Large scale genomic and transcriptomic profiles of rice hybrids revealed a novel universal mechanism underlying yield heterosis

jianyin Xie
China Agricultural University
Weiping Wang
Hunan Hybrid Rice Research Center
Quan Zhang
China agricultural university
Tao Yang
China agricultural university
Zhifang Zhang
China agricultural university
Xiaoyang Zhu
China agricultural university
Ni Li
Hunan Hybrid Rice Research Center
Linran Zhi
China agricultural university
Xiaoqian Ma
China agricultural university
Shuyang Zhang
China agricultural university
Yan Liu
National Institute of Biological Sciences, Beijing
Xueqiang Wang
Key Lab of Crop Heterosis and Utilization of Ministry of Education and Beijing Key Lab of Crop Genetic Improvement, China Agricultural University, Beijing, 100193, China.
Feimei Li
China agricultural university
Yan Zhao
Shandong Agricultural University
Xuewei Jia
China agricultural university https://orcid.org/0000-0003-3944-2968

Jieyu Zhou
China agricultural university
Ninjia Jiang
China agricultural university
Gangling Li
China agricultural university
Zhanying Zhang
China agricultural university
Jinjie Li
China Agricultural University
Ziding Zhang
China Agricultural University https://orcid.org/0000-0002-9296-571X
Zichao Li
China Agricultural University https://orcid.org/0000-0002-3186-1132
Hongliang Zhang (zhangl@cau.edu.cn)
Key Lab of Crop Heterosis and Utilization, China Agricultural University

Article

Keywords: heterosis, Liangyoupei 9, non-additive and additive factors, HoIIB model, hybrid rice breeding

Posted Date: August 17th, 2021

DOI: https://doi.org/10.21203/rs.3.rs-753426/v1

Large scale genomic and transcriptomic profiles of rice hybrids revealed a novel universal mechanism underlying yield heterosis

Jianyin Xie[1,3,4], Weiping Wang[2,4], Quan Zhang[1], Tao Yang[1], Zhifang Zhang[1], Xiaoyang Zhu[1], Ni Li[2], Linran Zhi[1], Xiaoqian Ma[1], Shuyang Zhang[1], Yan Liu[1], Xueqiang Wang[1], Fengmei Li[1], Yan Zhao[1], Xuewei Jia[1], Jieyu Zhou[1], Ningjia Jiang[1], Gangling Li[1], Zhanying Zhang[1], Jinjie Li[1], Ziding Zhang[3], Zichao Li[1]*, Hongliang Zhang[1]*

1 Key Laboratory of Crop Heterosis and Utilization, the Ministry of Education / Beijing Key Laboratory of Crop Genetic Improvement, China Agricultural University, Beijing, 100193, China;

2 State Key Laboratory of Hybrid Rice, Hunan Hybrid Rice Research Center, Changsha 410125, China;

3 State Key Laboratory for Agrobiotechnology, China Agricultural University, Beijing, 100193, China

4 These authors contributed equally to this work.

* To whom correspondence should be addressed: zhangl@cau.edu.cn or lizichao@cau.edu.cn

Abstract

The utilization of heterosis (or hybrid vigor) is a revolutionary technology in agricultural. However, its genetic mechanisms are still unclear in plants. Here we develop, sequence and record the phenotypes of 418 hybrids from crosses between two testers and a diverse mini core collection. Phenotypic analysis showed that heterosis is an extensive but not necessary phenomenon, which varied by combinations and environments. Evidence from both GWAS on the 418 hybrids and their parents and transcriptomics of the traditional rice hybrid Liangyoupei 9, indicated that dominance and overdominance are the main genetic contributions to heterosis. Furthermore, cumulation or complementation of repulsive genetic factors may account for 37.8% of the overdominant QTL and nearly half of the genes with overdominant expression pattern. We systematically compared non-additive and additive factors and observed a common phenomenon that non-additive factors are more sensitive to background than that of additive ones across species, phenotypes, QTLs and transcription levels, further evidence from both simulations and experiment demonstrated a novel universal molecular mechanism underlying heterosis, i.e. homo-insufficiency under insufficient background (HoIIB), which expounds that heterosis in most cases is not the heterozygote advantage but the homozygote disadvantage under the insufficient genetic background. The HoIIB model can explain most known hypotheses and phenomena about heterosis, thus provides a novel theory for future hybrid rice breeding.

Introduction

Hybrid breeding is a revolutionary technology in agricultural production and for food security. Due to their dramatic increased yield by tens of percent and even double compared to inbreds[1], hybrid have been the important and even the main variety type for agricultural plants and animals[2]. Rather different from traditional inbred breeding, which mainly exploit accumulation of homozygous beneficial alleles, hybrid breeding takes advantage of heterosis or hybrid vigor, which refers to the phenomenon that the hybrid from two genetically distantly related inbred lines show superior performance than their parents[3,4].

Although heterosis has been utilized extensively in agriculture, its mechanistic understanding is still fragmentary and challenging[5]. Regarding of its genetic basis, there are three classical hypothesis, including dominance[6,7], overdominance[8,9] and epistasis[10-12]. Through quantitative trait loci (QTL) mapping and genome-wide association studies (GWAS), previous study of multiple crops has identified a large number of genetic variant with various types of genetic effects[13-17]. Meanwhile, several well designed studies at transcriptome level has been carried out in plant hybrids such as *Arabidopsis*, maize, and rice, and many genes appear to be dominant, overdominant, or parent-specific in expression[18-20]. At the single gene level, genes with partial or complete dominance effect are commonly observed in many species, such as such as *PMA1* and *MSB2* in yeast[21], *PCSK9* in human heart diseases[22], *Dw3* in sorghum, and *GS3* and *Ghd7* in rice[23,24]. There are also several cases that one gene displayed overdominant effect, such as the *SFT* gene affecting fruit yield of tomato[25], the SHELL gene controlling the oil yield in oil palm[26], and the *FNS* gene impacting flower color in *Mimulus lewisii*[27]. The second typical view on heterosis suggested that the pleiotropic functions with compromise, balance[28,29], and complementation between two alleles or among factors, and gene-gene interactions at various levels[30-32], represent an important genetic mechanism underlying heterosis. The third explanation is that hierarchical effects at different levels or aspects contribute largely to heterosis, such as the multiplicative effect on yield by its component traits, where accumulation of partial dominance usually occurs[33,34]. However, the theories mentioned above are challenging to address this question: How does a single gene function as non-additive effect at the molecular level? whether there is a general mechanism for dominance or overdominance to occur?

Rice is one of the crops that successfully utilize heterosis in breeding. Numerous studies have been carried out to investigate genetic and molecular mechanisms of heterosis in rice. However, there is still no consensus on such mechanisms [35-37]. Early quantitative trait analysis in an *indica-japonica* hybrid suggested that dominance accumulation was the major cause of heterosis[35]. The subsequent genetic dissection of yield traits tended to support that epistasis and overdominance were the major genetic basis of heterosis in rice[35]. The decomposition of yield traits, based on an immortalized F_2 population from an *indica-indica* rice hybrid, indicated that relative contributions of dominant factors varied by traits and single-locus dominance has relatively small contributions in all traits[36]. Recent studies using 1,495 commercial hybrids and 10,074 F_2 individuals from 17

crosses demonstrated that the heterosis mainly attributes to accumulation of numerous rare superior alleles with positive dominance[37]. Nevertheless, according to the materials and methods of these studies adopted, two issues are still need to be addressed on heterosis in rice: one is that most of these studies mainly focused on commercial hybrids or their derived populations[37], such as the "immortalized F_2" derived from Zhenshan97 and Minghui63[36], and the BCF_1 population derived from Peiai64S and 9311[38]. These materials may not represent the overview of heterosis across both intra-subspecific and inter-subspecific rice hybrids. Thus, further extensive studies on combinations derived from a wider spectrum of rice germplasmic resources may provide mechanistic understanding of the heterosis. The second point is that although these researches have made great progress, most of them are still confined to the traditional concept and terms, and mainly focused on analyzing the proportions and contributions of various genetic component (including additive, dominance and overdominanc) to heretosis[39], insightful and overall studies on the mechanism of additive and non-additive are still lacking. Therefore, a comprehensive and systematic analysis of the internal mechanism of additive and non-additive is important to resolve the biological mechanism of heterosis in rice.

To get insight into rice heterosis, in the preset study, we generated 418 F_1 hybrids, which were from crosses between two testers (Nipponbare and 9311) and 265 diverse cultivated rice varieties collected from 35 countries[40] (**Supplementary Table 1, Supplementary Fig 1**). Using both phenotypic and genomic data collected from the hybrids and their parental lines, as well as transcriptomic profiling data from two sets of hybrid combinations, we then systematically identify genetic variants affecting heterosis of grain yield and yield related traits, followed by dissection of genetic effects including additive, dominance and overdominance **(Supplementary Fig 2**). Evidence from these analyses, as well as from simulations and experimental validation, indicated that there is a universal molecular mechanism underlying heterosis of single polymorphic locus in rice, that is, homo-insufficiency under insufficient background (HoIIB). The HoIIB model can explain the known hypotheses and phenomena about heterosis, thus provides a novel theory for future hybrid rice breeding.

Results

Genetic diversity and differentiation in the parental panel

We identified 4,625,141 SNPs in the parental panel (N = 267), after exclude SNPs with minor allele frequency (MAF) less than 5% and missing rate larger than 50%. According the neighbor-joing tree of 267 parental lines based on the SNPS and that of the 3,024 varieties of the rice genome based on 100,000 SNP[41](**Supplementary Fig 3a**), *japonica* lines in the panel can be classified into nine distinct sub-populations and *indica* lines can be classified into eight sub-populations. Regarding pairwise comparisons of intra- or inter-subpopulations, the differences of both intra- and inter-subpopulations in *japonica* were significantly less (215,949 or 435,510 on average) than those in *indica* (487,428 or 707,594 on average). Interestingly, even the inter-subpopulation differences in *japonica* were less than that of intra-subpopulation in *indica* **(Supplementary Fig 3b)**. These lines of evidence implied different potentials of intra-subspecific heterosis utilization and partially explained the fact that there are so fewer *japonica-japonica* hybrids than *indica-indica* ones in commercial rice production[42]. The real hybrid from four types of combination indicated diverse genome heterozygosity **(Supplementary Fig 4)**. As expected, the heterozygosity of *indica-japonica* hybrids was much higher than that of both *indica-indica* and *japonica-japonica* ones. The intra-subspecific hybrids of *Japonica*×Nipponbare (J×Nip) display distinctly lower heterozygosity than that of *Indica*×9311 (I×9311), due to the poor genetic diversity in *japonica*[43]. However, the inter-subspecific hybrids of *Indica*×Nipponbare (I×Nip) present higher heterozygosity than that of *Japonica*×9311 (J×9311), consistent with the fact that a small proportion of *japonica* genomic segments had been introgressed into the *indica* line 9311 **(Supplementary Fig 5)** and apparently lower genetic variation in *japonica* than that in *indica***.**

Heterosis is highly dependent on genetic background and environmental conditions

We phenotyped the 418 hybrids and their 267 parents in 2013 at respective Changsha (CS) (28°13′N, 112°58′E, a long-day environment) and Sanya (SY) (18°10′N, 109°28′E, a short-day

environment) of China. 6 yield related trait including grain weight per plant (GWP), spikelet number per panicle (SPP) and its two components traits (both primary and secondary branch numbers per panicle (PBP and SBP)), 1000-grain weight (KGW), and panicle number per plant (PNP) were investigated (**Supplementary Table 2a**).

