

# Genetic Basis of Tiller Dynamics of Rice Revealed by Genome-Wide Association Studies

**Shuyu Zhao**

Northeast Agricultural University

**Su Jang**

Seoul National University College of Agriculture and Life Sciences

**Yoon Kyung Lee**

Seoul National University College of Agriculture and Life Sciences

**Dong-Gwan Kim**

Sejong University College of Life Science

**Zhengxun Jin**

Northeast Agricultural University

**Hee-Jong Koh** (✉ [heejkoh@snu.ac.kr](mailto:heejkoh@snu.ac.kr))

Seoul National University College of Agriculture and Life Sciences <https://orcid.org/0000-0002-4357-7008>

---

## Research article

**Keywords:** rice tillering, tiller number, productive tiller number, heading date, phase transition, genome-wide association study

**Posted Date:** August 27th, 2020

**DOI:** <https://doi.org/10.21203/rs.3.rs-54023/v1>

**License:**  This work is licensed under a Creative Commons Attribution 4.0 International License. [Read Full License](#)

---

## Abstract

**Background:** Tiller number is the key determinant of rice plant architecture and panicle number and consequently controls grain yield. Thus, it is necessary to optimize the tiller number to achieve the maximum yield in rice. However, comprehensive analyses of the genetic basis of tiller number, considering the development stage, tiller type, and related traits, are lacking.

**Results:** We sequenced 219 Korean rice accessions and constructed a high-quality single nucleotide polymorphism (SNP) dataset. The tiller number at different development stages and heading traits involved in phase transitions were evaluated. By a genome-wide association study (GWAS), we detected 20 significant association signals for all traits. Five signals were detected in genomic regions near known candidate genes. Most of the candidate genes were involved in the phase transition from vegetative to reproductive growth. In particular, *HD1* was simultaneously associated with the productive tiller ratio and heading date, indicating that the photoperiodic heading gene directly controls the productive tiller ratio. Multiple linear regression models of lead SNPs showed coefficients of determination ( $R^2$ ) of 0.49, 0.22, and 0.41 for the tiller number at the maximum tillering stage, productive tiller number, and productive tiller ratio, respectively. Furthermore, the model was validated using independent japonica rice collections, implying that the lead SNPs included in the linear regression model were generally applicable to tiller number prediction.

**Conclusions:** We revealed the genetic basis of tiller number in rice plants during growth by a GWAS and formulated a prediction model by linear regression. Our results improve our understanding of tillering in rice plants and provide a basis for breeding high-yield rice varieties with the optimum tiller number.

## Background

Tiller number in rice is the major trait determining plant architecture and panicle number, consequently controlling grain yield. An excessive tiller number causes too many unproductive tillers, a reduced leaf area, and a reduction in photosynthetic efficiency by mutual shading. By contrast, too few tillers lead to low biomass production and a deficiency in the grain filling capacity and carbohydrate production [1]. It is important to minimize unproductive tillers and increase the sink size to improve the harvest index in rice, as in the case of China's super hybrid rice and the IRRI new plant type rice [2]. Tillers are formed from axillary buds, which are derived from the axillary meristems generated on the culm in the leaf axil [3]. Several genes directly regulate axillary meristem initiation, consequently affecting branching in rice. *LAX PANICLE1 (LAX1)*, which encodes a basic helix-loop-helix (bHLH) transcription factor, is the main regulator of axillary meristem formation in rice [4]. *lax1* mutant plants show reduced branching in both vegetative and reproductive growth, with suppressed tillering and panicle branches.

*MONOCULM1 (MOC1)* encodes a transcription factor in the GRAS family and is also involved in axillary bud regulation and rice tillering. *moc1* mutants show completely defective tillering, producing only one main culm [5].

Rice tillering is also influenced by genes involved in phytohormone signaling, such as auxins, cytokinins, strigolactone, and brassinosteroids. The auxin efflux carrier gene *OsPIN1* is involved in polar auxin transport. Transgenic plants with low *OsPIN1* expression show a reduced number of adventitious roots and a significantly increased number of tillers [6]. *Cytokinin oxidase/dehydrogenase 2 (OsCKX2)* encodes a cytokinin oxidase in rice. The suppression of *OsCKX2* expression by RNAi enhances growth and productivity by increasing the tiller number and grain weight [7]. *DWARF 53 (D53)* is a substrate of the SCF<sup>D3</sup> ubiquitination complex and acts as a repressor of strigolactone signaling [8]. In *d53* mutants, strigolactone signaling is blocked, leading to the dwarf and high-tillering phenotype. *Receptor-like cytoplasmic kinases 57 (OsRLCK57)* is involved in the modulation of brassinosteroid signaling. The repression of *OsRLCK57* results in reductions in tillers and panicle secondary branching [9].

In addition, growth phase transitions are a major determinant of rice tillering patterns. Axillary buds produce tillers during the vegetative stage. After the phase transition to the reproductive stage, the shoot apical meristem becomes an inflorescence meristem, and tiller development is delimited [10]. Late heading results in a long duration of vegetative growth, high biomass production, extended tillering, and a low harvest index [11, 12]. A short vegetative growth phase (i.e., early heading) results in insufficient biomass production [13]. Rice is a photoperiod-sensitive species with a short-day requirement for heading [14].

Previous studies have revealed several photoperiod-responsive genes. *Heading date 1* (*HD1*), an ortholog of *Arabidopsis CONSTANS* (*CO*), is a photoperiod-responsive flowering gene [15, 16]. *HD1* acts as an activator of *HEADING DATE 3a* (*HD3a*) under short-day conditions, resulting in heading [17]. *TIMING OF CAB EXPRESSION 1* (*OsTOC1*), an ortholog of *TOC1* in *Arabidopsis*, is an important circadian clock component and photoperiodic heading regulator [18]. *Hairy meristem 1* and *Hairy meristem 2* (*OsHAM1* and *OsHAM2*) regulate the vegetative to reproductive phase change [19]. *OsHAM1* and *OsHAM2* are regulated by *osa-miR171*. The down-regulation of these genes delays heading, thus increasing tiller numbers.

Genome-wide association studies (GWASs) using high-quality single nucleotide polymorphisms (SNPs) can be used to dissect the genetic basis of complex agronomic traits [20]. Loci associated with natural variation in tiller number have been identified by GWASs using several panels. Huang et al. conducted a GWAS using 671,355 SNPs and 373 accessions classified as the indica landrace, and detected eight association signals for tiller number at the full-ripe stage [21]. Another GWAS successfully identified 13 loci associated with tiller number at the booting stage using 4,136 SNPs detected in 469 indica accessions [22]. More recently, a GWAS using 700,000 SNPs revealed 23 loci associated with tiller number at the later tillering stage of 350 RDP-2 accessions [23]. However, these previous GWASs have focused on tiller number at a single stage, without considering alterations during different stages. To comprehensively understand the genetic basis of tiller number, it is necessary to consider developmental stages, tiller types, such as productive or unproductive tillers, and relationships between tiller number and other traits.

In this study, we sequenced 219 Korean rice accessions and constructed a high-quality SNP dataset. Then, we investigated natural variation in tiller number at different developmental stages and heading traits. By a GWAS, we identified the complex genetic regulation of tillering and interactions between tiller number and phase transitions in rice plants. We also revealed the contribution of each trait-associated loci to the maximum tiller number, productive tiller number, and ratio by a linear regression analysis. Lastly, using independent japonica accessions, we verified the applicability of the detected loci to molecular breeding aimed at optimizing tiller numbers.

## Results

### Population structure of Korean rice accessions

An SNP subset obtained by LD (linkage disequilibrium)-based pruning was used for analyses. A multi-dimensional scaling (MDS) analysis showed that population genetic structures of Korean rice accessions could be classified into two major groups, japonica and Tongil-type accessions, developed from inter-subspecific crosses between indica and temperate japonica (Fig. 1a). Japonica accessions were divided into two sub-groups, modern cultivars and landrace accessions. A neighbor-joining (NJ) tree revealed a similar pattern to that obtained by MDS plot (Fig. 1b).

### Variation and changes in tiller number among growth stages

The modern japonica cultivar group showed lower tiller numbers at the early tillering stage (TNE) and maximum tillering stage 1 (TNM1) than those of japonica landraces and Tongil-type rice (Fig. 1d and e). The tiller number at maximum tillering stage 2 (TNM2) was higher in the Tongil-type cultivar group than in the other groups (Fig. 1f). The productive tiller number (PTN) and productive tiller ratio (PTR) in all Korean rice accessions were mainly in the ranges of 10–11 and 0.65–0.7, respectively. The highest average PTN and PTR values were found in the landrace group (Fig. 1g–h). Tongil-type cultivars revealed the lowest phenotypic variation for all traits but showed the highest average values and variation in the heading interval (HDI) (Fig. 1i). The average heading date (HD) was lower in the landrace group than in the other groups (Fig. 1j).

Phenotypic relationships among all traits were investigated by correlation analysis (Fig. 2). Three tiller number traits at the vegetative stage (TNE, TNM1, and TNM2) were positively correlated. Positive correlation was found between HD and TNM2. HD was negatively correlated with HDI and PTR. PTR was negatively correlated with TNM2, while PTN were positively correlated with TNM2 and HDI.

