Australian Government
Australian Centre for International Agricultural Research

Final report

project	**Breeding for low chalk in rice**
project number	CIM/2016/046
date published	10th February 2017 – 31st December 2018
prepared by	Nese Sreenivasulu
co-authors/ contributors/ collaborators	Gopal Misra, Anindya Bandyopadhyay and Saurabh Badoni
approved by	Eric Huttner, Research Program Manager Crops
final report number	N/A
ISBN	N/A
published by	ACIAR GPO Box 1571 Canberra ACT 2601 Australia

Contents

1 Acknowledgments

2 Executive summary

Chalky areas appear as opaque region in translucent rice grains due to the likely impairment in seed storage processes causing loose starch packaging in rice grain. The increased proportion of percent grain chalk reduces the appearance quality of rice grains, which is also believed to lower milling quality by making grain predisposed to breakage. These reasons, in turn negatively affect the market value of rice and thus elevated percentage of chalk in breeding material is an undesirable trait.

To address this, we utilized the genetic resource of 320 landraces with 15x resequencing data to conduct Genome Wide Association Study (GWAS). We identified a highly significant ($-\log_{10}(P)=15$) fine-mapped genomic region on chromosome 5 mapped between 5.12 Mb to 5.43 Mb associated with the chalk trait. Utilizing Targeted Gene Association Studies (TGAS) we identified a novel unknown gene with important alleles/haplotypes underlying the chromosome 5 fine mapped genetic region identified by GWAS. Furthermore, upon scanning the key alleles underlying the hotspot region in the repertoire of key IRRI-bred varieties, the identified allelic variations which distinguish low chalk from high chalk groups have been confirmed.

This work, building on the legacy of successive projects, first started in 2007, has therefore **resulted in the identification of reliable molecular markers for the chalk trait**, now allowing breeders to efficiently select for low chalk. This work represents a successful example of results which can be obtained through the application of modern genetics and genomics methods, many of them unavailable at the onset of this research.

Development of biparental mapping population is under process in order to validate the marker-assisted mobilization of low chalk alleles into a high chalk cultivar. Functional proof of concept of the candidate gene identified in this work is also being undertaken.

This knowledge needs to be extrapolated to rice breeding programs in IRRI partner countries, who will be in a position to implement selection for lowering chalk in released rice varieties.

3 Background

Grain chalkiness is one of the key factors that determine the grain quality by affecting grain appearance, processing and milling, eating and cooking quality. Chalkiness is generally characterized as lower starch granule density in rice grain, resulting in opaque white discoloration in the endosperm, which in turn makes the grain less appealing in its appearance and more prone to breakage. Chalkiness in rice is due to the creation of air spaces between irregularly-shaped starch granules. This leads to the opaque white discoloration in the transparent rice endosperm. The increased proportion of grain chalk broadly affects the appearance and milling quality of rice grains[1,2]. Rice with a high proportion of broken grains is generally sold at a much lower price, often for animal feed. Owing to the lower standard and economic value of the grains, farmers/grower in turn do not get sufficient returns. Reduced availability of good quality rice also affects the food security of ever-increasing populations given rice is being consumed by more than half of global population. Furthermore, when chalky rice grains are cooked or boiled, traverse and longitudinal cracks occurs easily and form relatively harder texture than the normal grains that in turn badly affects its palatability. These attributes of chalky rice negatively impact the export quality of rice that ultimately leads to economic loss to growers and millers. Therefore, reducing the incidence of chalk in modern rice varieties would provide large benefits. This project aims at identifying the genetic regions associated with, and responsible for low chalk and developing germplasm with low chalk. Low chalk rice yields high level of head rice, reduce economic loss to grower caused by lower quality grains, and extend the consumer acceptability of the crop in domestic and international markets.

In general, chalkiness is influenced by genetic and environmental factors and can vary from very low (less than 2%) to very high (up to more than 20%)[3,4]. A previous ACIAR project, CIM-2006-176, mapped the genetic variation for the low-chalk trait despite a widely held belief that environmental variation was the main cause for the trait and that this would detract from finding a genetic basis for the trait. In this SRA project, we fine-mapped the genetic region on chromosome 5, which was linked to low chalk through genome-wide association studies. We showed the relevance of genetic markers from this region in IRRI's breeding material to ensure their suitability in marker assisted selection for reducing chalk. We also laid the foundation to functionally characterize the genes from this hot-spot on chromosome 5 as continued efforts.

4 Objectives

The Objectives of the project involves the validation of low chalk characters using marker-assisted selection and further functional validation of the identified candidate genes. The major objectives are as follows:

Objective 1: Validate the transfer of low-chalk characters into a high-chalk line through marker-assisted selection.

Objective 2: Functional validation of three candidate genes from the chromosome 5 fine-mapped target region through gene editing using Clustered Regularly Interspaced Short Palindromic Repeats (CRISPR) system to test their relevance in rice chalkiness.

Objective 3. Develop a business model for applying the results in partnership with National Agricultural Research and Extension Services and partnered private companies.

5 Methodology

Plant material and chalk phenotyping

A diversity panel consisting of 320 *indica* accessions selected from the 3,000 resequenced rice genomes (The 3000 Rice Genomes Project, 2014) were grown in wet and dry season by following standard crop management practices at the Experimental Station of the International Rice Research Institute (IRRI), Laguna, Philippines (14°N, 121°E). In addition, we selected the best IRRI's breeding material over the last 50 years and grew them under controlled condition to monitor the relevance of chalk.

A 50 g seed material of each accession with three independent replicates were acquired and equilibrated at room temperature (25°C). Grain chalkiness was determined using the SeedCount SC5000 Image Analyser, as Percent Grain with Chalkiness (PGC) using laboratory optimized methods. Protocols related to these methodologies have been published as book chapter (refer Santos et al., 2019).

Genome-wide and candidate-gene association studies (GWAS and TGAS)

Genome sequences were compared to references and filtered to retain Single Nucleotide Polymorphisms (SNPs) with a missing rate of not more than 5% and further filtered to ensure a minor allele frequency of ≥5%. This led to a total of 2,260,030 SNPs against Nipponbare reference genome. Furthermore, calling SNPs against two different *indica* reference genomes Minghui 63 (MH63; low chalk) and Zhenshan 97 (ZS97; high chalk) resulted in the identification of 802,117 and 450,604 SNPs, respectively. Mixed-linear model based EMMAX[5] was utilized to conduct GWAS. PGC data was transformed using WarpedLMM to fine-tune the phenotype distribution. Based on the P-value criterion of the SNPs ($-log_{10}P>5$), SNPs were tagged. The plot showing the loci with the variable P-values is mentioned as Manhattan plot, whereas the quantile-quantile (Q-Q) plot shows the distribution of expected association test statistics (X-axis) compared to the observed values (Y-axis) across the entire SNP set. All genomic positions and gene annotations were based on the Nipponbare (MSUv7), MH63 and ZS97 reference genomes.