It is obviously that the hybrid performance is substantially impacted by the parental background and heterosis, but mainly by the later in most case and varied among traits, environments and combinations **(Fig 1a; Supplementary Fig 6-7).** In general, parental background shapes the basic performance of hybrid. Taken SPP as an example, both *indica* and *japonica* hybrid with Nipponbare background showed relative lower SPP than those with 9311 background, no matter the environment is Changsha or Sanya **(Fig 1a)**. The other 5 traits exhibit the same phenomenon, except for PNP and GWP of *japonica* hybrids in Changsha and Sanya, respectively (**Supplementary Fig 6d and e**). More importantly, heterosis has a stronger influence on hybrid performance than the parents. Our observation indicated that the contribution of heterosis to hybrids was considerably higher than that of parental background, especially in the long-day environment **(Fig 1a; Supplementary Fig 7)**. In Changsha, the overall phenotypic contribution of heterosis to hybrids of both *japonica* and *indica* (63.2% and 67.7%) were much higher than that of the parental background (17.1% and 11.1%). In Sanya, the contribution of parental background to spikelet number related traits (PBP, SBP and SPP) and PNP in hybrid increased, as compared to Changsha with a long-day condition, while the contribution of heterosis to hybrids decreased for majority of the cases except for PNP. Comparing different traits in the same hybrid combination and environment, we found that the parental background contributions were much higher for KGW (19.2%) and spikelet number traits (26.4% for SPP, 27.1% for SBP, and 25.4% for PBP) than for PNP and GWP (just 6.2% and 2.5%, respectively) **(Supplementary Fig 7).** The lower heterosis contribution to hybrids in terms of spikelet number traits in short-day environment of Sanya suggested the necessity to reduce hybrid photosensitivity in hybrid breeding program, so as to maintain a certain period of growth duration under the short-day environment.

In order to investigate the degree of environment effect on yield traits of inbreds and hybrids, we performed the two-way analysis of variance (ANOVA) including environment as factor. The result showed that there were obvious environment effects (including genotype-environment interactions and environment effects) on both inbreds and hybrids **(Fig 1b).** Although different traits

displayed a similar pattern of residual effects on both inbreds and hybrids, that is, the highest for GWP, followed by PNP, SPP and SPP-related traits (PBP and SBP), and the lowest for KGW in both *japonica* and *indica* subspecies (**Supplementary Fig 8a; Supplementary Table 3 and 4**). However, the proportion of environment effect on hybrids was generally much higher than that on the corresponding inbreds, especially for SPP and its related traits **(Fig 1b; Supplementary Table 3 and 4)**. We used the same method to analyze the data of 266 maize hybrids and their parents in four environments, the results also showed that the effect of environment on hybrids was generally stronger than that of the inbred parents (**Supplementary Fig 8b-c**)[44]. Another recent maize panel also showed the similar result (data not shown)[45]. In summary, it is a common phenomenon that hybrids are more sensitive to the environment than their inbred parents.

Using our diverse and large scale of hybrids, investigation of the strength of heterosis in terms of different traits, combinations, and environments indicated that heteosis, especially the better-parent heterosis, are just the potentiality rather than the apodeictic result of hybridization or heterozygosity, despite which is predominant over the cases **(Fig 1c-d; Supplementary Fig 9a-h)**. It was clear that not all combinations showed hybrid vigor. On average 10.65 % of intra-subspecifc hybrids and 10.29% of inter-subspecific hybrids even displayed hybrid weakness **(Supplementary Fig9 a-h)**. The degree of middle parent heterosis (dHmp) from all of combination types, in terms of SPP related trait and GWP, ranged from 7.87%-70.13% with an average of 35.7% in Changsha, this was apparently higher than that in Sanya (-8.53%-53.30% with an average of 8.21%) **(Fig 1c-d; Supplementary Table 2b)**. However, For PNP and KGW, the dHmp of all combinations generally appeared to be no obvious difference between the two locations. Particularly, the proportion of positive overdiminant (POD) heterosis of all traits across the combinations in Changsha (averagely 63.19%) was much higher than that in Sanya (averagely 32.40%). And the decrease of POD from long-day to short-day environments was distinctly represented by spikelet number related trait (PBP, SBP and SPP), compared to the other traits, Compared to the intra-subspecific combinations, only PNP displayed stronger heterosis for the inter-subspecific combinations evaluated under the two environments. Regarding heterosis of the other traits, the differences between inter- and intra-subspecific combinations varied by traits, testers, subspecies, and environments (**Fig 1c-d**). Similarly, the proportion of POD heterosis of inter-subspecific combinations for all traits was apparently higher than that of intra-subspecific combinations in

Changsha, except for the PNP trait with consistent heterosis across the two environments. In contrast, for SPP related traits and GWP, the proportion of POD heterosis of inter-subspecific combinations was even lower than that of intra-subspecific combinations in Sanya (**Supplementary Fig 9e-h**). These lines of evidence indicated that the degree of heterosis is apparently dependent on environments, trait, testers, and combinations. Thus it is of significance to uncover the genetic basis and mechanism underlying inbreds, hybrids and especially heterosis so as to highlight the opportunity to produce strong heterosis and elite hybrids.

Genome-wide identification of QTLs affecting yield traits of rice hybrids

We carried out genome wide association studies (GWAS), using 120 sets of genetic and phenotypic data. The phenotypic data consist of three types of datum panels evaluated for six yield traits (PBP, SBP, SPP, KGW, PNP and GWP) under two environments (Changsha and Sanya) (**Supplementary Fig 10-21**). The three types of datum panels include 20 sets of data for each of the traits, i.e. (1) four sets from parents (both *indica* and *japonica* in two environments), (2) eight sets from F_1 (four types of combinations in two environments), and (3) eight sets of calculated middle-parent heterosis value (Hmp) **(Methods)**.

Totally, we identified 621 and 624 QTLs in Changsha and Sanya, respectively, from the parental datum panel (P_QTL), 828 and 895 QTLs from the F_1 datum panel (F_1_QTL), and 636 and 895 QTLs from the Hmp datum panel (Hmp_QTL) (**Supplementary Table 5-7)**. When comparing the two environments, P_QTLs appeared apparently to be more environment-stable (38.4% on average), compared to the F_1_QTLs (9.8% on average) and Hmp_QTLs (6.6% on average), regarding the traits related to grain number and grain size. As for PNP, the situation is combination-dependent. The three panels of QTLs related to grain weight per plant (GWP) were rather environment-specific (**Supplementary Fig 22**).

Comparing the shared QTLs from the three panels (P_QTL, F_1_QTL and Hmp_QTL), we found that the genetic architecture affecting hybrids synchronizes more with that impacting heterosis (24.09±21%), compared to that affecting inbred parent (12.28±10%), but there were some exceptions for some combinations, environments and traits (**Supplementary Table 8; Supplementary Fig 23**). The situation with more colocalized F_1_QTL and P_QTL than colocalized

222 F_1_QTL and Hmp_QTL was more in Sanya than Changsha, more for spikelet number than KGW
223 and PNP, more for 9311 combinations than Nipponbare ones. These results implied that the
224 improvement of hybrids was, in general, fulfilled mainly by use of heterosis under the genetic
225 background from elite inbred lines, but their respective contribution varied depending on the
226 combinations, environments and traits.

227 It is very surprising that few QTLs were able to be repeatedly identified across the four types
228 of combinations (from 0.00% to 12.51%, 2.45% on average), suggesting that different genetic basis
229 contributed to the different intra-subspecific and inter-subspecific hybrids[46]. The above results
230 indicated that heterosis of different combinations under different inbred backgrounds varied by
231 traits in response to different environments. These imply the complicated genetic basis of heterosis
232 and the essential relationship of heterosis with the genetic background within a combination.

233 Non-additive, which is more variable than additive, is the main contributor to
234 heterosis

235 We estimated both additive and non-additive effects of each QTL affecting grain yield of
236 hybrids, inbred parents, and calculated mid-parent heterosis (Hmp), in order to understand genetic
237 basis underlying heterosis. A QTL is referred to as overdominance preferred if the absolute ratio of
238 dominant effect to additive effects (|d/a|, degree of dominance) is no less than 1.5, and dominance
239 preferred if $0.5 \leq |d/a| < 1.5$ (including partial-dominance), and additive preferred if $|d/a| < 0.5$ (see
240 **Methods** for detail).

241 Among the 44 scenarios (five traits of four types of combinations under the two environments
242 plus GWP of two types of intra-subspecific combinations under the two environments), both
243 F_1_QTLs and Hmp_QTLs beared apparently more non-additive effects than did P_QTLs
244 (**Supplementary Fig 24**), except for primary branch number per panicle of J×9311 in Sanya.
245 Particularly, the majority of the F_1_QTLs and Hmp_QTLs displayed overdominant effects (69.27%
246 and 77.71%, respectively), with only a small portion of QTL represented additive effect (10.66%
247 and 7.55%, respectively). Conversely, the majority of P_QTLs demonstrated additive (44.16%) and
248 dominant (37.08%) effects, and only a small proportion (18.74%) showed overdominance.
249 Furthermore, proportions of the F_1_QTLs and Hmp_QTLs with non-additive effects varied by

traits, subspecific hybrids, and environments. Consistent with the observation that SPP related trait did not exhibit obvious heterosis in Sanya, fewer overdominant F_1_QTL and Hmp_QTL were identified in Sanya than that in Changsha. On average, 75.7% of the F_1_QTLs identified in Changsha expressed as overdominance for the SPP related traits, however, the proportion significantly reduced to 42.6% in Sanya. Compare the two subspecies, we found that the reduction is more remarkable in *japonica* hybrids (from 71.0% in Changsha to 22.3% in Sanya) than that in *indica* hybrids (from 80.4% in Changsha to 62.8% in Sanya) (**Supplementary Fig 24a-c**). On the other hand, the F_1_QTLs and Hmp_QTLs for KGW and PNP did not show such consistent changes between subspecies and between environments, varying by combinations (**Supplementary Fig 24d-e**). When comparing environmental stability among the QTLs of additive, dominant and overdominant ones, we found that a large proportion of additive QTLs showed environment-stable than non-additive QTLs, regarding all types of combinations for all traits except for PNP. In addition, a higher proportion of dominant QTLs indicated environment-stable than overdominant QTLs for most of the combinations and traits (**Supplementary Fig 25**). Thus, our results indicated that the higher magnitude of the dominant effect of a QTL, the stronger the environmental sensitivity of the hybrid. The distinctively larger proportion of unstable factors including overdominant and dominant QTLs identified in hybrids or heterosis than that in inbreds, consistent with the fact that the response of hybrid to environment was generally stronger than that of inbreds **(Fig 1b and Supplementary Fig 8b)**.

Dominance / partial-dominance cumulation and complementation are prevalent genetic basis of overdominance

The above mentioned evidence from both phenotyping and QTLs mapping indicated that hybrids and heterosis mainly attributes to the non-additive effects (dominance and over-dominance) and non-additive effects are more environmental sensitive than additive ones. It is worth noting that heterosis and non-additive effect for some loci may not be the necessary results of heterozygosity, but as a potential possibility depending on different subspecies, testers, traits and environments. Apparently, it is challenging to take advantage of the maximum effects of the loci, before we understand the molecular mechanism underlying the additive and non-additive effect. Currently,

there are three easily comprehensible genetic mechanisms that produce the non-additive effects and especially the overdominant effect through two or more repulsive factors, i.e. the multiplication of additive or dominant factors[10], the cumulation of dominant factors[37,47] and the complementation of two factors. In our observation, the effect of multiplication between/among repulsive additive factors contributed only a little to yield heterosis (**Supplementary Fig 26**). Therefore, we mainly focus on the last two genetic mechanisms at both the QTL and transcription level in this section.

To estimate the contribution of repulsive dominant allele (RDA) cumulation to non-additive effect and especially the overdominance, we calculated the repulsive degree of the SNP with the same direction (positive or negative) of dominant effect in each dominant or overdominant QTL (**see Methods**). As expected, the proportion of higher repulsive degree (0.2-0.4 and >0.4) in QTLs with overdominance (38.0%) was much higher than that in QTLs with dominance (14.57%), and this feature was prevalent across all traits, combinations and environments (**Fig 2a-c; Supplementary Fig 27a-c**). This phenomenon was also observed among the dominant and overdominant QTLs in 1086 three-line hybrids[34] (**Supplementary Fig 28-29**). When compared both the inter- and intra-subspecific combinations, the inter-subspecific combinations represent distinctly higher proportion (52.2%) of combinations with RDA averaged by the QTLs containing RDA than intra-subspecific combinations (25.6%) (**Fig 2d-f**; **Supplementary Fig 27d-f**). This indicated that the probability of RDA cumulation in the inter-subspecific hybrids is double of that in the intra-subspecific hybrids. Further, the probability of RDA cumulation in the J×J combinations was much lower than that in the I×I ones (**Fig 2d-f; Supplementary Fig 27d-f**). More interestingly, we found that the overdominant QTL with higher repulsive degree tended to be stable between two environments (**Supplementary Fig 30**). Given that additive effect was more stable across environments than dominant effect (**Supplementary Fig 25**), we may expect that the complementation of additive alleles with repulsive phase contributed apparent effect on the overdominant QTLs besides of the RDA cumulation. But we may not clearly distinguish the complementation of alleles from the RDA cumulation at the QTL level, and we thus estimated the contribution of complementation mainly at the transcription level (see below). Considering that the repulsive phase at short distance was not easy to be broken, we anticipated that the cumulation of dominant alleles or the complementation from repulsive phased alleles will continue to play an important role in heterosis utilization.