# Genome-wide association studies with factored spectrally transformed linear mixed models (FaST-LMM)

We detected 20 association signals ( $p \leq 1.0 \times 10^{-6}$ ) for all traits (Fig. 3; Table 1). Two significant associations were detected for each of TNE, TNM1, and TNM2 (Fig. 3a–c). For PTN and PTR, only one and two association signals were detected, respectively (Fig. 3d–e). The most significant association (chr01:8365187;  $p = 3.52 \times 10^{-16}$ ) was detected for HD (Fig. 3f). The largest number of loci (i.e., six) were associated with HDI (Fig. 3g; Table 1). We found a notable locus at 31.0–31.3 Mb on chromosome 4 where two lead SNPs, chr04:31093494 and chr04:31232808, were associated with TNM2 and PTN and with TNM1 and TNM2 (Fig. 4a–c). Although the two lead SNPs were separated by only 150 kb, the LD parameter  $r^2$  was  $< 0.4$ . Thus, each lead SNP was considered as independent association signals with the traits. chr04:31093494 and chr04:31232808 were both associated with TNM2 ( $p = 5.82 \times 10^{-10}$  and  $p = 2.88 \times 10^{-11}$ , respectively) (Fig. 4b). However, only chr04:31232808 was significantly associated with TNM1 ( $p = 1.36 \times 10^{-7}$ ; Fig. 4a), and chr04:31093494 was associated with PTN ( $p = 1.18 \times 10^{-7}$ ; Fig. 4c). The chr04:31232808 A allele was a positive (trait-enhancing) allele, causing higher TNM1 and TNM2 values, and was primarily detected in landrace accessions (Fig. 4e and g). Accessions with the chr04:31093494 A allele, mainly landrace accessions, showed higher TNM1 and TNM2 values than those of accessions with the T allele (Fig. 4d and h). Four haplotypes were constructed from two alleles for each lead SNP (Fig. 4i). Haplotype 1 (H1), consisting of two positive alleles of each lead SNP, showed the highest average TNM1, TNM2, and PTN. Most modern cultivars (96.7%) with H4 showed low TNM1, TNM2, and PTN values.

Table 1  
SNPs associated with tiller number and heading traits

Trait	Chr	Lead SNP	p-value	Candidate gene	Gene ID <sup>a</sup>	Description	Reference
TNE	1	chr01:42957568	6.11E-07	<i>OsRLCK57</i>	<i>Os01g0973500</i>	Receptor-like cytoplasmic kinase	[10]
TNE	3	chr03:9436356	1.47E-07				
TNM1	2	chr02:26906676	9.65E-07	<i>OsHAM1;</i> <i>OsHAM2</i>	<i>Os02g0662700;</i> <i>Os02g0663100</i>	Maintenance of shoot apical meristem indeterminacy;  Regulation of vegetative to reproductive phase change	[20]
TNM1	4	chr04:31232808	1.36E-07				
TNM2	2	chr02:24685790	7.67E-08	<i>OsTOC1/OsPRR1</i>	<i>Os02g0618200</i>	Circadian-associated rice pseudo response regulator;  Control of flowering time	[19]
TNM2	4	chr04:31093494	5.82E-10				
TNM2	4	chr04:31232808	2.88E-11				
PTN	4	chr04:31093494	1.18E-07				
PTR	6	chr06:8339606	3.52E-16	<i>HD1</i>	<i>Os06g0275000</i>	Zinc finger protein;  Control of photoperiodic flowering time	[16, 17]
PTR	11	chr11:17720543	9.10E-07				
HD	1	chr01:25634456	1.92E-08				
HD	1	chr01:34991240	3.30E-07				
HD	6	chr06:8365187	1.23E-20	<i>HD1</i>	<i>Os06g0275000</i>	Zinc finger protein;  Control of photoperiodic flowering time	[16, 17]
HD	6	chr06:21572894	2.37E-07				
HDI	1	chr01:29428424	4.54E-07				
HDI	2	chr02:22024430	6.07E-07				

Trait	Chr	Lead SNP	p-value	Candidate gene	Gene ID <sup>a</sup>	Description	Reference
HDI	2	chr02:33872721	7.49E-07				
HDI	6	chr06:29529906	2.72E-07				
HDI	9	chr09:18859830	7.24E-07				
HDI	10	chr10:18000023	5.13E-07				

<sup>a</sup> Gene ID for the Rice Annotation Project Database (RAP-DB).

Association signals near 8.3 Mb on chromosome 6 exhibited pleiotropic associations with PTR and HD, led by chr06:8339606 ( $P = 3.52 \times 10^{-16}$ ) and chr06:8365187 ( $p = 1.23 \times 10^{-20}$ ), respectively (Fig. 5a–b). The two lead SNPs were in high LD ( $r^2 > 0.94$ ) and were separated by about 25 kb. Thus, these lead SNPs were considered the same association signal. *HD1* was located within a region in moderate LD ( $r^2 > 0.6$ ) with lead SNPs approximately 1 Mb from association signals (Fig. 5a–b). *HD1* controls photoperiodic heading [15, 16]. In rice, heading results from a developmental switch from the vegetative to reproductive phase, and the repression of heading is involved in the maintenance of vegetative growth, including tiller development. Several studies have shown that *HD1* function is affected by whether the *HD1* haplotype includes functional or nonfunctional allele [24, 25]. Six *HD1* haplotypes were classified into functional (i.e., H1, H3, H5, and H6) and nonfunctional haplotypes (i.e., H2 and H4) (Fig. 5c). In Korean rice accessions, nonfunctional *HD1* was caused by two frame-shifting InDels, chr06:9338004 and chr06:9338220. Functional *HD1* showed a later HD and higher PTR than those for nonfunctional *HD1* (Fig. 5d–e).

Chr02:26906676, a lead SNP near 26.9 Mb, was significantly associated with TNM1 ( $p = 9.65 \times 10^{-7}$ ). We detected two candidate genes, *OsHAM1* and *OsHAM2*, approximately 62 and 52 kb from chr02:26906676 with  $r^2 > 0.8$  (Fig. 6a). These genes maintain shoot apical meristem indeterminacy and regulate the vegetative to reproductive phase change [19]. Five and four haplotypes were constructed for *OsHAM1* and *OsHAM2*, respectively, using variants in genic regions and the ~ 1.5 kb promoter region from the 5' UTR. In Korean accessions, H4 of *OsHAM1* was dominantly detected, accounting for 95% of accessions (Fig. 6b). In a comparison of average TNM1 values among haplotypes detected in more than two accessions, H2 showed a higher TNM1 than that of H4 (Fig. 6c). In *OsHAM2*, H4 was dominantly detected in Korean accessions, accounting for 93% of accessions. H1, only found in eight accessions, conferred the highest average TNM1 (Fig. 6d).

We detected association signals for TNE, represented by the lead SNP chr01:42957568 ( $p = 6.11 \times 10^{-7}$ ), at 42.9 Mb on chromosome 1. A candidate gene, *OsRLCK57*, in strong LD ( $r^2 > 0.97$ ) was located about 23 kb from chr01:42957568 (Fig. 6f). *OsRLCK57* is involved in the modulation of BR signaling and is required for the development of tillers and panicle secondary branching [9]. Three *OsRLCK57* haplotypes were constructed by three variants (Fig. 6g). H1, found in only four accessions, showed higher TNE values than those of other haplotypes (Fig. 6h).

Another locus at 24.5–24.8 Mb on chromosome 2, led by chr02:24685790 ( $p = 7.67 \times 10^{-8}$ ), was significantly associated with TNM2. *OsTOC1* was located about 113 kb from the lead SNP and was in LD with  $r^2 > 0.65$  (Fig. 6i). *OsTOC1* is an important circadian clock component and photoperiodic heading regulator [18]. Ten haplotypes were detected based on variants in the genic region of *OsTOC1*. H1 containing chr02:24572219, an InDel located at the junction between the coding region and 3'UTR, was only detected in landrace accessions, showing the highest TNM2 (Fig. 6j–k). H9 and H10 with the chr02:24571309 T allele showed relatively lower TNM2 values than those of the other haplotypes (Fig. 6j–k).

## Linear regression model for three tiller-related traits

The lead SNPs for TNE, TNM1, TNM2, and HD were included in a linear regression analysis of TNM2 (Fig. 7a). A linear regression model consisting of seven independent variables effectively explained variation in TNM2 ( $R^2 = 0.49$ ) (Table 2). As determined by estimates of relative importance (i.e., the contribution of an individual variable to  $R^2$ ), chr04:31232808 (23.6%) showed the largest contribution to the linear regression model for TNM2 (Fig. 7b). In a linear regression analysis of PTN, six lead SNPs were selected as independent variables, explaining 22.3% of the variance in PTN (Table 2). chr02:22024430 showed the highest relative importance of 38.1% (Fig. 7c). chr04:31232808 showed high relative importance for PTN in addition to TNM2, accounting for 21.5% of  $R^2$  (Fig. 7c). The linear regression model for PTR, consisting of six independent variables, exhibited an  $R^2$  of 0.41 (Table 2). chr06:21572894 and *HD1* showed relatively large contributions to the model, with relative importance values of 36.9% and 16.9% (Fig. 7d). Furthermore, an independent test population consisting of 47 japonica accessions collected from several countries was employed to verify the accuracy of the linear regression model (Additional file 2: Table S2). TNM2, PTN, and PTR were measured in the test population (Fig. 7e–g). Predicted values for TNM2, PTN, and PTR were calculated by applying the regression equations to genotype data for the test population, and were compared with observed values. The correlation coefficients ( $R$ ) between predicted and observed values of TNM2, PTN, and PTR were 0.73, 0.49, and 0.6, respectively (Fig. 7h–j). Linear regression models for TNM2, PTN, and PTR explained 52.7%, 23.8%, and 36.5% of phenotypic variation, respectively, indicating that the models consisting of lead SNPs consistently explained the three traits with similar accuracy in independent temperate japonica accessions (Fig. 7h–j).

