GWAS with principal component analysis identifies a gene comprehensively controlling rice architecture

Kenji Yano[a,b], Yoichi Morinaka[a], Fanmiao Wang[a], Peng Huang[a], Sayaka Takehara[a], Takaaki Hirai[a], Aya Ito[a], Eriko Koketsu[a], Mayuko Kawamura[a], Kunihiko Kotake[a], Shinya Yoshida[c], Masaki Endo[d], Gen Tamiya[b], Hidemi Kitano[a], Miyako Ueguchi-Tanaka[a], Ko Hirano[a], and Makoto Matsuoka[a,1]

[a]Bioscience and Biotechnology Center, Nagoya University, 464-8601 Nagoya, Japan; [b]Statistical Genetics Team, RIKEN Center for Advanced Intelligence Project, Nihonbashi, 103-0027 Tokyo, Japan; [c]Hyogo Prefectural Research Center for Agriculture, Forestry and Fisheries, Kasai, 679-0198 Hyogo, Japan; and [d]Institute of Agrobiological Sciences, National Agriculture and Food Research Organization, 305-8634 Tsukuba, Japan

Edited by Jiayang Li, Institute of Genetics and Developmental Biology (CAS), Beijing, China, and approved August 22, 2019 (received for review April 9, 2019)

Elucidation of the genetic control of rice architecture is crucial due to the global demand for high crop yields. Rice architecture is a complex trait affected by plant height, tillering, and panicle morphology. In this study, principal component analysis (PCA) on 8 typical traits related to plant architecture revealed that the first principal component (PC), PC1, provided the most information on traits that determine rice architecture. A genome-wide association study (GWAS) using PC1 as a dependent variable was used to isolate a gene encoding rice, SPINDLY (OsSPY), that activates the gibberellin (GA) signal suppression protein SLR1. The effect of GA signaling on the regulation of rice architecture was confirmed in 9 types of isogenic plant having different levels of GA responsiveness. Further population genetics analysis demonstrated that the functional allele of *OsSPY* associated with semidwarfism and small panicles was selected in the process of rice breeding. In summary, the use of PCA in GWAS will aid in uncovering genes involved in traits with complex characteristics.

plant architecture | PCA | GWAS | gibberellin | SPINDLY

Plant architecture, a collection of important agronomic traits that determine grain production in rice, is affected by various factors, including plant height, tillering, and panicle morphology (1–3). Rice breeders have attempted to improve plant architecture in compliance with grower demands. In over 100 years of Japanese rice cultivation, breeders have developed many rice varieties with plant architectures that can be briefly categorized into 2 categories: the panicle number type (large number and small-sized panicles with short plant stature) and the panicle weight type (small number and large-sized panicles with tall plant stature) (4) (*SI Appendix*, Fig. S1). Although varieties with both large panicle numbers and large size are desirable, this is difficult to achieve in practical breeding because of the complex correlation between components of plant architecture and the trade-off between number and size (5, 6).

Principal component analysis (PCA) is an effective means of extracting key information from phenotypically complex traits that are highly correlated while retaining the original information (7, 8). PCA can transform a set of correlated variables into a substantially smaller set of uncorrelated variables as principal components (PCs), which can capture most information from the original data (9). In this study, PCA was performed for rice architecture, and a genome-wide association study (GWAS) using PC scores was utilized to identify genetic factors regulating plant architecture. This approach was validated as effective in identifying causal genes associated with plant architecture.

Results

Phenotypic Analysis by PCA. A total of 169 *japonica* rice varieties grown in 2014 and 2015 were used in this study (Dataset S1). Eight traits related to plant architecture, including days-to-heading, culm length, panicle number, and 5 panicle-related traits (panicle length, rachis length, primary branch number per panicle, secondary branch number per panicle, and spikelet number per panicle) were measured (*SI Appendix*, Fig. S2). To investigate the relationships among trait variables and the factors underlying trait variation, PCA was performed for all 8 traits. For results in 2015, PC1 explained 62% of the trait variance (Fig. 1*A*). Except for the traits of days-to-heading and panicle number, the other 6 traits showed high positive loadings on PC1 (0.80–0.92), while panicle number showed negative loading (−0.54) (Fig. 1 *A* and *B*). This result suggested that plants with high PC1 scores exhibited long culms, large panicle sizes, and small panicle numbers, and vice versa. This corresponds to a trade-off relationship between panicle number and panicle weight. PC2 explained 16% of the total variance, and the loading on PC2 was high for days-to-heading (0.83) (Fig. 1*A*), suggesting that PC2 is representative of days-to-heading. This component was also loaded (0.55) with panicle number, which is consistent with the observation that prolonged vegetative growth due to late heading increases the number of panicles per plant. The PCA results for traits measured in 2014 were consistent with 2015 results (*SI Appendix*, Fig. S3), suggesting that PC1 and PC2 can be used as quantitative

Significance

Rice architecture is an important agronomic trait for determining yield; however, the complexity of this trait makes it difficult to elucidate the molecular mechanisms. This study applied a strategy of using principal components (PCs) as dependent variables for a genome-wide association study (GWAS). *SPINDLY* was identified to regulate rice architecture by suppressing gibberellin (GA) signaling. Further study using GA-signaling mutants confirmed that levels of GA responsiveness regulate rice architecture, suggesting that the utilization of a favorable *SPINDLY* allele will improve crop productivity. The strategy presented in this study of performing GWAS using PC scores will provide valuable information for plant genetics and will improve our understanding of complex traits at the molecular level.

Author contributions: K.Y., Y.M., G.T., H.K., M.U.-T., K.H., and M.M. designed research; K.Y., Y.M., P.H., S.T., T.H., A.I., E.K., M.K., K.K., and M.U.-T. performed research; K.Y., Y.M., P.H., S.T., T.H., K.K., and G.T. analyzed data; M.E. contributed new reagents/analytic tools; K.Y., Y.M., F.W., S.Y., M.U.-T., and M.M. wrote the paper; and S.Y. prepared the population material.