One novel universal molecular mechanism of dominance and overdominance - homo-insufficiency under insufficient background (HoIIB)

As mentioned in introduction and above, the heterosis and non-additive phenomenon at the phenotype level are often resulted from the integrated effects of multi-factors at various intermediate and fundamental levels (such as different genes, QTLs, gene expression, and physiological traits), thus it is challenging to investigate the molecular mechanism of heterosis at the phenotypic level. Transcription is such an intermediate step for a gene to perform its functions in development of complex phenotypes. Therefore, it is informative to explore gene expression patterns between parents and their F_1 hybrid, in order to understand molecular mechanisms underlying heterosis. Here we investigated transcriptome profile of young panicles from the hybrid LYP9 and its two parents PA64S and 9311. As a whole, 8,248 genes showed differential expressions between the two parents and their F_1 hybrid in at least one of three tissues (1 mm, 2 mm, 3 mm young panicles). Expression patterns can be classified as additive (A) (13%), dominant (D) (39%), and overdominant (OD) (48%) in at least one of three tissues (**Fig 3a, Supplementary Table 9**). The OD can be further grouped as negative and positive (NOD and POD), which mean expression level in hybrids is lower and higher than that in both parents, respectively. We identified that the NOD and POD accounted for 21.0% and 81.0%, respectively. Many of them belong to the directly observed complementary pattern, including the complementary negative or positive OD, that is, expression was observed only in both parents but absent in hybrids (CNOD (5.8%)), or vice versa (CPOD (36.7%)). The directly observed complementary effects (CNOD and CPOD) accounted for about 40% of the overdominant expression, and no more than 20% of the 7,248 non-additive expressed genes. These results strongly suggested that the complementarity may be involved in the mechanism of overdominance at transcriptional level.

When compared the stability of genes with additive and non-additive expression, we were surprised to find that the dominant and overdominant expression showed dramatically more variability across tissues than the additive expression (**Fig 3b**); in another word, non-additive or heterosis of expression is more tissue-specific and may be more background-dependent. This is consistent with the results mentioned above, where hybrids are more variable than the inbreds and non-additive QTLs are more variable than additive QTLs (**Fig 1b; Supplementary Fig 8 and 25**).

Although we repeatedly found that non-additive factors were more dependent on the backgrounds, it is still difficult to directly evaluate factors and their background at phenotypic and QTL level. Fortunately, it is possible to examine the relations between expressing genes and their transcription factors that directly regulate them, since many species had a well documented transcription factors and target gene annotations. In this sense, transcription factor can be considered as the background of its target expressing genes. Given these speculations, we carefully investigated the dependency between the types of genetic effects (additive, dominant, and overdominant) and the genetic background, by analyzing the correlation of expression levels between a gene and its transcription factors. The results indicated that the expression of genes with dominant and overdominant effects represented apparently stronger dependency on their transcription factors than those genes with additive effects (**Fig 3c**). The phenomenon was also observed in the transcriptomes of three *Arabidopsis thaliana* combinations (**Supplementary Fig 31)**[16]. These results implied an interesting phenomenon that expression of genes with non-additive effects is more sensitive to the dosage changes of its genetic background. Does the genetic background dependency of non-additive effects represent a universal molecular mechanism underlying heterosis? It is well known that no factor is absolutely independent in the biology system, and that ligand-receptor binding, including the binding of transcription factor to a target gene, is obviously the most common dependent relationship between molecules, where the ligand and receptor can be the genetic background of one another and their binding reaction is described by the Hill equation[48,49]. In order to investigate whether the insufficient genetic background could result in the target factor to be sensitive to the background and the possible internal relationship with the occurrence of dominance and overdominance in a system with diploid parents and their F_1, simulated genetic effects of one polymorphic site of one receptor were compared among the diploid parents and their F_1, according to the Hill equation with different ligand concentrations as the background. We here considered the following three major scenarios with the assumption that the ligand concentration keeps constant among parents and their F_1 for simplicity (**see Supplementary note for detail**).

Scenario 1: Null allele vs one functional allele of one polymorphic site under one genetic background, that is, one of two alleles of one polymorphic site of the receptor is loss-function and the other allele can be bound by one ligand as the background of the receptor (**Supplementary Fig**

32).

For the positive regulation, when the activator as the background is insufficient (smaller X/K) for the functional allele of the receptor, the receptor will express as positive (partial-)dominance. In contrast, when the background is sufficient (larger X/K), the receptor will express as additive effect (**Fig 4a and Supplementary Fig 33a-b**). Apparently, it is the insufficient ligand background that can only activate partial function of two homo-alleles in parents, but relatively full function of one allele in the F_1, which results in the positive (partial-)dominance. For the negative regulation, the performance is similar, but the receptor expresses as negative (partial-)dominance, when the insufficient ligand background can only suppress partial function of two homo-alleles in parents, instead relatively full function of one allele in the F_1 (**Supplementary Fig 33c-d**). It is common between positive and negative regulations that the reaction is dramatically more sensitive to the ligand (activator or repressor) concentration change under insufficient ligand background, where the (partial-)dominance is easy to be observed. It should be noted that there is no overdominance for this scenario if no synergistic effect were involved (when n is equals to 1).

Scenario 2: Two alleles of one polymorphic site under two independent backgrounds, that is, two alleles of one polymorphic site of the receptor can be bound by two respective and independent ligands as the backgrounds of the receptor (**Supplementary Fig 34-35**).

For the positive regulation, as expected in Scenario 1, the receptor easily appears positive dominance when the activator background for the allele with larger maximum function of the receptor is insufficient (smaller X/K). As different from Scenario 1, we can also observe positive overdominance under Scenario 2, when the receptor in F_1 can cumulate the effect from the (partial-)dominant allele with a larger function and that from the other allele with a smaller function. When both backgrounds of two alleles are sufficient (higher X/K), the receptor in both parents and F_1 can express the full function as two alleles and one allele, respectively, as a result the receptor expresses as additive (**Fig 4b and Supplementary Fig 36**). The performances of negative regulation are similar, but the receptor expresses as negative (partial-)dominance or overdominance under insufficient background (**Supplementary Fig 37**). It is common between positive and negative regulations that the reaction is dramatically more sensitive to the ligand (activator or repressor) concentration change under insufficient ligand background (smaller X/K), where the non-additive effect is easy to be observed.

Scenario 3: Two alleles of one polymorphic site with shared background, that is, two alleles of one polymorphic site of the receptor can be bound by the same ligand as the background of the receptor. But these two alleles may have different affinities (K) to the ligand and show different maximum functions (μ). Thus we considered two situations: (1) One allele has higher affinity and shows a larger maximum function, and the other has lower affinity and shows a smaller maximum function (abbreviated as HALF/LASF) (**Fig 4c and Supplementary Fig 38-39**); (2) One allele has higher affinity but shows smaller a maximum function, and the other has lower affinity but shows a larger maximum function (abbreviated as HASF/LALF) (**Fig 4d and Supplementary Fig 40**) . Before considering the above two situations, we found from the simulation that there is only additive effect if the ligand randomly and equally binds to two alleles (**see Supplementary note**). In spite of the positive or negative regulations, function and affinity are similar to scenario 2 that the reaction tends to appears non-additive under insufficient ligand background, especially for the allele with a larger maximum function. The insufficient ligand background renders the reaction dramatically more sensitive to the ligand (activator or repressor) concentration, compared to the sufficient ligand background (**Supplementary Fig 38-42**). But we can only observe the non-additive effect, when the background is dramatically insufficient under HALF/LASF. In addition, the degree of non-additive effect is apparent weaker in the HALF/LASF situation, compared to the HASF/LALF, because in the latter situation the background in F_1 can be reallocated to the allele with LALF from the allele with HASF when the background for latter has been saturated. Taken together, we suppose that overdominance results from the cumulation or compensation between the (partial-)dominance of the allele with a larger function and the effect of the other allele with a smaller function.

According to the above simulations, we put forward one model that explains a unique and important molecular mechanism underlying the non-additive effects and heterosis: homo-insufficiency under insufficient background (HoIIB) (**Fig 4e**). As indicated by HoIIB, it is the genetic background insufficient to maximize the function of two homo-alleles in parents but relatively or even completely sufficient to maximize the function of one-allele in F_1, thus resulting in the insufficient function of two homo-alleles in parents but the relatively or completely sufficient function of one-allele in F_1, that renders the target locus non-additive in effect, as contributing to heterosis. And there were three main features of this theoretic model, according to the simulation.

First, the background insufficiency of the allele with a larger function is the driving force for non-additive effects. What we see dominance and heterosis is not the consequence of a stronger F_1 hybrid, but the consequence of the lower down of the parent with homo-allele of larger function. In other word, the observable function of two homo-alleles is lower than their maximum function under insufficient background. Second, if there is no synergy (n = 1), the overdominance can only be found when both alleles are functional, which result from the cumulation or complementation between the (partial-)dominance of the allele with a larger function and the effect of the other allele with a smaller function. Third, we observed one universal phenomenon in the three scenarios mentioned above, that is, the reaction is dramatically more sensitive to the ligand (activator or repressor) concentration under insufficient ligand background, where the non-additive effect is easy to be observed.

The HoIIB model was supported by different levels of evidence

It is intrigue that in the observed experiments we have found extensive evidence that can represent the three features of the HoIIB model mentioned above. First, we observed the Homo-insufficiency of the allele with a large function and the cumulation or complementation from the allele with a smaller function at various levels including transcription, QTL, and traits (**Fig 5a-c**). Using transcriptome profile from the 1, 2, and 3 mm young panicles of 9311, PA64S, and their hybrids (LYP9), we investigated expression levels in the two parents for those genes with additive, positive dominant, and positive overdominant effects, respectively. The homo-insufficient expression was substantially observed in the higher parent for genes with dominant and overdominant transcription, compared to those with additive transcription (**Fig 5a; Supplementary Fig 43**). Meanwhile, the homozygous genotypes in lower parent showed increased expression for the positive overdominance in most cases (**Fig 5a; Supplementary Fig 43**). Then we compared the QTL with different types of genetic effects that were identified by our GWAS on the three main yield components (PNP, SPP and KGW). Apparently, the homozygous genotypes with lager effects of the dominant and overdominant QTLs represented decreased phenotype, compared to those of the additive QTLs (**Fig 5b; Supplementary Fig 44-46**). We also compared the QTL that were identified by the 278 immortal F_2 lines from the crosses between randomly selected RILs derived

from Minghui 63 and Zhenshan 97[36]. The four yield traits, which were investigated in 1998 and 1999, all indicated apparent HoIIB phenomenon for the dominant and overdominant QTLs, that is, the genotype with higher effect for dominant and overdominant QTLs represented decreased effect, compared to the additive QTLs (**Supplementary Fig 47**). We further investigated the distribution of the degrees of middle-parent heterosis for the five yield traits (SPP, PBP, SBP, PNP and KGW) among the MCC combinations evaluated under the two environments. The stronger heterosis tended to be found among the combinations whose higher parents show decreased phenotypes (**Fig 5c; Supplementary Fig 48**).

Second, under the HoIIB model, we may expect that the expression or the observable function of those genes with stronger heterosis are subject to more serious homo-insufficiency background and thus will show stronger response to the change of background, compared to those with weaker heterosis. The instability of the genes with (over-) dominant effects was reflected by their higher variance of expression levels across the three tissues, compared to the genes with additive effects (**Supplementary Fig 49**). Examining variance of the QTLs identified by the MCC GWAS or by the immortalized F_2 mapping panel, we observed that both homozygous and heterozygous genotypes of QTLs with (over-) dominant effects exhibited higher variability, compared to those with additive effects (**Fig. 5d**; **Supplementary Fig 50-51**)[33]. The combinations with higher degree of dominance also showed higher variability for most of the traits (**Supplementary Fig 52**).

