A natural tandem array alleviates epigenetic repression of *IPA1* and leads to superior yielding rice

Lin Zhang¹,²,³,⁎, Hong Yu⁴,⁎, Bin Ma⁵, Guifu Liu⁴, Jianjun Wang⁶, Junmin Wang⁶, Rongcun Gao⁷, Jijun Li⁷, Jiyun Liu¹, Jing Xu¹, Yingying Zhang¹, Qun Li¹, Xuehui Huang⁸, Jianlong Xu⁹, Jianming Li², Qian Qian⁹, Bin Han⁸, Zuhua He¹,³ & Jiayang Li³,⁴

Super hybrid rice varieties with ideal plant architecture (IPA) have been critical in enhancing food security worldwide. However, the molecular mechanisms underlying their improved yield remain unclear. Here, we report the identification of a QTL, qWS8/ipa1-2D, in the super rice Yongyou12 (YY12) and related varieties. In-depth genetic molecular characterization of qWS8/ipa1-2D reveals that this newly identified QTL results from three distal naturally occurring tandem repeats upstream of IPA1, a key gene/locus previously shown to shape rice ideal plant architecture and greatly enhance grain yield. The qWS8/ipa1-2D locus is associated with reduced DNA methylation and a more open chromatin state at the IPA1 promoter, thus alleviating the epigenetic repression of IPA1 mediated by nearby heterochromatin. Our findings reveal that IPA traits can be fine-tuned by manipulating IPA1 expression and that an optimal IPA1 expression/dose may lead to an ideal yield, demonstrating a practical approach to efficiently design elite super rice varieties.
Global food demand is expected to double by the year 2050 (ref. 1). Rice is a major cereal crop that feeds more than half of the world’s population. In the past 50 years, rice yield has increased massively due to the identification of semi-dwarf varieties and the development of hybrid rice2-3. However, rice production is stagnating and even collapsing in some regions, including China, threatened by the loss of arable farmland and growing population4-5. Therefore, it is urgent to develop new genetic resources and strategies to break the bottleneck and meet the increasing food demand.

The advances in high-quality rice genetics, genomics and molecular markers have accelerated the dissection of key genes for yield improvement. The first leap of rice productivity was attributed to the finding and application of the semi-dwarf locus sd1, which led to the Green revolution6, and sd1 was identified as a mutation in GA20ox-2 that results in reduced biosynthesis of the hormone giberrellin7. To further improve yield potential, the International Rice Research Institute raised the concept of new plant type or ideal plant architecture (IPA) for variety development, characterized by few unproductive tillers, more grains per panicle, and thick and sturdy spikes8. These traits show continuous phenotypic variation and are controlled by multiple quantitative trait loci (QTLs). Several QTLs related to the IPA definition have been cloned, including Gn1a, Ghd7, ipa1/WFP, SCM2, dep1 and SPIKE, all of which show great potential in improving rice yield9-15. In particular, IPA1 encodes the SBP-domain transcription factor OsSPL14. A point mutation in OsSPL14 relieves its repression by the miRNA OsmiR156 and affects all three IPA characteristics simultaneously9.

The utilization of hybrid vigour (heterosis) has greatly enhanced rice productivity, however, the underlying mechanisms have not been well-characterized. A global view of heterosis has been proposed by re-sequencing more than 1,000 hybrid varieties, which found that a number of superior alleles underlie the high yield potential of hybrids16. This finding suggests that heterosis can be achieved by combining key QTLs through an approach known as ‘gene pyramiding’. Indeed, this approach is used for breeding many super rice varieties with high productivity2. In particular, the widely commercialized super hybrid Yongyou12 (YY12) achieved a productivity of 14.5 t grains per hectare on average, and set a national record. YY12 and related varieties are intersubspecific hybrids (indica × japonica) exhibiting the IPA phenotypes, but we did not find the ipa1 allele in these varieties17, suggesting a novel mechanism contributing to the increased yield.

Here, we report the identification and characterization of a major QTL, qWS8/ipa1-2D, from the original breeding stock of YY12 which explains the major IPA phenotypes. YY12 and related varieties are homozygous for the wild-type allele ipa1, whereas Nipponbare (NIP) is homozygous for the null ipa1-2D allele. YYP1 bears extremely strong phenotypes, but we did not find the ipa1 allele in these varieties17, suggesting a novel mechanism contributing to the increased yield.

Results

Map-based cloning of the qWS8 locus. To uncover the new factors for the IPA traits in YY12 and related hybrid varieties that are widely grown in China, we generated an F2 mapping population by crossing the IPA breeding stock YYP1 with the japonica variety Nipponbare (NIP). YYP1 bears extremely strong culms, large panicles and modest tiller numbers compared to NIP (Fig. 1a-c and Supplementary Fig. 1). We identified several QTLs by linking genome-wide molecular markers with the stem traits (Supplementary Table 1). We focused on a major QTL on the long arm of chromosome 8 that showed the highest logarithm of the odd score and phenotypic contribution, which we named qWS8 (QTL of wide stem on chromosome 8, also named ipa1-2D, see below). To identify the gene underlying qWS8, we carried out fine mapping using backcross populations (BC2F2), with little interference of other loci (Supplementary Fig. 2). Linkage analysis confirmed the phenotypic effect of qWS8 (18.34 cM) and delineated it to a 15-kb region (Fig. 1d and Supplementary Fig. 3). A large BC2F2 population with 3,404 plants was then screened, and qWS8 was finally localized between markers SNP4 and SNP7 (Fig. 1e and Supplementary Fig. 3). This region covers a 3-kb interval of the NIP genome with no annotated genes, about 4 kb upstream of IPA1 (Fig. 1e). Phenotyping of key recombinants indicated that qWS8 has pleiotropic effects in promoting stem diameter and panicle number.
primary branch number but reducing tiller number (Fig. 1e–h), similar to \textit{ipa1} (ref. 9). However, more than 10 tillers were generated in homozygous \textit{qWS8} plants, far more than that of \textit{ipa1} and fulfilling the recent demands for IPA rice breeding\cite{18}, indicating that \textit{qWS8} only modestly reduces tillering capacity.

