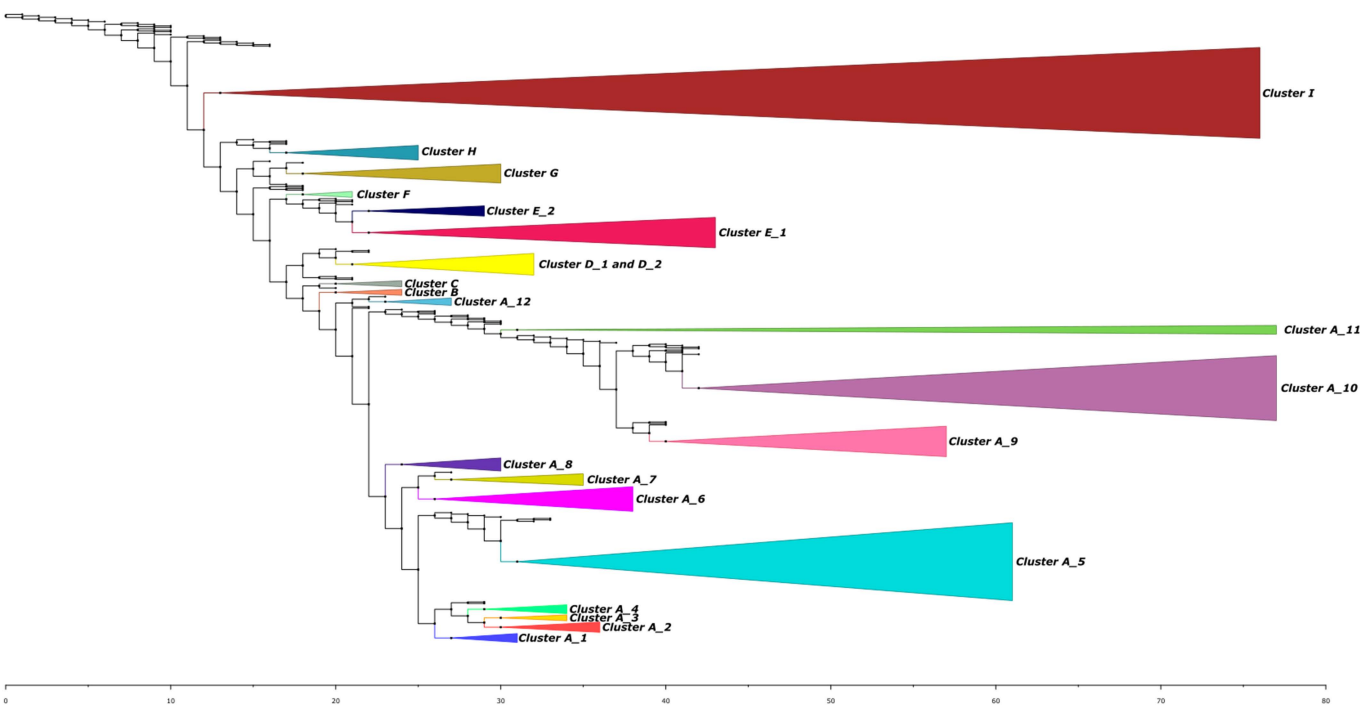


Fig. S1: A phylogenetic tree generated from the multiple sequence alignment (MSA) of the *NGR5* promoter and gene body regions, trimmed at a 5% gap threshold, showing the clustering of phylogenetically homogeneous core-set accessions. The x-axis depicts phylogenetic distance between clusters, and the triangle's size represents the number of accessions per cluster.