The HoIIB model was experimentally validated in yeast

To verify the HoIIB model, we designed an experiment to see whether we can manipulate the performance of heterosis of one gene by changing its background sufficiency within a living organism. In order to reduce the experimental complexity as possible as, we used the transcription level as the performance indicator of the target gene and the transcription factor as its background, and carried out the experiment in the simple organism, yeast. We screened the reported transcription factors and its target genes in yeast according to the following criteria: (1) the promoter region being bound by a transcription factor has been clearly validated; (2) there is strong and simple regulatory relationship between the transcription factor and its target gene. After investigating the co-expression of six pairs of genes (*WAR1* vs *PDR12*, *VHR1* vs *VHT1*, *VHR1* vs *BIO5*, *AZF1* vs

CLN3, *AFT1* vs *FIT3* and *FZF1* vs *SSU1*), we found that *SSU1* showed a strong co-expression with its transcription factor *FZF1* in strain BY4743 of *Saccharomyces cerevisiae* (R^2 = 0.88, **Supplementary Table 10b**). So we selected *FZF1* and its target gene *SSU1*. According to the reported binding features between two genes[50], we knocked out the *FZF1* recognition motif in *SSU1* promoter region, then constructed the heterozygous (SSU1/ssu1) and homozygous (ssu1/ssu1) knockout strain of *SSU1* in BY4743 (**Supplementary Fig 53a-c**). The ssu1/ssu1 genotype showed apparently decreased expression compared to wild genotype of *SSU1* (SSU1/SSU1), indicating the effective mutation. Gene *SSU1* indicated overdominant expression in the system comprising genotypes SSU1/SSU1, SSU1/ssu1 and ssu1/ssu1, implying that *FZF1* supply the insufficient background to *SSU1* in BY4743, and we may expect that we can decrease the dominance degree of *SSU1* if we can regulate up the expression of its background *FZF1* (**Fig 5e and Supplementary Table 10c)**. In the strains with overexpressed *FZF1*, we really observed dramatically decreased dominance degree among genotypes SSU1/SSU1, SSU1/ssu1 and ssu1/ssu1 of *SSU1* along with the increasing of expression level of *FZF1*, and *SSU1* even nearly transited into additive expression when the expression of *FZF1* upregulated more than 10 folds (**Fig 5e-f**). The results can be confirmed by a repeat experiment **(Supplementary Fig 53d-e and Supplementary Table 10d)**. Thus, our experiment clearly indicated that the dominance degree of downstream genes can be manipulated by changing the level of background sufficiency.

The systematic HoIIB phenomenon related to rice yield heterosis

The model and the results mentioned above revealed that insufficient background contributing to the homo-insufficiency is not only the limiting factor for (over-) dominant loci to reach their maximum function, but also the one that causes the instability of the target genes. Therefore, identification of (over-) dominant loci will provide us with a start point or hint to discover the key limiting factors along the genome, or gene regulatory network that impacts such important traits as yield, and thus guide the improvement of hybrids.

We performed candidate genes analysis for the identified SPP related trait QTLs, combining GWAS, transcription analysis. The candidate genes selected within the overdominant QTLs affecting SPP related traits include: (1) previously reported genes that controls SPP related traits;

(2) highly expressed genes in inflorescence less than 4 mm (http://ricexpro.dna.affrc.go.jp/)[51]; (3) genes with significantly differential expression between the two parents, and with expression levels in the hybrids significantly deviated from the middle expression level of the two parents, based on the transcriptome profiles of < 4 mm inflorescences from two combinations of PA64S×9311 and JBY×ZH100 (**Methods**). In total, we identified 3,983 candidate genes out of the 6,906 annotated genes relevant in F_1_QTLs and Hmp_QTLs for SPP and the related traits (**Supplementary Table 11**). Among the candidate genes, 33 genes were the cloned genes related to rice spikelet or grain number, 2,414 genes expressed in non-additive pattern in the two combinations, 1,678 genes expressed highly in inflorescence, and 32 genes exhibited as panicle specific expression pattern, which including the cloned genes of *OSH1*, *OSH3* and *FZP* (**Supplementary Table 11**). Among the cloned genes, *OSH1*, which acts as key regulatory factor in SAM development[53], exhibits the negative overdominant (NOD) expression pattern in the hybrids of JBY×ZH100, and associates with both F_1 and Hmp in I×9311 and I×Nip (**Supplementary Fig 54a-c**). *d35*, which regulates the panicle size as showed in two reports[54,55], encodes the gibberellin biosynthesis enzyme and acts as negative overdominance in J×Nip in both Changsha and Sanya. Further examination of non-synonymous SNP T/C in two-line and three-line hybrid combinations revealed that the homozygote of inferior allele C has been avoided in majority of commercial hybrids (**Supplementary Fig 54d-g**). Genes, such as *OsGLU1*, *FZP*, *ONAC106, OsGRF1, TGW6* and *XIAO,* were frequently identified as positive overdominance (**Supplementary Table 11-12**).

In order to investigate the possible systematic HoIIB factors impacting rice yield heterosis, we firstly compared the MCC GWAS QTLs identified from different combinations and environments, as well as the F_1 QTLs identified in 1086 three line hybrids[37], followed by gene set enrichment analysis using the candidate genes repeatedly identified by GWAS (**Supplementary Fig 55**). Results indicated that the Nipponbare combinations have apparently more colocalized non-additive QTLs than did the 9311 combinations, consistent with the fact that Nipponbare is less productive than 9311 and suggesting that Nipponbare may represent a more constrained background and thus easily result in non-additive effect in its F_1 hybrids compared to 9311. Regarding different subspecific combinations, negative overdominant QTLs were identified more frequently in *indica* combinations for traits related to SPP and PNP but in *japonica* ones for KGW; however, negative dominant and positive non-additive QTLs tended to be detected in *japonica* combinations for all

traits. Regarding different environments, the colocalized non-additive QTLs tended to be detected in Sanya compared to Changsha (**Supplementary Table 13**). These results indicated that the HoIIB appeared to be taxa- and environment-systematic to some degree, but mainly determined by two specific parents in the combination investigated. Secondly, the GO enrichment indicated that those genes within additive QTLs seldom show enrichment, but those genes within non-additive (dominant and overdominant) QTLs are frequently involved in many kinds of catalytic activities and binding functions (**Supplementary Fig 56, Supplementary Table 14**). Compared to those genes with non-additive performance in non-lethal deletion yeast strains grown in five different media[52], we also found that they enriched in the GO terms of catalytic activity (**Supplementary Fig 57**). The enrichment in catalytic activity for non-additive genes may be explained by the reports that most enzymes in organism usually operate at unsaturated substrate concentration[53,54], i.e. at the lower level of substrates, which may result in the insufficient background of these enzymes and thus their non-additive performance. Further checking those genes encoding rate-limiting enzymes (RLE) showed that the proportion of RLE genes in non-additive QTLs was generally higher than that in additive QTLs (**Supplementary Fig 58**), consistent with the fact that most of the RLE enzyme contain a relative large Kcat values relative to the physiological concentration of substrate, it normally not saturated with substrate and its activity will vary as the concentration of substrate varies [55]. These results suggested that the background/substrate of RLE and genes with additive effect may be the kind of important limiting factors that confer the systematically enriched non-additive catalytic activity in the pathway of these RLE genes. Thus, identifying and improving the upstream and downstream of these factors may provide the chance to make breakthroughs in future breeding of both inbreds and hybrids.

Discussion

HoIIB - a novel model revealed the universal molecular mechanism underlying heterosis of single polymorphic locus

Utilization of heterosis has been a revolutionary technology in plant and animal breeding for a century. Regarding genetic basis of heterosis, three common hypotheses, including dominance,

overdominance, and epistasis, are well noted. Nevertheless, the understanding of molecular mechanistic underlying heterosis is still limited. The following observations pushed us to rethink the possible mechanisms.

First, it has been recognized early that complex traits, such as grain yield in rice, often display observable degrees of heterosis[11,33]. It is essential to distinguish concepts of hybrid vigor and heterozygote advantage. As mentioned in the introduction, many lines of evidence have confirmed that hybrid vigor can be easily achieved by cumulation or complementation of a series of balanced and hierarchical additive factors. For instance, a recent study in rice indicated that additive multiplication of components traits are proven to be one source of heterosis in complex trait[34]. Similarly, yield heterosis in barley was predicted by the products of yield components, including ears/plant, kernels/ear, and average kernel weight[56]. Heterosis for plant height in snap bean could be well defined by multiplying internode length by internode number[30].

However, physiologic constrains often impact the total capacity of a biological system. When component traits are put together to form a more complicated trait, such as grain yield in rice, it is constrained by the "short" component(s), which display considerable difference when compared to their *per se* performance. Therefore, the degree of heterosis may be reduced in such traits as yield. Regarding the relationships between yield heterosis and multiplicative effect of the yield components, our results indicated that, if no dominant effect exist, the geometric multiplication of additive component traits did not contribute much to yield heterosis in rice (**Supplementary Fig 26**).

Second, non-additive genetic effects have been widely observed in F_1 hybrids of many plant species. In maize, dominance seems to be the main factors contributing to grain yield and its components, with a moderate role of overdominance and possible epistatic effect[57,58]. In *Arabdopsis*, epistasis was identified as the main factor contributing to heterosis of seven growth related traits[31]. Similarly, recent studies indicated that dominance and overdominance were the main factors affecting heterosis of flowering date, rosette diameter and rosette[15 15]. In rice, as summarized in the introduction, yield heterosis can result from any of the component traits governed by dominance, overdominance, or epistasis. Evidence from our present study indicated that dominance, dominance cumulation or complementation, and overdominance are the main genetic basis of heterosis of rice yield and its component traits. As high as 40% of the overdominant QTLs

may be interpreted by dominance repulsive linkage, including dominant cumulation and complementation (also refereed to as pseudo-overdominance). At the transcription level, we found that there were nearly 50% of the overdominant expression genes could be explained by complementary expression, which is thought as a common phenomenon[59,60]. As our observation and many reports suggested, dominant cumulation and complementation between loci may be the major genetic basis affecting rice yield heterosis (**Supplementary Fig 59-62**). But it should be noted that the dominant cumulation and complementation hypotheses were always challenged by one question, that is, why does heterosis not significantly decrease after pyramiding of superior alleles at dominant loci? Although this can be partially answered from the statistics point of view[1,61], we believe that our HoIIB model can give more persuasive answer as discussed below.

Despite the fact that multiplicative, cumulative, and complementary effects between two genes is the bases of heterosis, it is rarely reported that hybrid vigor can result from the heterozygote advantage at a single gene locus. Even in these few examples, most of them indicated that the gene is functionally pleiotropic, and at least one of its functions represents non-additive and the degree of non-additive usually varies by background[25,27]. For instance, *SFT* is a typical example that one single gene displays apparent overdominant effect on yield. This gene shows dominant effect on the lateral branching, but nearly additive effect on the sympodial shoot growth, resulting in overdominant effect on whole plant fruit yield in F_1 through the multiplicative effects between the lateral branching and the sympodial shoot growth. It was pointed out that the heterosis of *SFT* gets weaker under the background with functional *SP*[25,62]. To some extent, most of the studies on a single locus tend to explain how the non-additive model works at different levels, rather than why it works. In the present study, as evidenced by the results from yield heterosis of many rice hybrids, QTL mapping, and transcriptome profiling, as well as from the theoretical kinetic simulation, we propose one novel universal molecular mechanism underlying heterosis of single polymorphic locus, the homo-insufficiency under insufficient background (HoIIB).

The HoIIB model suggests that the non-additive effect is not the intrinsic feature of the gene under study, and heterosis is not the heterozygote advantage. Instead, the non-additive effect is a phenomenon that two alleles of the homozygote show insufficiency in function under the insufficient background, but under which one allele of the heterozygote shows relative sufficiency in function. Certainly, the heterozygote can even be overdominant, when the functions of its two

alleles can be accumulated, multiplied, or complementary. Under the HoIIB model, we can easily understand why one QTL appears non-additive in one combination, but additive in another, because different combinations can provide the gene with different genetic backgrounds. We also can explain why the non-additive QTLs are unstable across combinations and over environments, because the non-additive QTLs are under an insufficient background and thus easily subject to the changes of background. In the present study, we have found extensive evidence that supports the HoIIB model, at the levels of transcription, QTL, and phenotypes from several designed experiments. First, we observed apparent decrease in function for the homozygote of the allele with larger function under the situation of non-additive effect, compared to the situation of additive effect. Second, we observed accumulation or complementation between the allele with a smaller function that with a larger function under the situation of overdominance. Third, the stronger response of a gene showing non-additive expression to its transcription factors (TF), compared to that showing additive expression, suggested greater impact of the insufficient genetic background exerted by the TF to the expression of its target gene.