For candidate genes implicated by the most significant SNPs (either due to presence of SNP in their genic region or associated with a prominent SNP in Linkage Disequilibrium), targeted-gene association study (TGAS) was performed. TGAS considered only the SNPs falling in genic region and within 1 kb upstream of the start codon. These structural variants (reference and altered alleles of SNPs) were used to construct the respective haplotypes and are represented with the phenotypic variation in a form of boxplots.

Furthermore, the multi-locus (ML)-GWAS was undertaken to verify the regions identified using single-locus GWAS. A set of ML-GWAS methods – FASTmrEMMA[6], mrMLM[7] and ISIS EM-BLASSO [8] were performed, following Misra et al.[9].

6 Achievements against activities and outputs/milestones

Objective 1: Validate the transfer of low-chalk characters into a high-chalk line through marker-assisted selection

no.	activity	outputs/ milestones	completion date	comments
1.1	Whole-genome sequencing of low-chalk (<2%) and high-chalk lines (>20%) identified based on diagnostic haplotypes of chromosome 5	Whole genome sequencing of 320 diversity lines has been performed and same region was validated using two *indica* reference genomes. Lines exhibiting the haplotypes were resequenced.	December 2017	The selected low chalk and high chalk lines were used as parents for performing hybridization for the fulfilment of the next activity.
1.2	High-density F_2 mapping population developed between low-chalk and high-chalk lines and fine-mapping target genes to define the functional markers to breed low chalk.	F_2 mapping population exhibiting ~5000 plants was obtained.	June 2018	Resequencing of the pooled DNA of contrasting bulks was performed and bulk segregating analysis or QTL-Seq is underway to conduct QTL-mapping in order to identify the major QTLs/candidates underlying grain chalkiness in rice.

PC = partner country, A = Australia

Objective 2: Functional validation of three candidate genes using CRISPR from chromosome 5 fine-mapped targets to test their relevance in rice chalkiness

no.	activity	outputs/ milestones	completion date	comments
2.1	Transcriptome-wide association studies to identify *cis* factors on fine-mapped chromosome 5 and possible *trans* regulatory factors in other chromosomes that explain the genetic variation for low chalk	Lines containing contrasting haplotypes of fine mapped targets of chromosome 5 were subjected to transcriptome analysis using 60K microarray.	December 2018	Using the expression data of target genes identified within the fine mapped chromosome 5, we conducted transcriptome wide association studies. No cis-eQTLs were found for the targets.

2.2	The top three candidate genes would be functionally validated using CRISPR/Cas9 genome editing technology to prove that the target genes influence the chalk phenotype	Constructs for CRISPR-Cas9/Cpf1 were completed	December 2018	The knocks out cassettes (CRISPR-Cas9/Cpf1) for target genes were designed and undertaken Agrobacterium-based transformation into the IR64 rice plant. To ensure the progress of this work we need to secure future funding.

PC = partner country, A = Australia

7 Key results and discussion

Whole genome sequencing based GWAS of chalk

Utilizing whole-genome sequencing resource of three hundred twenty resequencing lines, we conducted single locus-GWAS (genome-wide association studies) and gene-set analysis. This approach identified a hotspot region on chromosome 5 strongly associated with grain chalkiness in *indica* subpopulation (Figure 1). The reference genome initially utilized for calling SNP was acquired from Nipponbare MSU version 7.0. GWAS conducted using SNPs acquired based on the alignment of whole genome sequencing information of 320 diverse *indica* lines with *japonica* reference genome, identified a highly significant ($-\log_{10}(P)$=15) fine-mapped genomic region on chromosome 5 mapped between 5.12 Mb to 5.43 Mb (Figure 1a), which was further confirmed with the multi-locus GWAS shown as red arrow. Notably, we annotated a putative novel candidate gene which encode for ~122 amino acid protein with unknown function, named in this study as *chalk5.1* (Figure 1d). TGAS of *Chalk5.1* gene with haplotype GCTCGCTGGGGGCCG (n=10) identified significant tagged SNP in the regulatory, exonic and intronic region. This gene is strongly associated with PGC, located ~3kb upstream region of the "grain width 5" locus LOC_Os05g09520. Notably, when we mapped the chalk hotspot region showing the collinearity in chromosome 5 region, we observed several breakpoints between high chalk (ZS97) and low chalk (Nipponbare) genome in the target hotspot region of PGC on chromosome 5 (Figure 1b). This indicated the existing structural variations in the hotspot region regulating grain chalkiness. The previously cloned *V-PPase* candidate gene of chalk[11] *chalk5* was observed distant from the target region we identified in the present study.

As the peak genetic region identified by GWAS was specific to *indica* subpopulation, GWAS was further employed using SNPs called against high quality *indica* reference genomes (Minghui 63 and Zhenshan 97)[10]. Subsequently, the hotspot genetic region was confirmed again using *indica* reference genomes with further validation from multi-locus GWAS methods[6,7,8] (Figure 2). Notably, Minghui 63 (MH63) and Zhenshan 97 (ZS97) *indica* reference genomes which possessed the extremely low and high chalkiness phenotype, respectively[11] were used as a source to fine map the target genes within the hot spot region of chromosome 5. Using genotyping data derived from ZS97 reference, GWAS analysis identified the same significant association on chromosome 5 with ~0.9 Mb region harbouring four major LD-blocks, as noted with *japonica* reference genome. Notably, the strong effect SNPs conferring low chalk phenotype found within candidate gene *chalk5.1* (unknown function) is present in a separate LD block, while *GW5*/*qSW5* (LOC_Os05g09520) regulating grain width is in another LD block. Reference allele of the topmost SNPs (snp_05_5361276) detected in the 5'-UTR of the newly predicted gene *chalk5.1*, matched with binding site for tri-helix protein. Other strongly associated-SNPs were mapped within the binding site of trans-regulatory elements including AP2, C2H2, MYB, TIFY-transcription factor and TBP (TATA-binding protein) identified in the upstream region of functionally un-annotated gene *chalk5.1*.