Table 2  
Summary of linear regression equations for TNM2, PTN, and PTR using lead SNPs

SNP	Coefficient	SD	<i>t</i> <sup>a</sup>	Allele <sup>b</sup>	
				0	1
<b>TNM2</b>					
chr03:9436356	4.51	1.49	3.034**	A	C
chr01:42957568	1.04	0.48	2.167*	A	G
chr02:26906676	4.21	1.02	4.122***	A	C
chr04:31232808	4.59	1.03	4.467***	T	A
chr02:24685790	2.80	1.17	2.389*	G	A
<i>HD1</i>	-2.37	0.67	-3.521***	F	N
chr06:21572894	-1.73	0.55	-3.132**	T	C
Intercept	16.53	0.34	47.949***		
<i>p</i>	2.20E-16				
<i>R</i> <sup>2</sup>	0.49				
<b>PTN</b>					
chr03:9436356	1.83	1.02	1.798	A	C
chr01:42957568	0.64	0.30	2.131*	A	G
chr04:31232808	2.40	0.67	3.581***	T	A
chr01:34991240	1.05	0.52	2.026*	C	A
chr02:22024430	4.13	1.00	4.149***	C	T
chr10:18000023	-1.36	0.64	-2.129*	C	T
Intercept	10.93	0.21	51.902***		
<i>p</i>	8.68E-06				
<i>R</i> <sup>2</sup>	0.22				
<b>PTR</b>					
chr03:9436356	-0.09437	0.05	-1.975	A	C
chr02:26906676	-0.0896	0.03	-2.768**	A	C
<i>HD1</i>	0.078779	0.02	3.446***	F	N
chr01:25634456	0.053531	0.02	2.883**	C	T
chr01:34991240	0.053239	0.03	1.872	C	A
chr06:21572894	0.087943	0.02	4.902***	T	C
Intercept	0.649476	0.01	66.599***		
<i>p</i>	7.68E-15				
<i>R</i> <sup>2</sup>	0.41				

SNP	Coefficient	SD	<i>t</i> <sup>a</sup>	Allele <sup>b</sup>				
				0	1			
<sup>a</sup> Values of <i>p</i> <0.05, 0.01, and 0.001 are denoted by *, **, and ***, respectively.								
<sup>b</sup> Two different alleles were designated 1 and 0.								

*R*<sup>2</sup>, Coefficient of determination; TNM2, Tiller number at maximum tillering stage 2; PTN, Productive tiller number; PTR, Productive tiller ratio; F, Functional allele; N, Nonfunctional allele.

## Discussion

Tiller number is a key factor affecting rice plant architecture and panicle number [2]. To identify the genetic mechanisms underlying tiller number, comprehensive investigations involving consecutive observations and accounting for relationships with other traits affecting tillering are necessary. We investigated relationships between tiller number and the heading traits, and performed a GWAS to dissect the genetic basis for tiller number.

## Phenotypic relationships

We detected a positive correlation between TNM2 and PTN, indicating that TNM2 could directly affect the potential to produce a panicle. TNM2 was also positively correlated with HD, which reflects the transition to reproductive development. A later HD is related to a longer duration of vegetative growth and the continued development of vegetative organs, causing an increase in TNM2. We detected positive correlations between TNM2 and HD, as well as between TNM2 and PTN, but not between HD and PTN. These results suggest that HD is not a direct determinant of PTN and might have an indirect effect via TNM2.

PTN is an important trait affecting grain yield as it directly determines the panicle number per plant. PTN showed a significant positive correlation with HDI (Fig. 2). A longer HDI represents a longer duration of reproductive growth and panicle development. These results suggest that panicle number, derived from PTN, could be increased by nonsynchronous flowering caused by a longer HDI.

Unproductive tillers are involved in discontinuing nutrient and carbohydrate translocation to the tillers from the mother stems; furthermore, they compete with reproductive tillers for nutrients in addition to light [26, 27]. Therefore, a high ratio of productive to unproductive tillers (PTR) is considered a desirable trait for high-yielding varieties [2]. We observed a strong negative correlation between HD and PTR (Fig. 2). As mentioned above, HD was positively correlated with TNM2 but was not significantly correlated with PTN. Additionally, PTN was not correlated with PTR. These results indicate that a higher PTR corresponding with an earlier HD could mainly be explained by a reduction in TNM2 rather than an increase in PTN.

## Association signals for tillering and heading

The tiller number at each growth stage is affected by the tillering capacity at earlier stages [28]. Tillering is controlled by the temporal expression of related genes at various development stages. However, some genetic factors are consistently associated with tillering at several stages [29]. In this study, two lead SNPs, chr01:42957568 and chr03:9436356, were associated with TNE (Fig. 3a). chr02:26906676 and chr02:24685790 were associated with TNM1 and TNM2, respectively, exhibiting stage specific-associations (Fig. 3b–c). Two lead SNPs, chr04:31093494 and chr04:31232808, were simultaneously associated with TNM2 and PTN and with TNM1 and TNM2, respectively. These results indicate that TNE was controlled by stage-specific QTLs detected only at certain stages. However, tiller numbers after the early stage, TNM1, TNM2, and PTN were affected by combinations of stage-specific and -nonspecific QTLs.

*HD1*, known to control photoperiodic heading, was a candidate gene for PTR and HD based on strong association signals within 8.33–8.37 Mb on chromosome 6, although *HD1* was located approximately 1 Mb away from this region (Fig. 3e–f). Nonfunctional *HD1* due to a frameshift results in early heading in rice [24, 25]. In the present study, nonfunctional *HD1* resulted

from two frame-shifting InDels, chr06:9338004 and chr06:9338220 (Fig. 5c). However, only the SNP subset excluding InDels was used for the GWAS, and this association may therefore have been missed. Thus, we also performed an additional GWAS for HD and PTR using all variants including InDels. However, the most significant association was still not detected in the *HD1* region. Instead, chr06:8329287 was the lead SNP ( $p = 2.14 \times 10^{-15}$ ; Additional file 3: Fig. S1). A similar pattern of association signals for HD have been reported in a previous GWAS using Japanese rice varieties [30] and a diverse collection [31]. The discrepancy between the strongest peak and *HD1* could be explained by the presence of several linked genes that contribute to heading across the region and/or allelic heterogeneity. Since a GWAS is based on independent comparisons of phenotypic variation for each polymorphic site, statistical significance is reduced when there are several causative alleles [30]. Similar to previous reports, we concluded that the most significant association signal was not detected in the *HD1* region due to allelic heterogeneity. It is notable that nonfunctional *HD1* leads to a higher PTR as well as early heading (Fig. 3e–f). This result supports that earlier heading date confers a higher PTR, as mentioned above, and suggests that PTR could be improved by the allele of *HD1* related to early heading.

Previous reports have shown that variation in tiller number is mainly explained by genes directly regulating axillary meristem formation [4, 5, 32] and genes involved in phytohormone signaling [6–9]. However, in the present study, we found three candidate genes (*OsHAM1*, *OsHAM2*, and *OsTOC1*) involved in the developmental phase transition near lead SNPs associated with TNM1 and TNM2 (Table 1). These results imply that natural variation in tiller number in Korean rice accessions is mainly modulated by genes involved in developmental phase transitions, rather than by genes directly regulating tiller development.

## Genetic determinants of TNM2, PTN, and PTR

To obtain a comprehensive understanding of the genetic basis of TNM2, PTN, and PTR, multiple linear regression models were estimated. The low  $R^2$  value for the linear regression model for PTN (0.22) indicated that a relatively small portion of the variance in PTN could be explained by genetic factors associated with tiller number and heading traits (Table 2). This result suggests that other genetic variants should be additionally considered to sufficiently explain PTN variation. Based on the relative importance of independent variables in linear regression models, TNM2 was largely contributed by chr04:3123280 (23.6%), the lead SNP for TNM1 and TNM2, followed by chr02:26906676 (17.9%), the lead SNP for TNM1 (Fig. 7b). For PTR, chr02:22024430, the lead SNP for HDI, was the major contributor (38.1%; Fig. 7c). Chr06:21572894, *HD1*, and chr01:34991240, which were associated with HD, were major contributors to PTR, accounting for 36.9%, 16.9%, and 14.6% of  $R^2$ , respectively (Fig. 7d). These results indicate that genetic variants associated with earlier stage traits and correlated traits could also influence TNM2, PTN, and PTR, together with lead SNPs for terminate traits (TNM2, PTN, and PTR).

In a previous study of progeny populations derived from various crosses, the heritability of tiller number was 37.8–97.7%, depending on the genetic background of the parents [33]. The heritability of tiller number in single segment substitution lines varied from 0–39.1% throughout plant growth from 7 to 63 days after transplanting (DAT), and the highest heritability was observed at 42 DAT [34]. In the USDA minicore rice diversity panel, SNP-based heritability of tiller number at 60 days after emergence was in the range from 20–25% [35]. These results indicate that the heritability of tiller number varies substantially depending on the growth stage, genetic population, and estimation method. In the present study, the heritability of tiller number varied from 39.2% (TNI; DAT 18) to 77% (TNM2; 42 DAT), showing substantial variation among growth stages (Additional file 4: Figure S2). Furthermore, the heritability of the other tiller number traits PTN and PTR were quite high (i.e., 58.1% and 70.1%, respectively), indicating that higher proportion of variation of tiller number trait could be explained by genetic factors. Linear regression models estimated using significant SNPs could explain 49%, 22%, and 41% of phenotypic variation in TNM2, PTN, and PTR, respectively (Table 2). Taken together, the small proportion of phenotypic variation explained by multiple linear regression models might be attributable to the use of significant SNPs in the GWAS, while heritability was estimated using all SNPs. Regardless, the multiple linear regression models and genetic variants used as independent variables could be effective molecular tools for the prediction of TNM2, PTN, and PTR in rice breeding programs.

## Conclusions

We dissected the basis of rice tiller number by analyzing relationships with heading traits. We also revealed the genetic basis of tiller alterations at different growth stages by GWASs and linear regression analyses. By detecting loci associated with tiller number and phase transitions, our results provide insight into the genetic basis of overall tillering dynamics. This information could be employed for improving rice plant architecture and for optimizing yield.