The authors declare no conflict of interest.

This article is a PNAS Direct Submission.

Data deposition: The sequence data reported in this study have been deposited in the DDBJ Sequence Read Archive (DRA) under accession numbers DRA004358 and DRA008452.

[1]To whom correspondence may be addressed. Email: makoto@agr.nagoya-u.ac.jp.

This article contains supporting information online at www.pnas.org/lookup/suppl/doi:10.1073/pnas.1904964116/-/DCSupplemental.

First published September 30, 2019.

A

Traits	PC1	PC2	PC3
Days-to-heading	0.37	0.83	0.18
Culm length	0.82	0.28	-0.16
Panicle length	0.85	0.24	0.25
Rachis length	0.91	0.16	0.07
Primary branch No.	0.80	-0.28	-0.33
Sec. branch No.	0.91	-0.17	-0.11
Spikelet No.	0.92	-0.23	-0.17
Panicle No.	-0.54	0.55	-0.59
Eigenvalue	4.98	1.31	0.62
Proportion of Variance	0.62	0.16	0.08
Cumulative Proportion	0.62	0.79	0.86

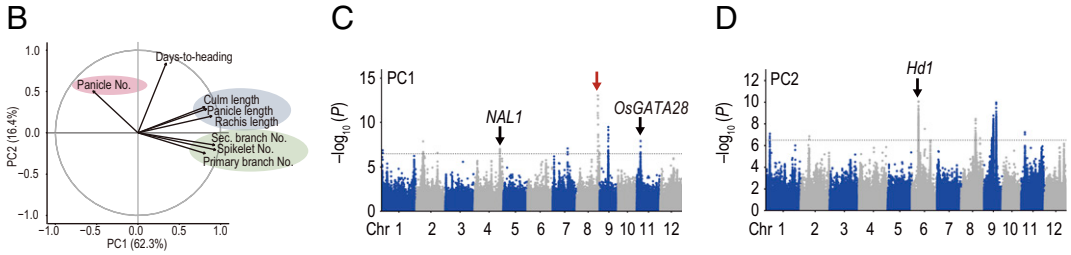

Fig. 1. PCA and GWAS for plant architecture. (*A*) Summary of the first 3 PCs (PC1, PC2, PC3) for 8 traits in the dataset of 169 *japonica* rice varieties. (*B*) Loading plot of PC1 and PC2. Red, blue, and green indicate clusters of panicle numbers, length-related traits (culm length, panicle length, rachis length), and number-related traits (secondary branch number, spikelet number, primary branch number), respectively. Circle corresponds to 100% of explained variance. Proportion of variances for PC1 and PC2 are shown in parentheses. (*C* and *D*) GWAS for PC1 (*C*) and PC2 (*D*) are shown using the data obtained in 2015. In Manhattan plots, horizontal dotted lines represent a significant threshold ($P = 3.2 \times 10^{-7}$). The positions of *NAL1* (*NARROW LEAF1*), *Hd1* (*HEADING DATE1*), and *OsGATA28* (*LOC_Os11g08410*) are indicated by black arrows; the red arrow (*C*) indicates a peak that was further analyzed.

indices to characterize plant architecture and heading date, respectively.

GWAS for PC Scores. The normality of 8 traits and PC scores was examined. The PCs (PC1 to PC3) displayed normal distribution and were suitable for GWAS; however, days-to-heading (2014, 2015), culm length (2014, 2015), panicle length (2015), and panicle number (2014, 2015) deviated significantly from normal distribution (shown in red in *SI Appendix*, Fig. S4). These results indicated that PCA can transform skewed data to a normal distribution, which consequently improves the statistical power for application to GWAS. Using a linear mixed model with correction of kinship bias, GWAS was conducted for PCs and all 8 traits (Fig. 1 *C* and *D* and *SI Appendix*, Figs. S5–S7). GWAS for PC1 and PC2 identified a total of 15 peaks ($P < 3.2 \times 10^{-7}$); however, significant peaks were not detected for PC3 (Fig. 1 *C* and *D* and *SI Appendix*, Fig. S7), which indicated that PC3 might be composed of nongenetic factors, such as variations caused by differences in the growth conditions of the plants. The significant peaks detected in GWAS for PC1 and PC2 are listed together with their phenotypic variance (*SI Appendix*, Fig. S7 *E* and *F* and Dataset S2). Interestingly, 2 peaks with an effect of 0.12–0.17 on PC1 contained 2 genes, *NAL1* and *OsGATA28*/*LOC_Os11g08410*; these were previously identified as causal genes involved in plant architecture (Fig. 1*C*) (10). Furthermore, 1 peak with an effect of 0.21–0.39 on PC2 contained the flowering-time controller gene *Hd1* (Fig. 1*D*). These results supported our hypothesis that PC1 and PC2 could be good indicators for plant architecture and heading date, respectively.

Next, we focused on the highest peak of PC1, which was located at the terminal end of the long arm of chromosome (Chr) 8 (red arrows in Fig. 1*C* and *SI Appendix*, Fig. S7*A*). Local Manhattan plots and linkage disequilibrium (LD) analysis revealed that this quantitative trait loci (QTL) was delineated to 26.0–28.4 Mb and consisted of 2 peaks: 26.2–26.7 Mb (peak 1) and 27.9–28.2 Mb (peak 2) (Fig. 2*A*). The GWAS for culm length showed results similar to the results for PC1 (*SI Appendix*, Fig. S8 *C* and *D*). It has been reported that 2 possibilities can account for multiple peaks in 1 LD region: Multiple causal variants in different peak regions are independently associated with phenotypic variation or only 1 peak contains causal variants accompanied by false-positive peaks due to an indirect synthetic association induced by a confounding genetic background (10–12). As the latter case was considered to be more likely, we conducted another GWAS for culm length using a different population consisting of 133 *japonica* rice varieties (Dataset S3). A significant peak was detected at the same region of Chr 8 ($P = 5.0 \times 10^{-7}$) (red arrow in *SI Appendix*, Fig. S8*A*); however, the local Manhattan plot contained only peak 2 (*SI Appendix*, Fig. S8*B*). This strongly suggested that peak 1 was caused by a spurious association with peak 2, which was due to the complex genome structure of the 169 accessions.

