAF). Both synonymous and non-synonymous SNPs were identified. Since the synonymous SNPs do not cause any alteration in the amino acid, the non-synonymous SNPs were enumerated further. Among the non-synonymous SNPs, SNP1 induces a base replacement of G→A at 16734725 positions of nucleotide sequence, causing R→W in the coded protein. SNP2 causes G→C substitution at 16734542 leading to R→G in the coded protein. SNP13 induces T→C base replacement at 16732870 positions, causing E→G substitution in the encoded protein. SNP15 causes T→G replacement of nucleotide sequence in 16732750 positions leading to T→P substitution of amino acid. SNP17 induces C→T transition at 16732732 positions causing G→D substitution in the coded protein.
### Table 1. List of genotypes used in the present study with its respective haplotype groups

| No. | Accessions     | H   | No. | Accessions     | H   | No. | Accessions     | H   | No. | Accessions     | H   |
|-----|----------------|-----|-----|----------------|-----|-----|----------------|-----|-----|----------------|-----|
| 1   | IRIS 313-9867  | H1  | 2   | IRIS 313-8699  | H1  | 3   | IRIS 313-10768 | H1  | 4   | IRIS 313-10497 | H1  |
| 5   | IRIS 313-10001 | H1  | 6   | IRIS 313-11790 | H1  | 7   | IRIS 313-10575 | H1  | 8   | IRIS 313-8850  | H1  |
| 9   | IRIS 313-8585  | H1  | 10  | IRIS 313-11423 | H1  | 11  | IRIS 313-8968  | H1  | 12  | IRIS 313-9696  | H1  |
| 13  | IRIS 313-9705  | H1  | 14  | IRIS 313-9492  | H1  | 15  | IRIS 313-8492  | H1  | 16  | IRIS 313-11398 | H1  |
| 17  | IRIS 313-11849 | H1  | 18  | IRIS 313-9551  | H1  | 19  | IRIS 313-8996  | H1  | 20  | IRIS 313-10374 | H1  |
| 21  | IRIS 313-8412  | H1  | 22  | IRIS 313-10775 | H1  | 23  | IRIS 313-12053 | H1  | 24  | IRIS 313-9482  | H1  |
| 25  | IRIS 313-9609  | H1  | 26  | IRIS 313-9758  | H1  | 27  | IRIS 313-11870 | H1  | 28  | IRIS 313-8994  | H1  |
| 29  | IRIS 313-11568 | H1  | 30  | IRIS 313-11052 | H1  |

### Table 2. Descriptive statistics for the studied population

| Statistics       | Whole population |
|------------------|------------------|
| Range            | 47-223           |
| Mean             | 113.34           |
| Median           | 107.33           |
| SE mean          | 2.53             |
| CV(%)            | 27.35            |
Shobica Priya et al., thus resulting in increased yield. The lowest grain number was recorded in PA KHENG and BORO with 47 and 51 grain numbers per panicle, respectively. Hence, these genotypes can be recommended as suitable donors for developing mapping populations in the crop improvement programme. Further, analysis is required to identify the causative SNPs.

Overall, this study resulted in unravelling the allelic variations in the An-1 gene with the help of 150 genotypes from 3K RG panel, and their grouping based on haplotyping identified a superior haplotype H1 and a further selection of genotypes from these haplotypes can be exploited in allele mining and plant breeding programmes.