Furthermore, evidence from our repeatedly detected non-additive genes, followed by GO enrichment analyses, indicated that enzyme catalytic activity may be a systematic HoIIB phenomenon that causes the non-additive effect, i.e. heterosis (**Supplementary Fig 56-57**)[63]. The result can be interpreted by many reports that substrate is usually under-sufficient in enzyme-catalyzed reactions, that is, most enzymes work at substrates at concentrations below saturation[53,64]. As indicated by our simulations, we can easily understand how heterosis is produced under the HoIIB model. In fact, all life phenomena and biological process are series of biochemical reactions, which can be explained by the Hill reaction. Thus we suspect that the HoIIB model is widely applicable to different biological processes and traits.

The HoIIB model can interpret most known mechanisms, models and phenomena about heterosis

To investigate generality of the HoIIB, we compared it with several known mechanisms, models, and phenomena about heterosis. The nonlinearity of the enzyme catalytic system was frequently described to explain heterosis[65,66]. But as our simulations indicated, nonlinearity is not

656 the absolute feature of enzyme catalytic activity, it only occurs when the substrate is insufficient to
657 support the full function of the enzyme. In fact, the gene related to the enzyme can also be linear or
658 additive when the substrate concentration is sufficient. BÄurger and Bagheri insisted that the
659 dominance can be evolutionally modified, after comparisons of the models proposed by Wright and
660 Kacser-Burns, respectively[67,68], they pointed out that the output gain curve will change from non-
661 linear to relative linear, and the dominance will transit to additive effect, if one mutation results in
662 a decrease in Kcat that leads to a lower saturation level (i.e. a status that substrate saturates the
663 enzyme at relative lower concentration level)[67]. For instance, genes with dl binding site are
664 activated or repressed by dl at low threshold levels when dl has a low Kcat[69]. If *dl* null mutations
665 possess higher Kcat, female flies heterozygous for *dl* null allele will expressed as dominant[70]. Of
666 course, according to our HoIIB model, the additive effect also can transit to dominance or
667 overdominance, when the background changes from sufficiency to insufficiency. In fact, sufficiency
668 and insufficiency are relative and dynamic. The transition from insufficiency to sufficiency for one
669 factor may cause new insufficiency for its counterpart factor. This may explain the challenge to
670 dominance and epistasis hypotheses, that is, why does heterosis not decrease along with the
671 pyramiding of superior alleles[61]

672 The balance between genes involved in a biological complex system is another important
673 hypothesis about heterosis. This hypothesis suggests that an imbalance in the concentration of the
674 subcomponents of a protein–protein complex / pathway / network can be deleterious. The typical
675 example for gene balance was reported by Balazs and coleagues[52,71]. Their studies indicated that
676 mutation of the subunit in a complex (or the factor in an interaction pair) can result in imbalance
677 and thus is harmful, which might indicate the impact of gene imbalance on dominance. However,
678 these studies did not consider the effects from the counterpart background. Thus we simulated the
679 effects of complex background on dominance. Our results indicated that the non-additive effect will
680 become weaker and even loss, when background of the considered subunit gets sufficient
681 (**Supplementary Fig 63a-c**). These lines of evidence indicated that heterosis is the result of low
682 function of homozygote (homo-insufficiency), rather than the heterozygote advantage.

683 There are plenty of other examples that indicate the dependency of heterosis on genetic
684 background and can be explained by the HoIIB. First, in the comparisons of functional categories
685 of enzymes, binding proteins, and transcription regulators, the proportion of haplosufficient genes

(i.e. dominant genes) is the highest among genes that encode proteins with enzymatic functions[63], which is highly consistent with the fact that most of the enzymes work in low saturation levels, due to insufficient substrate. Second, an increased gene dose or gene mutations lead to an enhanced affinity or function, the metabolic background often cannot synchronize with the target gene, which results in a more insufficient state of the background. A common observation is that, the increased dose or function is not harmful, while its potential is severely unrealized in parent, due to the insufficient background. Thus heterozygote is often observed as dominant[72,73]. Third, a decreased dosage or function often leads the background to a more sufficient state, compared to the original state, which frequently results in dosage sensitive phenomenon, including additive and haploinsufficient[69,70]. Fourth, regarding protein complexes or pathways composed of multiple factors, there is a stoichiometric equilibrium between the factors. Altering a member or subunit will affect assembly of the complexes that impacts functions. In such a system, insufficient or over-supplied bridging factors may cause relative insufficiency of the background for one of the factors, which theoretically increases the possibility of dominance or overdominance[29,74] (**Supplementary Fig 63d-e**). In summary, our HoIIB is a fundamental model and can interpret most models, hypotheses, and phenomena about heterosis. Of course, it is certain that there might be other situations beyond the scope of the HoIIB model, for instance, heterozygotes may have new functions or toxic protein alterations.

Implication of the HoIIB model to genetic improvement of hybrid rice

The HoIIB model may affect future utilization of heterosis in several aspects. First, our HoIIB model indicated that in most cases heterosis is not the consequence of heterozygote advantage, but homozygote disadvantage under insufficient background. This implies that current utilization of heterosis is not the best way to take advantage of maximum function of target genes[37]. Therefore, we need to identify and improve the constrained factor(s), or the target genes. In the present study, we extensively investigated yield QTLs or genes affecting parental lines, hybrids, and heterosis, followed by dissection of their genetic effect (**Supplementary Fig 24 and 55**). This may provide us with references to identify the limiting factors. In theory, the frequently detected additive factors may represent the systematic limiting factors that constrain the dominant or even overdominant

714 factors from maximizing their functions. When comparing the frequently detected additive vs
715 overdominant QTLs affecting yield traits, we observed that the candidate genes within the additive
716 QTLs displayed distinctly lower expression, compared to those within the overdominant ones
717 **(Supplementary Fig 64).** Of course, lower expression just represents one aspect of the insufficient
718 function of the genes, we may expect to observe the other aspects of insufficient functions, such as
719 enzyme activity and affinity. These results implied that the frequently detected additive factors,
720 rather than the dominant or overdominant factors, should be focused in future breeding programs.
721 For genes involved in a complex where the gene balance theory applies, we can take advantage of
722 their maximum functions in homozygote by improving the background rather than their
723 compromised functions in heterozygote in order to fit the insufficient background. Theoretically,
724 we can easily make use of homozygote that can maximize the functions of the target genes of
725 interest, which can be achieved by the improvement of the corresponding factors as the insufficient
726 genetic background of target genes.

727 Second, although the above discussion may illude us to think that hybrid breeding is not
728 necessary, our point is that utilization of heterosis will still be an important breeding strategy for a
729 long time and even forever. First, from the perspective of favorable alleles accumulation, even
730 though all insufficient factors can be improved to their maximum functions, followed by integration
731 into an inbred line in theory, it is impossible to be realized in a short time (**Supplementary Fig 65**).
732 Second, it is an extremely long process to construct the regulation network and thus clearly
733 understand the mutual dependency between genes. Third, the mechanism of HoIIB implies that
734 non-additive is common phenomenon in life system. The reason is simple, that is, it is not expected
735 that the factors in a system operate on the exactly required dependency each other. Thus, one most
736 insufficient factor will result in a batch of factors that present different degrees of homo-
737 insufficiency. So we may expect to find less additive factors than non-additive ones, including
738 partial-dominance, dominance and over-dominance. This is really consistent to our results. We
739 detected distinctly less additive QTLs (about 19%) than non-additive ones (including partial
740 domiance), and less genes with additive expression pattern (about 13%) than non-additive ones.
741 Fourth, genetic improvement is a dynamic process, involving the alleviation of insufficiency for
742 one factor, followed by induction of insufficiency for the counterpart factor, that is, breakdown of
743 old balance along with establishment of new unbalance, plus the background change under different

environments. For example, it has been proved that improvement of corn hybrids is mainly attributed to the improvement of their inbred parental lines. The high performance of inbreds did not decrease the degree of heterosis in hybrid corn breeding[1]. This phenomenon has also been observed in other organisms, such as cotton[61]. These results indicated that the dynamic breeding process contributes substantially for the continuous improvement of both inbreds and hybrids. Our results also indicated that we may consider different aspects, when we try to improve a variety. For example, we need to overcome the weakness of heterosis for SPP related traits under short-day environment (**Fig 1d**).

Third, the HoIIB model helps our understanding and utilization of general combining ability (GCA) and special combining ability (SCA), and provides guidance in breeding by genome selection. It was well known that additive effects contribute mainly to GCA, and non-additive effects, including dominance and epistasis, to SCA[75]. Our HoIIB model suggested that the additive factors are less background-sensitive, compared to the non-additive factors, which explain why SCA is more difficult to predict than GCA does. In addition, improvement of GCA through accumulation of additive superior alleles has proven to be an efficient strategy in hybrid breeding[76]. According the HoIIB model, the accumulation of additive superior alleles can definitely improve the genome background and thus release the potential functions of those limited factors. Our current study may suggest one possible and efficient strategy, in order to make breakthrough in hybrid rice breeding: (1) keeping on accumulation of superior alleles of the frequently identified additive factors, and try to improve them through both traditional and biotechnological methods, in order to continuously improve the genome background; (2) incorporating more subtle background effect into the model of genome selection in breeding for hybrids, in order to improve the prediction accuracy of special combining ability.

Author contributions

H.Z. conceived the project and its components. J.X. and H.Z. designed the studies and contributed to the original concept of the project. W.W. contributed to the phenotyping of the hybrid rice. Q.Z. and T.Y. contributed yeast gene related experiments. Z.Z., X.Z., N.L. and L.Z. performed the genome sequencing, X.M., S.Z., Y.L., X.W. and F.L. performed GWAS and QTL effect analysis,

Y.Z., X.J., J.Z., N.J. and G.L. contributed transcriptome data collection and analysis, Z.Z., J.L., Z.Z. and Z.L. contributed supervision, validation and visualization. J.X. and H.Z. analyzed the whole data and wrote the paper.

Acknowledgements

We are indebted to Lihuang Zhu from the Chinese Academy of Sciences for his generous gifts of the raw data of transcriptome for LYP9 combination, and to Kai Wang from China National Seed Group Company for his kindly supply of LYP9 seeds. We thank Zhongfu Ni, Zuofeng Zhu, Hai Wang and Xiangfeng Wang (China Agricultural University), Weibo Xie (Huazhong Agricultural University), and Renhua Li (National Institute of Health, USA), for their important and critical discussions. We also thank Huiqiang Lou and Lili Li (China Agricultural University) for kindly providing vectors and valuable suggestions in the design of yeast-related experiment. This work was supported by grants from the National Key Research and Development Program (2016YFD0100803), the National Natural Science Foundation (31171613), and the National Key Technology Support Program (2015BAD02B01).

Competing interests

The authors declare no competing interests.

Methods

Parental varieties and their F_1 population construction

We used 265 world-wide varieties from the mini-core collection (MCC) of cultivated rice as the parents to construct the F_1 population[40]. The F_1 population was constructed by the crossing between two testers (temperate *japonica* variety Nipponbare and *indica* variety 9311) as female parent and the varieties in MCC as male parent. It took us five seasons to generate

455 combinations by hand emasculation, and to exclude the false crossing, we documented the false hybrids by comparing the phenotype differences between hybrids and the corresponding female parent and further surveying the phenotypic segregation in the F_2 population of each combination. Finally, we used 418 combinations with at least 100 F_1 seeds for each combination in this study **(Supplementary Table2)**.

Resequencing and genotyping for parental varieties and their F_1 population

The parental varieties were re-sequenced as part of in the rice 3,000 rice genomes project[41]. Genomic DNA was prepared from the leaves of a single young plant for each variety by a modified CTAB method. After the quality control, at least 3 μg genomic DNA of each sample was randomly fragmented by sonication and size-fractionated by electrophoresis, and DNA fragments of approximately 500 bp were purified. Each sequencing library was sequenced in six or more lanes on the HiSeq2000 platform and 90 bp paired-end reads were generated. Subsequently, the reads from each sample were extracted based on their unique nucleotide multiplex identifiers as 83 bp reads (90 - 6 - 1, where 1 is the ligation base "T"). To ensure high quality, raw data was filtered by deleting reads having adapter contamination or containing more than 50% low quality bases (quality value ≤ 5).