We further evaluated the contribution of \textit{qWS8} using an F\textsubscript{2} population from YY12 (F\textsubscript{1}) bearing heterozygous \textit{qWS8} alleles. A significant difference was identified among segregating genotypes and the locus explains as much as 50\% of the phenotypic variance for panicle branch and 44\% for stem diameter (Supplementary Fig. 4), suggesting that \textit{qWS8} contributes substantially to the IPA traits of YY12. Therefore, cloning of \textit{qWS8} will have a large impact on future super rice breeding.

Next, we investigated the functional sequence variation at this locus. Besides the flanking SNPs (SNPM6 and SNPM7), we did not uncover sequence polymorphisms between YY12 and NIP by PCR-based sequencing of the mapping region. However, Southern blotting revealed that YY12 contains a large sequence insertion at the mapping region (Supplementary Fig. 5). Interestingly, digestion with XbaI showed that YY12 harbours a common band with NIP, as well as a specific ~3-kb band (Supplementary Fig. 5), suggesting a repetitive structure. We then sequenced a BAC containing the \textit{qWS8} locus and discovered that the \textit{qWS8} sequence in YY12 contains three tandem repeats of the 3,137-bp sequence matching the NIP region, making detection by simple PCR ineffective (Fig. 2a). We therefore postulated that the large tandem repeat region represents the functional polymorphism for the IPA trait determination in YY12.

\textit{qWS8} origin and allele frequency in natural populations. To dissect the origin of the tandem repeats in YY12, we sequenced the whole region covering \textit{IPA1} and \textit{qWS8} from both parents, and found only seven SNPs in addition to the repeats (Fig. 2a). Three SNPs are duplicated within the repeats, suggesting that they appeared before formation of the repeats. Using the seven SNPs as query, we extracted the SNP information from the 3,000 rice-SNP database\cite{19} to find varieties with the same SNPs. We manually excluded varieties with ambiguous and missing information, and obtained SNP types from a total of 2,464 accessions, which form 10 haplotypes (Supplementary Fig. 6). Interestingly, only five indic\textit{a} varieties from China share the SNPs identical to YY12, suggesting that \textit{qWS8} is a rare allele that recently arose in China. Three varieties, GENG77-4, Xiangai B and Jinxibai, were selected for further genotyping by Southern blotting (Fig. 2b). The tandem repeats can only be detected in GENG77-4 and Xiangai B but not in Jinxibai, though they share the same SNPs as YY12 (Fig. 2b and Supplementary Fig. 6), indicating that the repeats evolved recently from varieties related to Jinxibai during modern breeding. Moreover, Jinxibai exhibits a plant architecture similar to NIP, whereas GENG77-4 and Xiangai B exhibit the IPA phenotype (Fig. 2c–e). This result further supports the notion that the tandem repeats underlie the variation for \textit{qWS8}-mediated IPA traits.

To facilitate genotyping and breeding application, we designed PCR primers spanning the repeat boundary which amplify a 432-bp fragment only from YY12 but not NIP (Fig. 2a,b). We then genotyped a set of accessions showing IPA traits collected by breeders using a PCR-based approach. Interestingly, we found that more than half of the collections contain the \textit{qWS8} tandem repeats, and share the same SNPs as YY12 (Supplementary Table 2), suggesting a single origin of this allele. The IPA traits were evaluated and all the varieties with the \textit{qWS8} repeats showed wide stem and high panicle primary branch number (Supplementary Fig. 7). Therefore, the \textit{qWS8} allele has been enriched and adopted widely in recent rice breeding programs.

\textit{qWS8}/\textit{ipa1-2D} controls IM size by upregulating \textit{IPA1}. Given the proximity of the \textit{qWS8} non-coding region to \textit{IPA1}, we examined \textit{IPA1} expression in the progeny of heterozygous recombinants. We detected higher expression of \textit{IPA1} in plants with the YY12 allele of \textit{qWS8} compared to plants with the NIP allele (Supplementary Fig. 8). Therefore, \textit{qWS8} is a novel regulatory locus of \textit{IPA1}, which we then renamed \textit{ipa1-2D}. We also renamed the previous \textit{ipa1} allele\cite{20} that abrogates its miRNA-mediated repression as \textit{ipa1-1D} hereafter to uniform the nomenclature. We then developed high-quality near-isogenic lines (NILs) carrying an \textasciitilde{}80-kb segment with \textit{ipa1-2D} and \textit{IPA1} (NIP allele), and named them \textit{NIL\textsubscript{ipa1-2D}} and \textit{NIL\textsubscript{IPA1}} respectively.

We performed qRT-PCR to clarify the tissue-specific expression pattern of \textit{IPA1} and of two miRNAs thought to target \textit{IPA1} at different stages, miR156 and miR529 (ref. 20). \textit{IPA1} was expressed mainly in the inflorescence meristem (IM), and was hardly detected in the seedling, opposite to the pattern of the
two miRNAs (Fig. 3a–c) and consistent with the miRNAs determining the tissue-specific expression of IPA1. Further, we found that IPA1 expression in the IM was upregulated ~3–4 fold in NILipapa1-2D compared to NILIPA1 (Fig. 3a–c), and obtained similar results by RNA-seq and Northern blotting (Supplementary Fig. 9). These results indicate that the expression of IPA1 is controlled by both miRNAs and ipa1-2D, and that ipa1-2D employs a different mechanism to elevate the IPA1 expression. As IPA1 is expressed mainly in the IM, we compared its effect on IM development in NILipapa1-2D and NILIPA1. We found that the IM of NILipapa1-2D bears a larger size at early stages, and then formed more IM primary branch primordials than that of NILIPA1 (Fig. 3d–i and Supplementary Fig. 10). Therefore, we concluded that ipa1-2D determines the IPA phenotypes through tissue-specific up-regulation of IPA1. Increased IPA1 levels promote IM development in rice, laying the cellular foundation for strong culms and large panicles at the mature stage. A similar correlation between large IM size and subsequent branch increase with high productivity has been observed in maize and tomato21,22.

ipa1-2D alleviates position-dependent chromatin repression.