Table 3. Allelic diversity analysis for An-1 gene in 150 rice genotypes

| REGION | SNP POSITION | ALLELE     | TYPE     | MAF  |
|-------|--------------|------------|----------|------|
| EXON 1| 16734759-61  | 3bp        | INDEL1   | 4.00 |
| EXON 1| 16734725     | G/A        | SNP1     | 6.67 |
| EXON 2| 16734542     | G/C        | SNP2     | 10.67|
| EXON 2| 16734484     | C/G        | SNP3     | 8.00 |
| EXON 2| 16734476     | 1bp/0bp    | INDEL2   | 48.67|
| EXON 2| 16734442     | C/A        | SNP4     | 1.33 |
| INTRON 2| 16734402     | 0bp/1bp    | INDEL3   | 6.67 |
| INTRON 2| 16734378     | 1bp/0bp    | INDEL4   | 0.67 |
| INTRON 2| 16734332     | C/A        | SNP5     | 0.67 |
| INTRON 2| 16734244     | C/A        | SNP6     | 0.67 |
| EXON 3| 16734226-237 | 12bp/0bp   | INDEL5   | 0.67 |
| EXON 3| 16734204-218 | 15bp/0bp   | INDEL6   | 0.67 |
| INTRON 3| 16733811     | C/T        | SNP7     | 5.33 |
| INTRON 3| 16733609     | 1bp/0bp    | INDEL7   | 5.33 |
| INTRON 3| 16733608     | 1bp/0bp    | INDEL8   | 6.00 |
| INTRON 3| 16733607     | 1bp/0bp    | INDEL9   | 0.67 |
| INTRON 3| 16733493     | C/T        | SNP8     | 7.33 |
| INTRON 3| 16733434     | A/T        | SNP9     | 9.33 |
| EXON 4 | 16733302     | A/G        | SNP10    | 2.00 |
| INTRON 5| 16732975     | G/A        | SNP11    | 0.67 |
| EXON 6 | 16732882     | C/T        | SNP12    | 1.33 |
| EXON 6 | 16732870     | T/C        | SNP13    | 7.33 |
| EXON 6 | 16732841     | G/A        | SNP14    | 1.33 |
| EXON 6 | 16732750     | T/G        | SNP15    | 6.67 |
| EXON 6 | 16732748     | G/A        | SNP16    | 6.67 |
| EXON 6 | 16732732     | C/T        | SNP17    | 7.33 |
| EXON 6 | 16732701     | G/A        | SNP18    | 8.00 |
| INTRON 6| 16732590     | 1bp        | INDEL10  | 0.67 |
| INTRON 6| 16732582     | A/T        | SNP 19   | 5.33 |
| INTRON 6| 16732581     | T/A        | SNP 20   | 9.33 |

MAF - minor allele frequency

https://doi.org/10.37992/2021.1203.107
Apart from these missense mutations, even a nonsense mutation occurred at 16734476 positions of the nucleotide sequence (G deletion) leading to the frameshift mutation and production of the premature stop codon, which was analysed with ExPASy translate tool (http://expasy.org/tools/dna.html). The truncated protein consists of 97 amino acids and loses its functions by losing its bHLH domain (Fig. 3). In the first exon, a 3bp insertion causes insertion of alanine in the encoded protein. Since the An-1 gene is a transcription factor with transactivation activity, all these changes have an impact on the protein produced and finally the yield. Luo et al. (2013) findings state that a transposon-like indel in the promoter region of Nipponbare subhaplotype an-1(Tn+) and another sub-haplotype an-1(G-) with 1-bp nucleotide-G deletion in the second exon of An-1 led to frameshift mutation that enhances grain number per panicle comparatively and hence in our study the genotypes with G deletion in the second exon can be exploited further along with the consideration of non-synonymous SNPs.

Haplotype analysis of the An-1 gene (LOC_Os04g28280) divided all the genotypes (150) into four different haplotype groups (Table 1, Fig. 4). Vasumathy and Alagu, (2021) identified three haplotype clusters in the third hap-block for OsLG3 (grain yield and length) and twelfth hap-block for OsMFT1 in chromosome 3 and 6, respectively. Gouda et al. (2020) obtained five haplotype groups for the G11a gene while considering forty-eight genotypes.

Vasumathy and Alagu, (2021) identified three haplotype clusters in the third hap-block for OsLG3 (grain yield and length) and twelfth hap-block for OsMFT1 in chromosome 3 and 6, respectively. Gouda et al. (2020) obtained five haplotype groups for the Gm11a gene while considering forty-eight genotypes. OsSPL14 had 15 haplotypes whereas, OsMADS1 had only one haplotype across 3K RG panels (Abbai et al., 2019). Among the various SNPs of the An-1 gene, only seven significant SNPs were used for constructing four haplotypes using haploview as shown in Fig. 4. The Linkage Disequilibrium block depicts the alleles at nearby positions, that can occur on the same haplotype often than expected by chance (Fig. 5). Linkage Disequilibrium refers to the non-random association of alleles in a loci antagonist to Mendel's law of independent assortment. All pairwise combinations of SNPs yield high LD i.e. the linked SNPs have high LD. The SNP sites 10416732870...
Fig. 5. Linkage Disequilibrium (LD) block of An-1 gene for 150 genotypes. Values inside the block indicates LD $r^2$.