The 83-bp paired-end reads of 267 rice varieties were mapped to the temperate *japonica* Nipponbare reference genome (IRGSP-1.0) using the BWA software with default parameters except for "aln -m 10000 -o l -e 10 -t 4". The alignment results were then merged and indexed as BAM files[77]. SNP calling was based on the alignment using the Genome Analysis Toolkit 2.0-35(GATK) and Picard packages V1.71[78]. To minimize the number of mismatched bases for SNP and InDel calling, all reads from each accessions were further cleaned by (i) deleting the reads that unmapped to the reference in the alignment result, (ii) deleting duplicate reads, (iii) conducting alignment by the IndelRealigner package in GATK and (iv) recalibrating realignment using the BaseRecalibrator package in GATK.

SNP and InDel calling for each sample were conducted independently using the UnifiedGenotyper package in GATK with a minimum phred-scaled confidence threshold of 50, and a minimum phred-scaled confidence threshold for emiting variants at 10. SNP and

InDel calling at the population level was performed using the UnifiedGenotyper package in the GATK pipeline with 50 for the minimum phred-scaled confidence threshold for variant calling and 30 for the minimum phredscaled confidence threshold for variant emitting. Genotypes of the 267 rice varieties were called at the SNP sites. For the genotype datasets of all the accessions, SNPs with more than 50 % missing data and SNPs with MAF < 2% were excluded and 4,625,141 high quality SNPs were generated. For the genotype datasets in each subspecies, SNPs with MAF < 2% were excluded and finally 3562187 and 1649161 high quality SNPs for *indica* and *japonica* subspecies were generated respectively.

SNPs in coding regions were called on the basis of the gene models in MSU7 (release 7: http://rice.plantbiology.msu.edu/). The coding SNPs were annotated to be synonymous or non-synonymous, SNPs with large-effect variations were annotated and partitioned as SNPs that introduce stop codons, disrupt stop codons, disrupt initiation codons, or disrupt splice sites.

The F_1 genotypes for each combination were inferred by the genotypes of their parents.

Phenotyping of the parental varieties and their F_1 population

We Planted the 418 F_1 hybrids and their 267 parents in 2013 at respective Changsha (CS) (28°13′N, 112°58′E, a long-day environment) and Sanya (SY) (18°10′N, 109°28′E, a short-day environment) of China. One combined plot including the F_1 and the corresponding parents for each combination was planted with randomized complete block of two replicates in each environment. Each combined plot included five rows consisting of two testers (Nipponbare and 9311), F_1 and MCC parent in sequence. The row and plant distances were 29.5 cm and 16.7 cm respectively, with 10 plants in each row, being wider than the general field production so as to decreasing the interface among plants as much as possible.

The yield related traits were measured in two environments for each combination as following. Five plants in the middle of each row were used to measure six yield traits. The panicle number per plant (PNP) and grain weight per plant (GWP) was the average of all five plants. And we selected the main panicles of five plants to count the spikelet number per panicle (SPP), the secondary branch number per panicle (SBP) and primary branch number per panicle (PBP). The 1,000-grain weight (KGW) was rescaled by the grain weight of 300 grains selected

from five main panicles.

The middle parent heterosis value (Hmp) of each combination for each trait was measured as: F_1–(P_1+P_2)/2, i.e. the deviation of F_1 from middle parent performance, where F_1, P_1 and P_2 represent the phenotypic values of each trait in F_1, P_1 and P_2 respectively. In addition, we denoted the over higher parent heterosis (OHP) when the F_1 shows the phenotypic value over the higher parent, range between the two parents (RBP) when the F_1 shows the phenotypic value between two parents and below lower parent heterosis (BLP) when the F_1 shows the phenotypic value below the lower parent.

Population genetic analysis

The phylogenetic neighbor-joining tree and principal-component analysis were used to infer population structure of the parental panel. A pairwise distance matrix derived from the simple matching distance for all SNP sites was calculated to construct unweighted neighbor joining trees using the software MEGA5.0. According to the neighbor-joining tree of 267 varieties in parental panel and the tree of 3,024 rice varieties[41], we divided *japonica* and *indica* into nine and eight distinct sub-populations respectively **(Supplementary Fig. 3)**.

Estimation of Environment and genotypic variance.

For each variety, there are two environments and each environment has two replicate of phenotype data. Both inbred parents including *japonica* and *indica* subspecies and four types of hybrid combination (J×Nip, J×9311, I×Nip and I×9311) were used. The following linear model was fitted to the transformed data:

$$Y_{ijk} = \mu + G_i + E_j + G_i \times E_j + \varepsilon_{ijk}$$

Here Y_{ij} is the *ij*th phenotypic observation for the i th rice variety under j th environment. k represents two replications. μ is the overall mean, G*i* and E*j* is the genotypic and environmental effect. $G_i \times E_j$ the genotypic and environmental interaction effect, ε_{ijk} is the random residual effects.

Genome-wide association study (GWAS)

GWAS was conducted GAPIT using the compressed MLM[79]. The phenotype includes the trait value of parents in *indica* and *japonica* respectively, the F_1 trait value of four kind of combinations (*Japonica*×Nipponbare, *Japonica*×9311, *Indica*×Nip and *Indica*×9311) respectively and the Hmp of four kind of combinations (Nip×*Japonica*, Nip×*Indica*, 9311×*Japonica*, 9311×*Indica*) respectively. For the compressed MLM analysis, we used the equation[79,80]:

$$y = X\alpha + P\beta + K\mu + e$$

Here, y represents phenotype, X represents genotype matrix, P is the matrix of population structure and K is the matrix of genetic similarity between individuals. α and β represent fixed effects of genotype and population structure, and μ and e represent random effects of kinship and residuals. The first five principal components were used to estimate the population structure. The matrix of genetic similarity based on simple SNP matching coefficients was used to model the variance-covariance matrix of the random effect.

To avoid the over correction of the Bonferroni method, we used FDR to control overall errors as following. Permutation tests were used to estimate the FDR[81]. For each examined trait, we reshuffled the original phenotype data, and then performed association analysis using GAPIT with the same parameters. After 1000 permutations, we got 1000 association p value from permutation (p_per) for each SNP and we set the highest –log(p_per) as the FDR of that SNP. The SNP was denoted as significant association when the –log(p_GWAS) is no less than the highest –log(p_per), where p_GWAS is the association p value for each SNP for original phenotype data.

Estimation of additive and dominant effects for each significant SNP locus and QTL

Firstly, we estimated the additive and dominant effects for each significant SNP locus. We defined the tester's genotype (Nipponbare or 9311) of each SNP as A, The non-tester's genotype as B. When some varieties in the MCC parental panel show as A and the others show

as B, their F_1 will show the genotypes A and H (the heterozygous genotype). For the investigated trait, we set P_A as the mean phenotype of parents with genotype A, P_B as that of parents with genotype B, F_A as that of F_1 with genotype A and F_H as that of F_1 with genotype H. The additive effect of each SNP was half of the absolute difference between the two homozygotes, i.e. $a = abs(P_A–(P_A+P_B)/2)$. The traditional estimation for dominance effect was expressed as $d = F_H-(P_A+P_B)/2$, here we rescaled the dominance effects identified as $d = F_H-(F_A-P_A)-(P_A+P_B)/2$. In which $F_A - P_A$ means the background heterozygous effects.

Then, we delimitated those significant SNP with tight LD (linkage disequilibrium) as one QTL and selected the tagSNP of each QTL. We first constructed the LD block using GAB algorithm for all significant SNPs with $r^2 >= 0.8$ and selected out one tagSNP for each block[82]. If one block size was larger than 20kb and there were no less than 3 significant SNPs in the block, we defined the block as one QTL and the QTL was named by the position of its tagSNP. The additive and dominant effects of this QTL were estimated by the average additive and dominant effects of the significant SNPs in the QTL.

Finally, we defined the additive, dominant and over-dominant QTLs. A QTL is referred to as over-dominance preferred if the absolute ratio of dominant effect to additive effect ($|d/a|$, degree of dominance) is no less than 1.5, and (partial-) dominance preferred if $0.5 \leq |d/a| < 1.5$, and additive preferred if $|d/a| < 0.5$. The dominant and over-dominant QTLs can further be classified as positive ones when their $d > 0$ or negative ones when their $d < 0$.

Estimation of repulsive degree in one QTL

The repulsive effects of more than one locus within one QTL may be one mechanism resulting in non-additive effect[36]. So we estimated the repulsive degree in each QTL. If there are n significant SNPs showing the same type of dominant effect (positive or negative) in one QTL, then its maximum pairwise SNPs number is $N = C_n^2$. We denoted the SNP as S_{ref} if phenotype effect of its allele with genotype being same to tester, and S_{alt} otherwise, then the maximum possible pairwise repulsive SNPs number is $R = S_{ref} \times S_{alt}$. Thus, the repulsive degree of the QTL was $RD = R/N$. The RD is larger in one QTL, the higher potential the QTL involves repulsive effect.

Construction of *SSU1* mutants in yeast

In this study, diploid BY4743 was used as the wild type experimental strain. In order to knock out the recognition motif of *FZF* in the promoter of *SSU1* gene, PCR amplification of vectors Pfa6a-Leu1Mx (Leu) and Pfa6a-His3Mx6 (His) were performed using primers (HRR-SSU1-F and HRR-SSU1-R) to obtain recombinant components. The recombinant component was verified by sequencing and then transformed into strain BY4743. After verification of positive clones by electrophoresis and sequencing, the heterozygous mutant strains (SSU1/ssu1) contain single-strand DNA substitution (Leu or His) and diploid mutant strains (ssu1/ssu1) that contain both Leu and His substitution were successfully constructed **(Supplementary Fig 53a-c and Supplementary Table 10a).**

Over expression of *FZF1* gene in yeast

RNA was extracted from BY4743 strain, the coding sequence of *FZF1* was amplified from BY4743 cNDA and cloned into pAG416 vector by recombination methods. The constructed vector along with the empty vector pAG416GAL were transformed into *Saccharomyces cerevisiae* strain BY4743 (referred as SSU1/SSU1), two types of heterozygous mutant strain (Het-Leu and Het-His, referred as SSU1/ssu1) and diploid mutant strain (referred as ssu1/ssu1). The methods related to yeast cultures, transformations and growth assay mainly referred to Gietz et al[83]. Yeast cells were grown at 30°C in synthetic defined (SD) medium (0.67% yeast nitrogen base, Sigma) without amino acids, containing 2% (w/v) glucose or 2% (w/v) galactose (induction medium), supplemented with yeast synthetic dropout without Ura (Clontech, CA, USA), pH 5.8.

RNA extraction and Real-Time Quantitative PCR analysis

Total RNA isolated from fresh yeast cultures and reversed transcribed using a protocol as previously described. RT-qPCR analysis was performed using gene-specific primers listed in **Supplementary Table 10a** in the Supplement Material. *Saccharomyces cerevisiae* 18S RNA was used as reference genes to normalize the data.

Evaluation of heterosis of *SSU1* expression under different *FZF1* expression levels.