## Methods

### 1. Plant materials and phenotype analysis

A total of 266 rice accessions, including 219 Korean rice accessions for GWAS and 47 japonica accessions for the verification of linear regression models, were used. Korean rice accessions included 78 landraces, 130 modern cultivars, and 11 Tongil-type cultivars (Additional file 1: Table S1). Modern cultivars and Tongil-type cultivars, derived from a cross between temperate japonica and indica, were considered as separate groups. Forty-seven japonica accessions to verify linear regression models were from four countries: Japan, China, Taiwan, and the USA (Additional file 2: Table S2). A total of 32 accessions were provided by the National Agrobiodiversity Center (NAC), Rural Development Administration (RDA), South Korea. The other 234 accessions were conserved at the Agricultural Genetic Resource Center, Seoul National University (SNU), Suwon, South Korea. All plant materials were cultivated in an experimental field at SNU, Suwon, Republic of Korea (natural long-day conditions, latitude = 37°N). Thirty-day-old seedlings were transplanted into a paddy field under the following conditions: one plant per hill, 25 plants per row, 15 cm between plants in a row, 30 cm between rows, and three rows per accession. All phenotypes were measured in mid-row plants, excluding plants near other accessions and boarder plants. Tiller number traits, including tiller number at the early tillering stage (TNE), maximum tillering stage 1 (TNM1), maximum tillering stage 2 (TNM2), and PTN, were measured at 18 DAT, 35 DAT, 42 DAT, and 110 DAT, respectively (Fig. 1c). The PTR indicated the capacity for developing panicle-emerged tillers from whole tillers. PTR was calculated by the ratio of panicle-emerged tillers (productive tillers) to the maximum potential tillers (PTN/TNM2). The HD was defined as the time from the date of sowing to the date at which the first panicle emerged in plant. Panicle emergence was recoded when a tip of panicle was visible from the flag leaf sheath. The HDI was defined as the time from the date of emergence of the first panicle to the last panicle in a plant.

### 2. NGS analysis and genotyping

Total DNA was extracted from 90-day-old leaves of each accession by the CTAB method [36]. DNA was sheared into fragments of 450–500 bp and used for DNA library construction using TruSeq Nano DNA Library Prep kits (Illumina, San Diego, CA, USA) according to the manufacturer's protocol. The library size distribution was checked using the Agilent Technologies 2100 Bioanalyzer and a DNA 1000 chip (Santa Clara, CA, USA). Prepared libraries were quantified by qPCR according to the Illumina qPCR quantification protocol. Whole genome sequencing data were generated on the Illumina HiSeq X system to generate 2 × 150 bp paired-end reads with a sequencing depth of > 10 × per sample. Raw reads were processed to remove adaptors and low-quality bases using Trimmomatic v0.38 [37] with the parameters ILLUMINACLIP:2:30:10 SLIDINGWINDOW:4:15 MINLEN:50. Reads were aligned to the rice reference genome (Nipponbare, IRGSP v1.0) [38] using the BWA v0.7.17 MEM algorithm with default parameters [39]. Aligned reads were sorted using samtools v1.9 [40], and duplicates were removed using Picard v2.20.2 [41]. Nucleotide variants were called by the HaplotypeCaller function of GATK v4.1.2 [42] with the parameters –max-missing 0.95 –minQ 30 –minDP 5. In addition, nucleotide variants with proportions of heterozygous genotypes of > 0.05 were filtered using the vc.getHetcount command in GATK v4.12. Finally, 4,660,220 variants were obtained.

### 3. Population structure and genetic relationships

Linkage disequilibrium (LD)-based SNP pruning was performed using PLINK v1.9 [43] with the command –indep-pairwise 50 5 0.2. In total, 37,009 LD-pruned SNPs (minor allele frequency [MAF] > 0.05) were obtained for analyses of population structure and genetic relationships. Population structure was revealed by an MDS analysis conducted using the MDS function of PLINK v1.9. Neighbor-joining (NJ) trees were constructed to infer genetic relationships using MEGA v7 [44].

### 4. Genome-wide association mapping

Only 1,509,362 SNPs with MAF > 0.05 and bi-allelic genotypes were included in the GWASs. All GWASs were performed using linear mixed mode (LMM) implemented in FaST-LMM v2.07 [45]. The genetic similarities were used to estimate random effects, and the significance threshold for associated SNPs was set to a *p*-value of  $1.0 \times 10^{-6}$ . LD patterns between lead SNPs and the other SNPs were evaluated using PLINK v1.9 [43] with the -r2 command to calculate pairwise genotype correlations ( $r^2$ ). Lead SNPs were defined as the SNPs with the lowest *p*-value in loci including significant SNPs. Haplotypes were constructed using all variants including SNPs and InDels without consideration of MAF. Individuals containing at least one missing or/and heterozygous genotype were excluded from the haplotype analysis.

## 5. Statistical analysis

All statistical analyses were performed using R studio v1.2.5033 [46]. Pearson's correlation coefficients among all phenotypes were calculated without missing observations using the stats package [47]. A correlation network was constructed with  $r > |0.3|$  using the corrr package [48].

Multiple linear regression models for TNM2, PTN, and PTR were estimated using phenotypes as dependent variables and lead SNPs in the GWAS as independent variables. Independent variables consisted of lead SNPs not only for TNM2, PTN, and PTR, but also for earlier traits with correlations. Since these three terminal traits are affected by traits at the previous growth stage, the directions of relationships between traits were determined based on the sequence of growth stages (Fig. 2 and Fig. 7a). Variables for the best linear equation were selected based on Akaike information criterion (AIC) in a stepwise algorithm implemented in the stats package [47]. The relative importance of independent variables in the linear regression equation was estimated using the lmg method of the relaimpo package [49]. Observed values and predicted values were compared using the predict function of the stats package [47].

## 6. Estimation of heritability

GCTA v1.93 [50] was used to estimate the SNP-based heritability of traits as proportion of phenotypic variance of traits explained by SNP subset [51]. The genetic relationship matrix to estimate genetic relatedness between individuals was calculated using filtered 1,509,362 SNPs. Variance in phenotypes was calculated by the restricted maximum likelihood method with the genetic relatedness and eigenvectors from a principal component analysis.

## Declarations

### Ethics approval and consent to participate

Not applicable.

### Consent for publication

Not applicable.

### Availability of data and materials

The datasets used and/or analysed during the current study are available from the corresponding author on reasonable request.

### Competing interests

The authors declare that they have no competing interests.

### Funding

This study was supported by a grant from the Next-Generation BioGreen 21 Program (no. PJ013165) of the Rural Development Administration (RDA), Republic of Korea. The funding agency had no role in the experimental design, data collection and analysis, and preparation of the manuscript.

## Authors' contributions

SJ and SZ designed the experiment and prepared the manuscript. SZ conducted the field experiments and collected phenotype data. SJ and YKL prepared genotype data. SJ performed the computational analysis. HJK participated in the supervision of the overall work and contributed to the finalization of the manuscript. All authors read and approved the manuscript.

## Acknowledgments

Not applicable.

## References

1. Peng S, Cassman K, Virmani S, Sheehy J, Khush G. Yield potential trends of tropical rice since the release of IR8 and the challenge of increasing rice yield potential. *Crop Science*. 1999;39:1552–9.
2. Peng S, Khush GS, Virk P, Tang Q, Zou Y. Progress in ideotype breeding to increase rice yield potential. *Field Crops Research*. 2008;108:32–8.
3. Tanaka W, Ohmori Y, Ushijima T, Matsusaka H, Matsushita T, Kumamaru T, et al. Axillary Meristem Formation in Rice Requires the *WUSCHEL* Ortholog *TILLERS ABSENT1*. *Plant Cell*. 2015;27:1173–84.
4. Komatsu K, Maekawa M, Ujiie S, Satake Y, Furutani I, Okamoto H, et al. LAX and SPA: Major regulators of shoot branching in rice. *Proc Natl Acad Sci U S A*. 2003;100:11765–70.
5. Li X, Qian Q, Fu Z, Wang Y, Xiong G, Zeng D, et al. Control of tillering in rice. *Nature*. 2003;422:618–21.
6. Xu M, Zhu L, Shou H, Wu P. A PIN1 Family Gene, OsPIN1, involved in Auxin-dependent Adventitious Root Emergence and Tillering in Rice. *Plant and Cell Physiology*. 2005;46:1674–81.
7. Yeh S-Y, Chen H-W, Ng C-Y, Lin C-Y, Tseng T-H, Li W-H, et al. Down-Regulation of Cytokinin Oxidase 2 Expression Increases Tiller Number and Improves Rice Yield. *Rice*. 2015;8:36.
8. Jiang L, Liu X, Xiong G, Liu H, Chen F, Wang L, et al. DWARF 53 acts as a repressor of strigolactone signalling in rice. *Nature*. 2013;504:401–5.
9. Zhou X, Wang J, Peng C, Zhu X, Yin J, Li W, et al. Four receptor-like cytoplasmic kinases regulate development and immunity in rice. *Plant Cell Environ*. 2016;39:1381–92.
10. Wang L, Sun S, Jin J, Fu D, Yang X, Weng X, et al. Coordinated regulation of vegetative and reproductive branching in rice. *Proc Natl Acad Sci USA*. 2015;112:15504–9.
11. Cui K, Peng S, Ying Y, Yu S, Xu C. Molecular Dissection of the Relationships among Tiller Number, Plant Height and Heading Date in Rice. *Plant Production Science*. 2004;7:309–18.
12. Vergara BS, Tanaka A, Lilis R, Puranabhavung S. Relationship between growth duration and grain yield of rice plants. *Soil Science and Plant Nutrition*. 1966;12:31–9.
13. Yoshida S. Fundamentals of Rice Crop Science. Int. Rice Res. Inst.; 1981.
14. Pautler M, Tanaka W, Hirano H-Y, Jackson D. Grass Meristems I: Shoot Apical Meristem Maintenance, Axillary Meristem Determinacy and the Floral Transition. *Plant and Cell Physiology*. 2013;54:302–12.
15. Nemoto Y, Nonoue Y, Yano M, Izawa T. *Hd1*, a *CONSTANS* ortholog in rice, functions as an *Ehd1* repressor through interaction with monocot-specific CCT-domain protein Ghd7. *Plant J*. 2016;86:221–33.
16. Yano M, Katayose Y, Ashikari M, Yamanouchi U, Monna L, Fuse T, et al. Hd1, a major photoperiod sensitivity quantitative trait locus in rice, is closely related to the *Arabidopsis* flowering time gene *CONSTANS*. *Plant Cell*. 2000;12:2473–84.
17. Kojima S, Takahashi Y, Kobayashi Y, Monna L, Sasaki T, Araki T, et al. Hd3a, a rice ortholog of the *Arabidopsis* FT gene, promotes transition to flowering downstream of Hd1 under short-day conditions. *Plant Cell Physiol*. 2002;43:1096–105.
18. Murakami M, Ashikari M, Miura K, Yamashino T, Mizuno T. The Evolutionarily Conserved OsPRR Quintet: Rice Pseudo-Response Regulators Implicated in Circadian Rhythm. *Plant and Cell Physiology*. 2003;44:1229–36.