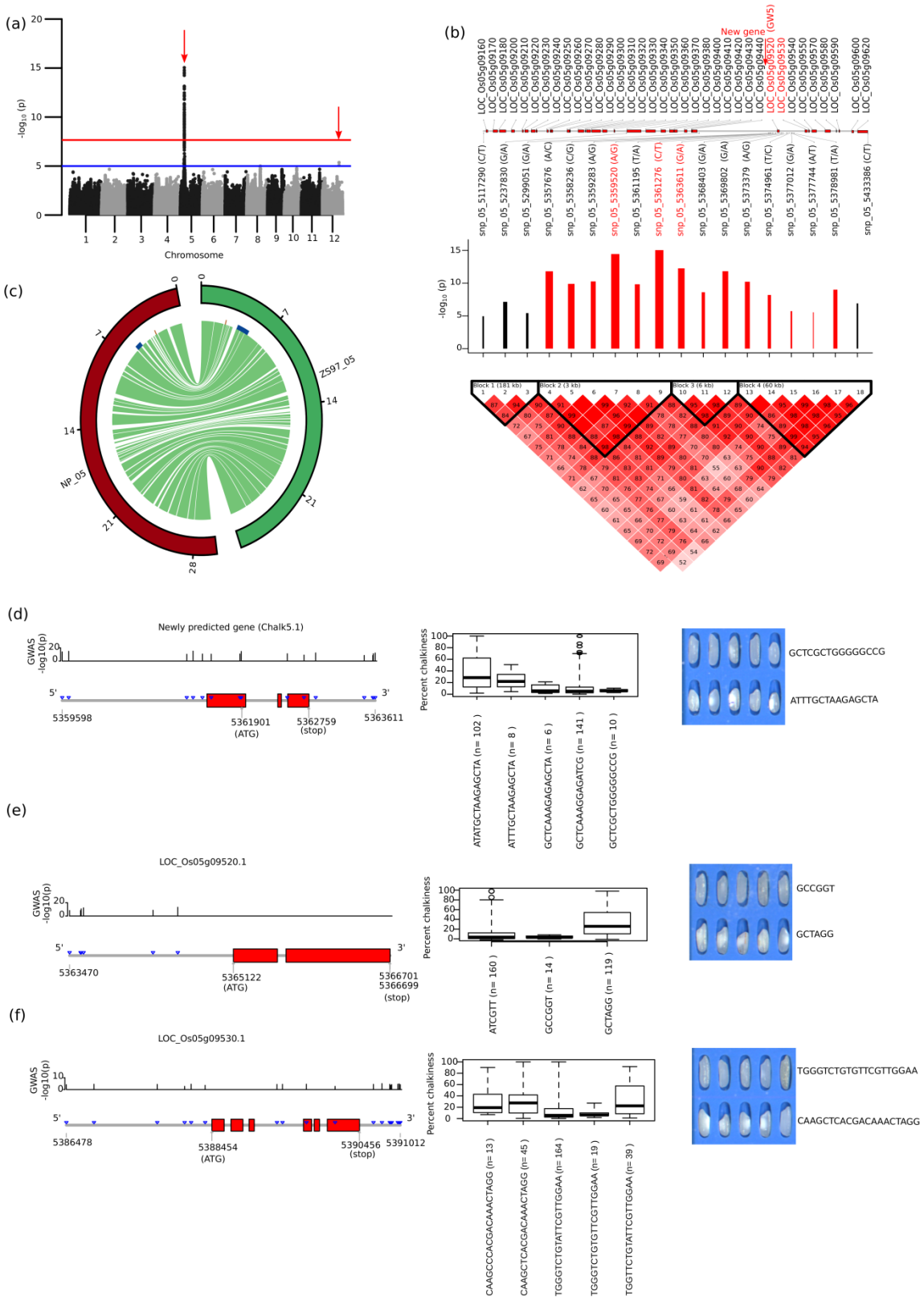


Figure1. Single-locus GWAS for grain chalkiness identified prominent loci on chromosomes 5 in 320 indica panel using japonica reference genomes. (a) Manhattan - plots of the GWAS on PGC revealed association of PGC with chromosome 5 hotspot region, which further conformed by multi-locus GWAS (indicated as red arrow). Horizontal red and blue line in Manhattan plots represents the genome-wide significant

threshold -log$_{10}$(P) value of 7.6 and 5, respectively; (b) linkage disequilibrium (LD) plot of the 18 tag SNPs significantly associated with PGC. A scaled and highly dense LD-based plot within the hotspot genomic region on the chromosome is represented. The positions of the 18 tagged SNPs are indicated with the log$_{10}$-scaled P-values (log$_{10}$(P)) and black/red bars reflecting their relative positive/negative effect sizes, respectively. The gene IDs further detected in TGAS, are highlighted in red color. (c) Circos showing the collinearity in chromosome 5 between Nipponbare and ZS97, especially within the hotspot region (marker as blue block) where collinearity break is evident, whereas earlier reported gene chalk5 regulating chalk was observed distant (small red block); (d-f) TGAS of the candidates with their gene structures and haplotypes, a newly detected hypothetical gene (termed as chalk5.1) (d); LOC_Os05g09520 (e); LOC_Os05g09530 (f), with -log$_{10}$(P)-values of SNPs (marked as blue triangle) present, boxplot showing constructed haplotypes and related phenotype distribution for PGC; 5 selected lines from each of two haplotypes showing extreme phenotype were represented in each case.

Using the reference genome of low chalk line MH 63, two regions of 181kb and 33kb, separated by 76kb region, were revealed to regulate the grain chalkiness in positive and negative manner, respectively (Figure 2). On the other hand, a relatively large region of ~0.9Mb was detected to significantly associate negatively with grain chalkiness, when reference genome of ZS 97 was deployed. This suggested the importance of the target genetic region invariably determining the chalk phenotype.

The diagnostic haplotypes identified are tested for its suitability for relating to chalk phenotypes using the resequenced genomic resources of breeding material. Furthermore, whole genome sequencing (WGS) of low chalk and high chalk varieties developed by IRRI during 1966-2015 was subjected to illumina sequencing with high density coverage (~30X). Alignment of genomic sequence of key hotspot region from low chalk and high chalk lines (from WGS lines and 3000 rice genomes) along with both *indica* references were undertaken to reveal the patterns of the key allelic variants and to relate the haplotypes for low chalk from the breeding material. Exploring the 10 significant SNPs identified in the unknown newly predicted candidate gene *chalk5.1* associated with contrasting chalk haplotypes identified four groups from the IRRI breeding lines (Fig. 3c, d). Notably, allelic variation in group-1 demonstrates narrower range of low chalk phenotype (median 2.8) which are represented by modern breeding material (IR156, IR155, IR154, IR152, IR151, IR150, IR149, IR143, IR142, IR140, IR136, IR135, IR134, IR131, IR28). Also group-2 lines were found to be low chalk (median 3.2) (Fig. 3d). The lines (IR5, IR8, IR111, IR113) which possess inferior haplotypes confer to be of high chalk. Notably, the distinct set of haplotypes were detected mostly in the recently IRRI bred varieties, which have further demonstrated the success of IRRI breeding programme to capture the beneficial alleles for lowering chalk in the modern rice varieties. The superior alleles identified for minimising chalk may potentially be deployed in expediting the rice breeding programme in order to improve the appearance and milling quality attributes, especially for the long slender rice varieties.