We using four strains including WT (referred as SSU1/SSU1), heterozygous mutant (referred as SSU1/ssu1) and diploid mutant (referred as ssu1/ssu1) which transferred the empty vector PAG416GAL as control. *FZF1* overexpression (FZF1-OE) strains were divided into 0-10, 10-20, and >20 groups according to the up-regulation ratio between *FZF1* expression levels in FZF1-OE and empty event, referred as OE (0-10), OE (10-20) and OE (>20) (**Fig 5e)**. For each group, the average expression level of all strains with the same genotype was used as the expression of that genotype (**Supplementary Table 10c**). In another repeat experiment, due to the fewer events of strain with high *FZF1* up-regulation ratio, we grouped them into 0-5,5-10 and >10 groups, referred as OE (0-5), OE (5-10) and OE (>10) (**Supplementary Table 10d; Supplementary Fig 53** d-e). The formula of additive and dominance of expression quantity is as follows:

Additive effect = abs(SSU1/SSU1- ssu1/ssu1)/2

Dominance effect = SSU1/ssu1 – (SSU1/SSU1+ ssu1/ssu1)/2

Degree of dominance d/a = Dominance effect /Aditive effect

Transcriptome in rice hybrid combinations and data analysis

One transcriptome was from the 2 mm young inflorescences of Jinbaoyin (JBY), Zihui-100 (ZH100) and their hybrid. And the total RNA was isolated by using Trizol reagent (Invitrogen). The mRNA sequencing libraries were constructed, and sequencing was performed using the Illumina HiSeq 2500 platform. RNA-seq reads of Jinbaoyin (JBY), Zihui-100 (ZH100) and their Hybrids were mapped to IRGSP1.0 http://rice.plantbiology.msu.edu/pub/data/Eukaryotic_Projects/o_sativa/annotation_dbs/pseudomolecules/version_7.0/all.dir/all.con using TopHat[84] software with parameters: minimum intron length of 20, maximum intron length of 10,000, and a maximum of two mismatches. Only unique mapped reads were extracted for the following analysis. The number of fragments per kilobase of exon model per million mapped reads (FPKM) for each gene was calculated

using Cufflinks[85], and transcripts per million reads (TPM) were finally used to measure the expression level. Differentially expressed genes among two parents and the hybrid were identified using the R package DEGseq[86]. For genes differentially expressed between two parents, when the expression level of a gene in the hybrid was significantly different from that in the low parent but not different from that in the high parent, then the gene was classified as 'high-parent expression' (HP), and if the expression level in the hybrid was significantly different from that in the high parent but not from that in the low parent, then the gene was classified as 'low-parent expression' (LP). If the expression level in the hybrid was significantly higher than that in the high parent, then the gene was classified as 'over higher-parent expression' (OHP). If the expression level in the hybrid was significantly lower than that in the low parent, then the gene was classified as 'below lower-parent expression' (BLP). If the expression level in the hybrid was not significantly different from two parents, then that gene was classified as 'middle-parent expression' (MP).

The original raw data of Jinbaoyin (JBY), Zihui-100 (ZH100) and Hybrids have been deposited in the Genome Sequence Archive (Genomics, Proteomics & Bioinformatics 2017) in National Genomics Data Center, China National Center for Bioinformation / Beijing Institute of Genomics, Chinese Academy of Sciences, under accession number CRA004341 that are publicly accessible at https://bigd.big.ac.cn/gsa.

The other transcriptome was from the 1 mm, 2 mm, 3 mm and 4 mm young panicles respectively in combination of hybrid LYP9 and its parents (9311 and PA64S). The raw RNA-seq data of 9311, PA64S and LYP9 were download from Genome Sequence Archive of Beijing Institute of Genomics, Chinese Academy of Sciences(gsa.big.ac.cn) under accession no PRJCA000131[38]. The quantification of genes expression level was the same to combination of JBY and ZH100. The expression patterns were determined as following. Firstly, the standard deviation of TPM for each gene was estimated according to three replicates in 4 mm young panicles of 9311, PA64S and LYP9, and two replicates in 3 mm young panicles of PA64S and LYP9. Secondly, the expression patterns were determined according to the significant different expression levels among 9311, PA64S, LYP9 and the middle parents at significance level $p = 0.01$. In detail, If LYP9 is significantly higher than the higher parent, the gene was classified as positive over-dominance (POD); if LYP9 is significantly lower than the lower parent, the

gene was classified as negative over-dominance (NOD); if LYP9 is significantly higher than the middle-parent but shows no significance from the higher parent, the gene was classified as positive dominance (PD); if LYP9 is significantly lower than the middle-parent but shows no significance from the lower parent, the gene was classified as negative dominance (ND); if LYP9 is significantly higher than the middle-parent and significantly lower than the higher parent, the gene was classified as positive partial dominance (PPD); if LYP9 is significantly lower than the middle-parent and significantly higher than the lower parent, the gene was classified as negative partial dominance (NPD); if LYP9 is not significantly different from the middle-parent, significantly lower than the higher parent and significantly higher than the lower parent, the gene was classified as middle parent (MP) or additive expression (A).

Arabidopsis transcriptome data analysis

The raw RNA-seq data of Col-0×Per-1, Col-0×Aa-0, Col-0×Ak-1, Col ×C24 and their parents were downloaded according to the information provided by the original literature[15,18]. Subsequent reads alignment, the quantification of genes expression and the identification of expression pattern are followed the method and process as mention above in rice combination JBY×ZH100.

Yeast mutant data analysis

Steinmetz *et al* measured growth rates of strains with precise deletions of each gene in the yeast genome using a parallel molecular bar-coding strategy[87]. We used their data (available at http://www-deletion.stanford.edu/YDPM/YDPM_index.html) for nonlethal gene deletion strains grown in YPD, YPG, YPDGE, YPE, and YPL media. Here, we consider only nonlethal mutations for which homozygous and heterozygous growth rate data are available on the media. For each media, we used the average performance of top 5% as normal wildtype (WT), homozygous deletion of the strain as Homozygous type (Hom), and the deletion gene in heterozygous strain recorded as Het. The additive effect, dominance effect and the degree of dominance was calculated as:

$$\text{Additive effect} = \text{abs}(WT - Hom)/2$$

$$\text{Dominance effect} = \text{Het} - (\text{WT}+\text{Hom})/2$$

$$\text{d/a} = \text{Dominance effect}/\ \text{Additive effect}$$

Here we defined the genes with |d/a| >0.5 as non-additive performance, only the genes identified as |d/a| >0.5 in 4 or more than 4 kinds of media were used to further GO enrichment analysis.

GO enrichment

GO analysis was performed using methods available at agriGO website[88].

Reference

1. Duvick, D.N. Heterosis: Feeding people and protecting natural resources. *Genetics and Exploitation of Heterosis in Crops*, 19-29 (1999).
2. Cheng, S.H., Zhuang, J.Y., Fan, Y.Y., Du, J.H. & Cao, L.Y. Progress in research and development on hybrid rice: A super-domesticate in china. *Annals of Botany* **100**, 959-966 (2007).
3. Darwin, C. *The Effects of Cross and Self Fertilisation in the Vegetable Kingdom*, (Murray, London, 1876).
4. Shull, G.H. The composition of a field of maize. *Journal of Heredity* **4**, 296-301 (1908).
5. Lippman, Z.B. & Zamir, D. Heterosis: revisiting the magic. *Trends in Genetics* **23**, 60-66 (2007).
6. Bruce, A.B. The Mendelian theory of heredity and the augmentation of vigor. *Science* **82**, 627–628 (1910).
7. Jones, D.F. Dominance of linked factors as a means of accounting for heterosis. *Genetics* **2**, 466-479 (1917).
8. Shull, G.H. The genotypes of maize. *American Naturalist* **45**, 234-252 (1911).
9. East, E.M. Heterosis. *Genetics* **21**, 375-397 (1936).
10. Minvielle, F. Dominance Is Not Necessary for Heterosis - a 2-Locus Model. *Genetical Research* **49**, 245-247 (1987).
11. Cockerham, C.C. & Schnell, F.W. Multiplicative vs. Arbitrary Gene Action in Heterosis. *Genetics* **131**, 461-469 (1992).
12. Jiang, Y., Schmidt, R.H., Zhao, Y. & Reif, J.C. A quantitative genetic framework highlights the role of epistatic effects for grain-yield heterosis in bread wheat. *Nat Genet* **49**, 1741-1746 (2017).
13. Xiao, J.H., Li, J.M., Yuan, L.P. & Tanksley, S.D. Dominance Is the Major Genetic-Basis of Heterosis in Rice as Revealed by Qtl Analysis Using Molecular Markers. *Genetics* **140**, 745-754 (1995).
14. Semel, Y. *et al.* Overdominant quantitative trait loci for yield and fitness in tomato. *Proceedings of the National Academy of Sciences of the United States of America* **103**, 12981-12986 (2006).
15. Seymour, D.K. *et al.* Genetic architecture of nonadditive inheritance in Arabidopsis thaliana hybrids. *Proc Natl Acad Sci U S A* **113**, E7317-E7326 (2016).

16. Yang, M. *et al.* Genomic architecture of biomass heterosis in Arabidopsis. *Proc Natl Acad Sci U S A* (2017).
17. Riedelsheimer, C. *et al.* Genomic and metabolic prediction of complex heterotic traits in hybrid maize. *Nature Genetics* **44**, 217-220 (2012).
18. Miller, M., Song, Q., Shi, X., Juenger, T.E. & Chen, Z.J. Natural variation in timing of stress-responsive gene expression predicts heterosis in intraspecific hybrids of Arabidopsis. *Nat Commun* **6**, 7453 (2015).
19. Shao, L. *et al.* Patterns of genome-wide allele-specific expression in hybrid rice and the implications on the genetic basis of heterosis. *Proc Natl Acad Sci U S A* **116**, 5653-5658 (2019).
20. Swanson-Wagner, R.A. *et al.* All possible modes of gene action are observed in a global comparison of gene expression in a maize F-1 hybrid and its inbred parents. *Proceedings of the National Academy of Sciences of the United States of America* **103**, 6805-6810 (2006).
21. Marti-Raga, M., Peltier, E., Mas, A., Beltran, G. & Marullo, P. Genetic Causes of Phenotypic Adaptation to the Second Fermentation of Sparkling Wines in Saccharomyces cerevisiae. *G3 (Bethesda)* **7**, 399-412 (2017).
22. Abifadel, M. *et al.* Mutations in PCSK9 cause autosomal dominant hypercholesterolemia. *Nat Genet* **34**, 154-6 (2003).
23. Fan, C.H. *et al.* GS3, a major QTL for grain length and weight and minor QTL for grain width and thickness in rice, encodes a putative transmembrane protein. *Theoretical and Applied Genetics* **112**, 1164-1171 (2006).
24. Xue, W. *et al.* Natural variation in Ghd7 is an important regulator of heading date and yield potential in rice. *Nat Genet* **40**, 761-7 (2008).
25. Krieger, U., Lippman, Z.B. & Zamir, D. The flowering gene SINGLE FLOWER TRUSS drives heterosis for yield in tomato. *Nature Genetics* **42**, 459-U138 (2010).
26. Singh, R. *et al.* The oil palm SHELL gene controls oil yield and encodes a homologue of SEEDSTICK. *Nature* **500**, 340-4 (2013).
27. LaFountain, A.M. *et al.* Molecular Basis of Overdominance at a Flower Color Locus. *G3-Genes Genomes Genetics* **7**, 3947-3954 (2017).
28. Allison, A.C. Protection afforded by sickle-cell trait against subtertian malareal infection. *Br Med J* **1**, 290-4 (1954).
29. Veitia, R.A., Bottani, S. & Birchler, J.A. Cellular reactions to gene dosage imbalance: genomic, transcriptomic and proteomic effects. *Trends in Genetics* **24**, 390-397 (2008).
30. Coyne, D.P. Component interaction in relation to heterosis for plant height in Phaseolus vulgaris L. variety crosses. *Crop Sci.* **5**, 17-18 (1965).
31. Melchinger, A.E. *et al.* Genetic basis of heterosis for growth-related traits in Arabidopsis investigated by testcross progenies of near-isogenic lines reveals a significant role of epistasis. *Genetics* **177**, 1827-37 (2007).
32. Tirosh, I., Reikhav, S., Levy, A.A. & Barkai, N. A Yeast Hybrid Provides Insight into the Evolution of Gene Expression Regulation. *Science* **324**, 659-662 (2009).
33. Richey, F.D. Mock-Dominance and Hybrid Vigor. *Science* **96**, 280-1 (1942).
34. Dan, Z. *et al.* Hierarchical additive effects on heterosis in rice (Oryza sativa L.). *Front Plant Sci* **6**, 738 (2015).
35. Li, Z.K. *et al.* Overdominant epistatic loci are the primary genetic basis of inbreeding depression and heterosis in rice. I. Biomass and grain yield. *Genetics* **158**, 1737-1753 (2001).