19. Fan T, Li X, Yang W, Xia K, Ouyang J, Zhang M. Rice osa-miR171c Mediates Phase Change from Vegetative to Reproductive Development and Shoot Apical Meristem Maintenance by Repressing Four OsHAM Transcription Factors. *PLoS ONE*. 2015;10:e0125833.
20. Huang X, Zhao Y, Wei X, Li C, Wang A, Zhao Q, et al. Genome-wide association study of flowering time and grain yield traits in a worldwide collection of rice germplasm. *Nat Genet*. 2012;44:32–9.
21. Huang X, Wei X, Sang T, Zhao Q, Feng Q, Zhao Y, et al. Genome-wide association studies of 14 agronomic traits in rice landraces. *Nat Genet*. 2010;42:961–7.
22. Lu Q, Zhang M, Niu X, Wang S, Xu Q, Feng Y, et al. Genetic variation and association mapping for 12 agronomic traits in indica rice. *BMC Genomics*. 2015;16:1067.
23. Jiang S, Wang D, Yan S, Liu S, Liu B, Kang H, et al. Dissection of the Genetic Architecture of Rice Tillering using a Genome-wide Association Study. *Rice*. 2019;12:43.
24. Takahashi Y, Teshima KM, Yokoi S, Innan H, Shimamoto K. Variations in Hd1 proteins, Hd3a promoters, and Ehd1 expression levels contribute to diversity of flowering time in cultivated rice. *Proceedings of the National Academy of Sciences*. 2009;106:4555–60.
25. Zhang J, Zhou X, Yan W, Zhang Z, Lu L, Han Z, et al. Combinations of the Ghd7, Ghd8 and Hd1 genes largely define the ecogeographical adaptation and yield potential of cultivated rice. *New Phytologist*. 2015;208:1056–66.
26. Nuruzzaman M, Yamamoto Y, Nitta Y, Yoshida T, Miyazaki A. Varietal Differences in Tillering Ability of Fourteen Japonica and Indica Rice Varieties. :12.
27. Ao H, Peng S, Zou Y, Tang Q, Visperas RM. Reduction of unproductive tillers did not increase the grain yield of irrigated rice. *Field Crops Research*. 2010;116:108–15.
28. Pawar S, Radhakrishnan V, K.V. M. The Importance of Optimum Tillering in Rice-An overview. *South Indian Journal of Biological Sciences*. 2016;2:125.
29. Xu YB, Shen ZT. Diallel analysis of tiller number at different growth stages in rice (*Oryza sativa* L.). *Theoret Appl Genetics*. 1991;83:243–9.
30. Yano K, Yamamoto E, Aya K, Takeuchi H, Lo P, Hu L, et al. Genome-wide association study using whole-genome sequencing rapidly identifies new genes influencing agronomic traits in rice. *Nat Genet*. 2016;48:927–34.
31. Zhao K, Tung C-W, Eizenga GC, Wright MH, Ali ML, Price AH, et al. Genome-wide association mapping reveals a rich genetic architecture of complex traits in *Oryza sativa*. *Nat Commun*. 2011;2:467.
32. Oikawa T, Kyozuka J. Two-Step Regulation of LAX PANICLE1 Protein Accumulation in Axillary Meristem Formation in Rice. *Plant Cell*. 2009;21:1095–108.
33. Roy SC, Shil P. Assessment of Genetic Heritability in Rice Breeding Lines Based on Morphological Traits and Caryopsis Ultrastructure. *Sci Rep*. 2020;10:7830.
34. Liu G, Zeng R, Zhu H, Zhang Z, Ding X, Zhao F, et al. Dynamic expression of nine QTLs for tiller number detected with single segment substitution lines in rice. *Theor Appl Genet*. 2009;118:443–53.
35. Qu M, Zheng G, Hamdani S, Essemene J, Song Q, Wang H, et al. Leaf Photosynthetic Parameters Related to Biomass Accumulation in a Global Rice Diversity Survey. *Plant Physiol*. 2017;175:248–58.
36. Murray MG, Thompson WF. Rapid isolation of high molecular weight plant DNA. *Nucleic Acids Res*. 1980;8:4321–5.
37. Bolger AM, Lohse M, Usadel B. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics*. 2014;30:2114–20.
38. Sakai H, Lee SS, Tanaka T, Numa H, Kim J, Kawahara Y, et al. Rice Annotation Project Database (RAP-DB): an integrative and interactive database for rice genomics. *Plant Cell Physiol*. 2013;54:e6.
39. Li H, Durbin R. Fast and accurate short read alignment with Burrows–Wheeler transform. *Bioinformatics*. 2009;25:1754–60.
40. Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, et al. The Sequence Alignment/Map format and SAMtools. *Bioinformatics*. 2009;25:2078–9.

41. Broad Institute. Picard Tools. 2018. <http://broadinstitute.github.io/picard/>. Accessed 28 May 2019.
42. McKenna A, Hanna M, Banks E, Sivachenko A, Cibulskis K, Kernytsky A, et al. The Genome Analysis Toolkit: A MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Res.* 2010;20:1297–303.
43. Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira MAR, Bender D, et al. PLINK: A Tool Set for Whole-Genome Association and Population-Based Linkage Analyses. *Am J Hum Genet.* 2007;81:559–75.
44. Kumar S, Stecher G, Tamura K. MEGA7: Molecular Evolutionary Genetics Analysis Version 7.0 for Bigger Datasets. *Mol Biol Evol.* 2016;33:1870–4.
45. Lippert C, Listgarten J, Liu Y, Kadie CM, Davidson RI, Heckerman D. FaST linear mixed models for genome-wide association studies. *Nature Methods.* 2011;8:833–5.
46. Allaire J. RStudio: integrated development environment for R. Boston, MA. 2012;537:538.
47. R Core Team. R: A language and environment for statistical computing. 2013. <https://www.R-project.org>.
48. Kuhn M, Jackson S, Cimentada J. corrr: Correlations in R. 2020. <https://CRAN.R-project.org/package=corrr>. Accessed 21 May 2020.
49. Groemping U, Matthias L. relaimpo: Relative Importance of Regressors in Linear Models. 2018. <https://CRAN.R-project.org/package=relaimpo>. Accessed 21 May 2020.
50. Yang J, Lee SH, Goddard ME, Visscher PM. GCTA: A Tool for Genome-wide Complex Trait Analysis. *Am J Hum Genet.* 2011;88:76–82.
51. Yang J, Bakshi A, Zhu Z, Hemani G, Vinkhuyzen AAE, Lee SH, et al. Genetic variance estimation with imputed variants finds negligible missing heritability for human height and body mass index. *Nature Genetics.* 2015;47:1114–20.