and 10416734395.1 had the highest LD $r^2$ value (90) indicating a high correlation between these two SNPs. From Table 4, the SNPs, their respective position and minor allele frequency could be easily identified. Among the four combinations, the H1 occupies the maximum number of genotypes (132) followed by H2 (10) and H3 (6). H4 occupies the least number of genotypes (2). It can be seen that the maximum number of grains per panicle falls under H1 (223) followed by H2 (124.67) and H3 (113) and the coefficient of variation is high for all the haplotypes except H4 (Table 5). Based on the median value from the descriptive statistics, H1 had 68 genotypes with more than 108 grains per panicle, whereas H2 and H3 had only two genotypes, which validates the superiority of H1 over other haplotypes. Further, Dunnett’s test revealed that there exists a significant difference between H2-H1, H3-H1 (p value < 0.05) but there is no significant difference between H4-H1. Although theoretically, there should be a significant difference between H4-H1, due to the least number of genotypes under this haplotype, it’s not significantly different (Table 6 & Fig. 6). In H1, the genotype SIGARDIS performed better with the highest number of grain numbers (223) and in H2 (124.67), GENIT yielded more grain numbers followed by DAMNOEUB KAUN KHMOM in H3 (113), thus resulting in increased yield. The lowest grain number was recorded in PA KHENG and BORO with 47 and 51 grain numbers per panicle, respectively. Hence, these genotypes can be

| Marker number | SNP site      | Position   | HW p-val   | MAF  | Alleles |
|---------------|---------------|------------|------------|------|---------|
| 6             | 10416732701   | 16732701   | 1.9491E-18 | 0.080| C:A     |
| 7             | 10416732732   | 16732732   | 2.7365E-17 | 0.074| C:T     |
| 12            | 10416732750   | 16732750   | 3.8441E-16 | 0.068| T:G     |
| 13            | 10416732870   | 16732870   | 2.5338E-17 | 0.074| T:C     |
| 14            | 10416732882   | 16732882   | 0.000069416 | 0.014| C:T     |
| 18            | 10416733302   | 16733302   | 1.1452E-06 | 0.020| A:G     |
| 64            | 10416734395   | 16734395   | 3.1186E-16 | 0.067| C:A     |
| 67            | 10416734541   | 16734541   | 2.7123E-22 | 0.108| G:C     |
| 72            | 10416734724   | 16734724   | 3.5835E-16 | 0.068| G:A     |

Table 4. Details of identified significant alleles.

Highlighted SNPs are responsible for haplotype formation
Table 5. Significant difference between haplotypes based on number of grains/panicle

| Statistics | H1 (132)       | H2 (10)        | H3 (6)       | H4 (2)       |
|------------|----------------|----------------|--------------|--------------|
| Range      | 75.67-223      | 51-124.67      | 47-113       | 72.3-94      |
| Mean       | 117.3          | 84.4           | 84.4         | 83.2         |
| Median     | 108.67         | 80.67          | 86.8         | 83.2         |
| SE mean    | 2.58           | 8.37           | 11.4         | 10.8         |
| CV (%)     | 25.3           | 31.36          | 33.05        | 18.42        |

H - haplotype

Table 6. Dunnett’s Simultaneous Tests for Level Mean-Control Mean

| Difference of levels | Mean difference | T-value | Adj. p-value |
|----------------------|-----------------|---------|--------------|
| H2-H1                | -32.9           | -3.42   | 0.002        |
| H3-H1                | -32.9           | -2.68   | 0.024        |
| H4-H1                | -34.1           | -1.63   | 0.281        |

Fig. 6. Box plot curve for different haplotypes indicating significant difference by Dunnett’s test.

REFERENCES

Abbai, R., Singh, V. K., Nachimuthu, V. V., Sinha, P., Selvaraj, R., Vipparla, A. K., Singh, A. K., Singh, U. M., Varshney, R. K. and Kumar, A. 2019. Haplotype analysis of key genes governing grain yield and quality traits across 3K RG panel reveals scope for the development of tailor-made rice with enhanced genetic gains. Plant Biotechnology Journal. 17(8):1612-1622. [Cross Ref]
Allen T. T. 2019. Introduction to Engineering Statistics and Lean Six Sigma (London: Springer) Software overview and methods review: Minitab 575-600. [Cross Ref]