For further isolation of causal genes, all polymorphisms in the peak 2 region were studied. There were 571 polymorphisms, with 34 mapping to the coding region of 14 genes (Dataset S4), which could induce the changes in gene function. Among these, we focused on one gene, *LOC_Os08g44510*, which is annotated as *N*-acetyl glucosamine transferase or rice *SPINDLY* (*OsSPY*). This gene functions as a negative regulator of gibberellin (GA) signaling and modulates plant growth and development (13–15). Two haplotypes were identified in the GWAS panel; these were designated haplotype I (Hap_I) in the Nipponbare (NP) reference genome and the alternative, Hap_II (Fig. 2*B*). Another haplotype, Hap_III, was predominant in *indica* genomes (see below). The polymorphisms induced 2 amino acid exchanges (S9T and R833L). The corresponding residue of S9 is not conserved in plant species, suggesting that S9T may not impact protein function (*SI Appendix*, Fig. S9). In contrast, R833 is located at the C-terminal end in a well-conserved region of the enzymatic domain (15).

The effect of R833L on the *O*-fucosyltransferase activity of OsSPY was examined using a rice DELLA protein, SLR1, as a substrate as described previously (15). A truncated OsSPY containing the 3 C-terminal repeats of tetratricopeptide and the enzymatic domain from residues 317–928 (with or without the replacement of R833L) was produced in *Escherichia coli* for enzyme assays (*SI Appendix*, Fig. S10). The 3TPR-OsSPY^{Hap_I} showed a significantly higher enzymatic activity than 3TPR-SPY^{Hap_II} (Fig. 2*C*), indicating that the R833L exchange partially diminished OsSPY activity.

The average PC1 score for varieties carrying Hap_II of OsSPY was significantly higher than for those carrying Hap_I (Fig. 2*D*); however, this was not the case for PC2 (Fig. 2*E*), confirming that *OsSPY* is only associated with PC1. Next, we examined the haplotype effects on 8 phenotypic traits. No difference between haplotypes was observed for days-to-heading (Fig. 2*F*). Compared to plants harboring Hap_I of *OsSPY*, Hap_II varieties showed increased length of culm/panicle/rachis, larger numbers of primary branch/secondary branch/spikelet per panicle, and reduced panicle numbers (Fig. 2 *G*–*M* and *SI Appendix*, Fig. S11).

Involvement of *OsSPY* in Plant Architecture. To confirm the effect of *OsSPY* on plant architecture, we attempted to knockout *OsSPY* by CRISPR-Cas9. However, with one exception none of the CRISPR/Cas9 knockout plants could grow; knockout plants with a 4-base pairs (bp) deletion in the 14th exon exhibited constitutively elongated internodes in the young seedling stage and immediately died (*SI Appendix*, Fig. S12). Next, we generated *OsSPY*_RNAi plants to examine the effect of *OsSPY*. The RNAi plants did not show seriously abnormal phenotypes at the