## Figures

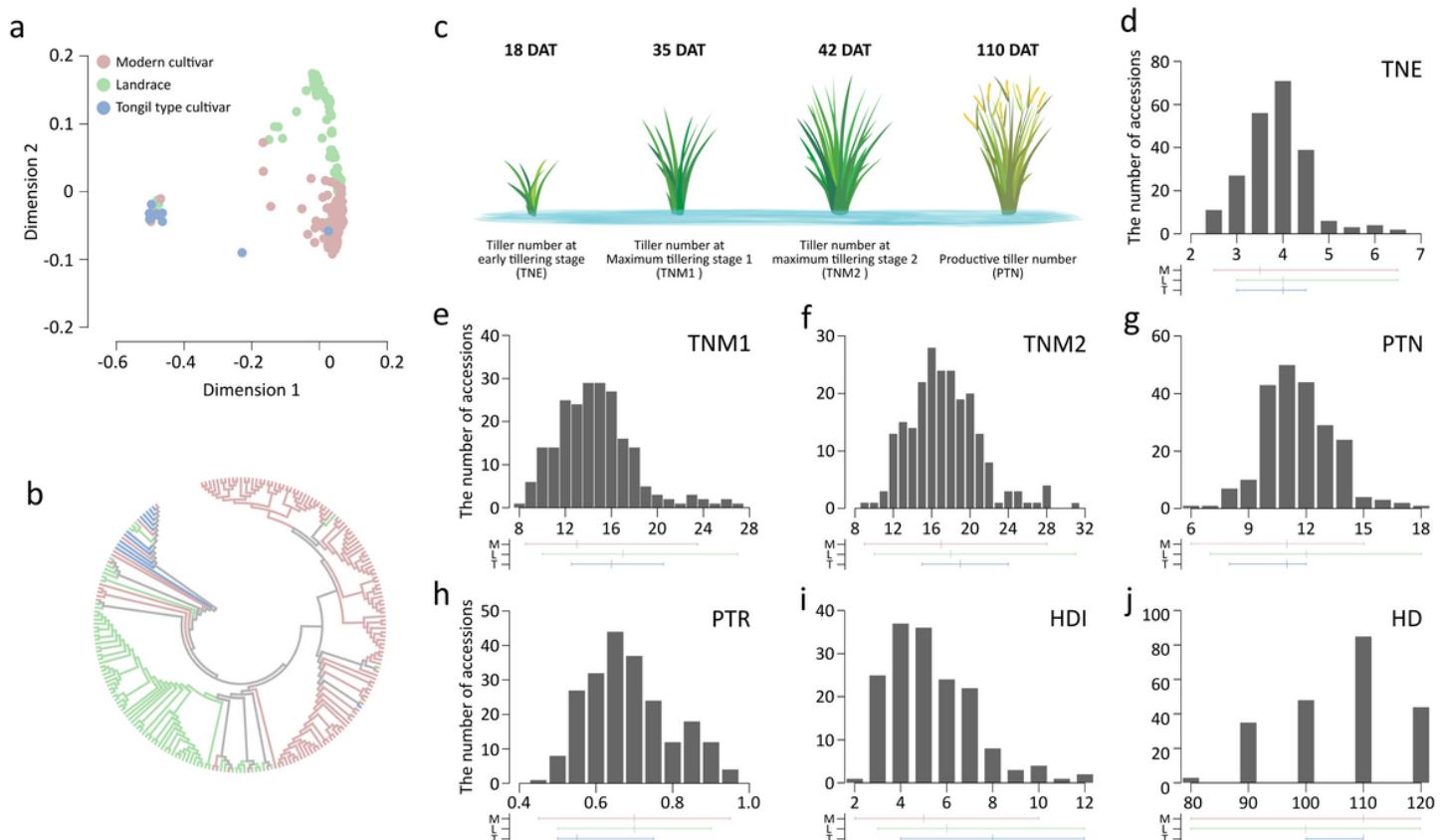
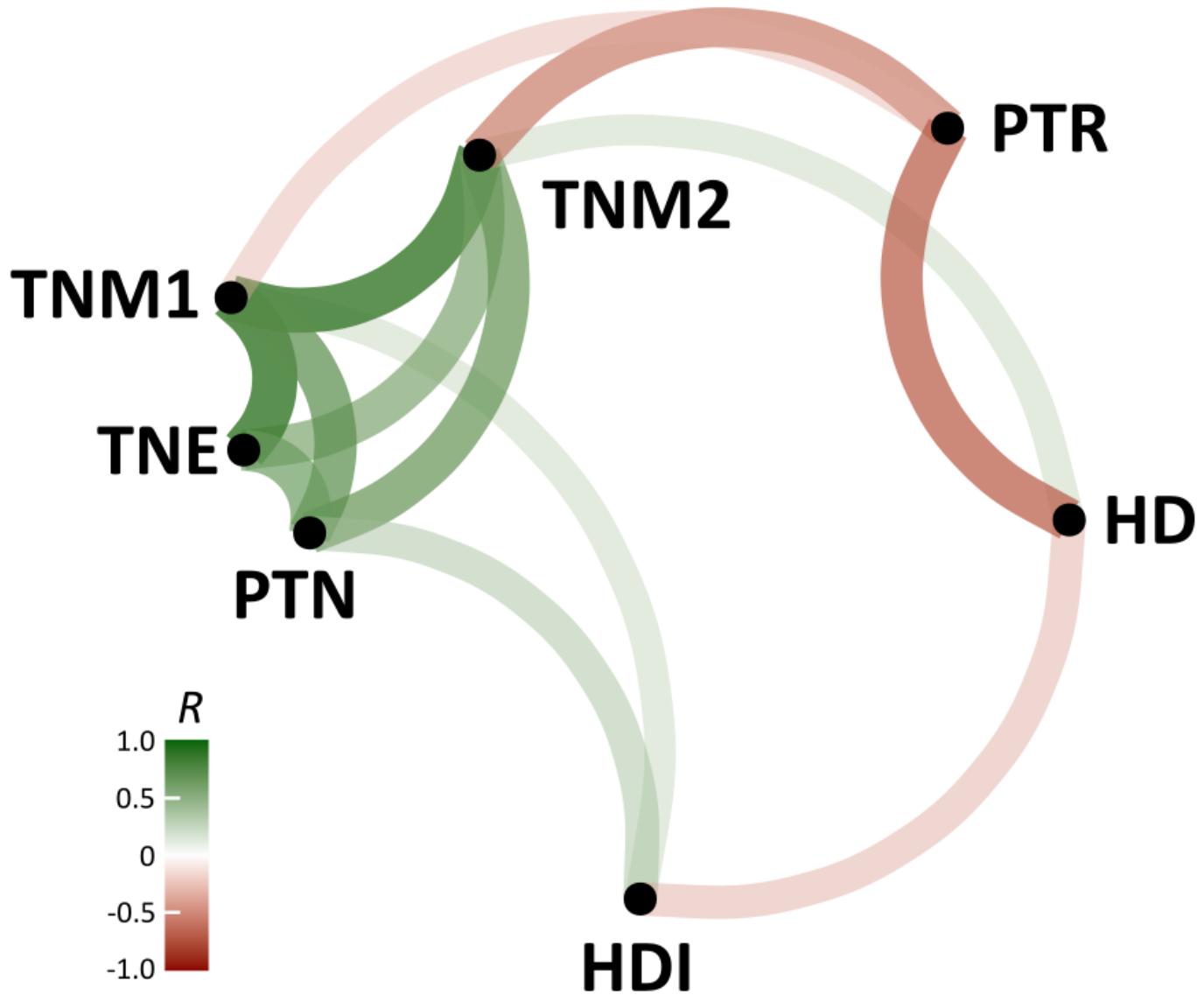


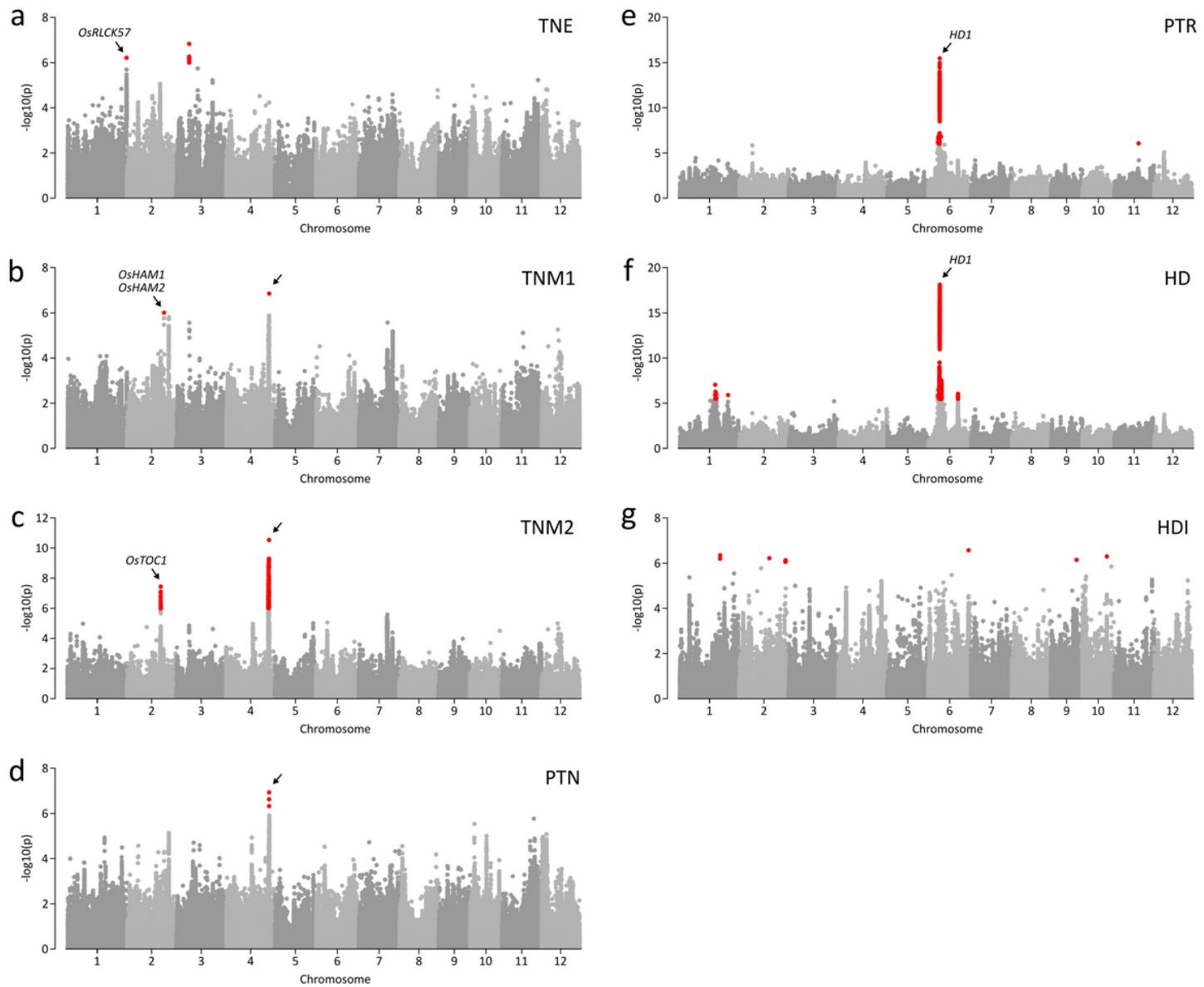
Figure 1

Population structure and phenotypic variation in Korean rice accessions. (a–b) An MDS plot (a) and NJ tree (b) were constructed using 37,009 LD-pruned SNPs. Red, modern cultivars; green, landrace; blue, Tongil-type cultivars. (c) Tiller number at different developmental stages. Tiller numbers were assessed at 18, 35, 42, and 110 days after transplanting (DAT). (d–j) Histogram showing distributions of TNE (d), TNM1 (e), TNM2 (f), PTN (g), PTR (h), HDI (i), and HD (j) across accessions. Horizontal and vertical lines below the histogram represent the range and average value, respectively. M, L, and T indicate modern cultivars, landrace, and Tongil-type cultivars, respectively.