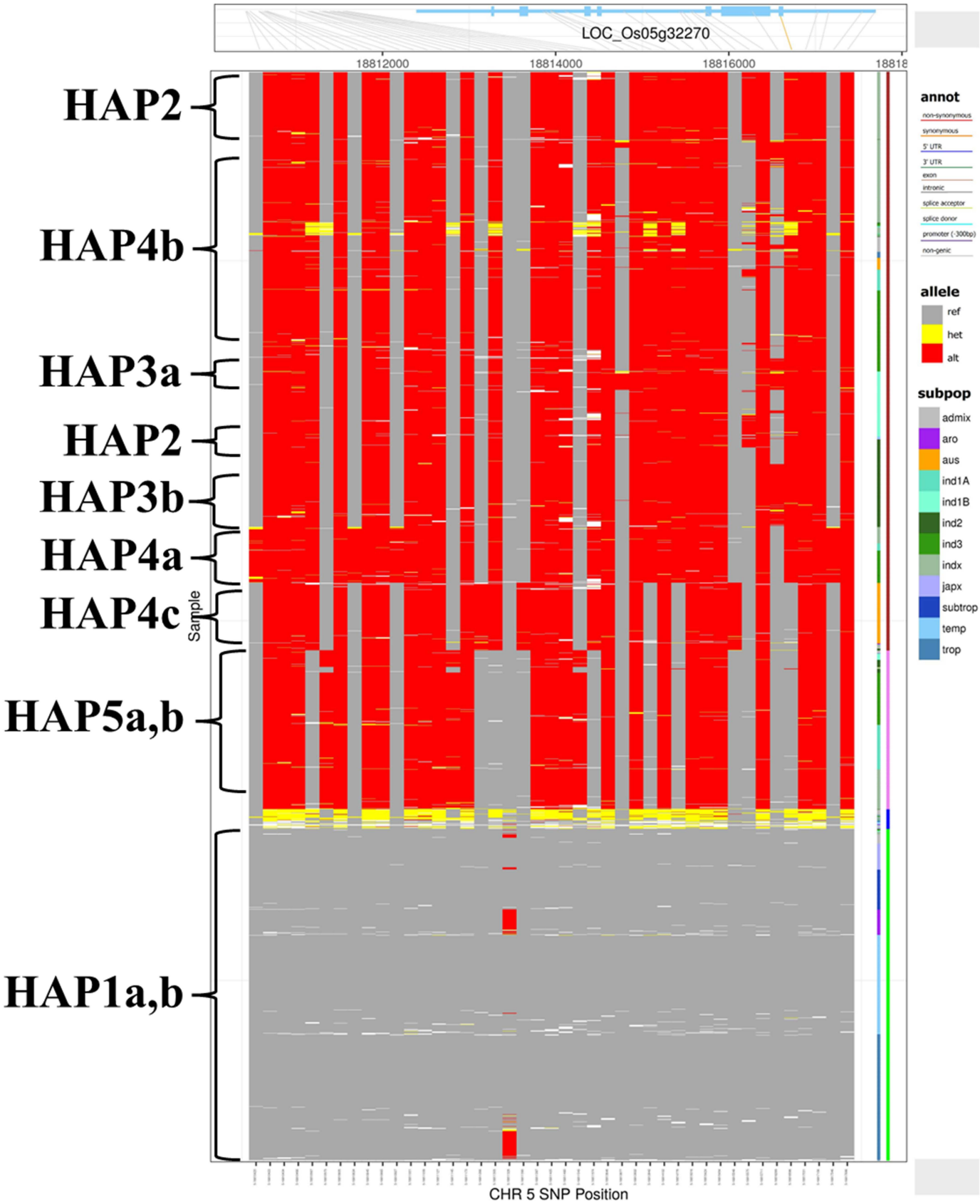
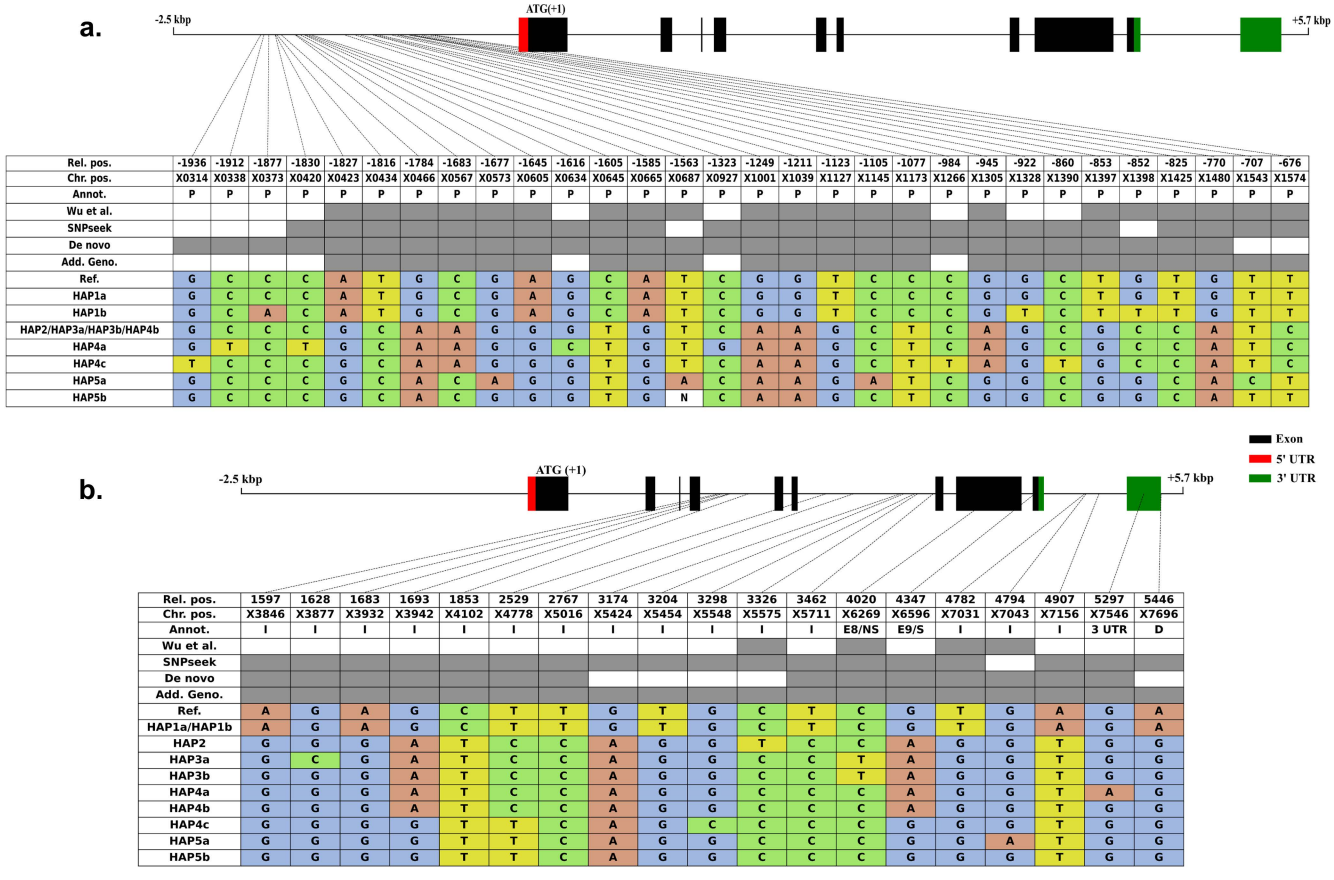


Fig. S2. Graphical representation of the *NGR5* haplotypes identified from the 3KRG SNP-Seek dataset. SNP positions across the *NGR5* gene body and promoter region are shown along the X-axis, while individual 3KRG accessions are arranged along the Y-axis. The top panel shows the *NGR5* gene structure and corresponding sequence variation, with blue boxes indicating exons, blue lines indicating introns, and the upstream, unfilled region representing the 2 kb promoter region. Grey, red, and yellow colours indicate reference alleles, alternate alleles, and heterozygous loci, respectively.