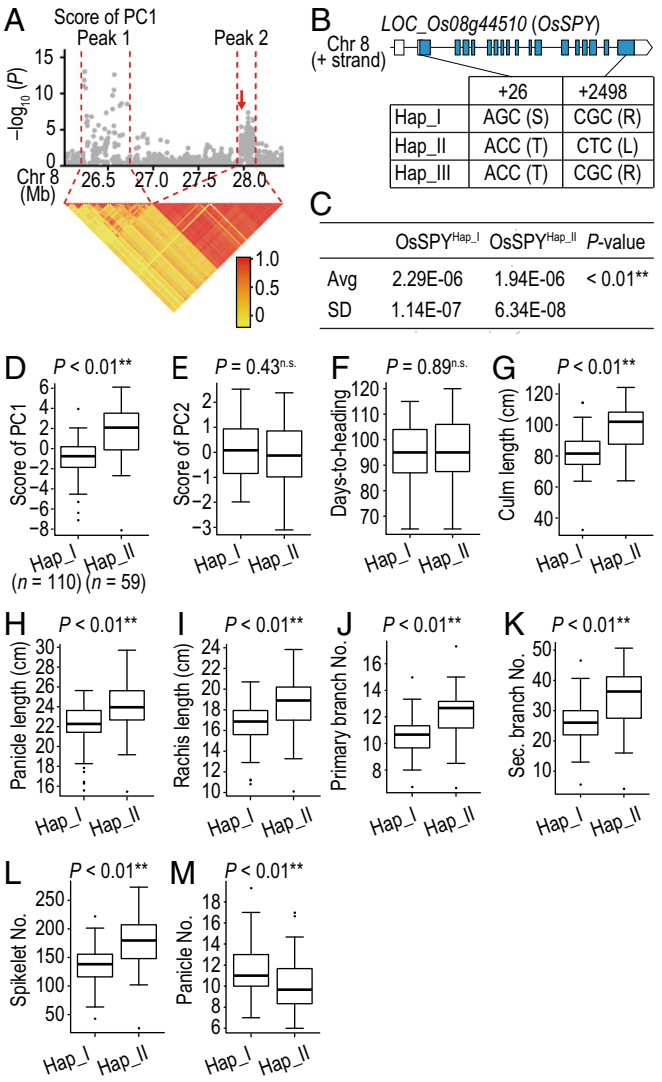

Fig. 2. Isolation of the causal gene of the highest peak on Chr 8 using the data obtained in 2015. (*A*) Local Manhattan plot (*Top*) and LD heat map (*Bottom*) for peaks on Chr 8. The red arrow (*Top*) indicates the position of *OsSPY*. The yellow to red gradient (*Bottom*) indicates the range of r^2 values. (*B*) Exon-intron structure and DNA polymorphism of *OsSPY*. (*C*) *O*-fucosyltransferase activity of OsSPY using SLR1 as a substrate. The activity of OsSPY produced by Hap_I or Hap_II was measured ($n = 3$). (*D*–*M*) Box plots for PC scores and the following 8 traits: PC1 (*D*); PC2 (*E*); days-to-heading (*F*); culm length (cm) (*G*); panicle length (cm) (*H*); rachis length (cm) (*I*); primary branch number per panicle (*J*); secondary branch number per panicle (*K*); spikelet number per panicle (*L*); panicle number per plant (*M*). Box edges represent the 0.25 and 0.75 quantiles, with the median values shown by bold lines. Whiskers extend to the most extreme point, which is no more than 1.5 times the interquartile range. Differences between haplotypes were statistically analyzed using Welch's *t* test (n.s., not significant; $^{**}P < 0.01$). Numbers of plants carrying Hap_I and Hap_II are shown in parentheses in *D*.

vegetative stage, but displayed a spindly phenotype with development of gravely aberrant panicles at the heading stage (Fig. S13). These studies confirmed that a decrease in *OsSPY* function results in GA overdose phenotypes, which has been previously reported (13).

We introduced the entire genomic region of Hap_I or Hap_II into NP and Omachi (OM), which carry Hap_I and Hap_II of OsSPY, respectively, and produced 4 different combinations of transformants. The NP plants transformed with Hap_I (NP_$OsSPY^{Hap_I}$) showed semidwarf or dwarf phenotypes; one plant showed severe dwarfing at the young seedling stage (*SI Appendix*, Fig. S14). At the heading stage, transgenic NP and OM plants showed essentially the same phenotypes (Fig. 3 and *SI Appendix*, Fig. S15). The culm length of plants transformed with Hap_II ($OsSPY^{Hap\_II}$) was not significantly different from that of control plants. However, $OsSPY^{Hap\_I}$ plants exhibited a shorter phenotype (Fig. 3 *A* and *B* and *SI Appendix*, Fig. S15 *A* and *B*), which was consistent with the observation that varieties carrying Hap_I exhibited shorter culm lengths than Hap_II plants (Fig. 2*G*). In general, $OsSPY^{Hap\_II}$ showed no significant differences from controls for panicle structural traits, but OM plants transformed with Hap_II displayed decreased secondary branch and spikelet numbers relative to controls (Fig. 3 *C*–*I* and *SI Appendix*, Fig. S15 *C*–*I*). In contrast, all $OsSPY^{Hap\_I}$ plants showed a significant decrease in panicle and rachis length, lower numbers of primary and secondary branches and spikelets, and an increase in panicle number compared with controls or $OsSPY^{Hap\_II}$ plants. These findings suggested that higher OsSPY activity suppressed GA signaling, resulting in lower culm length and smaller panicle size but larger panicle numbers.

GA Signaling Regulates Rice Plant Architecture. OsSPY functions as a negative regulator by enhancing the activity of SLR1, a DELLA protein that is degraded due to the interaction between GA and its receptor, GID1, resulting in GA-mediated responses

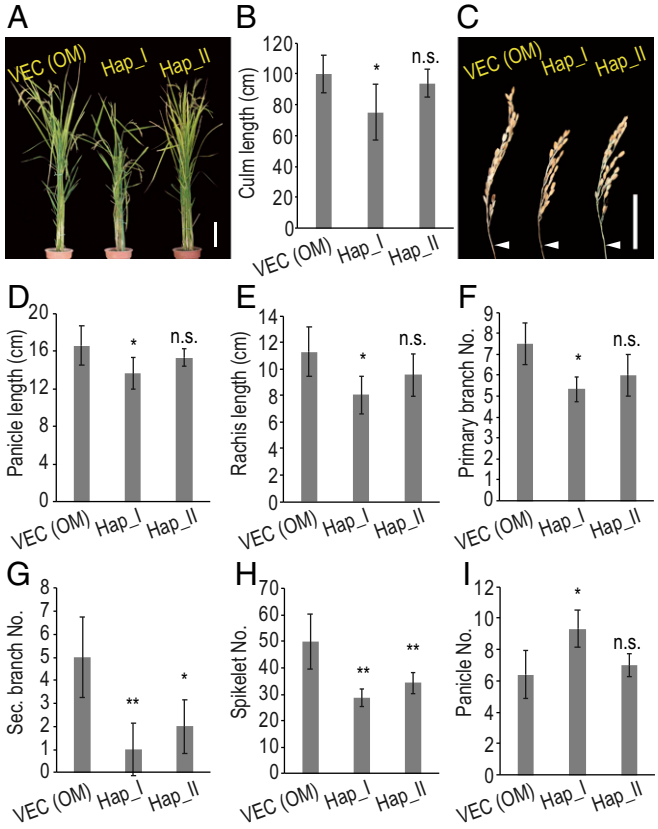

Fig. 3. Phenotypic analysis of plants transformed with Hap_I or Hap_II of *OsSPY*. OM carrying Hap_II was transformed with an empty vector (VEC), Hap_I, or Hap_II. (*A*) Gross morphology at the heading stage. (Scale bar, 15 cm.) (*B*) Culm length (cm). (*C*) Panicle morphology. (Scale bar, 5 cm.) (*D*) Panicle length (cm). (*E*) Rachis length (cm). (*F*) Primary branch number per panicle. (*G*) Secondary branch number per panicle. (*H*) Spikelet number per panicle. (*I*) Panicle number per plant. (Error bars, SD; $n = 10$). Asterisks indicate significant differences (Welch's *t* test) when compared to VEC plants (OM transformed with empty vector). $^{*}P < 0.05$; $^{**}P < 0.01$ (n.s., not significant).