**Figure 2**

Correlation network for tiller number traits and heading traits. A correlation network was built based on Pearson's correlation coefficients ( $R$ ) between traits. Each path indicates a correlation between two traits. The width and transparency of the line denote the strength of the correlation. Weak correlations with  $R$  between -0.3 and 0.3 are not shown. Green and red represent positive and negative correlations, respectively.



**Figure 3**

Genome-wide association scan using linear mixed models (LMM). (a–g) Manhattan plots of genome-wide association studies (GWASs) for TNE (a), TNM1 (b), TNM2 (c), PTN (d), PTR (e), HD (f), and HDI (g). Negative log<sub>10</sub>-transformed p-values from GWASs using LMM are plotted at positions on each of 12 chromosomes. Red dots indicate significant SNPs ( $p < 1.0 \times 10^{-6}$ ). Known candidate genes are marked above lead SNPs.

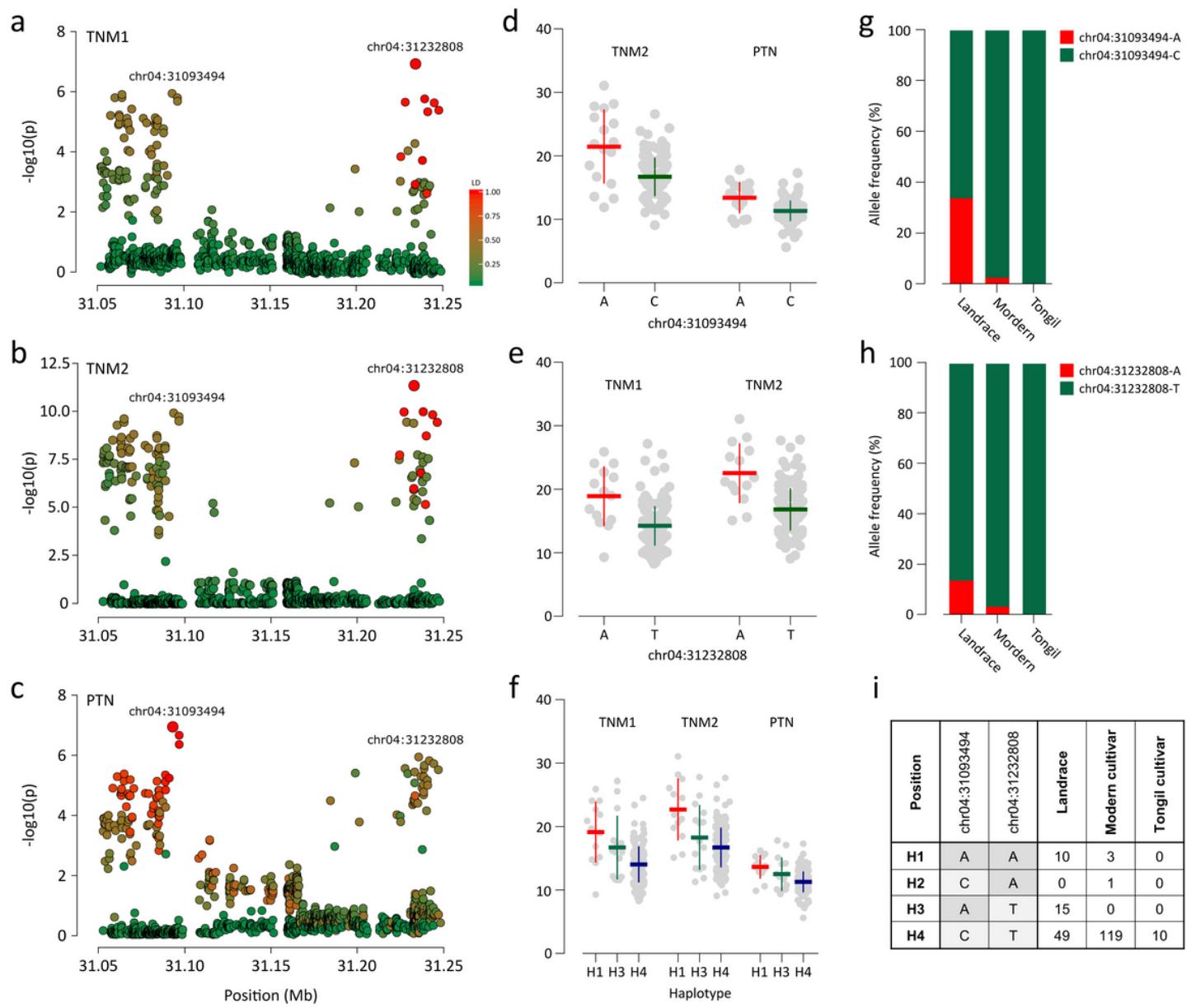
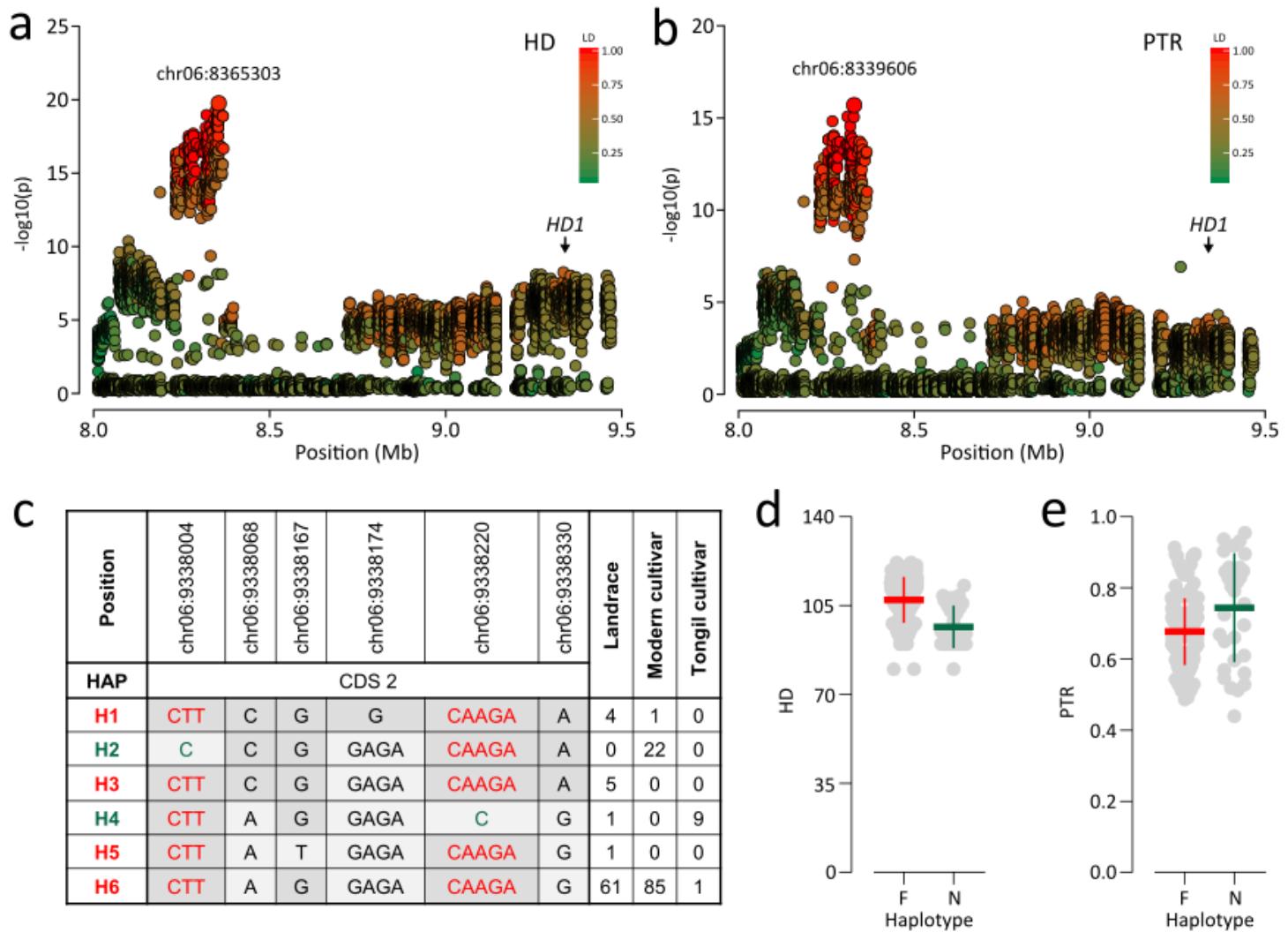


