Dynamic effects of interacting genes underlying rice flowering-time phenotypic plasticity and global adaptation

Tingting Guo[1], Qi Mu[1], Jinyu Wang[1], Adam Vanous[1], Akio Onogi[2], Hiroyoshi Iwata[3], Xianran Li[1,*], and Jianming Yu[1,*]

[1] Department of Agronomy, Iowa State University, Ames, IA 50011, USA
[2] Institute of Crop Science, National Agriculture and Food Research Organization, Ibaraki 305-8518, Japan
[3] Department of Agricultural and Environmental Biology, University of Tokyo, Tokyo 113-8657, Japan
[*] Corresponding authors. Emails: jmyu@iastate.edu; lixr@iastate.edu

Keywords: phenotypic plasticity, flowering time, genotype by environment interaction, reaction norm, genomics, haplotype

Supplemental Methods

Flowering-time genes underlying detected QTLs.

Four loci on chromosome 3, 6, 7, and 8 were consistently detected by all our mapping approaches. We reasoned that previously identified *Hd6, Hd1*, *Hd2*, and *Hd5* were responsible for the detected QTLs. First, although the confidence interval (on average of 6.16 cM) of the QTL covered many genes, previous research through map-based cloning strategy showed that only a single gene, was cloned from the corresponding interval with similar size for flowering time. This was the case for all four genes/QTL. Second, other known major flowering-time genes are located out of the intervals. For instance, four flowering-time genes (*Hd1*, *RFT1*, *Hd3a*, and *Hd17*) were cloned to be on chromosome 6 with populations derived from natural accessions, but only *Hd1* was within the QTL interval we detected (Hori et al. 2016).

Third, Koshihikari and Kasalath are two parental lines for deriving the mapping population in the present study. Besides this population, multiple genetic populations including either Koshihikari or Kasalath as the common parent have consistently detected these four QTLs for flowering time. The detected genes (*Hd1*, *Hd2*, *Hd5*, and *Hd6*) were declared as possessing large effects on flowering time in previous studies with 12 F_2 populations derived from crosses of the common parent Koshihikari with accessions originating from various regions in Asia (Matsubara and Yano 2018). Moreover, the four QTLs were co-localized with QTLs reported from multiple populations (F_2, RIL, NIL, etc) developed between Kasalath (one common parent) and Nipponbare. Finally, functional polymorphisms in two parental inbreds were identified from genome sequence information and agreed with previous QTL cloning findings.

Functional polymorphisms at flowering-time loci.

The two parents Kasalath and Koshihikari of the rice genetic population were included in the 3,000 Rice Genome Project, with unique IDs, CX227 and CX330, respectively (Wang et al. 2018a). The known genes underlying the QTLs were verified by checking the positions of significant markers and analyzing sequence polymorphisms between the two parents. The physical positions of

candidate genes were searched on Rice SNP-Seek Database and verified on Rice Genome Annotation Project. To identify DNA polymorphisms between two mapping parents, SNPs and structural variations within the coding regions of candidate genes were examined.

To determine the potential functional polymorphisms of the four flowering-time genes, we first tabulated the reported functional sites from the literature (Yano et al. 2000; Takahashi et al. 2001; Wei et al. 2010; Koo et al. 2013), then matched these characterized sites with the polymorphisms detected between the two parents. To verify the effects of the detected sites on flowering time, we further examined the differences of flowering time among the targeted Chromosome Segment Substitution Lines (CSSLs) and the two parents (Ebitani et al. 2005).

Sequence information of the 3,000 Rice Genomes Project.

The 3,000 Rice Genomes Project released sequencing data of 3,024 rice accessions from 89 countries (Wang et al. 2018b). The set of 3,010 genomes had an average mapping coverage of 92% and the estimated size of the genome was 375.1±20.9 Mb. Over 29 million single nucleotide polymorphisms (SNPs) in rice were discovered when aligned to the reference genome of the temperate japonica variety, Nipponbare (Version 7). Population structure analysis classified the 3,010 rice accessions into nine subpopulations, most of which could be connected to geographic origins. There were four XI clusters, three GJ clusters, and two single groups for South Asian cA and cB accessions.

Analysis of variance and estimation of variance components

The phenotypic value of genotype *i* when tested in replication *k* in environment *j* was modeled as $y_{ijk} = \mu + g_i + t_j + b_{k(j)} + (gt)_{ij} + e_{ijk}$, where μ is the population mean; g_i is the effect of genotype *i*; t_j is the effect of environment *j*; $b_{k(j)}$ is the block effect associated with replication *k* nested in the environment *j*; $(gt)_{ij}$ is the G × E effect associated with genotype *i* and environment *j*; and e_{ijk} is the error.

The analysis of variance was conducted by using R function "aov" which fits a model by a call to function "lm" for each stratum. The estimation of variance components was conducted by using R package "VCA" and function "anovaVCA".

Partitioning G × E into heterogeneity of genotypic variance and lack of genetic correlation

Following the steps laid out in previous publications (Yamada 1962; Cooper and Delacy 1994; Gibson and van Helden 1997), the G × E was partitioned into heterogeneity of genotypic variance among environments ($V = \frac{\sum_j (\sigma_{gj} - \bar{\sigma}_g)^2}{n_e - 1}$) and lack of genetic correlation ($L = \sigma_{ge}^2 - V$), where *j* is the j^{th} environment and n_e is the number of environments; σ_{ge}^2 is the variance of G × E; σ_{gj} is the genotypic variance in the j^{th} environment; and $\bar{\sigma}_g$ is the mean of genotypic variances in the individual environments.

Partitioning G × E into heterogeneity between regressions and error (F-W model)

Following the steps laid out in previous publications (Freeman 1971; Malosetti et al. 2013), the table for calculation of heterogeneity between regressions is listed as:

Source	Degree of Freedom	Mean Square
Line (Difference between genotypes)	(t-1)	$s\sum_i (d_i^{'})^2/(t-1)$
Environment (Joint regression)	(s-1)	$t\sum_j (\varepsilon_j^{'})^2/(s-1)$
Line × Environment	(t-1)(s-1)	
Heterogeneity between slopes	(t-1)	$\sum_i (\beta_i^{'})^2 \sum_j (\varepsilon_j^{'})^2/(s-1)$
Error	(t-1)(s-2)	$\sum_{ij} \delta_{ij}^2/(t-1)(s-2)$

where t is the number of lines; s is the number of environments; $d_i^{'}$ is the marginal effect of genotype i plus the overall mean; $\varepsilon_j^{'}$ is the marginal effect of environment j plus the overall mean; $\beta_i^{'}$ is the slope of regressing observed values on $\varepsilon_j^{'}$; and δ_{ij} is the regression leftover from G × E effect.

Partitioning G × E following the additive main effects and multiplicative interaction (AMMI) model

The additive main effects and multiplicative interaction model was written as: $y_{ij} = \mu + g_i + t_j + \sum_{k=1}^{K} b_{ij} Z_{jk} + \varepsilon_{ijk}$, where the phenotypic value is the mean of genotype i in environment j; $\mu + g_i$ is the mean of genotype i across all environments; t_j serves as the environmental index; the G × E is explained by K multiplicative terms (k = 1…K). Each multiplicative term is formed by the product of a genotypic sensitivity b_{ik} (genotypic score) and a hypothetical environmental characterization Z_{jk} (environmental score) (Vaneeuwijk 1995; Malosetti et al. 2013).

The analysis of AMMI was carried out through two steps: 1) ANOVA for generating G × E effect; 2) SVD (singular value decomposition) of G × E effect for generating principal component (PC) scores.

Multi-environment QTL mapping.

The following linear regression model (Li et al. 2008) was used for the combined analysis of nine environments, $y_{ih} = b_{oh} + \sum_{j=1}^{m+1} b_{jh} x_{ij} + \varepsilon_{ih}$, where y_{ih} is the phenotypic value of the i^{th} individual in the h^{th} environment; b_{oh} is the overall mean of linear model in the h^{th} environment; x_{ij} is the indicating variable for the j^{th} marker's genotype of the i^{th} individual, which is equal to 1 for parent 1 type or -1 for parent 2 type; b_{jh} is the partial regression coefficient of phenotype on the j^{th} marker in the h^{th} environment; and ε_{ih} is the residual random error in the h^{th} environment that is assumed to be normally distributed. Stepwise regression was conducted for phenotypic values in each environment to select significant markers.

Epistasis QTL mapping.

The following model (Li et al. 2015) was used in epistasis QTL mapping, $y_i = b_0 + \sum_{j=1}^{m+1} b_j x_{ij} + \sum_{j<k} b_{jk} x_{ij} x_{ik} + e_i$, where y_i is the trait phenotypic value of the i^{th} individual in the mapping population; b_0 is the overall mean of the linear model; x_{ij} is a dummy variable for the

genotype of the i^{th} individual at the j^{th} marker, taking value 1 for homozygote of marker type, and -1 for heterozygote; b_j is the partial regression coefficient of the phenotype on the j^{th} marker variable; b_{jk} is the partial regression coefficient of the phenotype on the multiplication variable of the j^{th} and k^{th} markers; and e_i is the residual random error which is assumed to be normally distributed.

Joint Genomic Regression Analysis (JGRA) for performance prediction.
The process of searching for an environmental index was the same for all three performance prediction scenarios: 1) predicting the performance of tested genotypes in untested environments; 2) untested genotypes in tested environments; and 3) untested genotypes in untested environments.

For the first scenario, the leave-one-environment-out cross-validation was conducted. 1) Let the j^{th} (j = 1, 2, 3…, m) environment be the untested environment, and the remaining as the training (tested) environments. 2) Search the environmental index by using environmental mean from the tested genotypes in the tested environments. 3) For the i^{th} (i = 1, 2, 3…, n) genotype, regress the observed phenotypes from the tested environments on the corresponding environmental index ($GDD_{9\text{-}50}$ for the genetic mapping population) to obtain intercept and slope estimates. 4) Predict the phenotype in the j^{th} untested environment by supplying the fitted linear models (regression models from step 3) with the value of the corresponding environmental index from the j^{th} untested (to be predicted) environment. 5) Repeat step 1 to 4 until each environment is predicted.