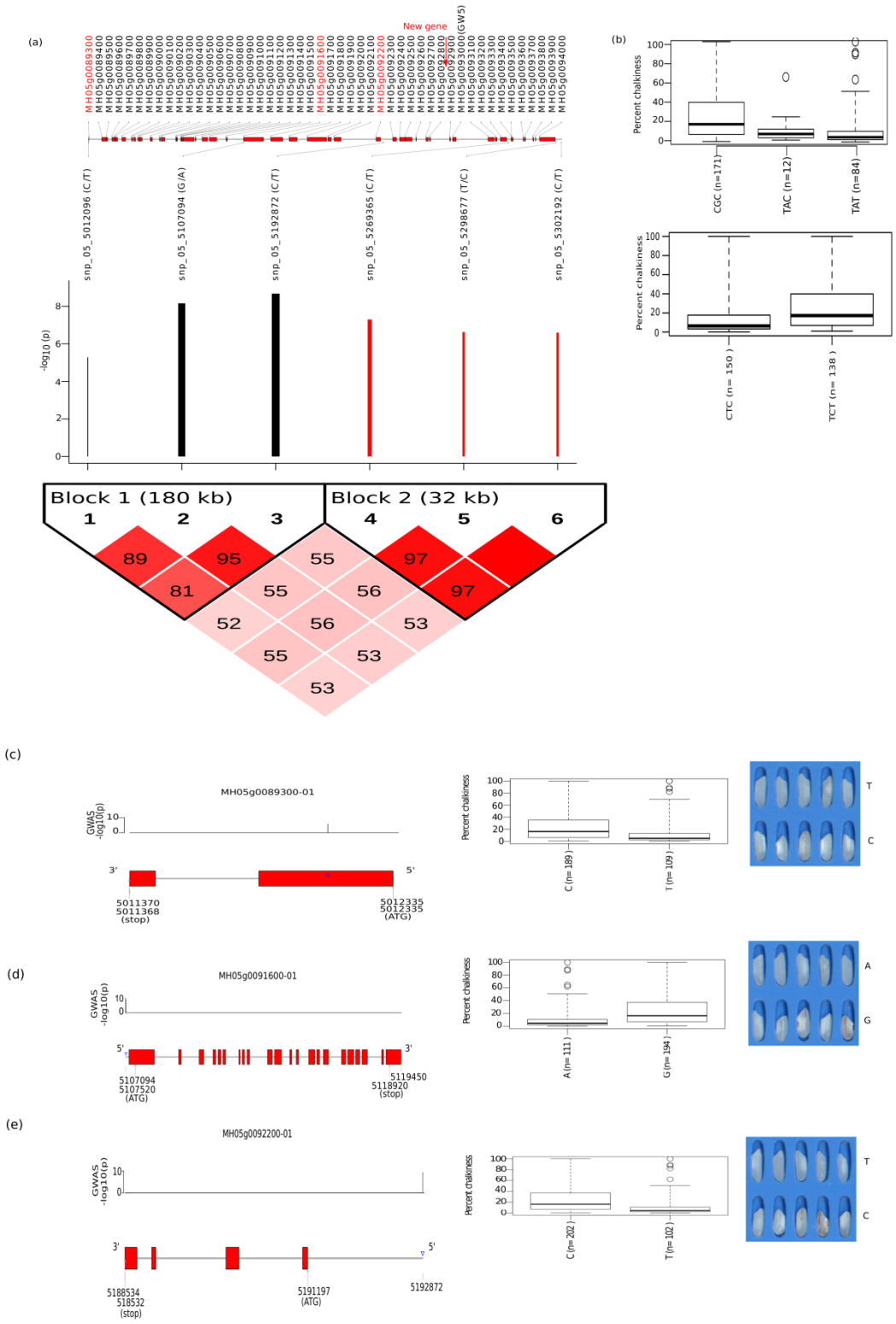


Figure 2: Single-locus GWAS for PGC identified highly associated chromosome 5 hotspot region using Minghui63 (indica) as a reference genome. (a) The linkage disequilibrium (LD) plot of the 6-tagged SNPs significantly associated with PGC. Scaled and highly dense LD-based plot within the hotspot genomic region are represented and the position of newly predicted gene is highlighted in red arrow. The positions of the 6 tagged-SNPs within two LD-blocks were indicated with the log_{10}-scaled P values (log_{10}(P)) and black/red bars reflecting their relative positive/negative effect sizes, respectively. (b) Haplotypes constructed within two LD-blocks and their respective phenotypic values for PGC are

represented as boxplot. (c-e) TGAS of selected genes MH05g0089300 (c), MH05g0091600 (d), MH05g0092200 (e) (highlighted as red in LD-plot) based on high effect SNPs with respective PGC values.