(Fig. 4*A*). Thus, the above findings indicate that GA signaling is a dominant regulator of rice architecture. To confirm this hypothesis, we produced 9 types of isogenic plant by stacking 2 mutant loci, *gid1-8* and *SLR1*^{A288V}, which induced different levels of responsiveness to GA signaling in the T65 genetic background (Fig. 4*B*). The *gid1-8* locus was isolated as a weak mutant of *GID1* (16); in contrast, *SLR1*^{A288V} was isolated as a revertant of *gid1-8*, although it could not completely reverse the dwarfism or GA insensitivity caused by *gid1-8* (*SI Appendix*, Fig. S16). Further analyses revealed that the revertant had a single nucleotide polymorphism (SNP) designated C863T in *SLR1*, where A288 was substituted with V in the LHR1 domain (*SI Appendix*, Fig. S17). This polymorphism suppressed the function of intact SLR1 in a semidominant manner, resulting in enhanced GA responsiveness and growth rate (*SI Appendix*, Fig. S18). The 8 traits related to plant architecture, as mentioned previously, were measured for the 9 isogenic plant types (4 plants per type) (*SI Appendix*, Fig. S19), and PCA was performed (Fig. 4 *D* and *E*). In general, the topology of the loading plot for the isogenic plants was very similar to that for the 169 varieties (Figs. 1*B* and 4*E*), confirming that GA signaling is a major regulator of rice architecture. The loading vector of days-to-heading showed different directions in the 2 loading plots: in PCA for the isogenic plants, it showed high negative loading (−0.81) on PC1 (Fig. 4*D*), whereas for the GWAS panel, it showed high positive loading (0.83) on PC2 (Fig. 1*A*). This can be explained by the positive regulation of flowering by GA signaling (17–19) in isogenic plants with different levels of GA responsiveness, whereas the GWAS panel contained other allelic differences in genes associated with days-to-heading that are independent of GA signaling.

Transition of *OsSPY* Haplotypes during Domestication and Breeding. The transitional processes of *OsSPY* haplotypes in domestication and modern breeding were studied by dividing varieties in the GWAS panel into the following 3 groups: 1) landrace and modern varieties developed 2) before and 3) after 1960 (Fig. 5*A*). The frequency of Hap_I increased with time, suggesting that *OsSPY*$^{Hap\_I}$ has been selected in modern breeding programs in Japan. We calculated the genome-wide Nei's genetic distance (20) between landraces and modern varieties and detected significant peaks (Fig. 5*B*), including the LD of *OsSPY* (*SI Appendix*, Fig. S8*E*). Furthermore, the haplotype frequency of *OsSPY* was studied for Chinese landraces and modern temperate *japonica* varieties in the 3,010 accessions (21). A new haplotype, Hap_III, was found in Chinese varieties which encodes an active SPY with R833 but contains a substitution of S9 with T (Fig. 2*B*). Similar to the trend in the Japanese breeding process, the frequency of Hap_I was increased in modern Chinese varieties (Fig. 5*C*).

We further compared *OsSPY* haplotype frequency among rice ecotypes using 2 public databases. With the 1,529 accessions by Huang et al. (22), we first performed a haplotype network analysis and predicted that Hap_I and Hap_II would be derived from Hap_III (Fig. 5*D*). In this process, the original amino acids, T9 and R833, in Hap_III could be replaced with S in Hap_I and L in Hap_II, respectively. This prediction was also supported by the comparison of amino acid sequences of OsSPY in the genus *Oryza* (23) (*SI Appendix*, Fig. S20). *Indica*, *aus*, and subgroups of *O. rufipogon* (except Or-IIIa) primarily contained Hap_III, whereas *japonica* and *O. rufipogon* subgroup Or-IIIa contained Hap_II (Fig. 5*E*). This analysis was also conducted using the 3,010 accessions (21), and the same results were obtained (*SI Appendix*, Fig. S21). We calculated an extended haplotype homozygosity (EHH) in temperate *japonica* around *OsSPY* and found that the EHH decay rate was 575.0 kb for Hap_I and 69.5 and 31.2 kb for Hap_II and Hap_III, respectively (Fig. 5*F*). To test the EHH results, we also measured the integrated haplotype score (iHS); the iHS of Hap_I (derived) and Hap_III (ancestral) was included in the top 1% of the empirical distribution on Chr 8 (*SI Appendix*, Fig. S22). These results confirmed that Hap_I had recently been subjected to positive selection in temperate *japonica* through plant breeding.

Discussion

PCA is an effective means of collecting information from complex, multiple traits that are highly correlated; furthermore, it is valuable for extracting underlying factors for traits by dimension reduction. A GWAS using PC scores as dependent variables is proposed as a strategy for performing efficient GWAS. First, this strategy can decrease the likelihood of a type I error rate by avoiding multiple testing (9, 24). Second, PC scores produced by PCA can transform the skewed original variables into approximate normal distribution, which results in robust, reliable GWAS results (25, 26). Third, GWAS using PC scores may detect genomic regions that could be overlooked by using individual traits, since PC scores represent integrated variables.

In this study, PCA on 8 architectural traits revealed that PC1 captured 62% of variations for most traits, whereas PC2 captured 16% of variations that primarily impacted days-to-heading (Fig. 1 *A* and *B*); thus, PC1 is a good indicator for plant architecture. Using the PC scores for GWAS, we identified significant peaks associated with PCs; these included genes previously reported for regulating plant architecture along with other peaks that were considered novel. The peak with the strongest effect on PC1, the most important index for plant height and panicle structure (Fig. 1*C*), was further investigated. Genetic studies confirmed that *OsSPY* is a causal gene for this peak and responsible for plant architecture. *OsSPY* functions as a negative regulator in GA signaling by enhancing the suppressive function