For the second scenario, we conducted leave-one-half-genotypes-out cross-validation. 1) Equally split n genotypes as the tested genotypes and untested genotypes. 2) Search the environmental index by using environmental mean from the tested genotypes. 3) Regress the observed phenotypes on the identified environmental index to obtain intercept and slope estimates for each tested genotype. 4) Treating intercept and slope as new "traits", run genomic prediction through "rrBLUP" (Endelman 2011a) to predict the intercept and slope for each untested genotype. 5) Predict phenotypes of the untested genotypes with the estimated intercept, estimated slope, and the environmental index value of each environment.

For the third scenario, we conducted leave-one-environment-and-one-half-genotypes-out cross-validation. 1) Let the j^{th} (j = 1, 2, 3…, m) environment be the untested environment, and the remaining as the training (tested) environments. 2) Equally split n genotypes in the tested genotypes and untested genotypes. 3) Search the environmental index by using environmental mean from the tested genotypes in the tested environments. 4) Regress the observed phenotypes on the identified environmental index to obtain intercept and slope estimates for each tested genotype. 5) Treating intercept and slope as new "traits", run genomic prediction through rrBLUP to predict the intercept and slope for each untested genotype. 6) Predict phenotypes of the untested genotypes in the untested environment with the estimated intercept, the estimated slope, and the environmental index value from the untested environment. 7) Repeat step 1 to 6 until each environment is processed.

Reaction norms at multiple levels.
The contribution of genotype by environment interaction to phenotypic variability was first defined as the "norm of reaction" by Richard Woltereck (Woltereck 1909; Woltereck 1928). Replicates of a specific genotypes (clones) may develop differently in different environments.

Different genotypes do not necessarily respond similarly in the same environment. In general, norm of reaction (or reaction norm) was defined as the pattern in phenotypes produced by a given genotype under different environmental conditions (Griffiths et al. 1996). Reaction norms have been observed by using a genetic population under varied environments (Li et al. 2018), but have not been dissected and showed using a particular single-nucleotide polymorphism genotype or a haplotype under different environments, although this was suggested (Walsh 2017).

Reaction norms of genotypes observed as individuals from a genetic population

Reaction norms of genotypes were described by visualizing two-dimensional data that use environmental index as *x*-axis, and phenotype as *y*-axis. We used linear lines in the reaction norm graph because the pattern of genotype responding to environment often showed strongly linear relationship as implemented in the joint regression analysis (Finlay and Wilkinson 1963). Each line represents an individual genotype. The lines were shown in two formats: lines connecting observed data points together or fitted regression lines with linear regression. Environmental mean was used first and then replaced by the environmental index ($GDD_{9\text{-}50}$). Identifying the environmental index enabled the calculation of the index value for a new environment, for which no environmental mean was observed.

Reaction norms of genotypes at the single-locus level

After QTL identification and candidate gene examination, we partitioned the genetic population into two groups by using the most significant marker for each QTL, each group representing one homozygous genotype (AA *versus* BB, as either of the two parents). In an individual environment, the genotypic value for AA was the average phenotypic values across all individuals having AA genotype at the locus, same for BB. Two lines were shown in the reaction norm graph, representing two different groups. Linear regression was applied to show the relationship between genotypic value and environment, also to show the fitted genotypic values across all environments.

Reaction norms of genotypes at the multi-locus combination level

Combinations of *4* loci resulted in 2^4 haplotypes, which represented homozygous genotypes across these loci. Similar to the calculation of phenotype for single locus, the phenotype for each class of haplotype was the average phenotypic values across all the genotypes having the same haplotype. There were 16 different phenotypes representing 16 haplotypes in each environment. Linear regression was applied to show the relationship between genotypic value and environment, also to show the fitted genotypic values across all environments.

Reaction norms of genome-wide marker effect continua

We implemented joint genomic regression analysis (JGRA) to model and predict flowering time. The approach of JGRA through genome-wide marker effect continuum was used for observing reaction norms of marker effects and combining marker effects for flowering time prediction. In each environment, the marker effects were estimated by using the mixed model in Package "rrBLUP" (Endelman 2011b). Linear regression was applied for each marker with marker effect as *y*-axis, and $GDD_{9\text{-}50}$ as *x*-axis.

Reaction norms of genetic effects at the single-locus level

We estimated genetic effects for flowering-time QTLs by using inclusive composite interval mapping in QTL IciMapping. As the QTL mapping was conducted under individual environments

and there were nine environments in total, each QTL was estimated to have nine genetic effects. The reaction norms of genetic effects to environmental index $GDD_{9\text{-}50}$ were exhibited by establishing single linear models.

Reaction norms of genetic effects at the single-locus level can also be established by using reaction-norm parameters of genotypes, intercept and slope. First, intercept and slope for each genotype were calculated by regressing flowering time on $GDD_{9\text{-}50}$. Second, linkage mapping was conducted by considering intercept and slope as two separate phenotypes. Genetic effects for detected QTLs were estimated. Third, the genetic effects from intercept and slope were combined into predicted genetic effects under individual environments. Forth, the reaction norms of genetic effects at the single-locus level were exhibited by using the line graph: predicted QTL effect as *y*-axis, and $GDD_{9\text{-}50}$ as *x*-axis.

References in Supplemental Methods:

Cooper M, Delacy IH. 1994. Relationships among analytical methods used to study genotypic variation and genotype-by-environment interaction in plant breeding multi-environment experiments. *Theor Appl Genet* **88**: 561-572.

Ebitani T, Takeuchi Y, Nonoue Y, Yamamoto T, Takeuchi K, Yano M. 2005. Construction and evaluation of chromosome segment substitution lines carrying overlapping chromosome segments of indica rice cultivar 'Kasalath' in a genetic background of japonica elite cultivar 'Koshihikari'. *Breeding Science* **55**: 65-73.

Endelman JB. 2011a. Ridge regression and other kernels for genomic selection with R package rrBLUP. *The Plant Genome* **4**: 250-255.

Endelman JB. 2011b. Ridge regression and other kernels for genomic selection with R package rrBLUP. *Plant Genome* **4**: 250-255.

Finlay K, Wilkinson G. 1963. The analysis of adaptation in a plant-breeding programme. *Australian journal of agricultural research* **14**: 742-754.

Freeman GH. 1971. Environmental and Genotype-Environmental Components of Variability .8. Relations between Genotypes Grown in Different Environments and Measures of These Environments. *Heredity* **27**: 15-&.

Gibson G, van Helden S. 1997. Is function of the Drosophila homeotic gene Ultrabithorax canalized? *Genetics* **147**: 1155-1168.

Griffiths AJ-F, Miller JH, Suzuki DT, Lewontin RC, Gelbart WM. 1996. An introduction to genetic analysis. W. H. Freeman, New York.

Hori K, Matsubara K, Yano M. 2016. Genetic control of flowering time in rice: integration of Mendelian genetics and genomics. *Theoretical and Applied Genetics* **129**: 2241-2252.

Koo BH, Yoo SC, Park JW, Kwon CT, Lee BD, An G, Zhang ZY, Li JJ, Li ZC, Paek NC. 2013. Natural Variation in OsPRR37 Regulates Heading Date and Contributes to Rice Cultivation at a Wide Range of Latitudes. *Mol Plant* **6**: 1877-1888.

Li H, Ribaut JM, Li Z, Wang J. 2008. Inclusive composite interval mapping (ICIM) for digenic epistasis of quantitative traits in biparental populations. *Theor Appl Genet* **116**: 243-260.

Li S, Wang J, Zhang L. 2015. Inclusive Composite Interval Mapping of QTL by Environment Interactions in Biparental Populations. *PLoS One* **10**: e0132414.

Li X, Guo T, Mu Q, Li X, Yu J. 2018. Genomic and environmental determinants and their interplay underlying phenotypic plasticity. *Proceedings of the National Academy of Sciences* **115**: 6679-6684.

Malosetti M, Ribaut JM, van Eeuwijk FA. 2013. The statistical analysis of multi-environment data: modeling genotype-by-environment interaction and its genetic basis. *Front Physiol* **4**: 44.

Matsubara K, Yano M. 2018. Genetic and molecular dissection of flowering time control in rice. In *Rice Genomics, Genetics and Breeding*, pp. 177-190. Springer.

Takahashi Y, Shomura A, Sasaki T, Yano M. 2001. Hd6, a rice quantitative trait locus involved in photoperiod sensitivity, encodes the alpha subunit of protein kinase CK2. *Proceedings of the National Academy of Sciences of the United States of America* **98**: 7922-7927.

Vaneeuwijk FA. 1995. Linear and Bilinear Models for the Analysis of Multi-Environment Trials .1. An Inventory of Models. *Euphytica* **84**: 1-7.

Walsh B. 2017. Crops can be strong and sensitive. *Nature Plants* **3**: 694-695.

Wang W, Mauleon R, Hu Z, Chebotarov D, Tai S, Wu Z, Li M, Zheng T, Fuentes RR, Zhang F et al. 2018a. Genomic variation in 3,010 diverse accessions of Asian cultivated rice. *Nature* **557**: 43-49.

Wang WS, Mauleon R, Hu ZQ, Chebotarov D, Tai SS, Wu ZC, Li M, Zheng TQ, Fuentes RR, Zhang F et al. 2018b. Genomic variation in 3,010 diverse accessions of Asian cultivated rice. *Nature* **557**: 43-+.

Wei XJ, Xu JF, Guo HN, Jiang L, Chen SH, Yu CY, Zhou ZL, Hu PS, Zhai HQ, Wan JM. 2010. DTH8 Suppresses Flowering in Rice, Influencing Plant Height and Yield Potential Simultaneously. *Plant Physiol* **153**: 1747-1758.