To explore the molecular mechanism of ipa1-2D-mediated regulation of IPA1, we first transformed the three tandem repeats into NIL, and analysed the phenotype of T1 generation plants. No obvious morphological difference was found between the transgenic and control plants (Supplementary Fig. 11), suggesting that the regulation of IPA1 by ipa1-2D may rely on the chromosomal position and act in cis. We further generated transgenic plants containing one or three repeats constructs fused to the CaMV35S mini promoter-glucuronidase (GUS) reporter and found that the fusion reporters did not show different GUS expression (Supplementary Fig. 12), suggesting that the tandem repeats do not have enhancer activity.

Genome annotation of NIP revealed that the upstream region of IPA1 contains many repeats and transposon elements (TEs), which are usually silenced by DNA methylation and form heterochromatin. Histone H3K9m2, a modification associated with heterochromatin, was also enriched in this region (Supplementary Fig. 13). It has been suggested that heterochromatin repression can spread to nearby genes23,24, suggesting a possible mechanism underlying ipa1-2D-mediated regulation of IPA1. To test this hypothesis, we first examined the methylation status of the region between the ipa1-2D repeats and the IPA1 coding region in the NILs by methylation sensitive Southern blotting (Fig. 4a,b). We found some sites are highly resistant to the digestion, coinciding with the character of hypermethylation in heterochromatin. Nevertheless, we found the region in NILipapa1-2D was more sensitive to the digestion of different methyl-sensitive restriction enzymes (Fig. 4b and Supplementary Fig. 14). The blotting assays suggest that the reduction in DNA methylation might happen on specific sites upstream of the IPA1 gene, and the hypomethylation might associate IPA1 up-regulation with ipa1-2D function.

To further clarify the relationship between DNA methylation and IPA1 regulation, we then performed bisulfite sequencing of the ~800-bp promoter region upstream of the IPA1 ATG, and found an uneven distribution of three cytosine contexts, with CG and CHG enriched distally, and CHH enriched proximally, which sharply distinguishes two regions of cytosine methylation (Fig. 4c). Importantly, we found a remarkable loss of cytosine methylation in the junction of two regions of NILipapa1-2D compared to NILIPA1. The CHH was particularly hypomethylated in NILipapa1-2D compared to NILIPA1 (Fig. 4d–f), suggesting a role for CHH demethylation in ipa1-2D-mediated up-regulation of IPA1. This is consistent with the recent notion that the CHH islands in the promoter coordinate gene expression near heterochromatin25.

More importantly, we found that this region contains a DNase I hypersensitive (DH) site (Fig. 4c) and several potentially important cis-elements (Supplementary Fig. 15) that could function as binding sites for transcription factors26. To test if the large repeats altered the chromatin organization of this region, we evaluated nucleosome organization by a mononuclease (MNase) digestion-PCR approach (Fig. 4g). We found that the region between the repeats and IPA1 is more sensitive to the MNase digestion, especially for the IPA1 promoter region (Fig. 4h,i), in NILipapa1-2D compared to NILIPA1, indicating that the repeats promote an open chromatin state. This result is consistent with the above notion that IPA1 is repressed by nearby heterochromatin in NILIPA1. To provide further support that the IPA1 repression is position-dependent, we transformed an 8-kb genomic DNA (gDNA) construct of IPA1 with its native promoter and coding region containing an artificial SNP to distinguish it from the native gene (Supplementary Fig. 16). Independent transgenic lines showed a higher transgenic mRNA/gDNA ratio than the endogenous gene, indicating a higher transcriptional activity of the transgene (Supplementary Fig. 16), confirming the position-dependent repression of IPA1. Collectively, we conclude that the repeat structure in NILipapa1-2D elevates IPA1 expression by creating an open chromatin structure which attenuates the epigenetic repression that would otherwise spread from nearby heterochromatin.
Optimum IPA traits through fine-tuning of IPA1 expression.

As the new IPA traits of *ipal-2D* are caused by tissue-specific up-regulation of IPA1, we hypothesized that optimal IPA traits can be achieved by manipulating IPA1 expression dosage. Supporting the hypothesis, varying IPA1 traits were observed for stable lines with different transgenic copy numbers of two different IPA1 gDNA constructs (Fig. 5a,b and Supplementary Fig. 17). As copy number increased, the tiller number decreased (Fig. 5c–e). We analysed the expression of the CaMV35S promoter, tillers (average 3–4) with strong culms could not be further reduced, suggesting a limit of IPA1 dosage effect in tillering control (Supplementary Fig. 18). Interestingly, the panicle branch number was not affected in the ectopic over-expression lines, possibly due to the CaMV35S promoter working inefficiently in the process of panicle branch initiation.

Noting that IPA1 expression has opposite effect on tiller number and panicle branches, two key components controlling yield, we hypothesize that the balance between these traits is a critical factor in shaping the optimal plant architecture for high yield potential. This balance can be obtained by optimizing IPA1 expression, where sub- and supra-optimal IPA1 expression will both lower yield potential. Based on this hypothesis, we compared the effect of the naturally occurring *ipal-2D* allele on yield components under two different growth conditions, Shanghai (eastern China) and Hainan (southern China). Three genotypes with 0, 1 (heterozygous) and 2 (homozygous) copies of the *ipal-2D* allele were analysed from the NIL population. The IPA traits with 0, 1 and 2 copies of the *IPAl* gene. The ratio of the NILipa1-2D allele in the Mn3 region after MNase digestion of nucleus from the NIL population was shown in Fig. 6a,b. Corresponding to higher tiller number were analysed from the NIL population. The IPA traits of NILipa1-2D and NILIPA1. The bands reflecting different methylation pattern are denoted by arrows. Note that the smallest band is a direct reflection of the repeat region respectively. Mn1–Mn4, positions of PCR detection after MNase digestion. (c) Methylation levels of CG (d), CHG (e) and CHH (f) in 800 bp-pair of the promoter region in ~800-bp promoter region of IPA1. Filled circles, methylated cytosine; empty circles, unmethylated cytosine. Two distinct regions are labelled by blue lines and the junction region is labelled by black frame. Note that the junction region presents obvious methylation difference especially for CHH and overlaps with the DH site labelled by black double arrows. (d–f) Methylation levels of CG (d), CHG (e) and CHH (f) in 800 bp-windows of the promoter region shown in c. Blue, NILipa1-2D; Red, NILIPA1. Dark yellow shades highlight the junction region. (g) Schematic map of the approach to detect open chromatin. Chromatin with loose nucleosome occupation is more susceptible to mononuclease (MNase) digestion and inefficiently amplified by PCR. (h) Sensitivity of four regions (Mn1–Mn4) to increasing dosage of MNase digestion between NILipa1-2D and NILIPA1. The position of each region is labelled in a. Note that the Mn3 region was more sensitive to the digestion in NILipa1-2D. (i) Ratio of the NILipa1-2D allele in the Mn3 region after MNase digestion of nucleus from heterozygous NIL. Two alleles are distinguished by the SNP at position –419 shown in a.