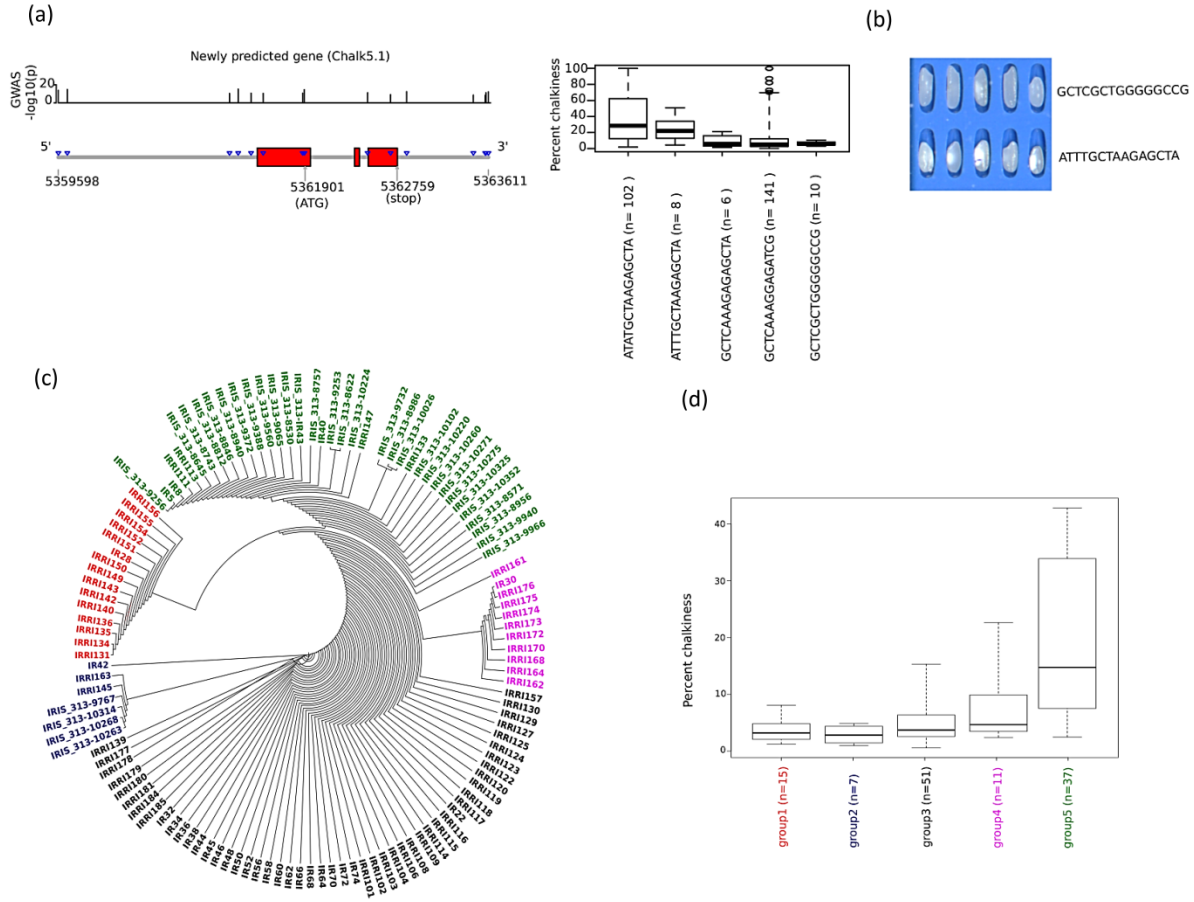


Figure 3: Identification of new genic region through GWAS and its distribution in breeding lines. (a) Predicted gene structure of newly predicted gene chalk5.1 and significant SNPs with log P values constructing the superior haplotypes conferring low chalkiness. (b) Representative figure of grain type exist in respective haplotypes. (c-d) Dendrogram (left) showing the distribution of key alleles of unknown gene chalk5.1 in 92 breeding lines and selected low chalk landraces in form of groups/clusters, with their corresponding phenotypic values (right side). In the boxplot, chalk is represented as the median values.

Development of bi-parental mapping population for chalk

Simultaneously, an effort has been made to develop bi-parental mapping population by hybridizing the low-chalk (IRGC 15577; <2%) and intermediate chalk (IRGC 6550; 23%) parental genotypes to develop F_1, and subsequently selfed to provide a ~3700 F_2 mapping population. The leaf samples from entire population were collected. The progenies of high-density F_2 mapping population exhibited high phenotypic variability for grain chalkiness (Figure 4). Subsequently, samples exhibiting the extreme chalk phenotypes were bulked and have been used for pooling the DNA. Whole genome resequencing of the pooled DNA

of contrasting bulks were performed in high coverage to conduct bulk segregating analysis or QTL-Seq based QTL-mapping[12] to identify the major QTLs/candidates governing grain chalkiness in rice (Figure 4). Presently, the analysis is underway to dissect the underlying genetic basis for the trait.

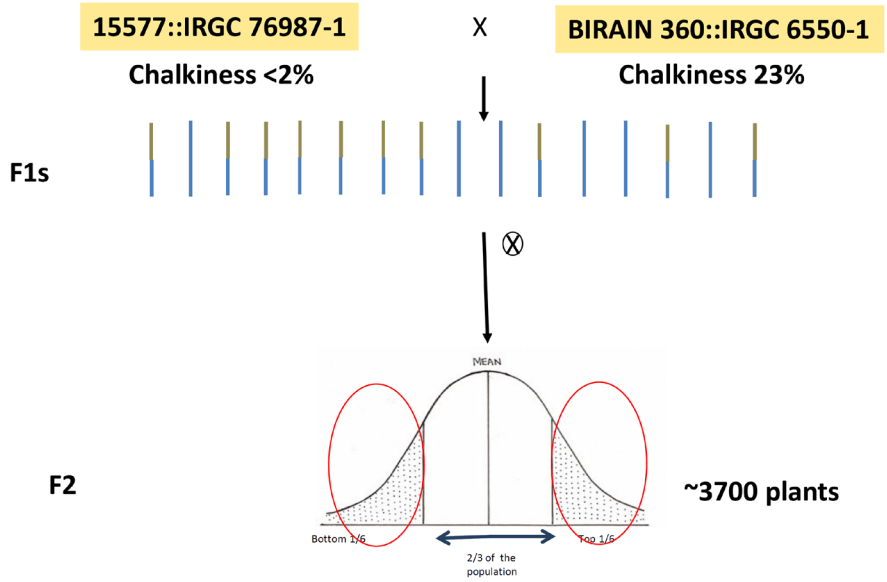


Figure 4: Schematic diagram for the development of bi-parental population. At F_2 stage, both bulks were constituted from the lines with extreme trait phenotype (red ovals) from the entire F_2 progenies.

Functional validation of candidate genes using CRISPR from chromosome 5 fine-mapped targets

In order to functionally validate the candidates identified through GWAS and gene-set analysis (LOC_Os05g09510, LOC_Os05g09520, newly annotated gene and LOC_Os05g09530) we opted for CRISPR (Clustered Regularly Interspaced Short Palindromic Repeats), CRISPR-associated (Cas9) and CRISPR from *Prevotella* and *Francisella* 1 (CPF1), as a gene editing tool. In the year 2, we designed the CRISPR targets against potential candidates LOC_Os05g09510, LOC_Os05g09520, and LOC_Os05g09530, and developed the knock-out constructs for genome editing in the target genes (Figure 5). For designing the knock-out constructs, we adopted the backbones based on CRISPR-cas9 and most advanced CRISPR-cpf1 (especially LbCpf1 from *Lachnospiraceae bacterium* and FnCpf1 from *Francisella tularensis*) system which possesses rather small endonuclease (Figure 5, 6a). In addition, the Cpf1 was identified to leave its target with staggered ends (as opposed to the blunt end cuts of Cas9), said to overcome certain limitation of Cas9 to have more efficiency[13]. We designed the oligos based on our gene targets (preferably three targets per candidate gene to ensure efficiency) and successfully able to clone candidate targets into the designated constructs, further confirmed through Sanger's sequencing (Figure 6b,c). Using the existing set of cassette, no successful events were identified. We are presently revisiting the transformation process, using another Cas9 cassette targeting the unknown gene.