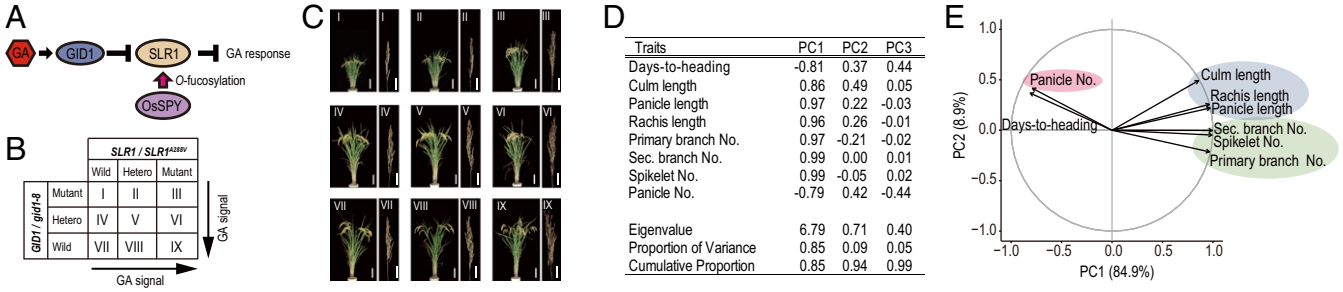

Fig. 4. GA signaling levels impact rice morphology. (*A*) Model of the GA signal transduction pathway in rice (see text for details). (*B*) Combinations of 2 mutations, *gid1-8* and *SLR1*^{A288V}, in 9 isogenic plants. Arrows indicate the strength of GA signaling. (*C*) Plant and panicle morphology of 9 isogenic plants. Type I and type IX plants show the lowest and highest GA signaling, respectively. (Scale bar, 15 and 5 cm in plant and panicle images, respectively.) (*D* and *E*) PCA of 9 isogenic plants with different GA signaling levels. Results are presented as described in Fig. 1 *A* and *B*.

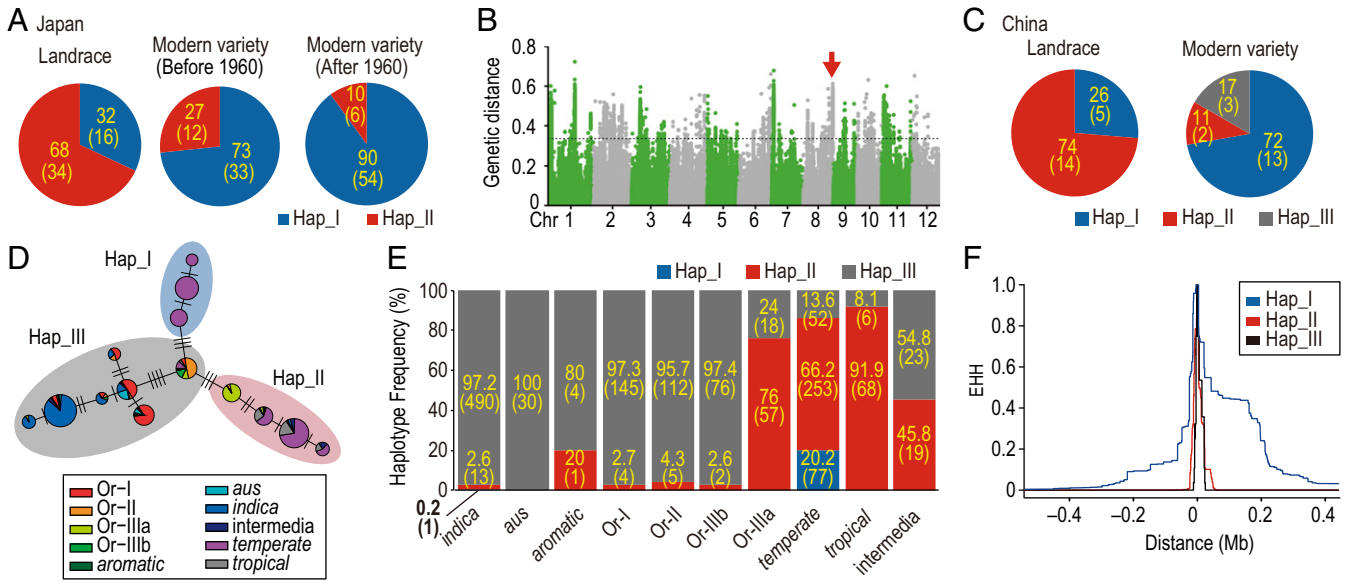

Fig. 5. Transition and selection of *OsSPY* haplotypes during rice domestication and breeding. (*A*) Haplotype frequency in Japanese landraces and modern varieties before and after 1960. Percentages and numbers in parentheses represent percentages of haplotypes and number of varieties, respectively. (*B*) Genome-wide Nei's genetic distance. The red arrow indicates the position of *OsSPY*. (*C*) Haplotype frequency in Chinese landraces and modern varieties. (*D*) Haplotype network of *OsSPY* using genotype data (1,529 rice accessions of *O. rufipogon* and *O. sativa*) (22). Blue, red, and gray ellipses indicate Hap_I, Hap_II, and Hap_III, respectively. (*E*) Haplotype frequency in different ecotypes of the 1,529-rice panel. Numbers in bars represent percentage of haplotypes and number of accessions (parentheses). (*F*) EHH decay of *OsSPY* in temperate *japonica* of the 3,010-accession panel (21). Zero indicates the position of *OsSPY*. Blue, red, and gray lines represent EHH decay of Hap_I, Hap_II, and Hap_III, respectively. The right side of the *x*-axis corresponds to the terminal end of Chr 8.

of DELLA proteins (13, 15). Thus, we considered GA signaling to be a major mechanism regulating plant architecture. This was confirmed by studies using 9 isogenic plant types showing different levels of GA signaling. In general, PCA for these isogenic plants was very similar to the GWAS panel except for the days-to-heading trait; this might have been caused by allelic variation in heading genes that are only present in the GWAS panel.

It is well-known that GA is an important regulator for plant height. To enhance lodging resistance, breeders have developed semidwarf varieties with lower levels of GA accumulation or signaling. However, reductions in plant height may have negative effects on panicle size and crop productivity; furthermore, our results show that GA signaling also regulates plant architecture, especially panicle structure. Recently, Wu et al. (27) reported that high levels of GA accumulation by GA20ox1 are desirable for increasing crop yield, although the effects of this gene on plant architecture were not discussed.