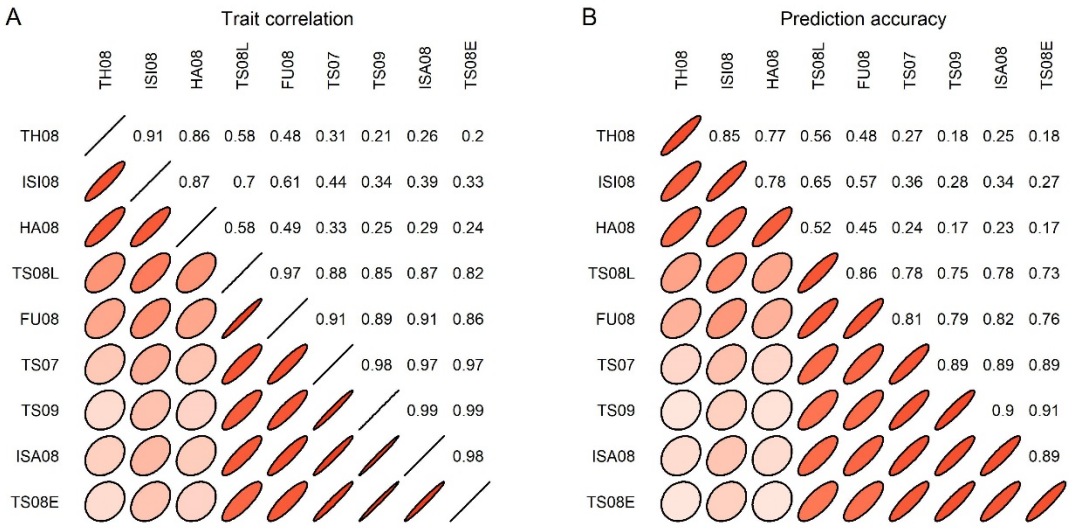
Woltereck R. 1909. Weitere experimentelle Unter-suchungen uber Artveranderung, speziell uber das Wesen quantitativer Artunterschiede bei Daphniden. *Verh D Tsch Zool Ges* **1909**: 110-172.

Woltereck R. 1928. Bemerkungen über die Begriffe “Reaktionsnorm” und “Klon”. *Biol Zent Bl* **48**: 167-172.

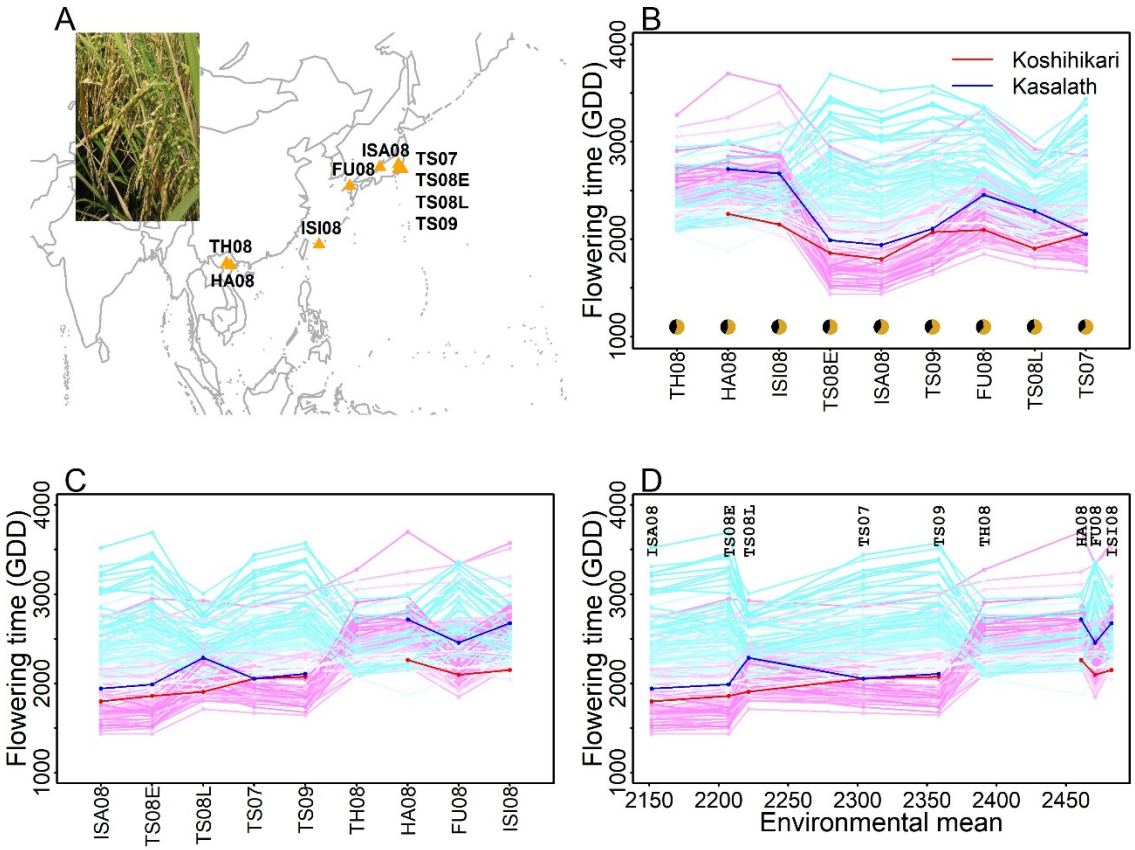
Yamada Y. 1962. Genotype × environment interaction and genetic correlation of the same trait under different environments. *Janpanese Journal of Genetics* **37**: 498-509.

Yano M, Katayose Y, Ashikari M, Yamanouchi U, Monna L, Fuse T, Baba T, Yamamoto K, Umehara Y, Nagamura Y et al. 2000. Hd1, a major photoperiod sensitivity quantitative trait locus in rice, is closely related to the arabidopsis flowering time gene CONSTANS. *Plant Cell* **12**: 2473-2483.

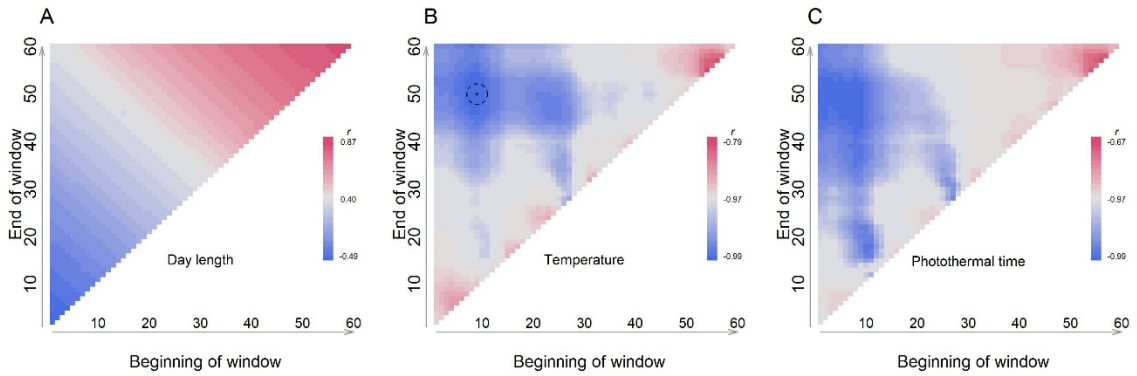
Supplemental Figures



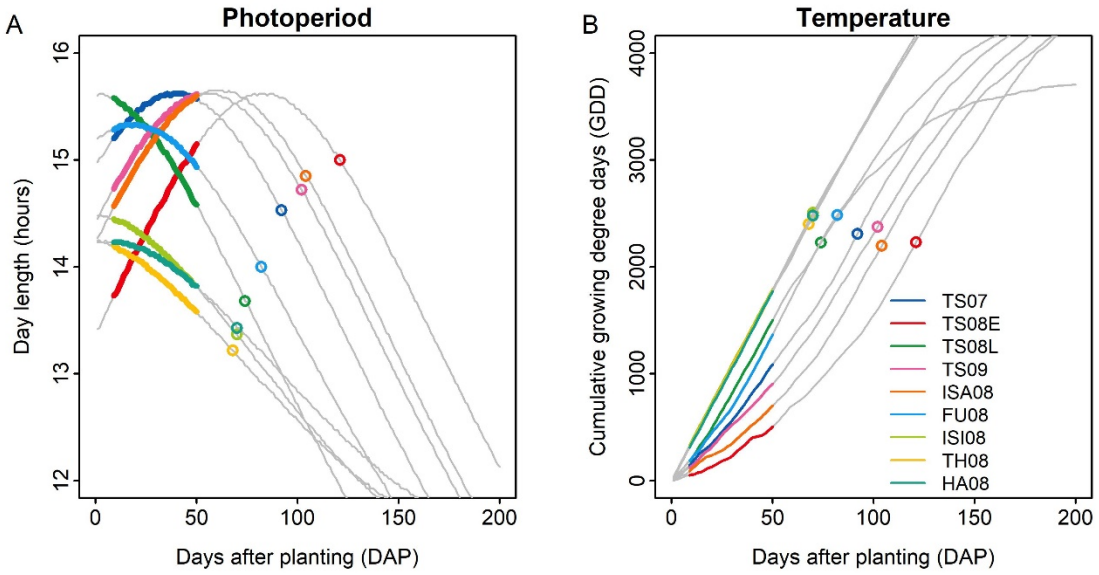
Supplemental Figure S1. Trait correlation and prediction accuracy between environments. (**A**) Flowering time correlations between each pair of environments. Two clusters are detected with the first three environments grouping together and the other six grouping together (**B**) Prediction accuracy within individual environments (diagonal) and between two environments (off diagonal and row to column). Prediction accuracy is the correlation between predicted values and observed values. Lower left and upper right are visualizations and values, respectively.