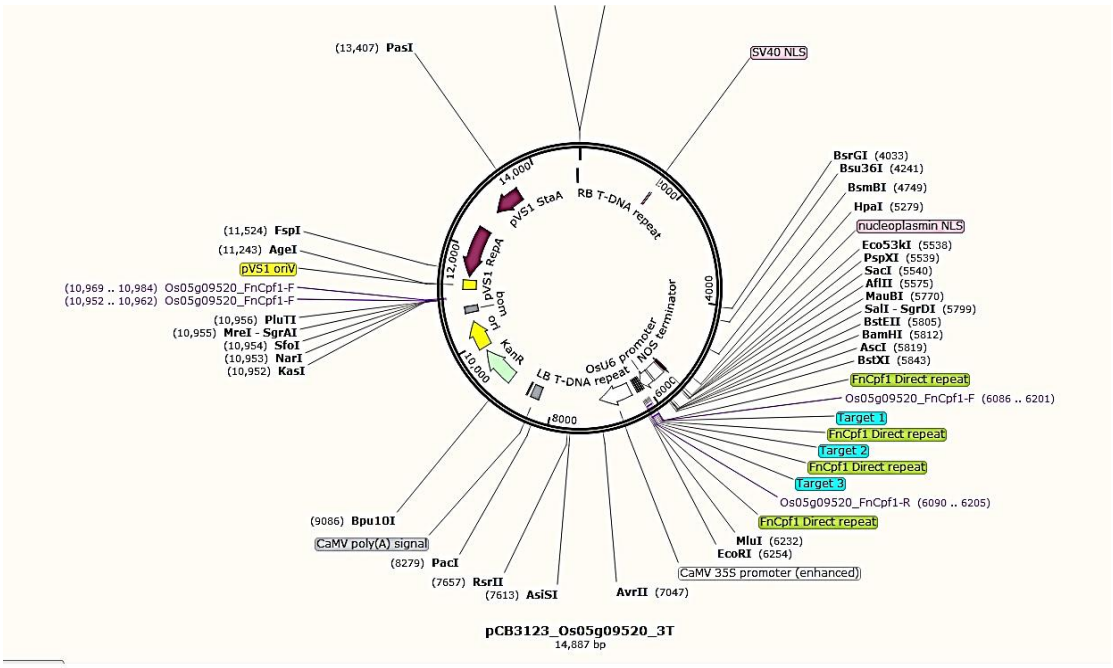
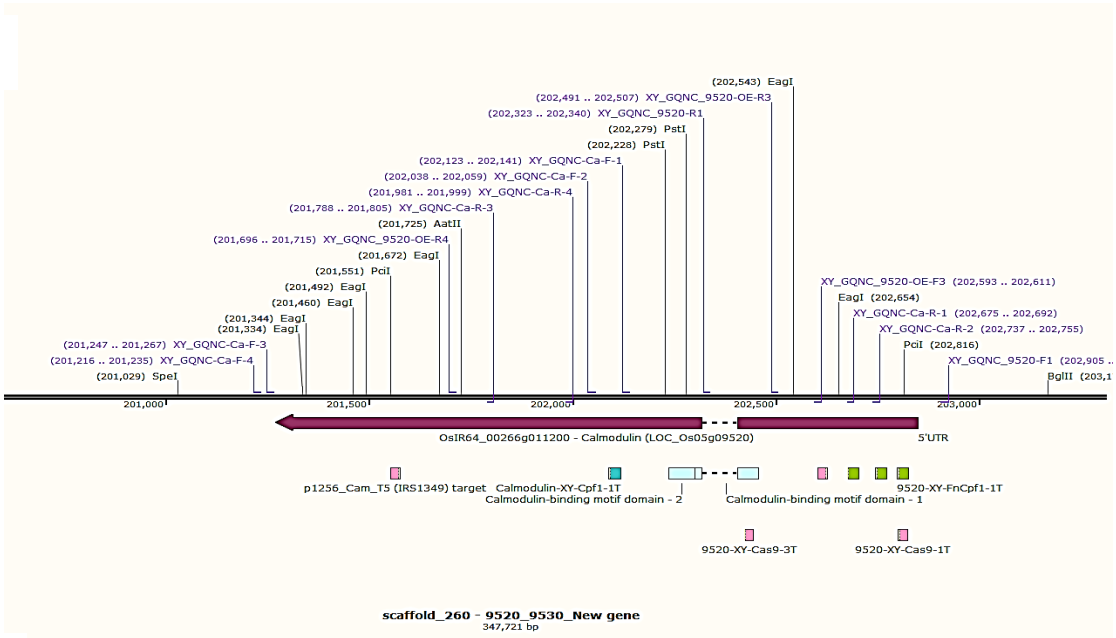
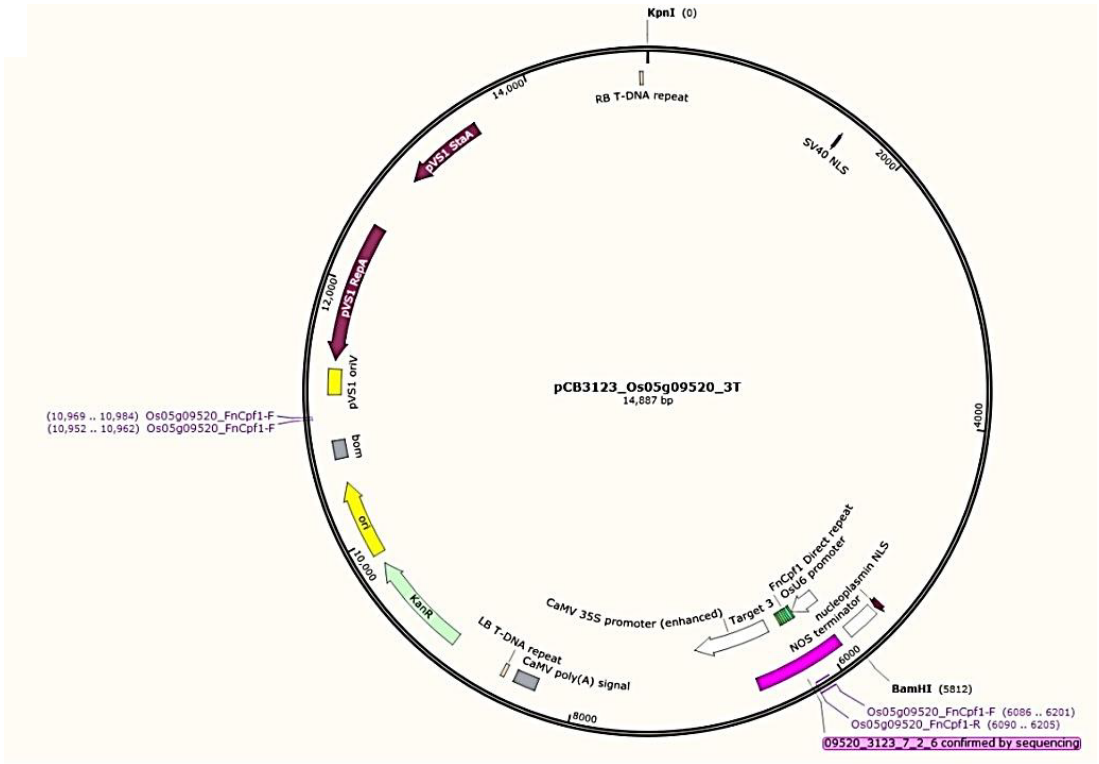


Figure 5: A representative construct designed for one of the candidate (LOC_Os05g09520) identified through GWAS.