Rel. pos.	-1936	-1912	-1877	-1830	-1827	-1816	-1784	-1683	-1677	-1645	-1616	-1605	-1585	-1563	-1323	-1249	-1211	-1123	-1105	-1077	-984	-945	-922	-860	-853	-852	-825	-770	-707	-676
Chr. pos.	X0314	X0338	X0373	X0420	X0423	X0434	X0466	X0567	X0573	X0605	X0634	X0645	X0665	X0687	X0927	X1001	X1039	X1127	X1145	X1173	X1266	X1305	X1328	X1390	X1397	X1398	X1425	X1480	X1543	X1574
Annot.	P	P	P	P	P	P	P	P	P	P	P	P	P	P	P	P	P	P	P	P	P	P	P	P	P	P	P	P	P	P
Wu et al.																														
SNPseek																														
De novo																														
Add. Geno.																														
Ref.	G	C	C	C	A	T	G	C	G	A	G	C	A	T	C	G	G	T	C	C	C	G	G	C	T	G	T	G	T	T
HAP1a	G	C	C	C	A	T	G	C	G	A	G	C	A	T	C	G	G	T	C	C	C	G	G	C	T	G	T	G	T	T
HAP1b	G	C	A	C	A	T	G	C	G	A	G	C	A	T	C	G	G	T	C	C	C	G	T	C	T	T	T	G	T	T
HAP2/HAP3a/HAP3b/HAP4b	G	C	C	C	G	C	A	A	G	G	G	T	G	T	C	A	A	G	C	T	C	A	G	C	G	C	C	A	T	C
HAP4a	G	T	C	T	G	C	A	A	G	G	C	T	G	T	G	A	A	G	C	T	C	A	G	C	G	C	C	A	T	C
HAP4c	T	C	C	C	G	C	A	A	G	G	G	T	G	T	C	A	A	G	C	T	T	A	G	T	G	C	C	A	T	C
HAP5a	G	C	C	C	G	C	A	C	A	G	G	T	G	A	C	A	A	G	A	T	C	G	G	C	G	G	C	A	C	T
HAP5b	G	C	C	C	G	C	A	C	G	G	G	T	G	N	C	A	A	G	C	T	C	G	G	C	G	G	C	A	T	T

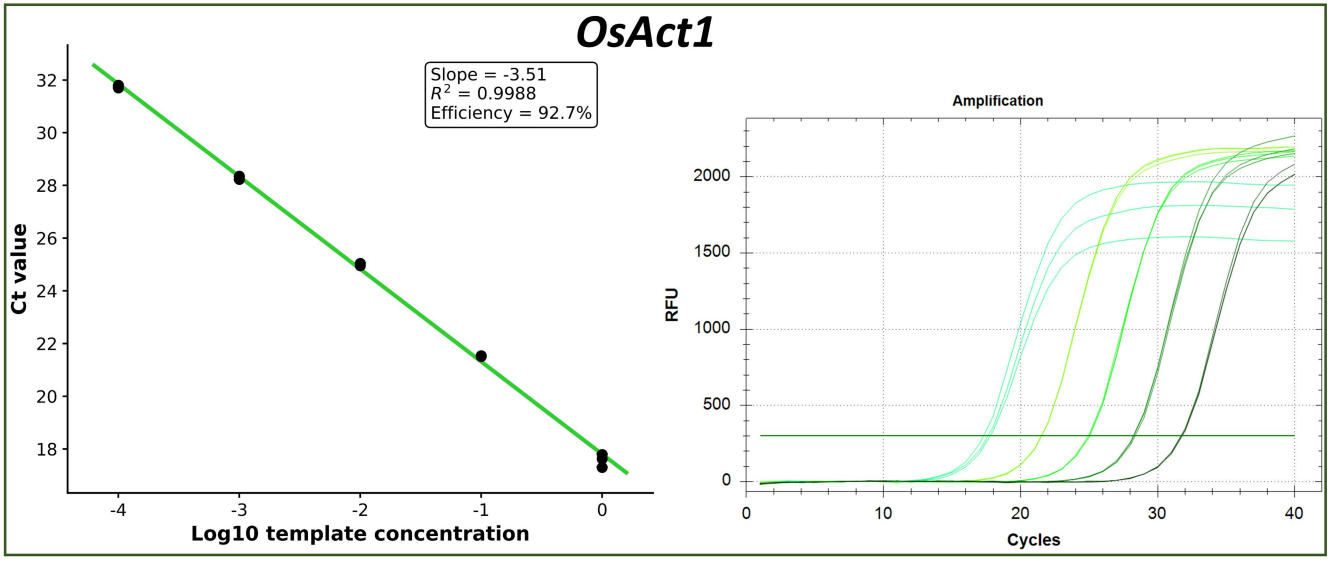
Rel. pos.	1597	1628	1683	1693	1853	2529	2767	3174	3204	3298	3326	3462	4020	4347	4782	4794	4907	5297	5446
Chr. pos.	X3846	X3877	X3932	X3942	X4102	X4778	X5016	X5424	X5454	X5548	X5575	X5711	X6269	X6596	X7031	X7043	X7156	X7546	X7696
Annot.	I	I	I	I	I	I	I	I	I	I	I	I	E8/NS	E9/S	I	I	I	3 UTR	D
Wu et al.																			
SNPseek																			
De novo																			
Add. Geno.																			
Ref.	A	G	A	G	C	T	T	G	T	G	C	T	C	G	T	G	A	G	A
HAP1a/HAP1b	A	G	A	G	C	T	T	G	T	G	C	T	C	G	T	G	A	G	A
HAP2	G	G	G	A	T	C	C	A	G	G	T	C	C	A	G	G	T	G	G
HAP3a	G	C	G	A	T	C	C	A	G	G	C	C	T	A	G	G	T	G	G
HAP3b	G	G	G	A	T	C	C	A	G	G	C	C	T	A	G	G	T	G	G
HAP4a	G	G	G	A	T	C	C	A	G	G	C	C	C	A	G	G	T	A	G
HAP4b	G	G	G	A	T	C	C	A	G	G	C	C	C	A	G	G	T	G	G
HAP4c	G	G	G	G	T	T	C	A	G	C	C	C	C	G	G	G	T	G	G
HAP5a	G	G	G	G	T	T	C	A	G	G	C	C	C	G	G	A	T	G	G
HAP5b	G	G	G	G	T	T	C	A	G	G	C	C	C	G	G	G	T	G	G