**Figure 4 | Alleviation of epigenetic repression at the IPA1 promoter by qWS8/ipal-2D.** (a) Schematic map showing the sites of two methylation sensitive enzymes (HpaII and MspI) and probe for methylation detection by Southern blotting. The blue and orange double arrows indicate the mapping region and the single repeat region respectively. Mn1–Mn4, positions of PCR detection after MNase digestion. (b) Southern blot analysis of DNA methylation differences between NILipa1-2D and NILIPA1. The bands reflecting different methylation pattern are denoted by arrows. Note that the smallest band is a direct reflection of different methylation at sites −2,545 and −2,974. M, DNA markers. (c) Distribution of three cytosine contexts and methylation pattern in ~800-bp promoter region of IPA1. Filled circles, methylated cytosine; empty circles, unmethylated cytosine. Two distinct regions are labelled by blue lines and the junction region is labelled by black frame. Note that the junction region presents obvious methylation difference especially for CHH and overlaps with the DH site labelled by black double arrows. (d–f) Methylation levels of CG (d), CHG (e) and CHH (f) in 800 bp-windows of the promoter region shown in c. Blue, NILipa1-2D; Red, NILIPA1. Dark yellow shades highlight the junction region. (g) Schematic map of the approach to detect open chromatin. Chromatin with loose nucleosome occupation is more susceptible to mononuclease (MNase) digestion and inefficiently amplified by PCR. (h) Sensitivity of four regions (Mn1–Mn4) to increasing dosage of MNase digestion between NILipa1-2D and NILIPA1. The position of each region is labelled in a. Note that the Mn3 region was more sensitive to the digestion in NILipa1-2D. (i) Ratio of the NILipa1-2D allele in the Mn3 region after MNase digestion of nucleus from heterozygous NIL. Two alleles are distinguished by the SNP at position −419 shown in a.
number per panicle, and weak effect in reducing tiller number and grain weight (Fig. 6d–h). Consequently, the yield per plant increased greatly, to as high as 37% in Shanghai and 41% in Hainan (Fig. 6i). The trait variation is consistent at the two stations, indicating the genetic stability of the \( \text{ipa1-2D} \) allele for yield improvement in multiple environments. In addition, NIL\( \text{ipa1-2D} \) showed a wider diameter of all internodes and better resistance against both mechanical bending and breaking force than NIL\( \text{IPA1} \), which even did not lodge from strong typhoons in Shanghai in 2013 (Supplementary Fig. 19).

It is worth noting that the yield of the heterozygous plant (\( \text{ipa1-2D/IPA1} \)) inclines toward the level of homozygous plants (\( \text{ipa1-2D/ipa1-2D} \)) and produces mid-parent heterosis of the hybrid of NILs (Supplementary Table 3), in particular in secondary branch number and spikelet number per panicle, suggesting that large panicle produced by \( \text{ipa1-2D} \) is a critical factor both for high yield and heterosis, supporting the notion that the heterozygous \( \text{IPA1/ipa1} \) allele showed strong heterosis in the \( \text{indica} \times \text{japonica} \) cross\(^{17} \). As the yield improvement smoothened in homozygous \( \text{ipa1-2D} \) plants (Fig. 6i), we inferred that this allele could have achieved close to optimal \( \text{IPA1} \) expression, balancing yield components and achieving the superior yield potential. Considering the fruitful effect of \( \text{ipa1-2D} \), it is rational to expect that the allele will be applied as a primary target for breeding practice.

Application of \( \text{ipa1-1D} \) and \( \text{ipa1-2D} \) alleles in breeding. As \( \text{ipa1-1D} \) and \( \text{ipa1-2D} \) had different effects in shaping the IPA traits, especially for tiller number, we clarified their effects on yield potential in actual breeding. By molecular assisted selection, we \textit{de novo} designed and developed four hybrid varieties. Two of them (JYZK-6 and JYZK-33) bear \( \text{ipa1-1D} \) and another two (JYZK-3 and JYZK-4) bear \( \text{ipa1-2D} \). Field tests showed that the new varieties, either with \( \text{ipa1-1D} \) or \( \text{ipa1-2D} \), greatly improved yield performance in comparison with the control variety at two locations of different seasons (as high as 31% in the Zhejiang station and 24% in the Hainan station; Supplementary Table 4), meeting the standard of Chinese super rice. More importantly, the \( \text{ipa1-2D} \) allele could further improve productivity with more tillers in comparison with the \( \text{ipa1-1D} \) allele (Supplementary Table 4). In particular, JYZK-4 showed better yield performance with more productive tillers than the NIL variety JYZK-6, while both of them bear strong culms and large panicle (Fig. 7a,b). Using the \( F_2 \) population of two hybrids, we accurately determined the contribution of the two alleles in tillering. We found that \( \text{ipa1-2D} \) had a less obvious effect on tillering, whereas \( \text{ipa1-1D} \) caused a significant decrease in tiller number, explaining 34.7% of the total variation (Fig. 7c,d). These findings confirm that \( \text{ipa1-2D} \) confers less of a reduction in tiller number in the varieties.

To compare \( \text{IPA1} \) expression associated with the two alleles in parallel, we obtained plants with \( \text{ipa1-1D/ipa1-2D} \) heterozygous alleles and compared the mRNA levels of two alleles in young panicles (YPs) and seedlings (Fig. 7e,f). The sequencing result showed that the \( \text{ipa1-1D} \) allele was predominantly expressed in both tissues, far more than that of \( \text{ipa1-2D} \) allele (Fig. 7e,f). Similar results were obtained by qRT-PCR analysis of three genotypes in the progeny, and much higher \( \text{IPA1} \) expression was generated in either homozygous or heterozygous \( \text{ipa1-1D} \) plants.