(a)


(202,543) EagI
(202,491 .. 202,507) XY_GQNC_9520-OE-R3
(202,323 .. 202,340) XY_GQNC_9520-R1
(202,279) PstI
(202,228) PstI
(202,123 .. 202,141) XY_GQNC-Ca-F-1
(202,038 .. 202,059) XY_GQNC-Ca-F-2
(201,981 .. 201,999) XY_GQNC-Ca-R-4
(201,788 .. 201,805) XY_GQNC-Ca-R-3
(201,725) AatII
(201,696 .. 201,715) XY_GQNC_9520-OE-R4
(201,672) EagI
(201,551) PciI
(201,492) EagI
(201,460) EagI
(201,344) EagI
(201,334) EagI
(201,247 .. 201,267) XY_GQNC-Ca-F-3
(201,216 .. 201,235) XY_GQNC-Ca-F-4
(201,029) SpeI
XY_GQNC_9520-OE-F3 (202,593 .. 202,611)
EagI (202,654)
XY_GQNC-Ca-R-1 (202,675 .. 202,692)
XY_GQNC-Ca-R-2 (202,737 .. 202,755)
PciI (202,816)
XY_GQNC_9520-F1 (202,905 ..
BglII (203,1
201,000
201,500
202,000
202,500
203,000
OsIR64_00266g011200 - Calmodulin (LOC_Os05g09520)
5'UTR
p1256_Cam_T5 (IRS1349) target
Calmodulin-XY-Cpf1-1T
Calmodulin-binding motif domain - 2
Calmodulin-binding motif domain - 1
9520-XY-FnCpf1-1T
9520-XY-Cas9-3T
9520-XY-Cas9-1T
scaffold_260 - 9520_9530_New gene
347,721 bp

(b)


KpnI (0)
RB T-DNA repeat
SV40 NLS
pVS1 StaA
pVS1 RepA
pVS1 oriV
pCB3123_Os05g09520_3T
14,887 bp
(10,969 .. 10,984) Os05g09520_FnCpf1-F
(10,952 .. 10,962) Os05g09520_FnCpf1-F
bom
ori
KanR
LB T-DNA repeat
CaMV poly(A) signal
CaMV 35S promoter (enhanced)
Target 3
FnCpf1 Direct repeat
OsU6 promoter
nucleoplasmin NLS
NOS terminator
BamHI (5812)
Os05g09520_FnCpf1-F (6086 .. 6201)
Os05g09520_FnCpf1-R (6090 .. 6205)
09520_3123_7_2_6 confirmed by sequencing
2000
4000
6000
8000
10,000
12,000
14,000

(c)

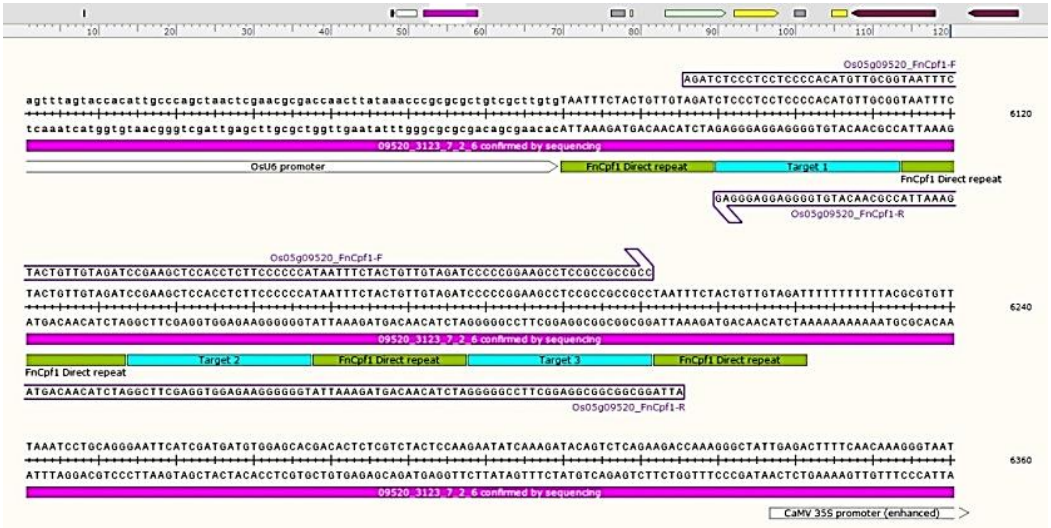


Figure 6: A schematic representation of gene targets of a candidate. (a) Representative CRISPR targets for the candidate (LOC_Os05g09520) with cas9 or cpf1 (mentioned as green, sky blue and pink colour blocks) targets; (b) Construct showing the target region confirmed through Sanger's sequencing, and (c) a zoomed-in view.

8 Impacts

8.1 Scientific impacts – now and in 5 years

An invited methodology article highlighting the screening techniques addressing milling quality and its stability, as well imaging techniques related to chalk has been discussed in the two articles, which have been published in Methods in Molecular Biology (Springer) book series.

In addition, we have finalized the research article of deciphering the molecular mechanisms of chalk and highlighting the importance of genetic region to reduce chalk in modern breeding programs. The compiled submitted manuscript is presently under review in *Journal of Experimental Botany*.

The lecture and the workshop session on unravelling milling potential and chalk have been highlighted in International Hybrid Rice Symposium 2017. In addition, we have interacted with various private stake holders to create awareness about the importance of chalk and milling quality in hybrid programs through hybrid rice consortium.

8.2 Capacity impacts – now and in 5 years

IRRI has established a Centre of excellence for rice value addition with grain quality in Varanasi, India. In addition, the state of art technologies with genomics focus will remain in IRRI headquarters. The knowledge we have generated to improve breeding programs to enhance quality can be better tapped in future efforts. Until now we have mainly addressed strategies to reduce chalk. On top of this we need to combine the genetic regions which contribute to enhance the percent head rice yield (% of intact grains after milling) with reduced chalk. These set of traits are valued by millers and consumers in the rice value chain.

8.3 Community impacts – now and in 5 years

The direct beneficiaries of the outcome of this small research activity are breeders and rice value chain stake holders. The identified genetic region identified on chromosome has been fine mapped to specific gene targets. The gene based haplotypes conferring differential chalk phenotypes have been validated from the independent years and found to be relevant in both dry and wet season. We also confirmed these haplotypes from more than 100 breeding material through sequencing. This will help set up the value of markers to breed low chalk material. Once used by breeding programs to release low chalk varieties, the markers' benefits will be received by growers and consumers.