Fig. S3: Diagram showing haplotype-specific allelic variation in the *NGR5* promoter (a.) and gene body (b.) individually. The allelic variation presented here is a compilation of SNPs identified using three different approaches (as described in Figure 3). The 1st row (Rel. pos.) indicates relative chromosomal position with respect to the translation start site (TSS), while the 2nd row (Chr. pos.) indicates the actual chromosomal coordinates of the *NGR5* locus based on the IRGSP-1.0 reference genome. The first four digits of all positions (1881) are identical and therefore abbreviated as 'X'. The 3rd row (Annot.) indicates the genomic annotation of each SNP site, where P: promoter, I: intron, E: exon, UTR: untranslated region, and D: downstream. The result of an SNP in an exon is classified as a NS (non-synonymous) or S (synonymous) mutation in the amino acid sequence. Rows 4–7 indicate SNP sites reported by Wu et al. (2020) and those identified using the different approaches used in this study. Row 8 (Ref.) indicates the nucleotide present at each position in the Nipponbare IRGSP-1.0 reference genome. Rows 9–15 (a.) and row 9-17 (b.) indicate the nucleotides present at the corresponding positions in the haplotypes.

a.

Primer Name	Sequence (5'-> 3')	Amplicon size (bp)
OsActin1_F	CTTCATAGGAATGGAAGCTGCGGGTA	197
OsActin1_R	CGACCACCTTGATCTTCATGCTGCTA	
NGR5_CDE6.1_F	GCAGTGCCTTTTCCAGGGGTTTAC	137
NGR5_CDE8.1_R	CCCATCCAGCGTGATGTCCTTC	

b.



c.

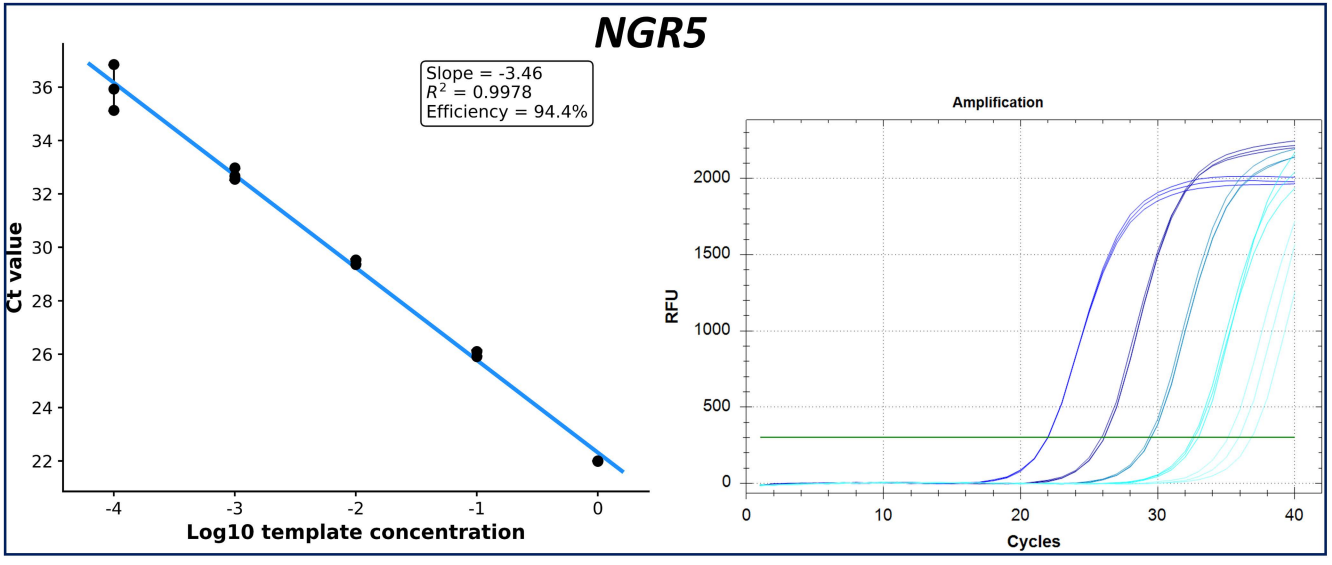


Fig. S4. Standard curve of qPCR primers. a. The table provides the primer sequence used in qPCR. Standard curve analysis of b. *OsAct1* and c. *NGR5* used in qPCR analysis