![Figure 5](image-url) **Figure 5 | IPA1 dosage affects IPA characteristics.** (a) Morphology of plants carrying different IPA1 gDNA copy number. 1C, 2C, 3C, and 4C represent plants with 1, 2, 3 and 4 insertions of the transgenic construct, compared with the wild-type (WT) NIP. Scale bar, 15 cm. (b) Copy number detection of four lines by Southern blotting with WT as control. A probe from hygromycin phosphotransferase gene was used to detect the transgene. M, DNA markers. (c-e) IPA traits of plants with increasing IPA1 copy number, including tiller number (c), stem diameter (d) and panicle primary branch diameter (e). Values are means ± s.d. (n = 12). Different letters at top of each column indicate a significant difference at \( P < 0.05 \) determined by Tukey’s HSD test. (f-h) Plots of relative IPA1 expression normalized to rice Actin with trait performance in different lines, including tiller number (f), stem diameter (g) and panicle primary branch number (h). Curves fitting the trait change are calculated by quadratic equation with \( R^2 \) values.
plants, which is otherwise heavily methylated in the wild-type conditions (Fig. 7h). Moreover, pyramiding environments and genetic backgrounds, fitting diverse cultivation provide efficient approaches to generate IPA rice in different backgrounds (Fig. 7h). Together with the highest yield potential in the tested field condition and genetic combination of tiller number and panicle size to reach the ipa1-2D strong allele, and in turn results in a decrease in methylation level of repetitive sequences that mediates gene regulation. We provided evidence that the rice IPA gene is under the control of a non-coding region via a novel enhancer activity, presenting a novel case for non-coding RNAs, and the phenotype can be regenerated by transforming the repeat sequence only. Most recently, we also found the direct release the position-dependent repression of IPA1 that the tandem repeats, a recently occurring is under the control of a non-coding region via a novel epigenetic mechanism. The tandem repeats, a recently occurring weak allele for ipa1-2D, a naturally occurring heterozygous allele is denoted by grey columns respectively. Values are means ± s.d. (n = 67, 124, 62 in Shanghai and 39, 78, 42 in Hainan for three genotypes). Different letters at top of each column indicate a significant difference among genotypes in the respective locations at P < 0.05 determined by Tukey’s HSD test.

than homozygous ipa1-2D plants (Fig. 7g). Together, our data demonstrate that ipa1-2D is a naturally occurring weak allele for IPA1 expression compared to ipa1-1D, a naturally occurring strong allele, and ipa1-2D generates plants with the optimal combination of tiller number and panicle size to reach the highest yield potential in the tested field condition and genetic backgrounds (Fig. 7h). Together with ipa1-1D and IPA1 alleles, six different allele combinations will form and certainly provide efficient approaches to generate IPA rice in different environments and genetic backgrounds, fitting diverse cultivation conditions (Fig. 7h). Moreover, pyramiding ipa1-2D and qPL6, a major QTL for panicle length that we previously identified, shaped a better plant architecture with high yield potential (Supplementary Fig. 20). Thus, rice yield potential could be further improved by functionally pyramiding these additive QTLs in future super rice breeding.

Discussion

Breeding rice with new plant type or IPA to break through the yield ceiling has been proposed since 1990s (ref. 8). However, the molecular mechanisms that generate IPA are not completely understood. In the current study, we reveal that the rice IPA gene IPA1 is under the control of a non-coding region via a novel epigenetic mechanism. The tandem repeats, a recently occurring regulatory sequence identified in the super hybrid rice YY12 and related varieties, may enhance an open chromatin configuration, and in turn results in a decrease in methylation level of repetitive sequences nested in the proximal promoter of IPA1 in ipa1-2D plants, which is otherwise heavily methylated in the wild-type (normal) IPA1 allele. This epigenetic modification may represent a universal mechanism of de-repression for genes proximal to heterochromatin as suggested by recent genome-wide studies. Several QTLs underlying important agronomic traits have been mapped to non-coding regions that modify nearby gene transcription. However, the molecular mechanisms of gene transcriptional regulation by these non-coding regions are different, either in cis or in trans, as indicated by the studies of the maize loci tb1 and b1 (refs 30–32). Both tb1 and b1 are distant from the functional ORFs (~60 kb for tb1 and ~100 kb for b1), but tb1 confers up-regulation whereas b1 confers down-regulation of the downstream ORF, respectively. It has been shown that tb1 covers a functional transposon insertion that up-regulates downstream ORF transcription. In contrast, b1 was identified as tandem repeats that exert function via trans-acting RNAs, and the phenotype can be regenerated by transforming the repeat sequence only. Most recently, we also found the direct copy number variation of the rice grain length gene GL7 leads to up-regulation of GL7 and down-regulation of its nearby negative regulator, resulting in an increase in grain length and improvement of grain quality. Interestingly, the three tandem repeats of ipa1-2D exert neither trans-acting activity nor enhancer activity, presenting a novel case for non-coding sequence that mediates gene regulation. We provided evidence that the ipa1-2D repeats induce an open chromatin status to release the position-dependent repression of IPA1 in wild-type varieties, probably through an unrecognized chromatin complex. Nevertheless, the mechanisms need to be clarified in depth with advanced techniques in future work.
The application of rice hybrid vigour or heterosis beginning in the 1970s enabled a substantial yield leap in rice; however, the underlying molecular basis remains largely unknown. Our study indicated that IPA1 functions as a pleiotropic regulator and determines different IPA traits in a dosage-dependent manner, which provides an alternative mechanistic explanation of heterosis. As IPA1 expression or copy number increases in transgenic lines, the panicle size increases to a plateau rapidly, whereas the tiller number continues to decrease, indicating that an optimal IPA1 dosage will confer the best yield potential. Thus, the F1 plants with big panicles and moderate productive tillers can be generated by crossing plants with contrasting IPA1 dosage, which even forms single-gene heterosis with specific allele combination (Fig. 7h). Coincidentally, dosage sensitivity of the florigen pathway gene also result in yield heterosis in tomato by fine-tuning shoot architecture, and the maximum yield can be generated by combining heterozygous mutations of multiple genes related to floral signal. All these results proved the concepts that the dosage effect of gene products is important for the expression of quantitative traits and optimum performance. This is consistent with our recent findings that several heterosis-related genes act through partial dominance for yield components owning to allele dosage and that heterozygous IPA1 could explain 48.1% of heterosis advantage.