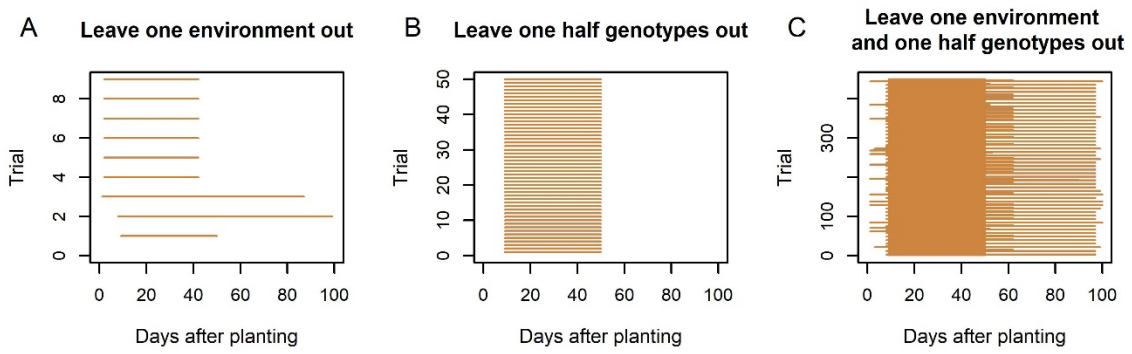
Supplemental Figure S2. Phenotypic plasticity in rice when flowering time was expressed as growing degree days (GDD). (**A**) Nine natural field environments obtained from six field sites. (**B**) Reaction norm for flowering time based on a categorical order of day length. (**C**) Reaction norm based on a categorical order of environmental mean. (**D**) Reaction norm based on a numerical order of environmental mean. Flowering time expressed as GDD does not show a pattern that can be readily modeled, unlike when flowering time expressed as days after planting (Figure 1 and Figure 2).



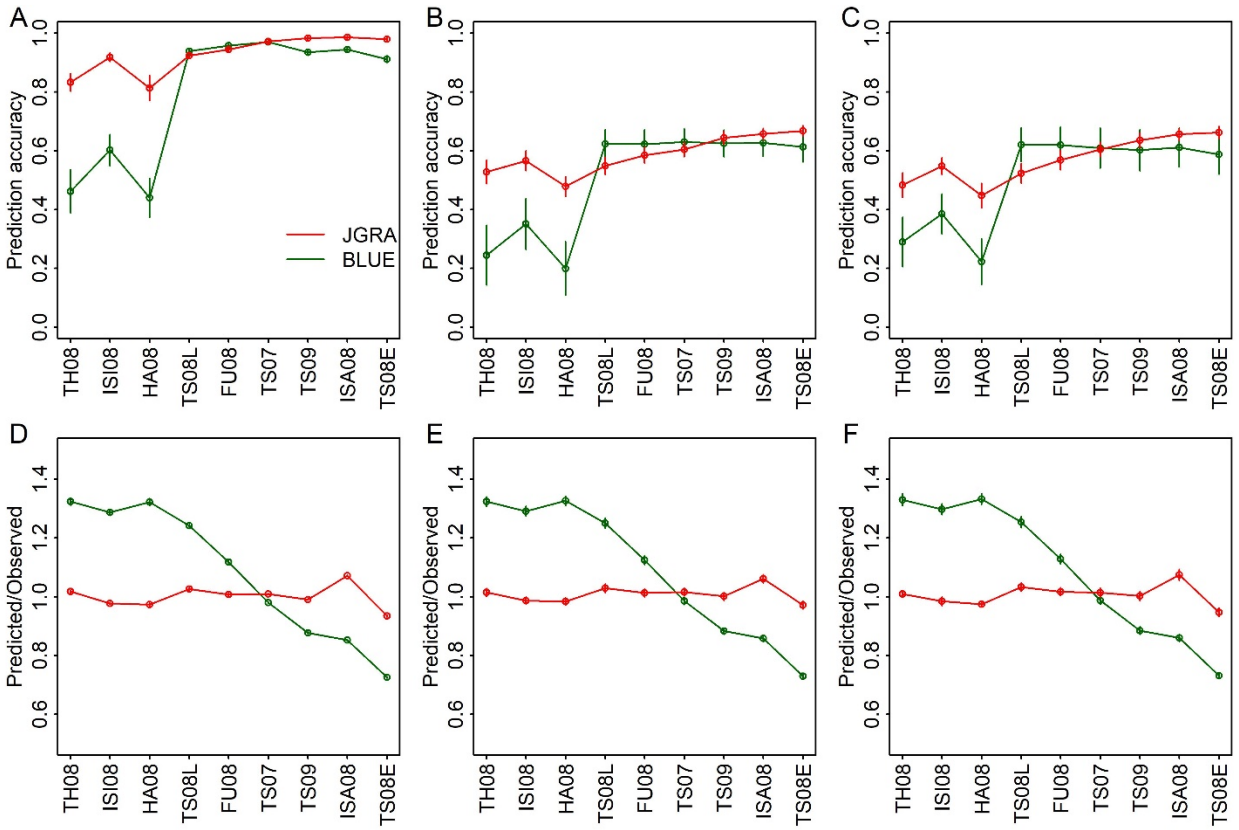
Supplemental Figure S3. Environmental index search across different windows. (**A**) Correlation between environmental mean and day length (DL). (**B**) Correlation of environmental mean with temperature (GDD). (**C**) Correlation of environmental mean with photothermal time (PTT = GDD × DL). Temperature within the window of 9-50 days after planting was chosen as the environmental index and denoted as $GDD_{9\text{-}50}$. PTT was not chosen due to the negligible increase in correlation strength when considering the additional environmental factor (photoperiod) beyond temperature to obtain PTT, and that the general overlap between the windows from GDD (**B**) and PTT (**C**).



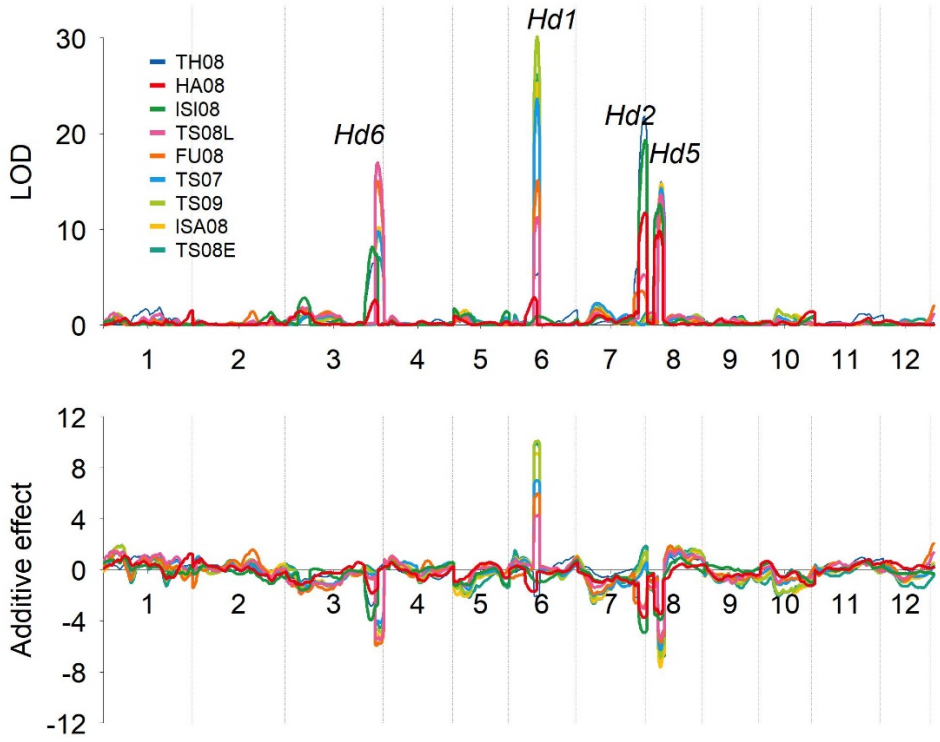
Supplemental Figure S4. Photoperiod (Day length) and temperature (cumulative GDD) profiles of the nine environments. (**A**) Photoperiod profiles of individual environments. All nine environments are long-day length (LD) environments with the classification threshold of 13.5 hour. (**B**) Cumulative temperature profiles of individual environments. The bold segments of the lines represent the critical window (9-50 days after planting) when environmental inputs were sensed by rice plants to determine the flowering time. The open circles represent the average flowering time of the population (*i.e.*, environmental mean).



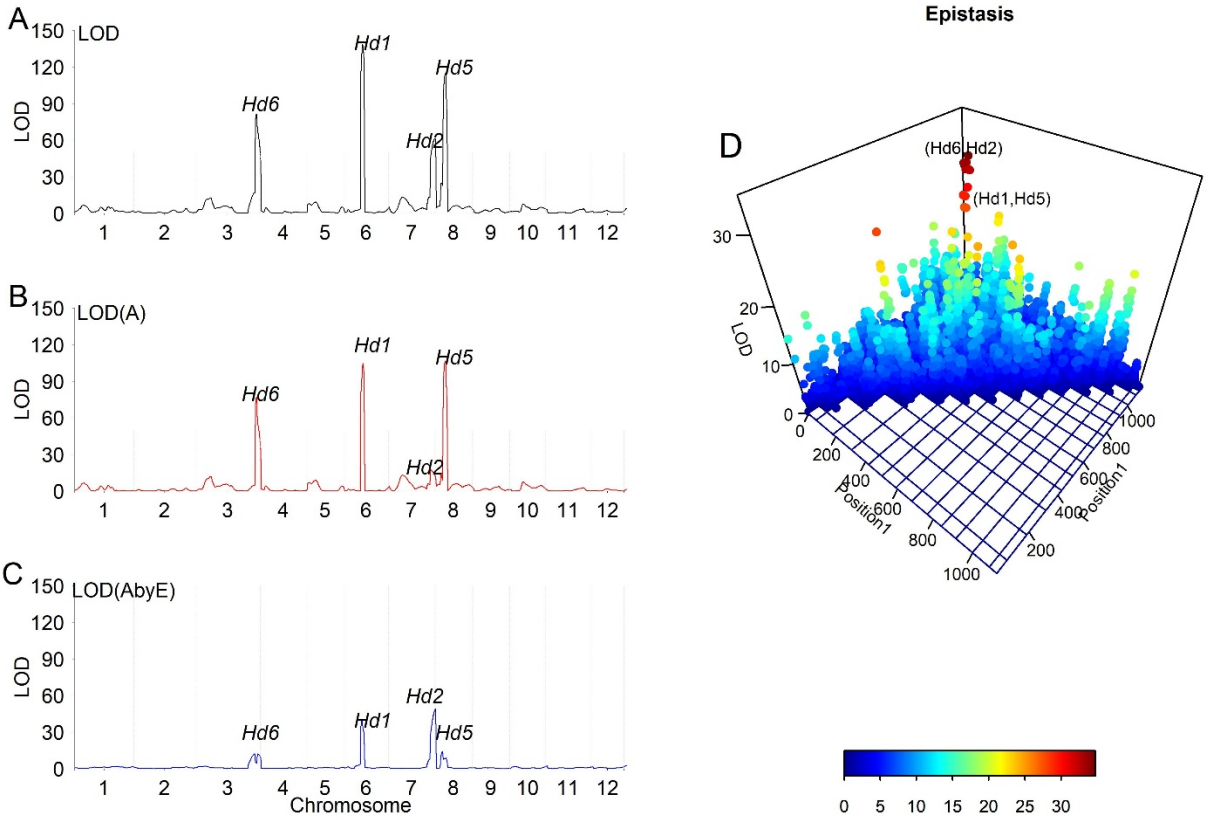
Supplemental Figure S5. Subsampling analysis to verify the consistency of the 9-50 days-after-planting window that gave the environmental index ($GDD_{9\text{-}50}$) identified through searching with overall environmental mean (from all genotypes) of flowering time across 9 environments. Three sets of simulation experiments were conducted, (**A**) Leave-one-environment-out. Each line segment is the best growth window obtained from 8 other environments after omitting that particular environment. Trial 1-9 stand for TS07, TS08E, TS08L, TS09, ISA08, FU08, ISI08, TH08, and HA08. (**B**) Leave-one-half-genotypes-out. Within each trial, 50% of genotypes were used to search for the growth window that gave the best correlation between GDD and environmental mean. (**C**) Leave-one-environment-and-one-half-genotypes-out. The combinations of A and B (450 trials = 9 environments × 50 times of different 50% genotypes).