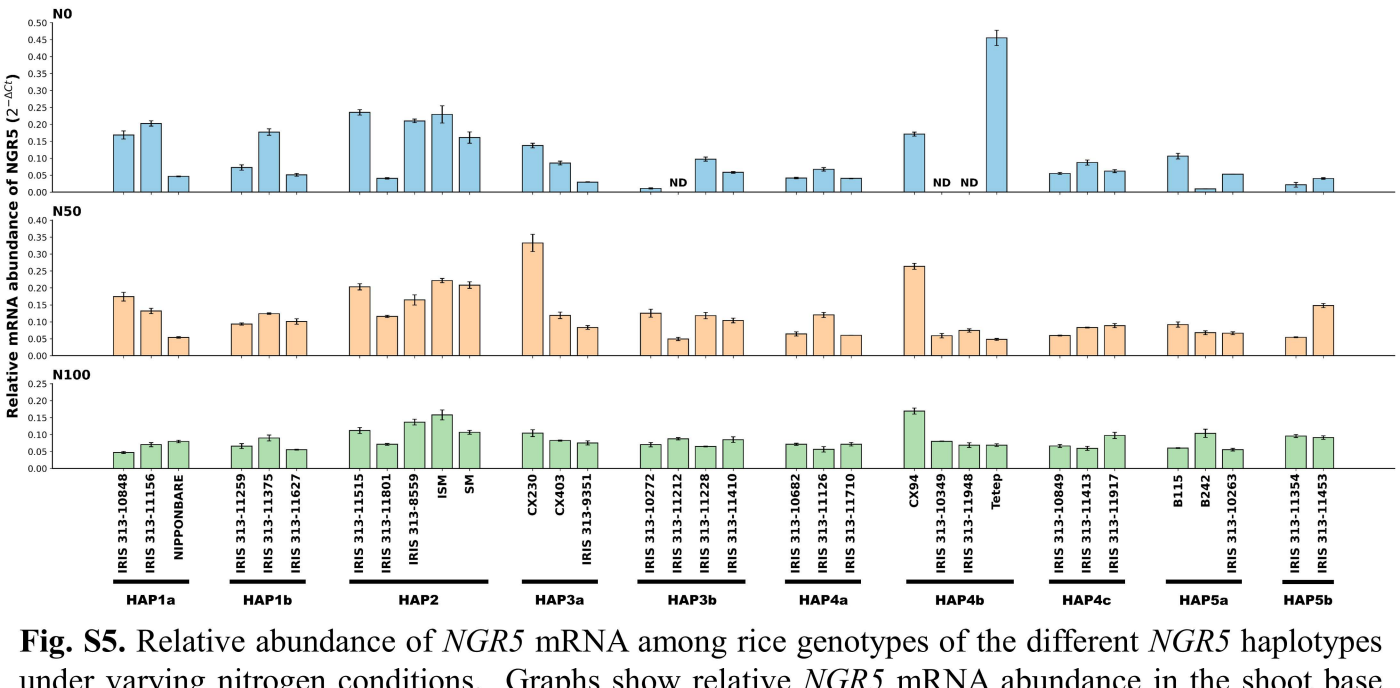
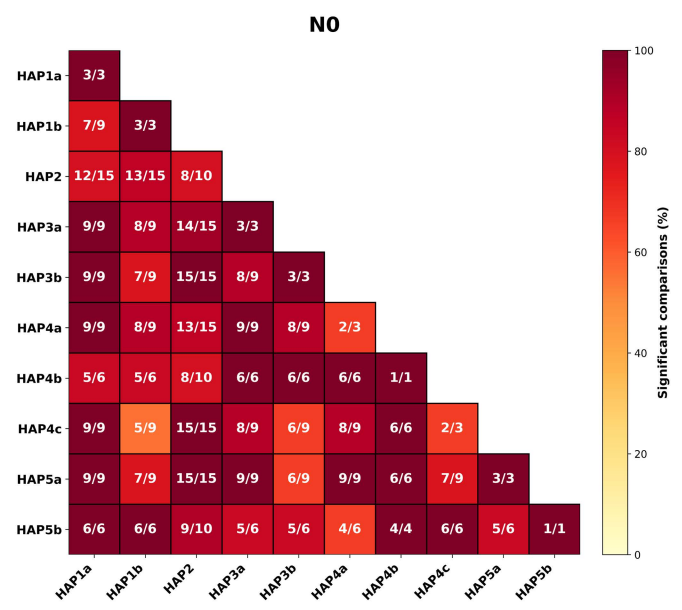
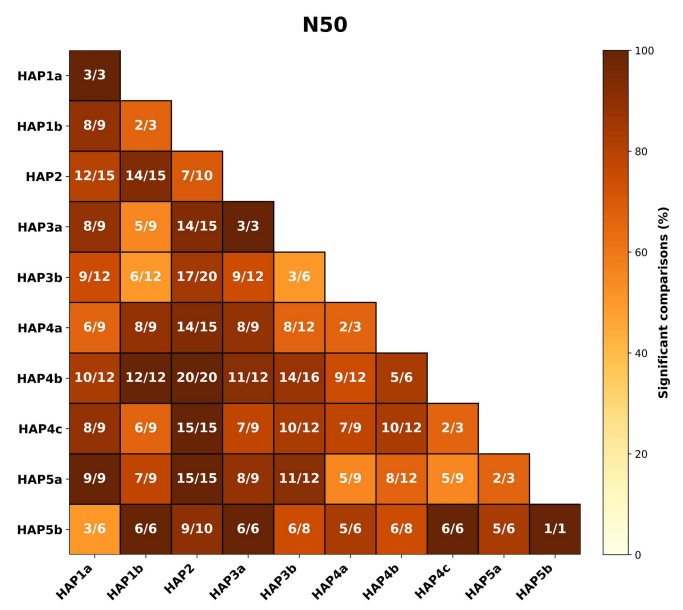