IPA has been selected in rice breeding since it was first conceived. However, few super rice varieties with IPA traits have been found containing the ipa-1D allele, probably due to its large effect in reducing tillers. In contrast, we found that many super rice varieties carry the ipa-2D allele, indicating that ipa-2D has been preferred by breeders in diverse breeding programs. Compared to ipa-1D, ipa-2D is a weak allele for up-regulation of IPA1 expression, which confers a large effect on panicle size but little effect on tiller number, resulting in a better overall performance with the increase of grain production of per unit area. Nevertheless, our breeding practice confirmed that ipa-1D could also be a good candidate for hybrid rice development with high yield potential. As higher yield potential could be achieved by integrating the ipa-2D allele with other yield QTLs, we propose that the tiller reduction in both ipa-1D and ipa-2D could be improved by independent high-tiller genes, which will further increase the yield potential. Interestingly, the non-coding repeats of ipa-1D still do not affect the tissue-specific expression of IPA1. It raises the possibility to identify novel promoter alleles with better IPA1 expression pattern to balance panicle size and tiller number either by mutagenesis or genome editing technology. Also, the markers associated with the tandem repeats developed in this study promise to greatly improve rational design breeding in the future.

### Methods

#### Plant materials

YY12 is the well-known commercial hybrid varieties developed and widely grown in China, and YYP1 is the original IPA breeding stock of YY12 with indica background. For mapping of qWS8/ipa-1D, YYP1 was crossed with the japonica variety NIP, and the resulting F1 plants were selfed to produce F2 generation and were backcrossed with NIP two or three times to generate BC1F1 and BC2F1. The F2, BC2F2, and BC3F2 populations were used for QTL mapping. A single plant with double recombination covering a ~80-kb heterozygous introgression with ipa-2D was selected from the BC2F1 generation, and its sibling lines in BC3F2 generation with homozygous allele of ipa-2D and IPA1 were selected as NIL/ipa-2D and NIL/ipa-2D by successive selection of single heterozygous plants in each generation followed by a final determination of homozygous alleles. In addition, a BC2F1 plant with homozygous ipa-1D and ipa-2D was selected to generate the gene pyramiding line ipa-2D/ipa-2D/ipa-2D and other combination lines of homozygous alleles.

By searching the SNP database of the 3,000 rice genome, GENG77-4, Xiangqin and Jinxiabai with the same qWS8/ipa-1D SNPs were ordered and planted for further genotyping and trait evaluation. Since the combination of IPA breeding strategy, rice accessions with varying IPA character has been developed and enriched in China, and we grew 188 such accessions in the Zhejiang...
station and Hainan station for genotyping and phenotyping. JYZK-3, JYZK-4, JYZK-6 and JYZK-33 are four hybrids, and their restorer lines were bred by introgressing *ipa1*-2D or *ipa1*-1D alleles from the donor plants assisted by molecular makers. Note that the restorer lines of JYZK-4 and JYZK-6 bear similar background and were crossed with the same sterile lines to facilitate trait comparison.

**Trait measurement.** Plants were grown at paddy field with the same management for agronomic traits comparison. Stem diameter was determined by measuring the third internodes of main culms at mature stage using a slide caliper, and data are presented as means of the major axis and minor axis. Tiller number is counted as all fertile panicles in one plant. Panicle primary branch number, secondary branch number and spikelet number were counted manually from main panicles. Grain yield per plant was measured with all grains from each plant. Maximum bending resistance and bending distance were detected by a digital force tester when mature plants were bent to 45° and broken respectively. For yield test of JYZK hybrid varieties, total grain weight of plants in a plot of 13.34 m² with three replications was calculated and normalized to yield per mu (667 m²). Tiller number was counted as all the fertile panicle in 13.34 m² region and normalized to number per mu (667 m²).

**Linkage analysis and QTL mapping.** For QTL detection, 190 plants of F₂ population derived from YY12 and JYZK-4/6, the derived F₂ population was genotyped by tightly linked marker, BC3F3 progeny, and more markers were developed to discriminate the recombinants, and more markers were discriminated to detect the recombination sites (Supplementary Table 5). Using tightly linked marker, BC3F2 progeny of key recombinants were genotyped to select BC3F2 homozygous sibling plants, and traits of both generations were analysed to validate the phenotype. The SNP markers flanking mapping region were also used to genotype different IPA varieties to obtain their SNP haplotypes.

To determine the contribution of *qWS8/ipa1*-2D or *ipa1*-1D alleles in hybrid rice of YY12 or JYZK-4/6, the derived F₂ population was performed by single marker analysis module. Progenies of eight BC3F₂ recombinants were planted to clarify the mapping region with valid phenotype. A larger BC3F₂ population with 3,404 plants was genotyped using the markers RMA339P and RM23246 to find additional recombinants, and more markers were developed to discriminate the recombination sites (Supplementary Table 5). Using tightly linked marker, BC3F₂ progeny of key recombinants were genotyped to select BC3F2 homozygous sibling plants, and traits of both generations were analysed to validate the phenotype. The SNP markers flanking mapping region were also used to genotype different IPA varieties to obtain their SNP haplotypes.

**Southern blot analysis.** Genomic DNA was extracted by the CTAB method and digested by restriction enzymes in 400 μl volumes (10–15 μg gDNA, 40 μl 10 × buffer and 200 U enzymes) for 18 h. The digested DNA was extracted using phenol:chloroform treatment followed by ethanol precipitation and transferred onto Hybond N membranes (Amersham) after agarose gel electrophoresis. Probes were synthesized by DIG PCR labelling system and hybridization was performed by chemiluminescent imaging using a Chemiluminescent Imaging System (Tanon Science & Technology).