8.3.1 Economic impacts

These diagnostic markers will be of high interest to both public and private breeding institutions to ensure selections of low chalk material in rice. The functional markers developed under this small research activity will be used to reduce the incidence of chalk in inbreds and hybrids. Reducing the incidence of chalk can significantly increase market value and improve milling yield, which translates to higher income for the farmer and lower the market price of rice for the consumers.

8.3.2 Social impacts

Rice has been one of the most economically important crops in Asia which faces key challenges of poor milling quality and high chalk. Conversely, the rice varieties with low chalk and high milling quality improves the livelihood of the farmers or growers by giving them higher returns of their harvest. Therefore, reducing chalk in the breeding programs has been an important issue, which needs to be tackled to ensure success of breeding programs.

8.3.3 Environmental impacts

Future low chalk rice varieties are not expected to have any sort of negative environmental effect like current rice varieties. No change is expected from the agronomy point of view.

8.4 Communication and dissemination activities

The validated markers lowering chalk in the breeding material will be communicated to national partners. This information will be communicated to various stake holders to initiate public and private partnership. In addition, the top low chalk breeding lines identified from the breeding programs will be provided to national partners, various value chain stake holders and millers to ensure higher acceptability.

9 Conclusions and recommendations

9.1 Conclusions

In brief, the fine mapped candidates and haplotypes of chromosome 5 identified for reducing chalk have been successfully demonstrated in IRRI breeding programs. To provide economic benefits to the farmers, superior alleles and haplotypes from the fine mapped *chalk 5.1* gene need to be deployed in rice breeding programs to improve appearance and milling quality attributes for newly released *indica* varieties.

9.2 Recommendations

The direct beneficiaries of the outcome of this small research activity are consumers, farmers, and millers. Up to 50% of rice is lost because of postharvest and grain processing losses. The formation of chalk can indirectly contribute to these losses, as a consequence of the decreased head rice and milled rice yields of high chalk varieties. To create impact, we need to develop a holistic based solution to develop climate resilient varieties with high milling yield and lower chalk through pyramiding the identified genetic regions and testing the performance of lines in the regional programs following best post-harvest practices. The recommendation as per business models are the need of combining traits to enhance head rice yield and reduce chalk in the hybrid rice programs to enhance the better acceptability. There is a need to establish public-private partnership to support breeding programs of inbreds and hybrids and connect with value chain stake holders to meet the requirements of export market. Private companies signified their interest in conducting collaborative research on the rice chalkiness trait and its association with minimizing milling yield losses, if we manage to bring additional resources from ACIAR.

10 References

10.1 References cited in report

References

1 Del Rosario, A. R., Briones, V. P., Vidal, A. J. & Juliano, B. O. Composition and endosperm structure of developing and mature rice kernel. *Cereal Chem.* **45**, 225–235 (1968).

2 Lisle, A. J., Martin, M. & Fitzgerald, M. A. Chalky and translucent rice grains differ in starch composition and structure and cooking properties. *Cereal Chem.* **77**, 627–632 (2000).

3 Zhao, X. & Fitzgerald, M. Climate change: implications for the yield of edible rice. *PLoS One* **8**, e66218, doi:10.1371/journal.pone.0066218 (2013).

4 Siebenmorgen, T. J., Grigg, B. C. & Lanning, S. B. Impacts of preharvest factors during kernel development on rice quality and functionality. *Annual review of food science and technology* **4**, 101-115 (2013).

5 Kang, H. M. *et al.* Variance component model to account for sample structure in genome-wide association studies. *Nature Genet.* **42**, 348-354, doi:10.1038/ng.548 (2010).

6 Wen, Y.-J. *et al.* Methodological implementation of mixed linear models in multi-locus genome-wide association studies. *Briefings in bioinformatics*, bbw145 (2017).

7 Wang, S.-B. *et al.* Improving power and accuracy of genome-wide association studies via a multi-locus mixed linear model methodology. *Scientific reports* **6** (2016).

8 Tamba, C. L., Ni, Y.-L. & Zhang, Y.-M. Iterative sure independence screening EM-Bayesian LASSO algorithm for multi-locus genome-wide association studies. *PLoS computational biology* **13**, e1005357 (2017).

9 Misra, G. *et al.* Whole genome sequencing-based association study to unravel genetic architecture of cooked grain width and length traits in rice. *Scientific reports* **7**, 12478 (2017).

10 Zhang, J. *et al.* Extensive sequence divergence between the reference genomes of two elite indica rice varieties Zhenshan 97 and Minghui 63. *Proceedings of the National Academy of Sciences, USA* **113**, E5163-5171, doi:10.1073/pnas.1611012113 (2016).

11 Li, Y. *et al.* Chalk5 encodes a vacuolar H+-translocating pyrophosphatase influencing grain chalkiness in rice. *Nat Genet* **46**, 398–404, doi:10.1038/ng.2923 (2014).

12 Takagi, H. *et al.* QTL-seq: rapid mapping of quantitative trait loci in rice by whole genome resequencing of DNA from two bulked populations. *Abstract workform* **74**, 174-183 (2013).

13 Ledford, H. Beyond CRISPR: A guide to the many other ways to edit a genome. *Nature News* **536**, 136 (2016).

10.2 List of publications produced by project

- Lapis JR, Cuevas RPO, Sreenivasulu N, Molina L. 2019. Measuring Head Rice Recovery in Rice. *Methods Mol Biol*.; 1892:89-98. doi: 10.1007/978-1-4939-8914-0_5.
- Butardo VM Jr, Sreenivasulu N. 2019. Improving Head Rice Yield and Milling Quality: State-of-the-Art and Future Prospects. *Methods Mol Biol.*; 1892:1-18. doi: 10.1007/978-1-4939-8914-0_1.
- Santos MV, Cuevas RPO, Sreenivasulu N, Molina L. 2019. Measurement of Rice Grain Dimensions and Chalkiness, and Rice Grain Elongation Using Image Analysis. *Methods Mol Biol.*; 1892:99-108. doi: 10.1007/978-1-4939-8914-0_6.
- Sreenivasulu, N., Butardo Jr, V.M., Misra, G., Cuevas, R.P., Anacleto, R. and Kavi Kishor, P.B., 2015. Designing climate-resilient rice with ideal grain quality suited for high-temperature stress. *Journal of Experimental Botany*, 66(7), pp.1737-1748.
- Misra, G., Anacleto, R., Badoni, S., Butardo Jr, V., Molina L., Graner A., Demont M., Morell M.K and Sreenivasulu N. Dissecting the genome wide genetic variants of milling and appearance quality traits in rice. (*Under revision in the Journal of Experimental Botany*).