Due to trade-off effects, enhancing GA effectiveness would increase lodging risks and decrease panicle numbers per plant. This trade-off effect is not limited to the GA response mechanism. In fact, several genes that induce large panicle size, such as *SCM3* (28), *NAL1* (29, 30), *IPA1* (31–33), and *OsOTUB1* (34), simultaneously decrease panicle numbers per plant. Therefore, to breed rice varieties with large panicle size and number, it is essential to identify novel genetic mechanisms that can disrupt or attenuate trade-off relationships (5). The present study showed that PCA provides useful information on potential mechanisms for breaking trade-off relationships. Our PCA results revealed that variations in panicle number were equally divided into PC1 and PC2 (Fig. 1*B*), indicating the presence of unknown factors regulating panicle number without affecting PC1 score. Peak regions for PC2 could contain genetic factors that regulate panicle number independent of GA signaling.

In order to study the transmission of *OsSPY* alleles, haplotype frequency was calculated for various rice ecotypes and 3 haplotypes were identified for *OsSPY*. In wild rice (*O. rufipogon*), ~75% of Or-IIIa subtypes were Hap_II, while other subtypes (Or-I, Or-II, Or-IIIb) were Hap_III (Fig. 5*E*). In contrast to *indica*, *aus*, and *aromatic*, Hap_II was predominant in *japonica* rice (Fig. 5*E* and *SI Appendix*, Fig. S21*B*). These observations agreed with the model of rice domestication proposed by Huang et al. (22). According to their model, *japonica* was domesticated from Or-IIIa in southern China and was subsequently crossed to Or-I in southeast and southern Asia, resulting in the generation of *indica*. On the basis of this model, we can discuss the transmission of the *OsSPY* haplotype as follows. Hap_II was generally a minor haplotype in *O. rufipogon*, but was dominant in the subtype Or-IIIa. Hap_II was transmitted into ancient *japonica* from Or-IIIa, the ancestor of domesticated rice, in the process of *japonica* domestication, and allele frequency was maintained in *japonica*. During the domestication of *indica*, the haplotype frequency of Hap_III in Or-I was also maintained in *indica*. Thus, *OsSPY* was not targeted by artificial selection in the process of *japonica* or *indica* domestication. In contrast, Hap_I, a haplotype present only in temperate *japonica*, has been selected in the modern breeding process. Hap_I frequency in modern temperate *japonica* varieties was much higher than Japanese and Chinese landraces (Fig. 5 *A* and *C*). Furthermore, EHH statistics showed that the EHH decay rate was lower in Hap_I than in Hap_II or Hap_III (Fig. 5*F* and *SI Appendix*, Fig. S22).

Interestingly, the *OsSPY* transition from Hap_II to Hap_I in Japanese cultivated rice occurred early in the 20th century (Fig. 5*A*), which corresponds to a rapid increase in nitrogen input during agricultural production (35). It is well known that *sd-1* (36) was the causal mutation for the rice "Green Revolution" during the 1950s and 1960s. Although the selection of Hap_I of *OsSPY* occurred earlier than Green Revolution, both events were essentially identical in terms of the underlying driving forces and mechanisms. That is, both depended on the introduction of semidwarf varieties to enhance lodging resistance during increased fertilization, and the causal genes were involved in GA synthesis or signaling (37). It is very likely that breeders

independently used GA-related genes, *sd-1* and *OsSPY*, to develop new varieties adapted to high nutrient conditions. From the current viewpoint of sustainable and environmentally friendly agriculture, new rice varieties with moderate crop yield and limited nutrient input are more desirable (38). In this context, the replacement of Hap_I with Hap_II in *OsSPY* could be an approach to maintaining crop yield in low nutrient conditions, along with the utilization of genes controlling panicle number and nutrient use efficiency.

Materials and Methods

Detailed descriptions of plant materials, population genetic analyses, and molecular methods can be found in the *SI Appendix*.

ACKNOWLEDGMENTS. We thank Dr. T. Akagi (Okayama University) for suggestions on population genetics, and Y. Hattori (Nagoya University) for technical assistance. This work was supported by the Grant-in-Aid for Advanced Integrated Intelligence Platform Project, JSPS Fellows (Grant 16J08722), Young Scientists (B) (Grant 17K15209), Scientific Research (A) (Grant 17H01458), Postdoctoral Fellowships (Grant 19F19103), and Scientific Research on Innovative Areas (Grants 16H06464 and 16H06468).