Barrett, J. C., Fry, B., Maller, J. and Daly, M. J. 2005. Haploview: analysis and visualization of LD and haplotype maps. Bioinformatics. 21: 263–265. [Cross Ref]

Deshmukh, R., Singh, A., Jain, N., Anand, S., Gacche, R., Singh, A., Gaikwad, K., Sharma, T., Mohapatra, T. and Singh, N. 2010. Identification of candidate genes for grain number in rice (Oryza sativa L.). FunctIntegr Genomics. 10(3):339–347. [Cross Ref]

Fess, T. L., Kotcon, J. B. and Benedito, V. A. 2011. Crop breeding for low input agriculture: a sustainable response to feed a growing world population. Sustainability. 3(10):1742–1772. [Cross Ref]

Gouda, G., Gupta, M. K., Donde, R., Kumar, J., Vadde, R., Mohapatra, T. and Behera, L. 2019. Computational approach towards understanding structural and functional role of cytokinin oxidase/dehydrogenase 2 (CKX2) in enhancing grain yield in rice plant. J Biomol Struct Dyn. 38(4):1158-1167. [Cross Ref]

Gouda, G., Gupta, M. K., Donde, R., Kumar, J., Parida, M., Mohapatra, T., Dash, S. K., Pradhan, S. K. and Behera, L. 2020. Characterization of haplotypes and single nucleotide polymorphisms associated with Gn1a for high grain number formation in rice plant. Genomics. 112(3):2647-2657. [Cross Ref]

Hall, T. 2011. Bio Edit: an important software for molecular biology. GERF Bull Biosci. 2: 6.

Jonathan P. Weeks 2010. plink: An R package for linking mixed-format tests using IRT-based methods. Journal of Statistical Software. 35(12):1-33. [Cross Ref]

Li, J. Y., Wang, J. and Zeigler, R. S. 2014. The 3,000 rice genomes project: new challenges and challenges for future rice research. Gigascience. 3(1):2047-217X. [Cross Ref]

Luo, J., Liu, H., Zhou, T., Gu, B., Huang, X., Shangguan, Y., Zhu, J., Li, Y., Zhao, Y., Wang, Y., Zhao, Q., Wang, A., Wang, Z., Sang, T., Wang, Z. and Hana, B. 2013. An-f encodes a basic helix-loop-helix protein that regulates awn development, grain size, and grain number in rice. The Plant Cell. 25:3360-3376. [Cross Ref]

Singh, S. K., Pandey, V., Mounika, K., Singh, D. K., Khaire, A. R., Habde, S. and Majhi, P. K. 2020. Study of genetic divergence in rice (Oryza sativa L.) genotypes with high grain zinc using Mahalanobis’ D2 analysis. Electronic Journal of Plant Breeding, 11(02):367-372. [Cross Ref]

Umadevi, M., Shanthi, P. and Saraswathi. R. 2019. Characterization of rice landraces of Tamil Nadu by multivariate analysis. Electronic Journal of Plant Breeding, 10(3):1185-1193. [Cross Ref]

Varshney, R.K., Singh, V.K., Kumar, A., Powell, W. and Sorrells, M.E. 2018. Can genomics deliver climate-change ready crops? CurrOpin Plant Biol. 45:205-211. [Cross Ref]

Vasumathy, S. K. and Alagu, M. 2021. SSR Marker based genetic diversity analysis and snp haplotyping of genes associating abiotic and biotic stress tolerance, rice growth and development and yield across 93 Rice Landraces. Mol Bio. Reports. 48(8):5943-5953. [Cross Ref]

Vemireddy, L. R., Kadambari, G., Reddy, G. E., Rao, Kola, V. S. R., Ramireddy, E., Puram, V. R. R., Badri, J., Eslavath, S. N., Bollineni, S. N., Naik, B. J., Chintala, S., Pottepaleni, R., Akkareddy, R., Nagireddy, R., Reddy, L. V. B., Bodanapu, R., Lekkala, S. P., Chakravartty, N. and Siddiq, E. A. 2019. Uncovering of natural allelic variants of key yield contributing genes by targeted resequencing in rice (Oryza sativa L.). Scientific Reports. 9:8192. [Cross Ref]

Xing, Y. and Zhang, Q. 2010. Genetic and molecular bases of rice yield. Annu. Rev. Plant Biol. 61:421-442. [Cross Ref]