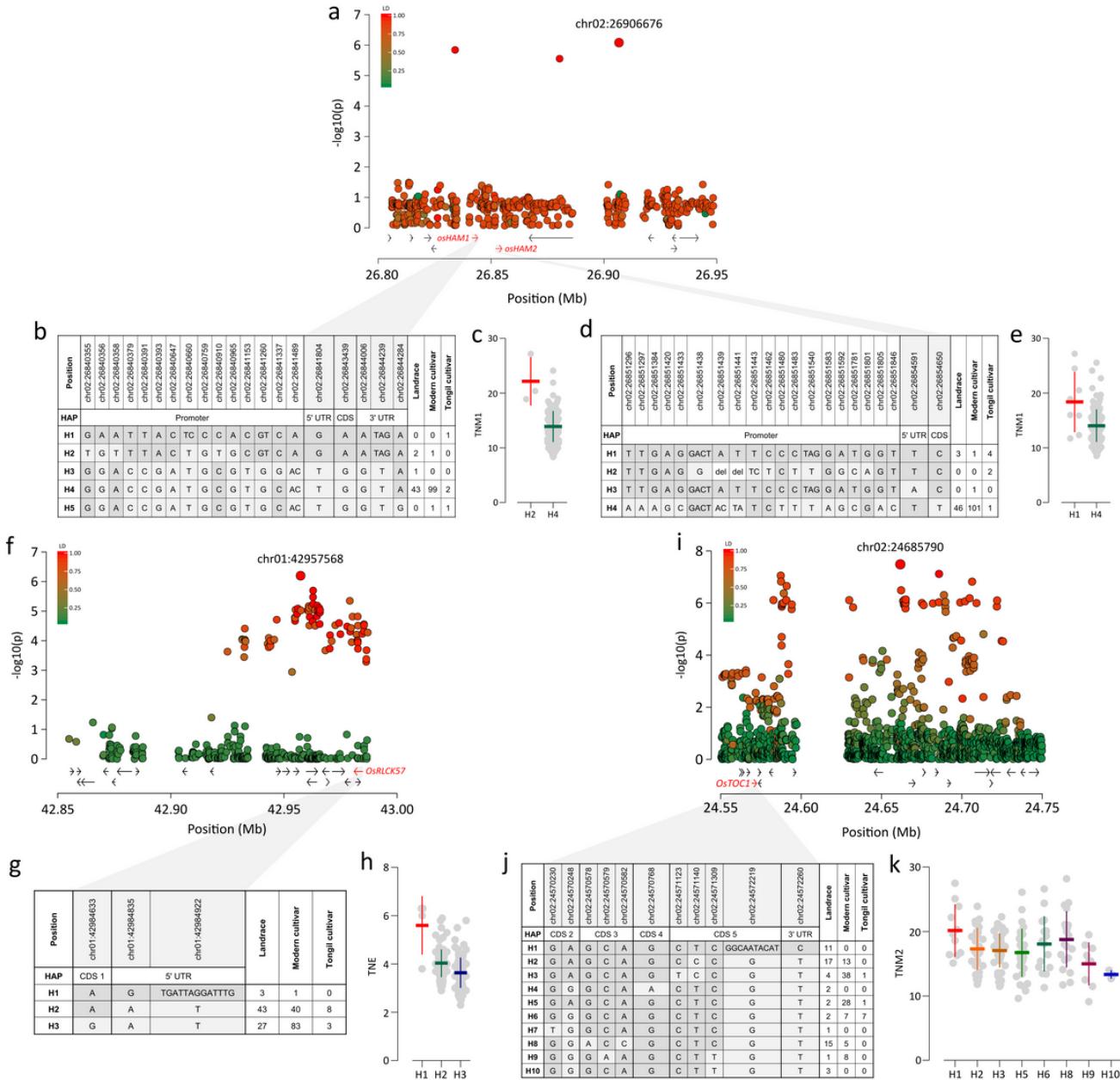
Figure 4

Strong association signal on chromosome 4 for TNM1, TNM2, and PTN. (a–c) Loci on chromosome 4 near the lead SNP (i.e., the SNP with highest significance) for TNM1 (a), TNM2 (b), and PTN (c). The color of each SNP indicates the  $r^2$  value for the correlative with the lead SNP. Red and green color intensities indicate stronger and weaker LD (0 to 1). (d–f) Dot plots for differences in phenotypes by alleles of chr04:31093494 (d) and chr04:31232808 (e), and the haplotype constructed from two lead SNPs (e). (g–h) Allele frequencies for (g) chr04:31093494, (h) chr04:31232808, and (i) haplotypes constructed from the two lead SNPs.



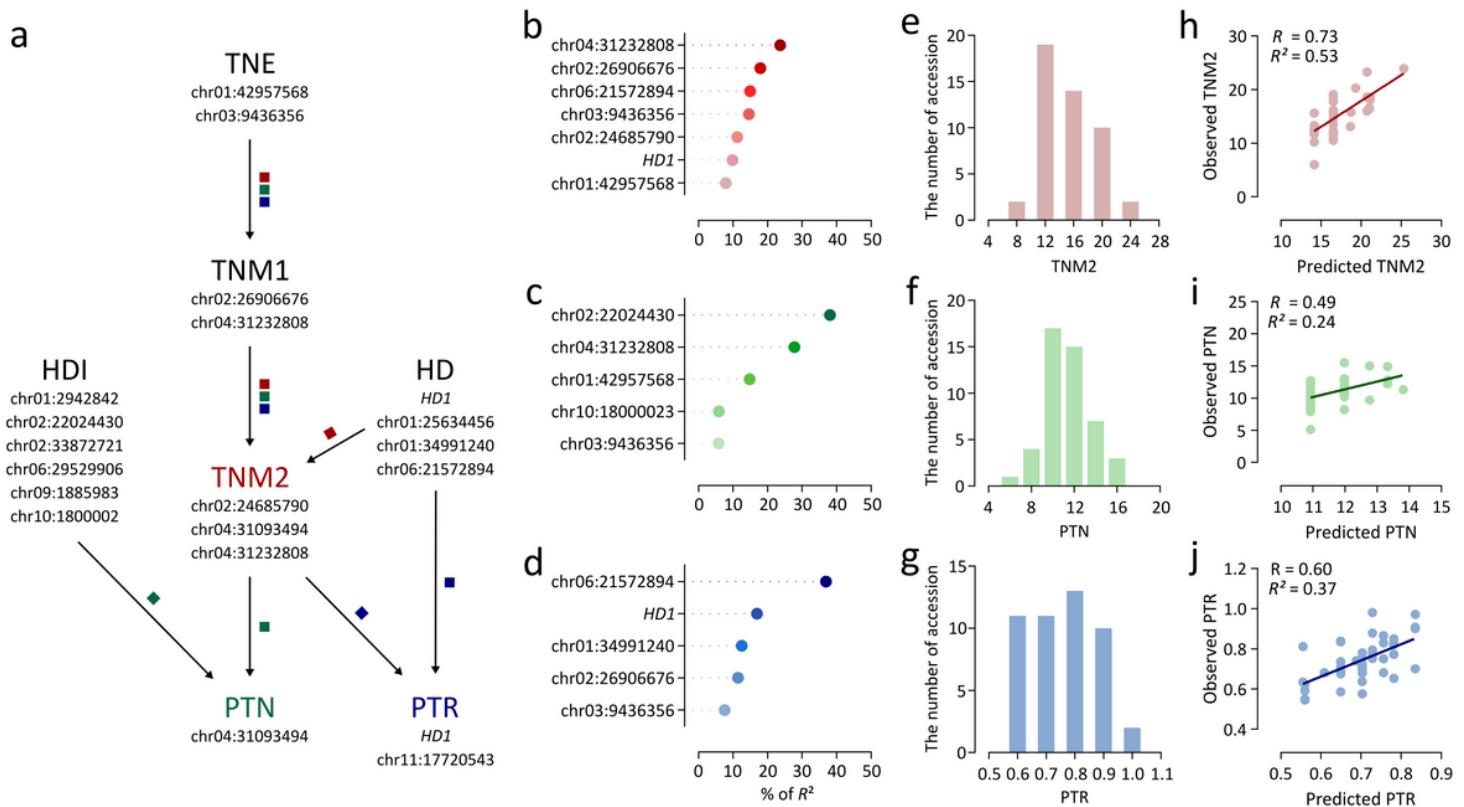
**Figure 5**

The Loci showing strong association signals for HD and PTR. (a–b) Genomic regions showing strong association signals for HD (a) and PTR (b). The color of each SNP indicates  $r^2$  for the correlation with the lead SNP. Red and green color intensities indicate stronger and weaker LD (0 to 1). (c) HD1 haplotypes. Sequence variants in green are frame-shifting InDels, causing a premature stop codon. Haplotypes in green include nonfunctional alleles led by a premature stop codon. Haplotypes in red denote include functional alleles. (d–e) Dot plots for differences in HD (d) and PTR (e) between functional and nonfunctional HD1. F, functional alleles; N, Nonfunctional alleles.



**Figure 6**

Loci showing strong association signals near candidate genes. Candidate genes located near the lead SNPs (a) chr02:26906676, (f) chr01:42957568, and (i) chr02:24685790. Haplotypes of candidate genes and differences in phenotype among haplotypes. (b–c) OsHAM1. (d–e) OsHAM2. (g–h) OsRLCK57. (j–k) OsTOC1. The color of each SNP indicates the  $r^2$  value for the correlation with the lead SNP. Red and green color intensities indicate stronger and weaker LD (0 to 1). The phenotypic differences were compared for haplotypes detected in more than three accessions.



**Figure 7**

Multiple linear regression models for TNM2, PTN, and PTR. (a) Relationships among traits and variants as independent variables for linear regression models. The paths marked by red, green, and blue squares indicate the traits and trait-associated variants involved in TNM2, PTN, and PTR as terminal traits, respectively. (b-d) Percent of variation explained by independent variables in linear regression models for TNM2 (b), PTN (c), and PTR (d). (e-f) Phenotypic variance in 47 independent japonica accessions. (e) TNM2. (f) PTN. (g) PTR. (h-j) Verification of linear regression models using 47 independent japonica accessions. (h) TNM2. (i) PTN. (j) PTR.

## Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- AdditionalFile2.xlsx
- AdditionalFile1.xlsx
- AdditionalFile4.pptx
- AdditionalFile3.pptx