1. Y. Wang, J. Li, The plant architecture of rice (Oryza sativa). *Plant Mol. Biol.* **59**, 75–84 (2005).
2. Y. Wang, J. Li, Molecular basis of plant architecture. *Annu. Rev. Plant Biol.* **59**, 253–279 (2008).
3. S. Bai, S. M. Smith, L. Jiayang, "Rice plant architecture: Molecular basis and application in breeding" in *Rice Genomics, Genetics and Breeding*, T. Sasaki, M. Ashikari, Eds. (Springer, Singapore, 2018), pp. 129–154.
4. H. Morishima, H. Oka, T. T. Chang, Analysis of genetic variations in plant type of rice. I. Estimation of indices showing genetic plant types and their correlations with yielding capacity in a segregating population. *Jpn. J. Breed.* **17**, 73–84 (1967).
5. R. L. Ordonio, M. Matsuoka, New path towards a better rice architecture. *Cell Res.* **27**, 1189–1190 (2017).
6. S. Takeda, M. Matsuoka, Genetic approaches to crop improvement: Responding to environmental and population changes. *Nat. Rev. Genet.* **9**, 444–457 (2008).
7. A. Ishikawa, T. Namikawa, Mapping major quantitative trait loci for postnatal growth in an intersubspecific backcross between C57BL/6J and Philippine wild mice by using principal component analysis. *Genes Genet. Syst.* **79**, 27–39 (2004).
8. M. Ringnér, What is principal component analysis? *Nat. Biotechnol.* **26**, 303–304 (2008).
9. L. N. He *et al*., Genomewide linkage scan for combined obesity phenotypes using principal component analysis. *Ann. Hum. Genet.* **72**, 319–326 (2008).
10. K. Yano *et al*., Genome-wide association study using whole-genome sequencing rapidly identifies new genes influencing agronomic traits in rice. *Nat. Genet.* **48**, 927–934 (2016).
11. S. Atwell *et al*., Genome-wide association study of 107 phenotypes in Arabidopsis thaliana inbred lines. *Nature* **465**, 627–631 (2010).
12. E. Kerdaffrec *et al*., Multiple alleles at a single locus control seed dormancy in Swedish Arabidopsis. *eLife* **5**, e22502 (2016).
13. A. Shimada *et al*., The rice SPINDLY gene functions as a negative regulator of gibberellin signaling by controlling the suppressive function of the DELLA protein, SLR1, and modulating brassinosteroid synthesis. *Plant J.* **48**, 390–402 (2006).
14. N. E. Olszewski, C. M. West, S. O. Sassi, L. M. Hartweck, O-GlcNAc protein modification in plants: Evolution and function. *Biochim. Biophys. Acta* **1800**, 49–56 (2010).
15. R. Zentella *et al*., The Arabidopsis O-fucosyltransferase SPINDLY activates nuclear growth repressor DELLA. *Nat. Chem. Biol.* **13**, 479–485 (2017).
16. M. Ueguchi-Tanaka *et al*., Molecular interactions of a soluble gibberellin receptor, GID1, with a rice DELLA protein, SLR1, and gibberellin. *Plant Cell* **19**, 2140–2155 (2007).
17. J. Moon *et al*., The SOC1 MADS-box gene integrates vernalization and gibberellin signals for flowering in Arabidopsis. *Plant J.* **35**, 613–623 (2003).
18. H. Yu *et al*., Floral homeotic genes are targets of gibberellin signaling in flower development. *Proc. Natl. Acad. Sci. U.S.A.* **101**, 7827–7832 (2004).
19. C. Dai, H. W. Xue, Rice early flowering1, a CKI, phosphorylates DELLA protein SLR1 to negatively regulate gibberellin signalling. *EMBO J.* **29**, 1916–1927 (2010).
20. M. Nei, Genetic distance between populations. *Am. Nat.* **106**, 283–292 (1972).
21. W. Wang *et al*., Genomic variation in 3,010 diverse accessions of Asian cultivated rice. *Nature* **557**, 43–49 (2018).
22. X. Huang *et al*., A map of rice genome variation reveals the origin of cultivated rice. *Nature* **490**, 497–501 (2012).
23. S. Ge, T. Sang, B. R. Lu, D. Y. Hong, Phylogeny of rice genomes with emphasis on origins of allotetraploid species. *Proc. Natl. Acad. Sci. U.S.A.* **96**, 14400–14405 (1999).
24. C. J. Holberg *et al*., Factor analysis of asthma and atopy traits shows 2 major components, one of which is linked to markers on chromosome 5q. *J. Allergy Clin. Immunol.* **108**, 772–780 (2001).
25. D. I. Boomsma, C. V. Dolan, A comparison of power to detect a QTL in sib-pair data using multivariate phenotypes, mean phenotypes, and factor scores. *Behav. Genet.* **28**, 329–340 (1998).
26. L. Goh, V. B. Yap, Effects of normalization on quantitative traits in association test. *BMC Bioinformatics* **10**, 415 (2009).
27. Y. Wu *et al*., The QTL GNP1 encodes GA20ox1, which increases grain number and yield by increasing cytokinin activity in rice panicle meristems. *PLoS Genet.* **12**, e1006386 (2016).
28. K. Yano *et al*., Isolation of a novel lodging resistance QTL gene involved in strigolactone signaling and its pyramiding with a QTL gene involved in another mechanism. *Mol. Plant* **8**, 303–314 (2015).
29. D. Fujita *et al*., NAL1 allele from a rice landrace greatly increases yield in modern indica cultivars. *Proc. Natl. Acad. Sci. U.S.A.* **110**, 20431–20436 (2013).
30. G. H. Zhang *et al*., LSCHL4 from Japonica Cultivar, which is allelic to NAL1, increases yield of indica super rice 93-11. *Mol. Plant* **7**, 1350–1364 (2014).
31. K. Miura *et al*., OsSPL14 promotes panicle branching and higher grain productivity in rice. *Nat. Genet.* **42**, 545–549 (2010).
32. Y. Jiao *et al*., Regulation of OsSPL14 by OsmiR156 defines ideal plant architecture in rice. *Nat. Genet.* **42**, 541–544 (2010).
33. J. Wang *et al*., Tissue-specific ubiquitination by IPA1 INTERACTING PROTEIN1 modulates IPA1 protein levels to regulate plant architecture in rice. *Plant Cell* **29**, 697–707 (2017).
34. S. Wang *et al*., Non-canonical regulation of SPL transcription factors by a human OTUB1-like deubiquitinase defines a new plant type rice associated with higher grain yield. *Cell Res.* **27**, 1142–1156 (2017).
35. Y. Hayami, Demand for fertilizer in the course of Japanese agricultural development. *Am. J. Agric. Econ.* **46**, 766–779 (1964).
36. A. Sasaki *et al*., Green revolution: A mutant gibberellin-synthesis gene in rice. *Nature* **416**, 701–702 (2002).
37. M. Ashikari *et al*., Loss-of-function of a rice gibberellin biosynthetic gene, *GA20 oxidase* (*GA20ox-2*), led to the rice 'green revolution'. *Breed. Sci.* **52**, 143–150 (2002).
38. S. Li *et al*., Modulating plant growth-metabolism coordination for sustainable agriculture. *Nature* **560**, 595–600 (2018).