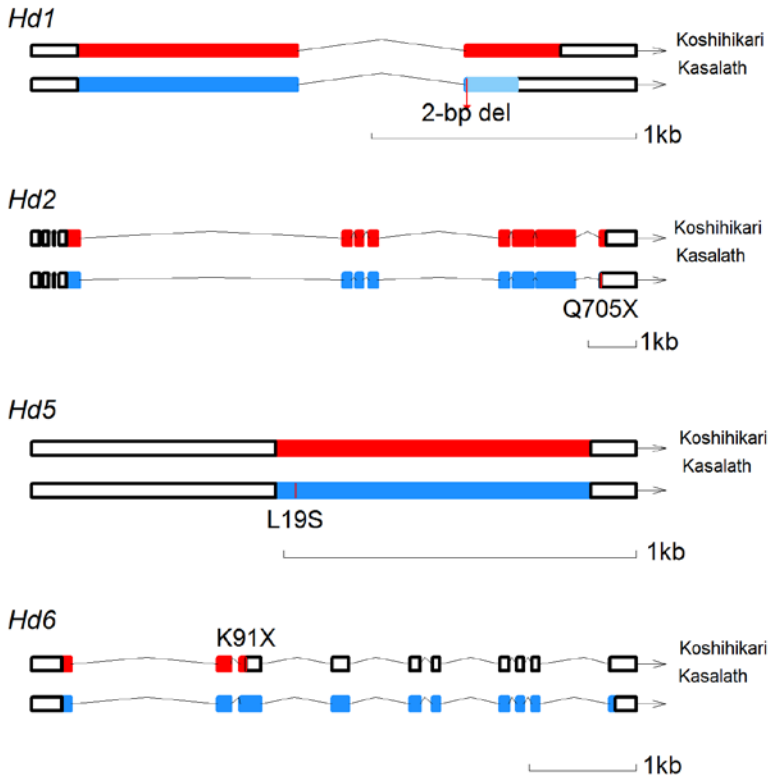
Supplemental Figure S6. Prediction accuracy and on-target assessment of joint genomic regression analysis (JGRA) through reaction-norm parameters compared with fixed predictions relying on averages across tested environments (best linear unbiased estimation, BLUE). (**A**) Prediction accuracy for tested genotype under untested environment. (**B**) Prediction accuracy for untested genotypes under tested environments. (**C**) Prediction accuracy for untested genotypes under untested environments. (**D**) Predicted/observed ratio for tested genotypes under untested environments, (**E**) Predicted/observed ratio for untested genotypes under tested environments. (**F**) Predicted/observed ratio for untested genotypes under untested environments. Mean and standard deviation across 50 runs were obtained by randomly sampling 50% genotypes for each performance prediction scenario.



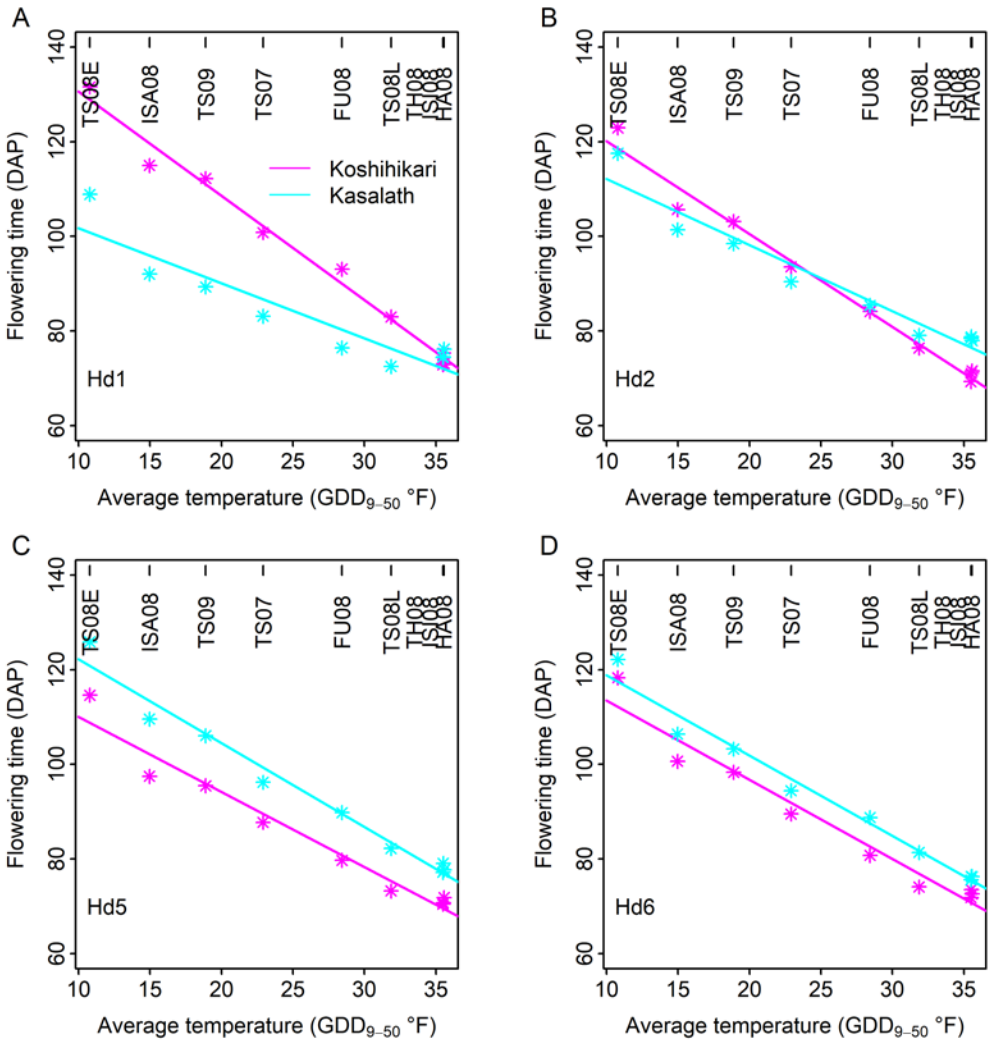
Supplemental Figure S7. QTL mapping in individual environments detected *Hd1*, *Hd2*, *Hd5*, and *Hd6*. Loci detected from individual environment analysis show varied effects across environments. Upper panel shows the LOD scores and lower panel shows the additive effects. Significance threshold of LOD = 2.91 was based on permutation. There is a direction change for additive effects of *Hd2* and *Hd1*.