Fig. S5. Relative abundance of *NGR5* mRNA among rice genotypes of the different *NGR5* haplotypes under varying nitrogen conditions. Graphs show relative *NGR5* mRNA abundance in the shoot base tissue of rice genotypes grown under nitrogen-deficient (N0; 0 mM NH_4NO_3), moderate nitrogen (N50; 0.715 mM NH_4NO_3), and high nitrogen (N100; 1.43 mM NH_4NO_3) conditions. Data are shown as mean ± SD of three technical replicates. Relative mRNA abundance was calculated as $2^{-\Delta Ct}$ using *OsActin1* as an internal control.





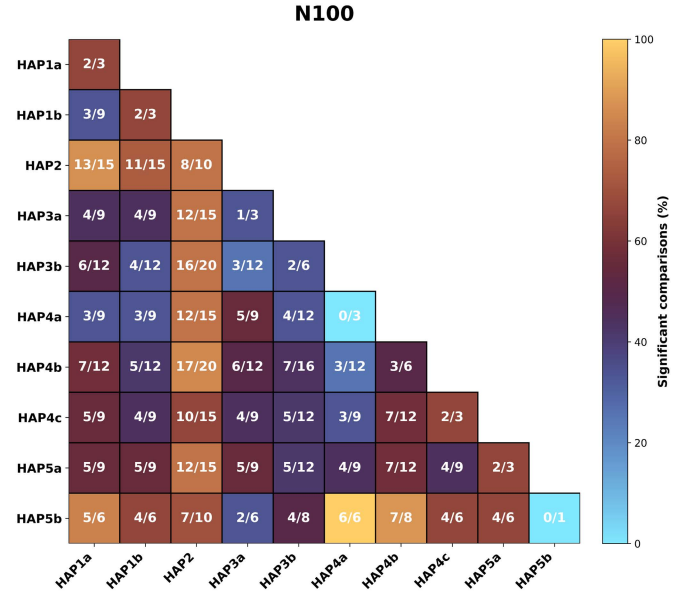


Fig. S6. Pairwise comparisons of *NGR5* expression levels reveal statistically significant differences both within and among haplotypes. The diagrams show pairwise comparisons of *NGR5* expression levels between genotypes across all haplotypes under a. N0, b. N50 and c. N100 conditions.