**Determination of IPA1 transcription and microRNAs.** Sequential stages of IM development were obtained by sampling the IM every 2 days starting from the booting stage (time of IM initiation) for RNA preparation. Total RNAs were prepared using a TRIzol kit according to the user’s manual (Invitrogen). For detection of IPA1 transcripts, 1 μg of total RNAs was used for cDNA synthesis with a reverse transcription kit (TaKaRa). qRT-PCR was performed in a 20 μl volume with 2 μl cDNA, 0.5 μM gene specific primers (Supplementary Table 5) and 10 μl 2 × Master Mix (TaKaRa) were subjected to 20 μl by water on an ABI 7900 real-time PCR machine according to the manufacturer’s instruction (Applied Biosystems). The rice *Actin* gene (*LOC_Os03g50885*) was used as the internal control. Northern blot analysis was carried out to confirm IPA1 transcript levels with the rice *Actin* probe as loading control. For detection of miR156 and miR529, the same RNA were reverse-transcribed by the miRNA RT-PCR Kit (TaKaRa) and qRT-PCR was performed following the same procedure described above, and universal adaptor primers (supplied by the Kit) and Uni-miRNA primers were used with 5 μl RNA as internal control (Supplementary Table 5).

**Bisulfite sequencing-based DNA methylation analysis.** Genomic DNA from IM tissues was extracted using CTAB method. Bisulfite treatment was performed using the EpiTect Bisulfite kit (Qiagen). Bisulfite-treated DNA was then used to amplify IPA1 promoter regions from the different genotypes by PCR using primers listed in Supplementary Table 5. Amplified PCR fragments were cloned into pGEM-T easy vector (Promega) and sequenced. Sequences of 17 colonies from each genotype were analysed with Kismeth software 38 to obtain the percentage of methylated sites for three cytosine contexts. Results were confirmed with more than 3 independent experiments.

**IPA1 transgene expression ratio detection.** For transgene ratio detection, the genomic DNAs and cDNAs from IM tissues of transgenic plants (T₄) harbouring the IPA1 construct with an artificial SNP (see below) were amplified separately using primers in Supplementary Table 5, and cloned into pGEM-T Easy vector, and individual colonies were sequenced to determine the SNP types (transgenic or native IPA1). The number of colony with respective SNP type was subject to the ratio calculation.

**Open chromatin detection by MNase digestion.** Open chromatin was detected using a published protocol 39. In brief, 2-week-old seedlings of NILs frozen in liquid nitrogen were ground to powder and suspended in 10 ml of ice cold nuclei isolation buffer (1 M hexylene glycol, 20 mM PIPES-KOH [pH 7.6], 10 mM MgCl₂, 1 mM EGTA, 15 mM NaCl, 0.5 mM spermidine, 0.15 mM spermine, 0.5% Triton X-100, 10 mM β-mercaptoethanol and 1 × protease inhibitor cocktail (Roche)) with gentle rotation for 15 min. The suspension was filtered through 30-nm CellTrics, and the elute was centrifuged for 10 min at 1,500 g at 4°C. The pellet was resuspended as crude nuclei extract with digestion buffer (40 mM Tris-HCl [pH 7.9], 0.3 M Suc, 10 mM MgOAc, 1 mM CaCl₂ and 1 × protease inhibitor cocktail (Roche)). Equal aliquot of crude nuclei extract was subject to digestion with increasing MNase amounts at 30°C for 20 min and DNA was extracted using phenol:chloroform treatment followed by ethanol precipitation and used for PCR analysis.

**Constructs for genetic transformation.** The region containing all the three repeats of *qWS8/ipa1*-2D was cloned into binary vector PCMv3101 and transformed into NIP for analysis of possible function of the repeats. The same region was then fused with CaMV35S mini promoter driving the GUS reporter for reporter analysis. The single repeat from NIP was also cloned and fused with the reporter construct. These constructs were transformed in NIP callus to get independent transfectants, which were grown in medium supplemented with X-Gluc to detect reporter activity. Three genomic fragments containing IPA1 with size of 8, 11.9 and 13.9 kb from the NIP BAC AP060049 were cloned into pCAMb1301, respectively, and an artificial SNP was introduced into the 8-kb construct by site-directed mutagenesis. The IPA1 cDNA AK170191 was also cloned into pCAMb1301 with the CaMV35S promoter to obtain ectopic IPA1 over-expression construct. All the gDNA and cDNA constructs of IPA1 were transformed into NIP to generate more than 15 independent transgenic lines (T₄ generation) were used for gene expression and phenotyping analysis.

**Infrorescence meristems imaging and measurement.** IM tissues at different stages were sampled and fixed in the FAA solution (45% ethanol, 5% acetic acid, 5% formaldehyde). For scanning electron microscopy, the samples were critical-point dried with liquid CO₂ and coated with gold, followed by visualization with a scanning electron microscope (ISM-6300LV, JEOL). For histological sections, samples were embedded in paraffin or resin, and transverse sections were made at 10 μm for paraffin samples and 2 μm for resin samples. Sections were examined microscopically (BX51, Olympus) and photographed. The IM size was measured from pictures of scanning electron microscopy using ImageJ software (NIH).

**Data availability.** The authors declare that all data supporting the findings of this study are available within the manuscript or its supplementary files or are available from the corresponding authors upon request.