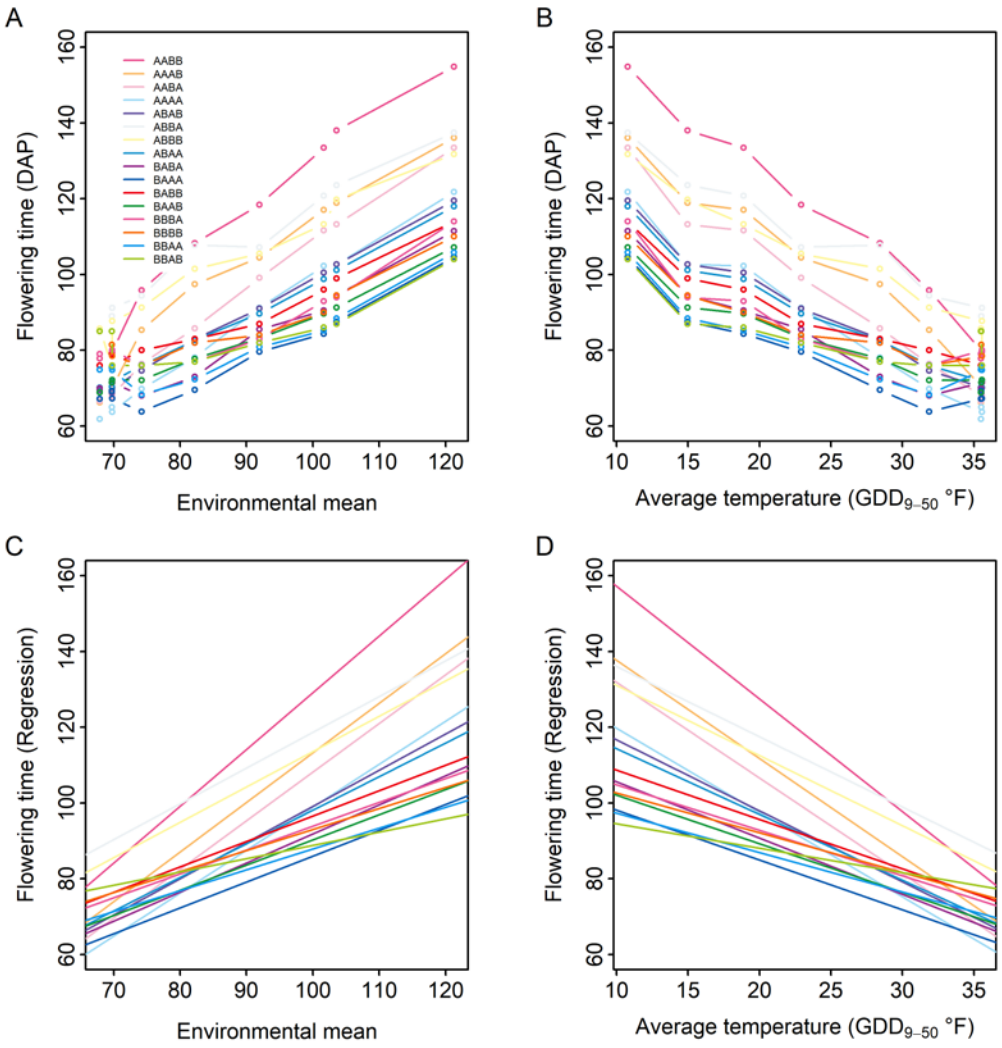
Supplemental Figure S8. Multi-environment QTL mapping to identify additive and epistasis QTLs. (**A**) LOD score resulted from both additive effect and additive effect × environment. (**B**) LOD score resulted from only additive effect. (**C**) LOD score resulted from only additive effect × environment. (**D**) Epistasis QTL mapping. The highest peak is from interaction between *Hd2* and *Hd6*, and the third highest peak is from interaction between *Hd1* and *Hd5*. Significance thresholds for all analyses were based on permutation: 6.91 for additive effect detection and 9.10 for epistasis detection.



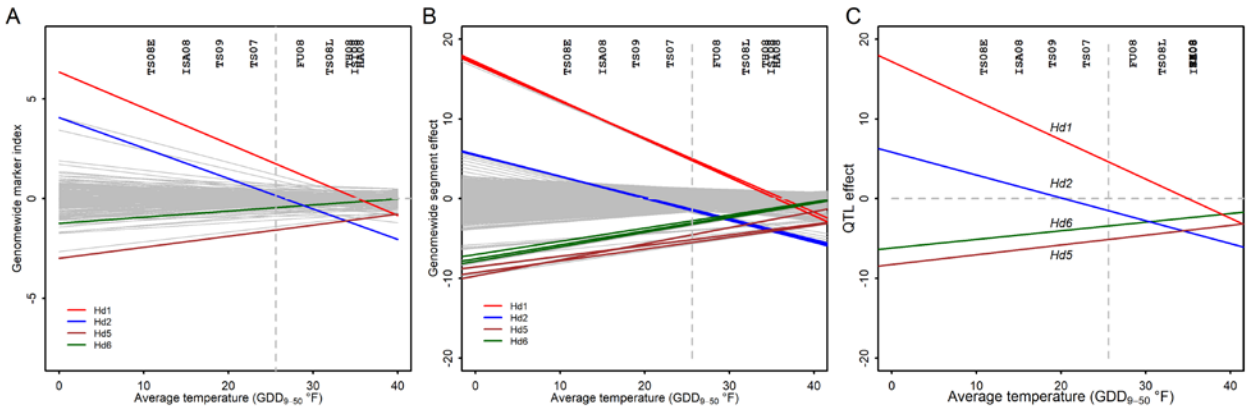
Supplemental Figure S9. Functional polymorphisms between two parents in four flowering-time genes (*Hd1*, *Hd2*, *Hd5*, and *Hd6*). For *Hd1* on chromosome 6, the Kasalath allele has a 2-bp insertion, leading to a frame shift and a premature stop codon. For *Hd2* on chromosome 7, the Kasalath allele has a SNP substitution resulting a premature stop codon. For *Hd5* on chromosome 8, the Kasalath allele has an amino acid substitution from L to S at the 19th amino acid. For *Hd6* on chromosome 3, the Koshihikari allele has a SNP substitution resulting a premature stop codon. Note: This figure is the same as the main Fig. 4C. The purpose of having it here is to describe the polymorphisms in detail.



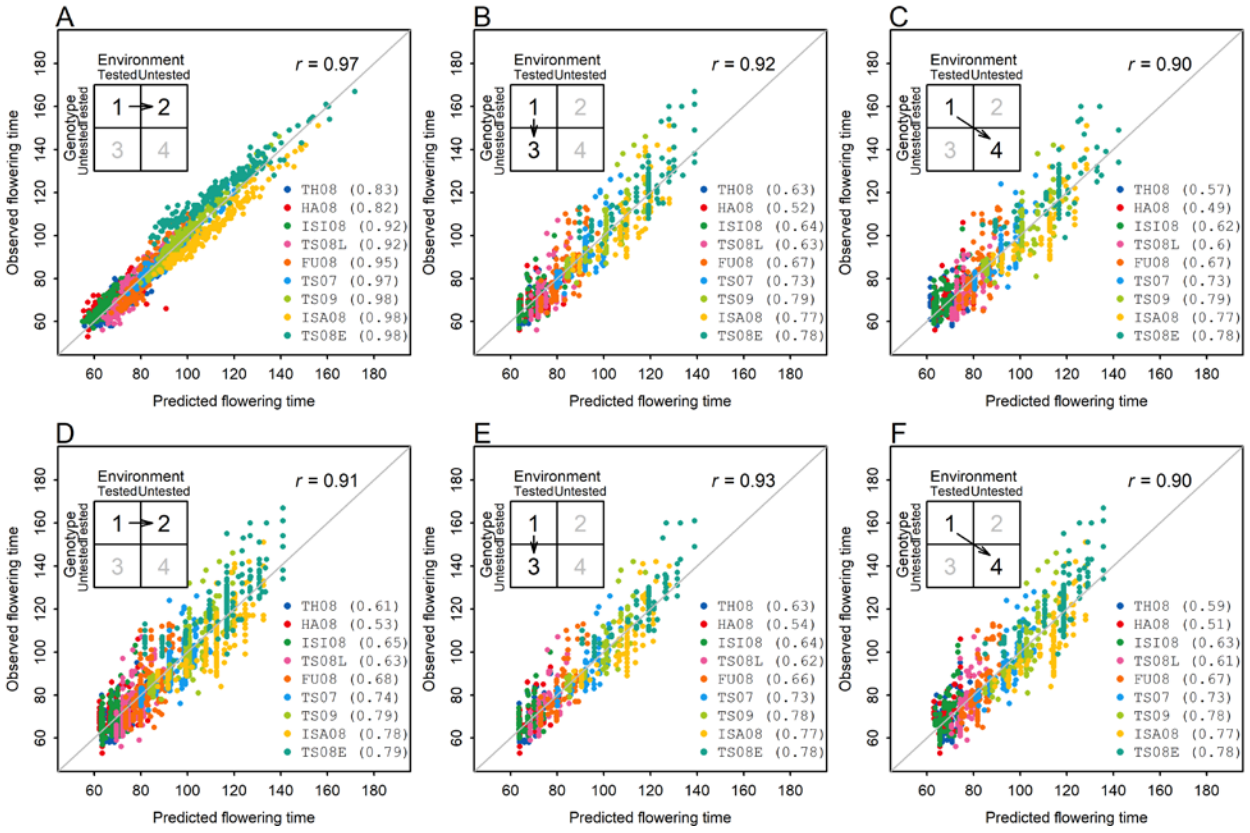
Supplemental Figure S10. Reaction norms of genotypes at the single-locus level to temperature changes. Two genotypes are represented by two alleles for each gene: *Hd1* (**A**), *Hd2* (**B**), *Hd5* (**C**), and *Hd6* (**D**). Allele from parent Koshihikari is represented by magenta line and dots, while allele from parent Kasalath is represented by cyan line and dots.