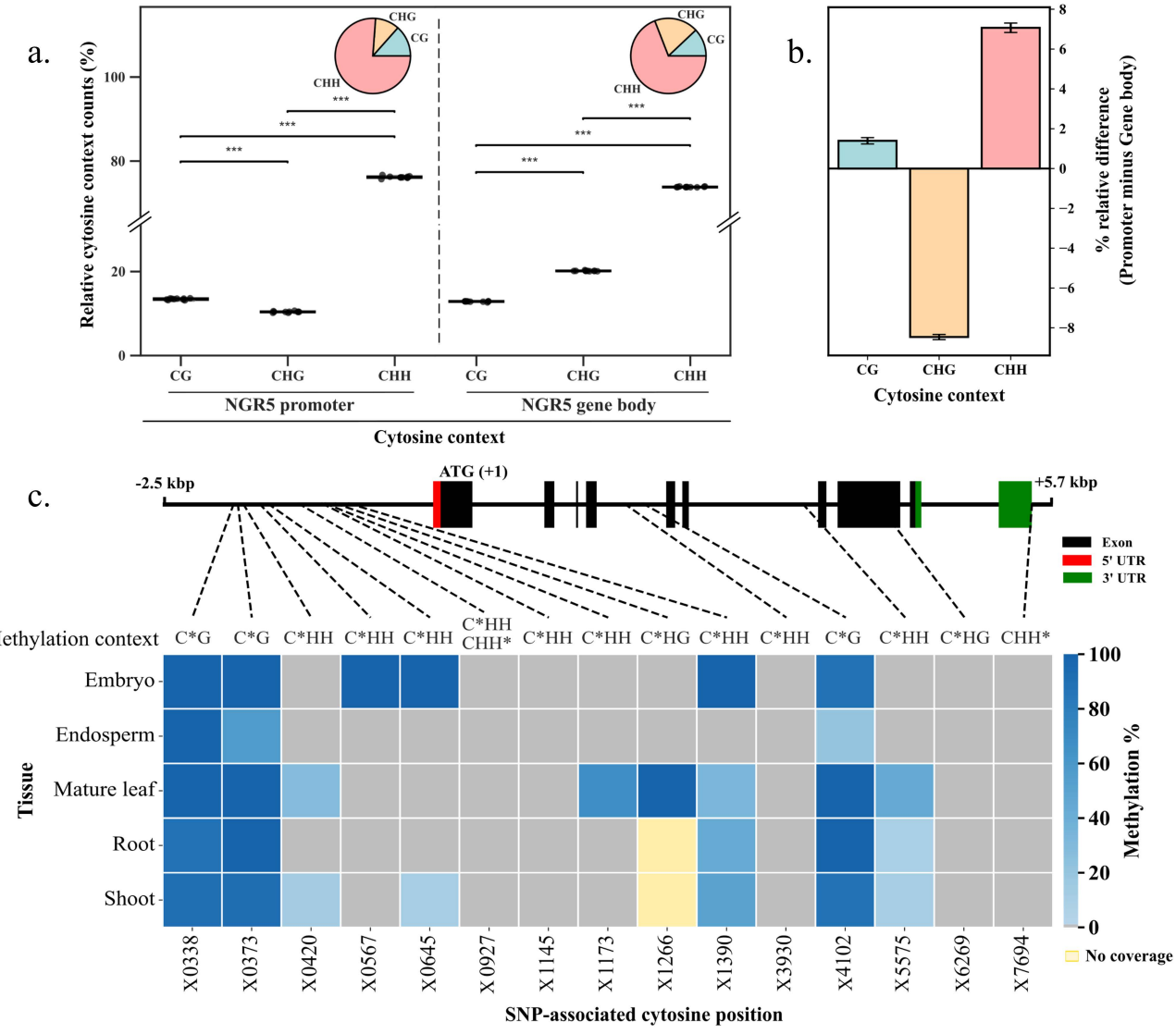


Fig. S7. Distribution and methylation status of haplotype-associated cytosine contexts in *NGR5*. a. Distribution of cytosine methylation contexts (CG, CHG, and CHH) across the promoter and gene body regions of the ten *NGR5* haplotypes. Boxplots indicate the relative abundance of each cytosine context among haplotypes, while inset pie charts represent the average proportion of CG, CHG, and CHH contexts in the promoter and gene body regions. b. Percentage difference in the abundance of CG, CHG, and CHH contexts between the promoter and the gene body regions, calculated as promoter minus gene body. c. Tissue-specific methylation status of cytosine positions associated with haplotype-defining SNPs based on publicly available bisulfite sequencing data of Nipponbare. The heatmap depicts methylation percentages across different tissues. The x-axis labels indicate the chromosomal coordinates of the *NGR5* locus based on the IRGSP-1.0 reference genome. The common first four digits (1881) are abbreviated as 'X'.