**References**

1. Tilman, D., Balzer, C., Hill, J. & Befort, B. L. Global food demand and the sustainable intensification of agriculture. *Proc. Natl Acad. Sci. USA*, 108, 20260–20264 (2011).
2. Qian, Q., Guo, L. B., Smith, S. M. & Li, J. Y. Breeding high-yield superior-quality hybrid super-rice by rational design. *Nat. Commun.*, 7, 6156 (2016).
3. Ray, D. K., Ramankutty, N., Mueller, N. D., West, P. C. & Foley, J. A. Recent patterns of crop yield growth and stagnation. *Nature*, 560, 421–424 (2016).
4. Grassini, P., Eskridge, K. M. & Cassman, K. G. Distinguishing between yield advances and yield plateaus in historical crop production trends. *Nat. Commun.*, 4, 2918 (2013).
5. Ray, D. K., Ramankutty, N., Mueller, N. D., West, P. C. & Foley, J. A. Recent patterns of crop yield growth and stagnation. *Nature*, 560, 421–424 (2016).
6. Khush, G. S. Green revolution: the way forward. *Nat. Rev. Genet.*, 2, 815–822 (2001).
7. Sasaki, A. et al. Green revolution: A mutant gibberellin-synthesis gene in rice. *Science*, 301, 741–744 (2003).
8. Khush, G. S. Breaking the yield frontier of rice. *Geological Journal*, 35, 329–332 (1995).
9. Jiao, Y. et al. Regulation of OsSPL14 by OsMIR156 defines ideal plant architecture in rice. *Nat. Genet.*, 42, 541–544 (2010).
10. Miura, K. et al. OsSPL14 promotes panicle branching and higher grain productivity in rice. Nat. Genet. 42, 545–549 (2010).
11. Xue, W. Natural variation at the DEPI locus enhances grain yield in rice. Nat. Commun. 1, 132 (2010).
12. Huang, X. Z. et al. Natural variation at the DEPI locus enhances grain yield in rice. Nat. Genet. 41, 494–497 (2009).
13. Fujita, D. et al. NAT1 allele from a rice landrace greatly increases yield in modern indica cultivars. Proc. Natl Acad. Sci. USA 110, 20431–20436 (2013).
14. Miura, K. et al. Quantitative trait locus analysis and fine mapping of the Ghd7 locus. Theor. Appl. Genet. 132, 1160–1163 (2011).
15. Li, Q. et al. Massive analysis of rice small RNAs: mechanistic implications of regulated microRNAs and variants for differential target RNA cleavage. Plant Cell 23, 4185–4207 (2011).
16. Xue, W. et al. Genomic analysis of hybrid rice varieties reveals numerous superior alleles that contribute to heterosis. Nat. Commun. 6, 6258 (2015).
17. Huang, X. H. et al. Genomic architecture of heterosis for yield traits in rice. Nature 537, 629–633 (2016).
18. Peng, S. B., Kushn, G. S., Virk, P., Tang, Q. Y. & Zou, Y. B. Progress in ideotype breeding to increase rice yield potential. Field Crops Res. 108, 32–38 (2008).
19. Zhang, L. et al. SNP-Seek database of SNPs derived from 3000 rice genomes. Nucleic Acids Res. 43, D1023–D1027 (2015).
20. Jeong, D. H. et al. Massive analysis of rice small RNAs: mechanistic implications of regulated microRNAs and variants for differential target RNA cleavage. Plant Cell 23, 4185–4207 (2011).
21. Nagasawa, N. S., Nagazawa, N. S. & Jackson, D. Quantitative variation in maize cleavage. Plant Cell 23, 594–597 (2006).
22. Zhang, X. et al. Quantitative trait locus analysis and fine mapping of the qPL6 locus for panicle length in rice. Theor. Appl. Genet. 128, 1151–1161 (2015).
23. Li, Q. et al. RNA-directed DNA methylation enforces boundaries between heterochromatin and euchromatin in the maize genome. Proc. Natl Acad. Sci. USA 112, 14728–14733 (2015).
24. Swinnen, G., Goossens, A. & Pauwels, L. Lessons from domestication: targeting cis-regulatory elements for crop improvement. Trends Plant Sci. 21, 506–515 (2016).
25. Stam, M., Belele, C., Dorweiler, J. E. & Chandler, V. L. Differential chromatin structure within a tandem array 100 kb upstream of the maize b1 locus is associated with paramutation. Genes Dev. 16, 1906–1918 (2002).
26. Clark, R. M., Wagler, T. N., Quijada, P. & Doebely, J. A distant upstream enhancer at the maize domestication gene tb1 has pleiotropic effects on plant and inflorescence architecture. Nat. Genet. 38, 594–597 (2006).
27. Studer, A., Zhao, Q., Ross-Ibarra, J. & Doebely, J. Identiﬁcation of a functional transposon insertion in the maize domestication gene tb1. Nat. Genet. 43, 1160–1163 (2011).
28. Belele, C. L. et al. Specific tandem repeats are sufﬁcient for paramutation-induced trans-generational silencing. PLoS Genet. 9, e1003773 (2013).
29. Wang, Y. et al. Copy number variation at the GL7 locus contributes to grain size diversity in rice. Nat. Genet. 47, 944–948 (2015).
30. Jiang, K., Liberatore, K. L., Park, S. J., Alva, E. P. & Lippman, Z. B. Tomato yield heterosis is triggered by a dosage sensitivity of the florigen pathway that fine-tunes shoot architecture. PLoS Genet. 9, e1004043 (2013).
31. Park, S. J. et al. Optimization of crop productivity in tomato using induced mutations in the florigen pathway. Nat. Genet. 46, 1337–1342 (2014).
32. Birchler, J. A., Johnson, A. F. & Veitia, R. A. Kinetics genetics: Incorporating the concept of genomic balance into an understanding of quantitative traits. Plant Sci. 245, 128–134 (2016).
33. Gent, J. I. et al. Kismet: Analyzer of plant methylation states through bisulﬁte sequencing. BMC Bioinf. 9, 371 (2008).
34. Shu, H., Gruissem, W. & Hennig, L. Measuring Arabidopsis chromatin accessibility using DNase I-polymerase chain reaction and DNase I-chip assays. Plant Physiol. 162, 1794–1801 (2013).

Acknowledgements
We thank Linyou Wang, Lixia Zhang and Yonghui Qi for rice growth and yield field tests, Zhikang Li for providing rice germplasms, Ertao Wang for valuable suggestions. This work was supported by grants from Ministry of Science and Technology of China (2013ZX08009-003-001, 2016YFD0101800, 2014AA10A600), the Chinese Academy of Sciences (XDA0801 and XDA0803), the National Natural Science Foundation of China (31600990).

Author contributions
J.L., B.H., Q.Q., Z.H. and J.L. conceived and designed the research. L.Z. and H.Y. designed experiments and analysed the data. L.Z., H.Y., B.M., G.L., J.W., J.W., R.G., J.L., J.L., J.X., Y.Z., Q.L., X.H. and J.X. conducted the experiments. J.L. and Z.H. oversaw the entire study. L.Z., H.Y., Z.H. and J.L. wrote the manuscript.

Additional information
Supplementary Information accompanies this paper at http://www.nature.com/...