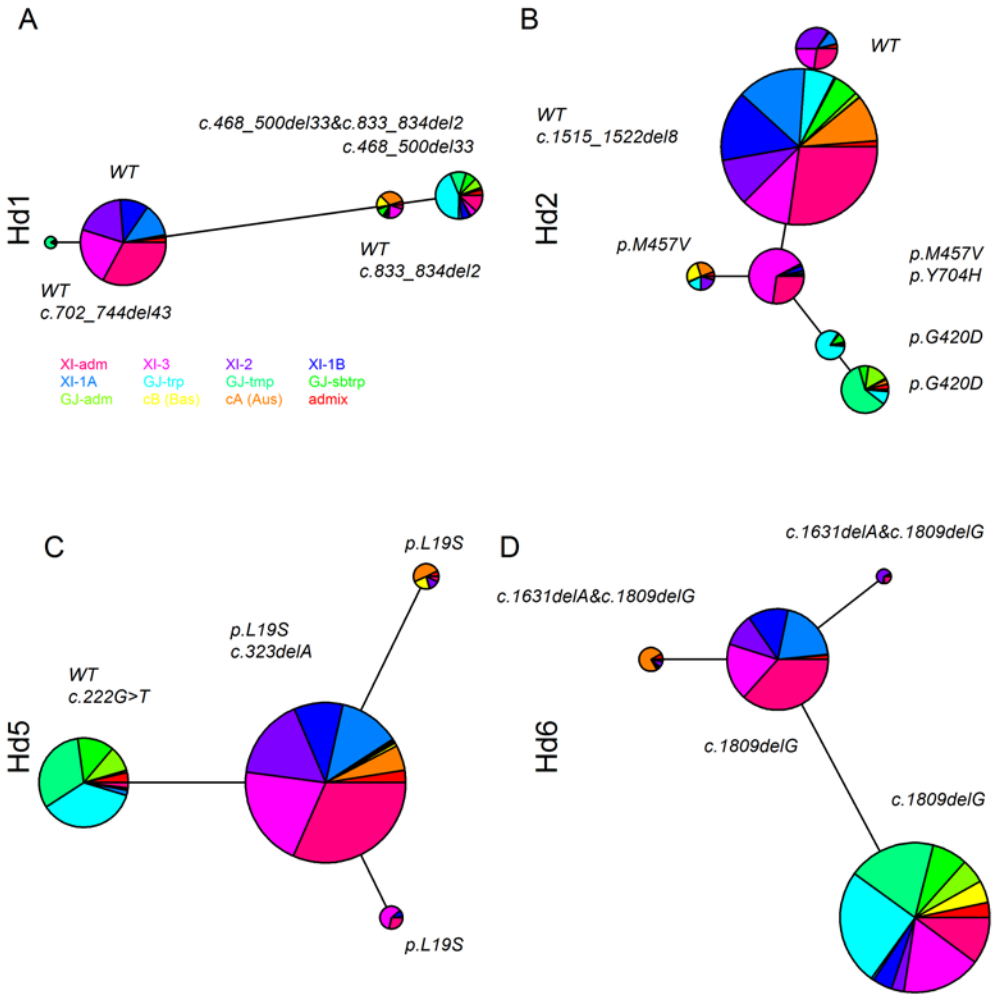
Supplemental Figure S11. Reaction norms of genotypes at the multi-locus combination level under different environmental conditions. Sixteen genotypes are represented by four genes, each having two alleles, A from Koshihikari and B as Kasalath. Allelic combinations are listed in the order of *Hd1*, *Hd2*, *Hd5*, and *Hd6*. (**A**) Reaction norms for flowering time to different environmental mean values. (**B**) Reaction norms for flowering time to different temperature values. (**C**) Fitted reaction norms for flowering time to different environmental mean values. (**D**) Fitted reaction norms for flowering time to different temperature values.



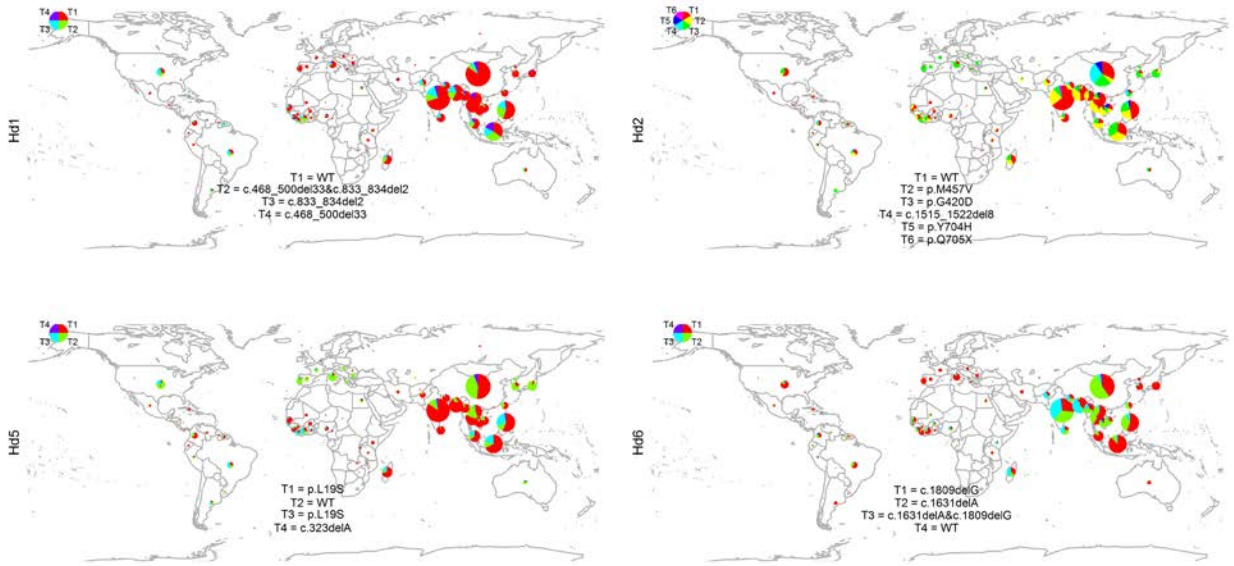
Supplemental Figure S12. Fitted reaction norms at the genome-wide marker effect level and for the major flowering-time loci. (**A**) Fitted reaction norms of marker effect continua along the environmental index by temperature ($GDD_{9\text{-}50}$). The closest flanking markers to the loci are highlighted in color and their distances to the QTL peaks are within 1 cM. (**B**) Fitted reaction norms of genome segment effects along the environmental index by temperature ($GDD_{9\text{-}50}$). The whole genome is partitioned into 1,120 segments of 1 cM. The segments including the gene position with ±1 cM are highlighted. (**C**) Fitted reaction norms of QTL effects along the environmental index by temperature ($GDD_{9\text{-}50}$).



Supplemental Figure S13. Performance prediction of flowering time with four gene loci to leverage environmental index and genomic prediction. (**A-C**) JGRA using four gene loci reaction-norm parameters. (**D-F**) JGRA using marker effects for four gene loci. The three scenarios are: predicting performance for tested genotypes in untested environments (A, D), predicting untested genotypes in tested environments (B, E), and predicting untested genotypes in untested environments (C, F). Prediction accuracy within each individual environment (in parentheses) and across all environments (*r*) are indicated; the diagonal line indicates the exact match between observed and predicted values. Note: panel A is the same as the panel A in Fig. 3 because this is the case where effects are not partitioned to markers.



Supplemental Figure S14. Haplotype determination for four flowering-time genes across rice diverse accessions. Haplotype networks of *Hd1* (**A**), *Hd2* (**B**), *Hd5* (**C**), and *Hd6* (**D**) were constructed, each haplotype matching with the corresponding functional sites. Size of haplotype is proportional to the total number of accessions from XI-adm, XI-3, XI-2, XI-1B, XI-1A, GJ-trp, GJ-tmp, GJ-sbtrp, GJ-adm, cA (Aus), cB (Bas), and Admix.



Supplemental Figure S15. Geographic distribution of major haplotypes observed in the 3,010 diverse rice accessions. The different colors represent different haplotypes of *Hd1* (**A**), *Hd2* (**B**), *Hd5* (**C**), and *Hd6* (**D**). The relative size of each pie indicates the percentage of accessions sampled from a country.

Supplemental Tables
(All tables are also in Excel files)

Supplemental Table S1 Extract: Flowering time in the multi-environment trial with two replications in each of nine environments. (**Full Supplemental Table S1 available for download with this paper**)

Line Code	Replicate	Environment	Flowering Time (days) [a]
Line1	1	TH08	73
Line1	1	HA08	72
Line1	1	ISI08	NA
Line1	1	TS08L	77
Line1	1	FU08	79
Line1	1	TS07	90
Line1	1	TS09	102
Line1	1	ISA08	103
Line1	1	TS08E	117
Line1	2	TH08	73
Line1	2	HA08	73
Line1	2	ISI08	74
Line1	2	TS08L	76
Line1	2	FU08	82
Line1	2	TS07	91
Line1	2	TS09	101
Line1	2	ISA08	102
Line1	2	TS08E	118

Note: [a], missing values are represented by NA.

Supplemental Table S2: Summary of analysis of variance for flowering time from the combined analysis of nine environments.

Source	Degree of Freedom	Sum of Squares	Mean Square	F value	Pr (>F)
Environment	8	947911.750	118488.969	97085.932	< 2.2e-16 ***
Rep (Environment)	9	267.755	29.751	24.622	< 2.2e-16 ***
Genotype	175	268770.291	1535.830	1258.396	< 2.2e-16 ***
Genotype × Environment	1369	115113.727	84.086	68.897	< 2.2e-16 ***
Residual	1509	1841.666	1.220		

Supplemental Table S3: Estimates of variance components from the combined analysis of nine environments.

Variance Component	Estimate	Standard Error	Percent of Total
Environment	347.029	18.629	73.285
Rep (Environment)	0.159	0.398	0.034
Genotype	83.203	9.122	17.571
Genotype × Environment	41.920	6.475	8.853
Residual	1.220	1.105	0.258
Entry-mean based heritability	0.946		

Supplemental Table S4: Estimates of variance components and parameters for the combined analysis of flowering time tested in nine environments.

Variance Component and Parameter	Estimate	Ratio to G × E
Genotypic variance component (σ_g^2)	83.203	1.985
G × E interaction variance component (σ_{ge}^2)	41.920	-
Heterogeneity of genotypic variance	8.812 (21%)	-
Lack of genetic correlation	33.108 (79%)	-
Error variance component (σ_ε^2)	1.220	0.029
Entry-mean based heritability (h^2)	0.946	-
Pooled genetic correlation (r_g)	0.734	-

Supplemental Table S5: Analysis of variance for flowering time following the regression on the mean model (Finlay-Wilkinson model).

Source [a]	Degree of Freedom	Sum of Squares	Mean Square	F value	Pr (>F)
Environment	8	489154.908	61144.364	4754.464	< 2.2e-16 ***
Genotype	175	136066.095	777.521	60.458	< 2.2e-16 ***
Heterogeneity of slopes	175	43245.647	247.118	19.215	< 2.2e-16 ***
ε	1225	15754.005	12.860		
Total	1583	684220.654			

Note: [a], analysis was done for the means of genotypes in individual environments.