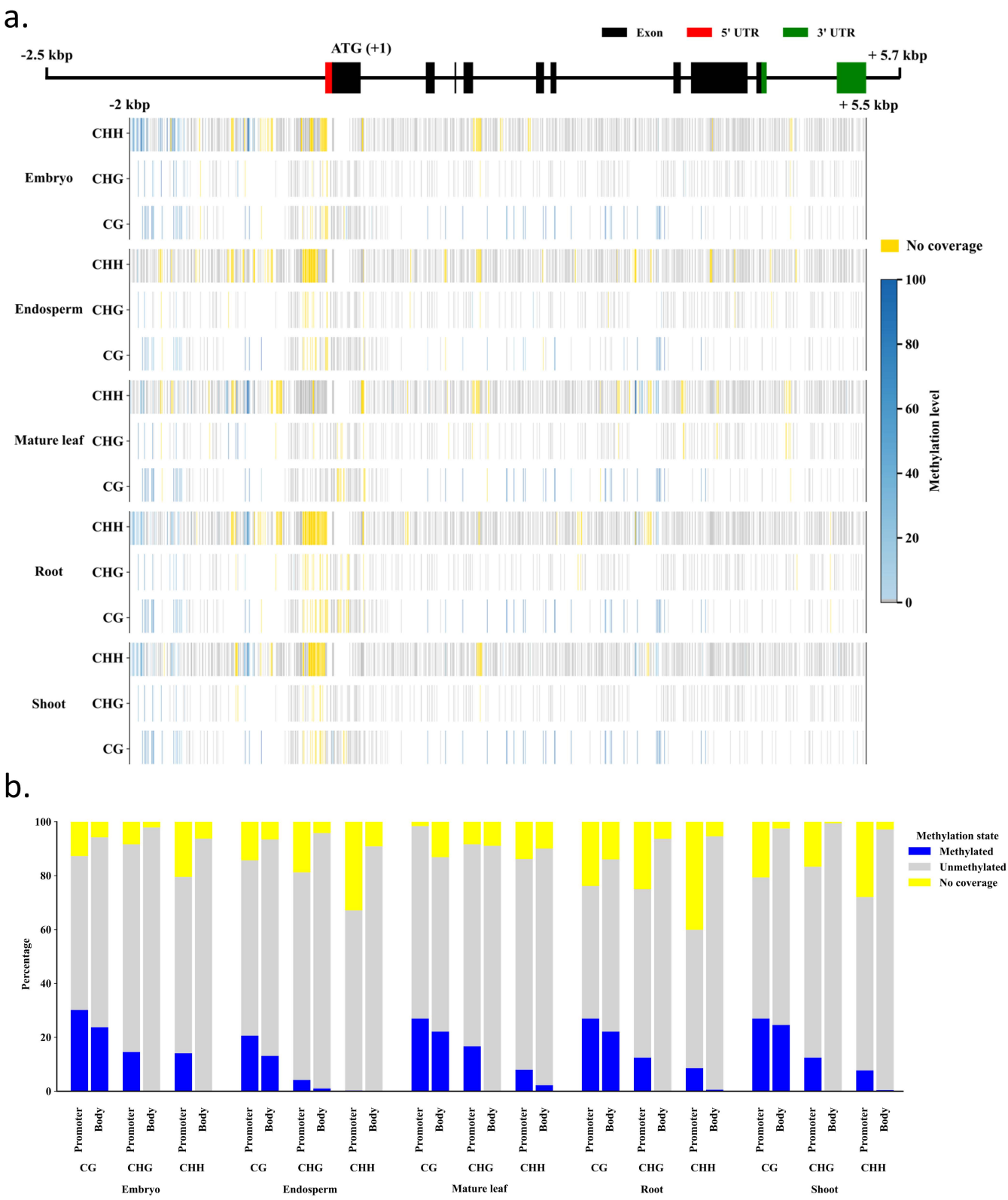


Fig. S8. Cytosine methylation landscape across the *NGR5* promoter and gene body.
a. Heatmap depicts the methylation landscape of all possible cytosine contexts (CG, CHG, and CHH) in the 2 kb upstream promoter region and the *NGR5* gene body (~5.5 kb downstream from the ATG (+1) site) across different tissues Nipponbare rice tissues. The upper panel represents the *NGR5* gene structure, where black boxes indicate exons, black lines indicate introns, and red and green regions represent the 5′ and 3′ untranslated regions (UTRs), respectively. Methylation levels are represented as a blue intensity gradient, where darker blue indicates higher methylation percentage and grey indicates absence of methylation, while yellow denotes positions lacking sequencing coverage. Methylation heat map was generated using publicly available bisulfite sequencing data b. Bar graph representing the overall percentage distribution of methylated, unmethylated, and the non-covered cytosine sites across the promoter and gene body regions in different cytosine contexts across different tissues.

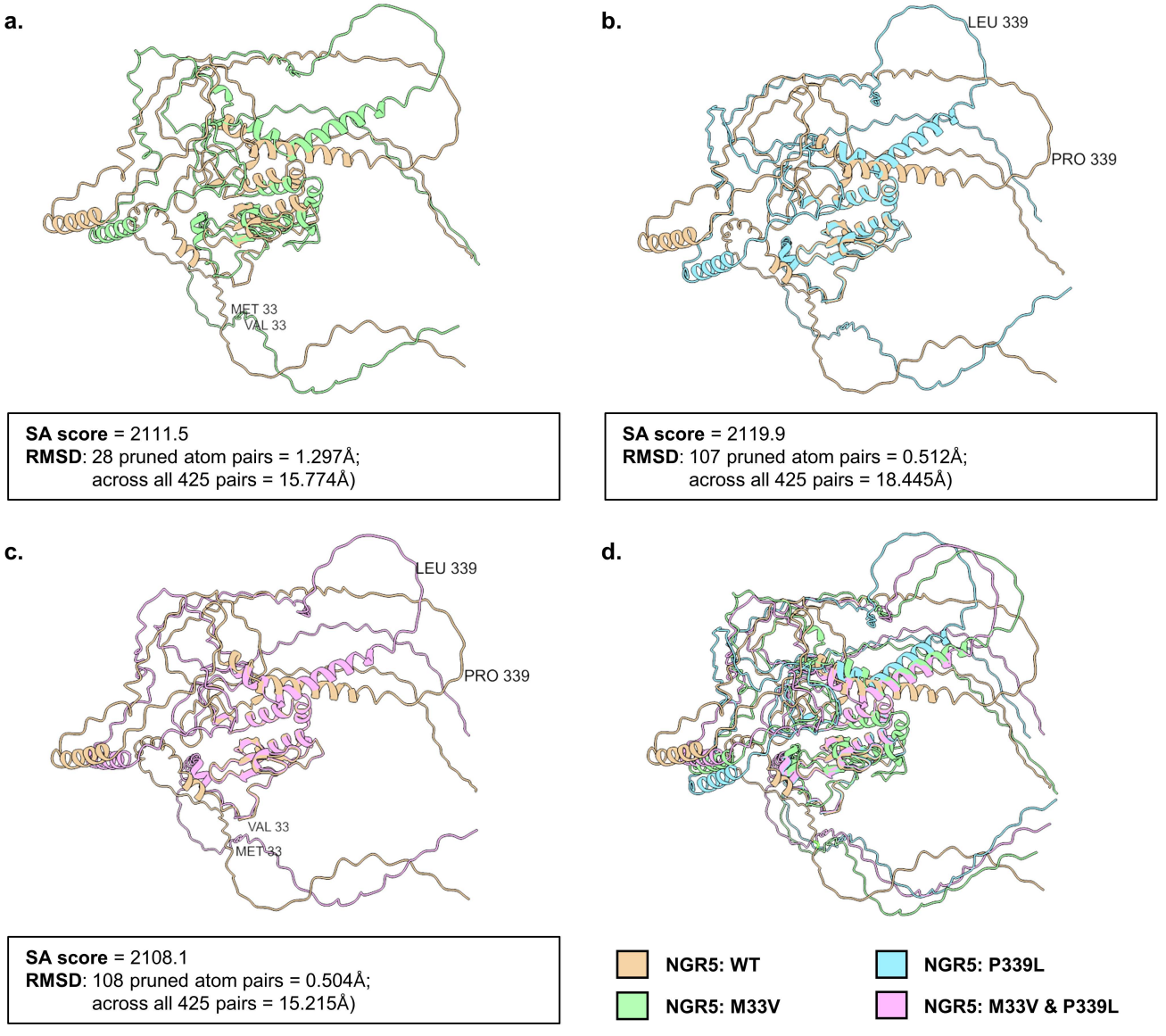


Fig. S9. Predicted three-dimensional (3D) structural variation among NGR5 protein variants carrying haplotype-associated amino acid substitutions. Diagrams show AlphaFold predicted structures of NGR5 protein variants, superimposed on the predicted 3D structure of NGR5 of Nipponbare as the reference: a. M33V variant b. P339L variant c. M33V and P339L variants d. The superimposed structures of all the variants and the reference NGR5 protein. The structural similarity among the predicted models was evaluated using root-mean-square deviation (RMSD) and structure alignment (SA) scores.