Supplemental Table S6 Extract: Weather data for the multi-environment trial. (**Full Supplemental Table S6 available for download with this paper**)

Environment	Date	Max Temp [a]	Min Temp [b]	Day Length	GDD	Photothermal Time
TS07	5/14/2007	70.7	60.26	14.98	15.48	231.8904
TS07	5/15/2007	71.78	57.11	15.02	14.445	216.9639
TS07	5/16/2007	73.4	53.96	15.03	13.68	205.6104
TS07	5/17/2007	74.39	59.72	15.05	17.055	256.6778
TS07	5/18/2007	75.38	59.63	15.08	17.505	263.9754
TS07	5/19/2007	73.04	58.01	15.12	15.525	234.738
TS07	5/20/2007	76.91	58.01	15.15	17.46	264.519
TS07	5/21/2007	72.59	56.39	15.18	14.49	219.9582
TS07	5/22/2007	76.01	62.69	15.2	19.35	294.12
TS07	5/23/2007	80.87	64.4	15.22	22.635	344.5047
TS07	5/24/2007	80.87	66.2	15.25	23.535	358.9087
TS07	5/25/2007	81.725	65.03	15.27	23.3775	356.9744
TS07	5/26/2007	82.58	63.86	15.3	23.22	355.266
TS07	5/27/2007	82.94	59.18	15.32	21.06	322.6392
TS07	5/28/2007	67.46	59.18	15.33	13.32	204.1956
TS07	5/29/2007	72.05	54.5	15.37	13.275	204.0367
TS07	5/30/2007	70.79	57.02	15.38	13.905	213.8589

Note: [a], Maximum temperature; [b], Minimum temperature.

Supplemental Table S7: Analysis of variance for predicted flowering time values.

Source	Degree of Freedom	Sum of Squares	Mean Square	Estimate of Variance	Percent of Total	Standard Deviation	Coefficient of Variation (%)
Predictions for tested genotypes in untested environments (1 to 2)							
Environment	8	476528.285	59566.036	338.242	73.600	18.391	21.218
Genotype	175	141437.060	808.212	85.861	18.683	9.266	10.690
G×E/Residual	1400	49646.482	35.462	35.462	7.716	5.955	6.870
Predictions for untested genotypes in tested environments (1 to 3)							
Environment	8	513170.776	64146.347	364.412	86.262	19.090	21.969
Genotype	175	77615.969	443.520	48.186	11.406	6.942	7.989
G×E/Residual	1400	13787.929	9.849	9.849	2.331	3.138	3.612
Predictions for untested genotypes in untested environments (1 to 4)							
Environment	8	500835.869	62604.484	355.648	85.594	18.859	21.754
Genotype	175	79598.093	454.846	49.374	11.883	7.027	8.105
G×E/Residual	1400	14677.688	10.484	10.484	2.523	3.238	3.735

Supplemental Table S8 Extract: Functional sites and flowering time effect of *Hd1*, *Hd2*, *Hd5* and *Hd6*. (**Full Supplemental Table S8 available for download with this paper**)

Gene	Name	Chromo-some	Start Position	End Position	Functional Site	Functional Site Position
Hd1	Os06g0275000	6	9336359	9338643	33 bp del	9337002
Hd1	Os06g0275000	6	9336359	9338643	43 bp del	9337236
Hd1	Os06g0275000	6	9336359	9338643	2 bp del	9338004
Hd2	Os07g0695100	7	29616705	29629223	SNP	29623803
Hd2	Os07g0695100	7	29616705	29629223	SNP	29627021
Hd2	Os07g0695100	7	29616705	29629223	SNP	29627212
Hd2	Os07g0695100	7	29616705	29629223	8 bp del	29627357
Hd2	Os07g0695100	7	29616705	29629223	SNP	29628481
Hd2	Os07g0695100	7	29616705	29629223	SNP	29628484
Hd2	Os07g0695100	7	29616705	29629223	SNP	29628500
Hd5	Os08g0174500	8	4332106	4334829	SNP	4334714

Supplemental Table S9 Extract: Collected information for the 3,000 Rice Genomes accessions. (**Full Supplemental Table S9 available for download with this paper**)

ID	Country	Region	Admixture	Group	Longitude	Latitude
B001	China	EAS	GJ-tmp	GJ	104.195397	35.86166
B002	China	EAS	GJ-tmp	GJ	104.195397	35.86166
B003	China	EAS	GJ-adm	GJ	104.195397	35.86166
B004	Japan	EAR	GJ-tmp	GJ	138.252924	36.204824
B005	Japan	EAR	GJ-tmp	GJ	138.252924	36.204824
B006	Viet Nam	SEA	XI-adm	XI	108.277199	14.058324
B007	Viet Nam	SEA	XI-adm	XI	108.277199	14.058324
B008	Viet Nam	SEA	GJ-tmp	GJ	108.277199	14.058324
B009	Viet Nam	SEA	XI-1A	XI	108.277199	14.058324
B010	Malaysia	SER	XI-1A	XI	101.975766	4.210484
B011	India	SAC	XI-adm	XI	78.96288	20.593684
B012	India	SAC	XI-2	XI	78.96288	20.593684
B013	Sri Lanka	IOC	XI-adm	XI	80.771797	7.873054
B014	Uzbekistan	WAS	GJ-tmp	GJ	64.585262	41.377491
B015	Romania	EUR	XI-1A	XI	24.96676	45.943161
B016	Hungary	EUR	GJ-tmp	GJ	19.5033041	47.162494
B017	Bulgaria	EUR	GJ-tmp	GJ	25.48583	42.733883

Supplemental Table S10: Haplotypes determination based on functional sites within genes.

Gene	Representative Accession	Haplotype Name 1 [a]	Haplotype Name 2 [b]
Hd1	B130	Nipponbare(Hd1)	*WT*
Hd1	IRIS_313-11946	Kasalath(Hd1)	*c.468_500del33&c.833_834del2*
Hd1	IRIS_313-9239	Hap2	*c.833_834del2*
Hd1	CX282	Hap1	*c.468_500del33*
Hd2	B147	PRR37-1	*WT*
Hd2	IRIS_313-11651	PRR37-2b	*p.M457V*
Hd2	IRIS_313-10689	PRR37-2	*p.G420D*
Hd2	B062	PRR37-1a	*c.1515_1522del8*
Hd2	B076	PRR37-1e	*p.Y704H*
Hd2	IRIS_313-10534	PRR37-1c	*p.Q705X*
Hd5	IRIS_313-8643	Type3-5	*p.L19S*
Hd5	IRIS_313-10656	Type1	*WT*
Hd5	IRIS_313-11946	Type2	*c.222G>T*
Hd5	IRIS_313-12016	Type8	*c.323delA*
Hd6	CX282	Hap2	*c.1809delG*
Hd6	B130	Hap1	*c.1631delA*
Hd6	IRIS_313-10002	Hap3	*c.1631delA&c.1809delG*
Hd6	B195	Hap4	*WT*

Note: [a], Haplotype name from references; [b], Haplotype name from recommendations for the description of DNA changes (Mutation nomenclature).

Supplemental Table S11 Extract: Multi-gene haplotype determination. (**Full Supplemental Table S11 available for download with this paper**)

Accession	*Hd1*	*Hd2*	*Hd5*	*Hd6*	Hap_4_Genes	No. of Multi-Gene Haplotype
B007	Nipponbare(Hd1)	PRR37-1	Type3-5	Hap1	2	267
B001	Nipponbare(Hd1)	PRR37-2	Type1	Hap2	38	257
B030	Nipponbare(Hd1)	PRR37-1	Type3-5	Hap3	3	159
CX110	Nipponbare(Hd1)	PRR37-2b	Type3-5	Hap2	24	147
B086	Nipponbare(Hd1)	PRR37-1	Type3-5	Hap2	1	139
CX144	Nipponbare(Hd1)	PRR37-2b	Type3-5	Hap1	25	109
B010	Nipponbare(Hd1)	PRR37-1a	Type3-5	Hap1	44	88
CX106	Kasalath(Hd1)	PRR37-2	Type2	Hap2	90	67
B036	Hap1	PRR37-2	Type1	Hap2	149	59
CX149	Nipponbare(Hd1)	PRR37-2b	Type3-5	Hap3	26	53
B134	Nipponbare(Hd1)	PRR37-1	Type1	Hap2	5	48
B006	Nipponbare(Hd1)	PRR37-1e	Type3-5	Hap1	54	47
B020	Hap2	PRR37-2	Type2	Hap2	124	37
B025	Kasalath(Hd1)	PRR37-2	Type1	Hap2	88	35
B002	Nipponbare(Hd1)	PRR37-2b	Type1	Hap2	28	34
CX368	Hap2	PRR37-1	Type3-5	Hap3	101	27
CX151	Kasalath(Hd1)	PRR37-2b	Type2	Hap2	83	26
B180	Kasalath(Hd1)	PRR37-1	Type3-5	Hap2	69	25
CX59	Hap1	PRR37-2b	Type3-5	Hap2	141	24
CX226	Hap2	PRR37-1	Type3-5	Hap2	99	21
B043	Kasalath(Hd1)	PRR37-1	Type2	Hap2	75	20

Supplemental Table S12: Overall F-statistics for population genetic differentiation based on SNPs within the gene.

	Hs [a]	Ht [b]	Gst_Nei [c]
Hd1	0.384	0.672	0.429
Hd2	0.521	0.831	0.373
Hd5	0.370	0.675	0.452
Hd6	0.390	0.669	0.418
Overall	0.416	0.712	0.415

Note: [a], heterozygosity of sub populations; [b], heterozygosity of the total populatoin; [c], Nei's Gst, (Ht - Hs)/Ht. Reference: Nei M, Chesser RK. (1983). Estimation of fixation indices and gene diversities. Annals of Human Genetics. 47: 253-259.

Supplemental Table S13: Functional sites and documented flowering time effect for *Hd1*, *Hd2*, *Hd5* and *Hd6* in two parents: Kasalath and Koshihikari.

ID	Parent Name	*Hd1*	*Hd2*	*Hd5*	*Hd6*
CX227	Kasalath	c.468_500del33&c.833_834del2	p.Q705X	p.L19S	Wildtype
CX330	Koshihikari	Wildtype	Wildtype	Wildtype	p.K91X
Koshihikari allele [a]					
Short day (10-h light)		Early	Late	Early	Late
Long day (14.5-h light)		Late	Early	Early	Early

Note: a, flower time data were obtained from reference: Ebitani et al. (2005) Construction and evaluation of chromosome segment substitution lines carrying overlapping chromosome segments of indica rice cultivar 'Kasalath' in a genetic background of japonica elite cultivar 'Koshihikari'. Breeding Science 55: 65-73.