36. Zhou, G. *et al.* Genetic composition of yield heterosis in an elite rice hybrid. *Proceedings of the National Academy of Sciences of the United States of America* **109**, 15847-15852 (2012).
37. Huang, X.H. *et al.* Genomic analysis of hybrid rice varieties reveals numerous superior alleles that contribute to heterosis. *Nature Communications* **6**(2015).
38. Li, D.Y. *et al.* Integrated analysis of phenome, genome, and transcriptome of hybrid rice uncovered multiple heterosis-related loci for yield increase. *Proceedings of the National Academy of Sciences of the United States of America* **113**, E6026-E6035 (2016).
39. Birchler, J.A., Yao, H., Chudalayandi, S., Vaiman, D. & Veitia, R.A. Heterosis. *Plant Cell* **22**, 2105-12 (2010).
40. Zhang, D.L. *et al.* Genetic structure and differentiation of Oryza sativa L. in China revealed by microsatellites. *Theoretical and Applied Genetics* **119**, 1105-1117 (2009).
41. Wang, W. *et al.* Genomic variation in 3,010 diverse accessions of Asian cultivated rice. *Nature* **557**, 43-49 (2018).
42. Deng, H.F., He, Q., Shu, F., Zhang, W.H. & Yang, F. Status and Technical Strategy on Development of *Japonica* Hybrid Rice in China. *Hybrid rice* **21**, 1-6 (2006).
43. Huang, X.H. *et al.* A map of rice genome variation reveals the origin of cultivated rice. *Nature* **490**, 497-+ (2012).
44. Flint-Garcia., S.A., Buckler., E.S., Tiffin., P., Ersoz., E. & Springer., N.M. Heterosis Is Prevalent for Multiple Traits in Diverse Maize Germplasm. *Plos One* **4**(2009).
45. Xiao, Y. *et al.* The genetic mechanism of heterosis utilization in maize improvement. *Genome Biol* **22**, 148 (2021).
46. Huang, X.H. *et al.* Genomic architecture of heterosis for yield traits in rice. *Nature* **537**, 629-+ (2016).
47. Li, X., Li, X.R., Fridman, E., Tesso, T.T. & Yu, J.M. Dissecting repulsion linkage in the dwarfing gene Dw3 region for sorghum plant height provides insights into heterosis. *Proceedings of the National Academy of Sciences of the United States of America* **112**, 11823-11828 (2015).
48. HIll, A.V. The possible effects of the aggregation of the molecules of haemoglobin on its dissociation curves. *J. Physiol* **40**, IV-VIII (1910).
49. Goutelle, S. *et al.* The Hill equation: a review of its capabilities in pharmacological modelling. *Fundam Clin Pharmacol* **22**, 633-48 (2008).
50. Averam, D., Leid, M. & Bakalinsky, A. Fzf1p of Saccharomyces cerevisiae is a Positive Regulator of SSU1 Transcription and Its First Zinc Finger Region is Required for DNA Binding. *YEAST*, 473-480 (1999).
51. Sato, Y. *et al.* RiceXPro version 3.0: expanding the informatics resource for rice transcriptome. *Nucleic Acids Res* **41**, D1206-13 (2013).
52. Papp, B., Pal, C. & Hurst, L.D. Dosage sensitivity and the evolution of gene families in yeast. *Nature* **424**, 194-197 (2003).
53. Kruse, J., Rennenberg, H. & Adams, M.A. Steps towards a mechanistic understanding of respiratory temperature responses. *New Phytologist* **189**, 659-677 (2011).
54. Affourtit, C., Krab, K. & Moore, A.L. Control of plant mitochondrial respiration. *Biochimica Et Biophysica Acta-Bioenergetics* **1504**, 58-69 (2001).
55. Zhao, M., Chen, X., Gao, G., Tao, L. & Wei, L. RLEdb: a database of rate-limiting enzymes and their regulation in human, rat, mouse, yeast and E. coli. *Cell Res* **19**, 793-5 (2009).
56. Grafius, J.E. Heterosis in barley. *Agronomy Journal* **51**, 551-554 (1959).

57. Mezmouk, S. & Ross-Ibarra, J. The pattern and distribution of deleterious mutations in maize. *G3 (Bethesda)* **4**, 163-71 (2014).
58. Guo, T. *et al.* Genetic basis of grain yield heterosis in an "immortalized F(2)" maize population. *Theor Appl Genet* **127**, 2149-58 (2014).
59. Zhang, H.Y. *et al.* A genome-wide transcription analysis reveals a close correlation of promoter INDEL polymorphism and heterotic gene expression in rice hybrids. *Mol Plant* **1**, 720-31 (2008).
60. Paschold, A. *et al.* Complementation contributes to transcriptome complexity in maize (Zea mays L.) hybrids relative to their inbred parents. *Genome Res* **22**, 2445-54 (2012).
61. Campbell, B.T., Bowman, D.T. & Weaver, D.B. Heterotic Effects in Topcrosses of Modern and Obsolete Cotton Cultivars. *Crop Science* **48**, 593-600 (2008).
62. Vrebalov, J. *et al.* A MADS-box gene necessary for fruit ripening at the tomato ripening-inhibitor (Rin) locus. *Science* **296**, 343-346 (2002).
63. Kondrashov, F.A. & Koonin, E.V. A common framework for understanding the origin of genetic dominance and evolutionary fates of gene duplications. *Trends Genet* **20**, 287-90 (2004).
64. Kruse, J. & Adams, M.A. Sensitivity of respiratory metabolism and efficiency to foliar nitrogen during growth and maintenance. *Global Change Biology* **14**, 1233-1251 (2008).
65. Kacser, H. & Burns, J.A. The molecular basis of dominance. *Genetics* **97**, 639-66 (1981).
66. Fievet, J.B., Nidelet, T., Dillmann, C. & de Vienne, D. Heterosis Is a Systemic Property Emerging From Non-linear Genotype-Phenotype Relationships: Evidence From in Vitro Genetics and Computer Simulations. *Front Genet* **9**, 159 (2018).
67. Bürger, R. & Bagheri, H. Dominance and Its Evolution. *Encyclopedia of Ecology* (2008).
68. Wright, S. Physiological and evolutionary theories of dominance. *American Naturalist* **68**, 24-53 (1934).
69. Jiang, J. & Levine, M. Binding affinities and cooperative interactions with bHLH activators delimit threshold responses to the dorsal gradient morphogen. *Cell* **72**, 741-52 (1993).
70. St Johnston, D. & Nusslein-Volhard, C. The origin of pattern and polarity in the Drosophila embryo. *Cell* **68**, 201-19 (1992).
71. Birchler, J.A. & Veitia, R.A. Gene balance hypothesis: connecting issues of dosage sensitivity across biological disciplines. *Proc Natl Acad Sci U S A* **109**, 14746-53 (2012).
72. Oliner, J.D., Kinzler, K.W., Meltzer, P.S., George, D.L. & Vogelstein, B. Amplification of a gene encoding a p53-associated protein in human sarcomas. *Nature* **358**, 80-3 (1992).
73. Chartier-Harlin, M.-C. *et al.* α-synuclein locus duplication as a cause of familial Parkinson's disease. *The Lancet* **364**, 1167-1169 (2004).
74. Gregori, C., Schuller, C., Frohner, I.E., Ammerer, G. & Kuchler, K. Weak organic acids trigger conformational changes of the yeast transcription factor War1 in vivo to elicit stress adaptation. *J Biol Chem* **283**, 25752-64 (2008).
75. Sprague, G.F. & Tatum, L.A. Citation-Classic - General Vs Specific Combining Ability in Single Crosses of Corn. *Journal American Society Agronomy*, 923–932 (1942).
76. Russell, W.A. Genetic Improvement of Maize Yields. in *Advances in Agronomy*, Vol. 46 (ed. Sparks, D.L.) 245-298 (Academic Press, 1991).
77. Li, H. & Durbin, R. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* **25**, 1754-1760 (2009).
78. McKenna, A. *et al.* The Genome Analysis Toolkit: A MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Research* **20**, 1297-1303 (2010).

79. Zhang, Z.W. *et al.* Mixed linear model approach adapted for genome-wide association studies. *Nature Genetics* **42**, 355-U118 (2010).
80. Yu, J.M. *et al.* A unified mixed-model method for association mapping that accounts for multiple levels of relatedness. *Nature Genetics* **38**, 203-208 (2006).
81. Churchill, G.A. & Doerge, R.W. Empirical Threshold Values for Quantitative Trait Mapping. *Genetics* **138**, 963-971 (1994).
82. Barrett, J.C., Fry, B., Maller, J. & Daly, M.J. Haploview: analysis and visualization of LD and haplotype maps. *Bioinformatics* **21**, 263-265 (2005).
83. Gietz, R.D. & Woods, R.A. Transformation of yeast by lithium acetate/single-stranded carrier DNA/polyethylene glycol method. *Methods Enzymol* **350**, 87-96 (2002).
84. Kim, D. *et al.* TopHat2: accurate alignment of transcriptomes in the presence of insertions, deletions and gene fusions. *Genome Biology* **14**(2013).
85. Trapnell, C. *et al.* Differential analysis of gene regulation at transcript resolution with RNA-seq. *Nat Biotechnol* **31**, 46-53 (2013).
86. Wang, L.K., Feng, Z.X., Wang, X., Wang, X.W. & Zhang, X.G. DEGseq: an R package for identifying differentially expressed genes from RNA-seq data. *Bioinformatics* **26**, 136-138 (2010).
87. Phadnis, N. & Fry, J.D. Widespread correlations between dominance and homozygous effects of mutations: Implications for theories of dominance. *Genetics* **171**, 385-392 (2005).
88. Tian, T. *et al.* agriGO v2.0: a GO analysis toolkit for the agricultural community, 2017 update. *Nucleic Acids Research* **45**, W122-W129 (2017).

Figures and legends

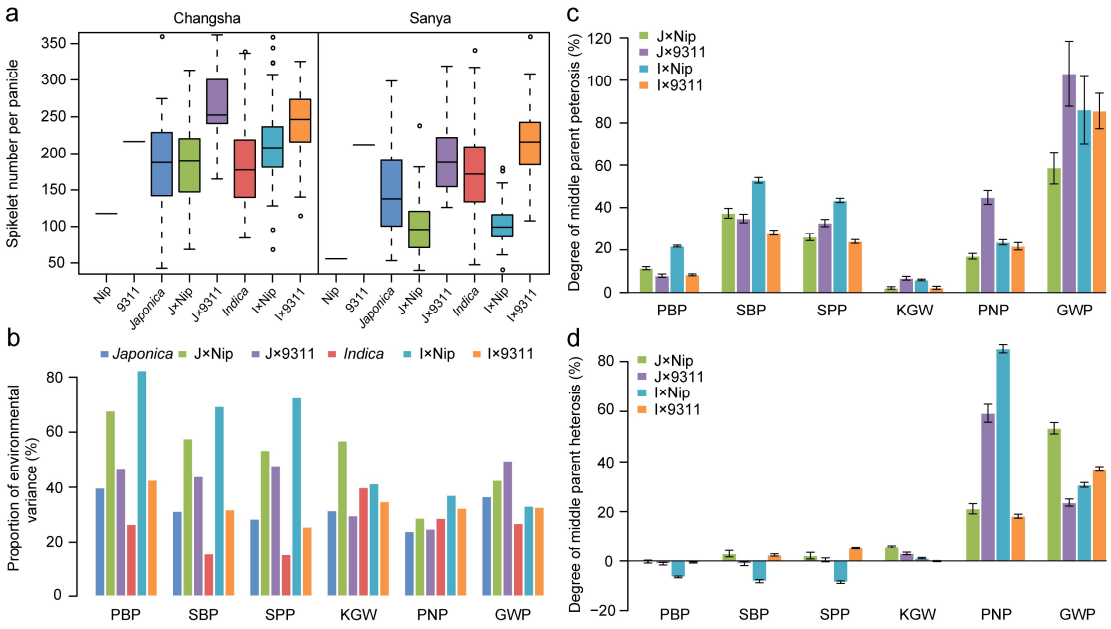

Figure 1: Architecture of yield traits and heterosis among 418 combinations. (a) Spikelet number per panicle (SPP) of inbred parents and their hybrids in Changsha and Sanya. (b) Proportion of environment variance (including environment-additive and interaction of genetic by environment variance) for yield related traits in panels of inbred parents and hybrids. (c)

Middle-parent heterosis of four types of combinations for yield related traits in Changsha. (d) Middle-parent heterosis of four types of combinations for yield related traits in Sanya. J×Nip, J×9311, I×Nip and I×9311 represent the four types of combinations for *Japonica* × Nip, *Japonica*×9311, *Indica*×Nip and *Indica*×9311, respectively. PBP, SBP, KGW, PNP and GWP represent primary branch number per panicle, secondary branch number per panicle, 1000-grain weight, panicle number per plant and grain weight per plant, respectively.

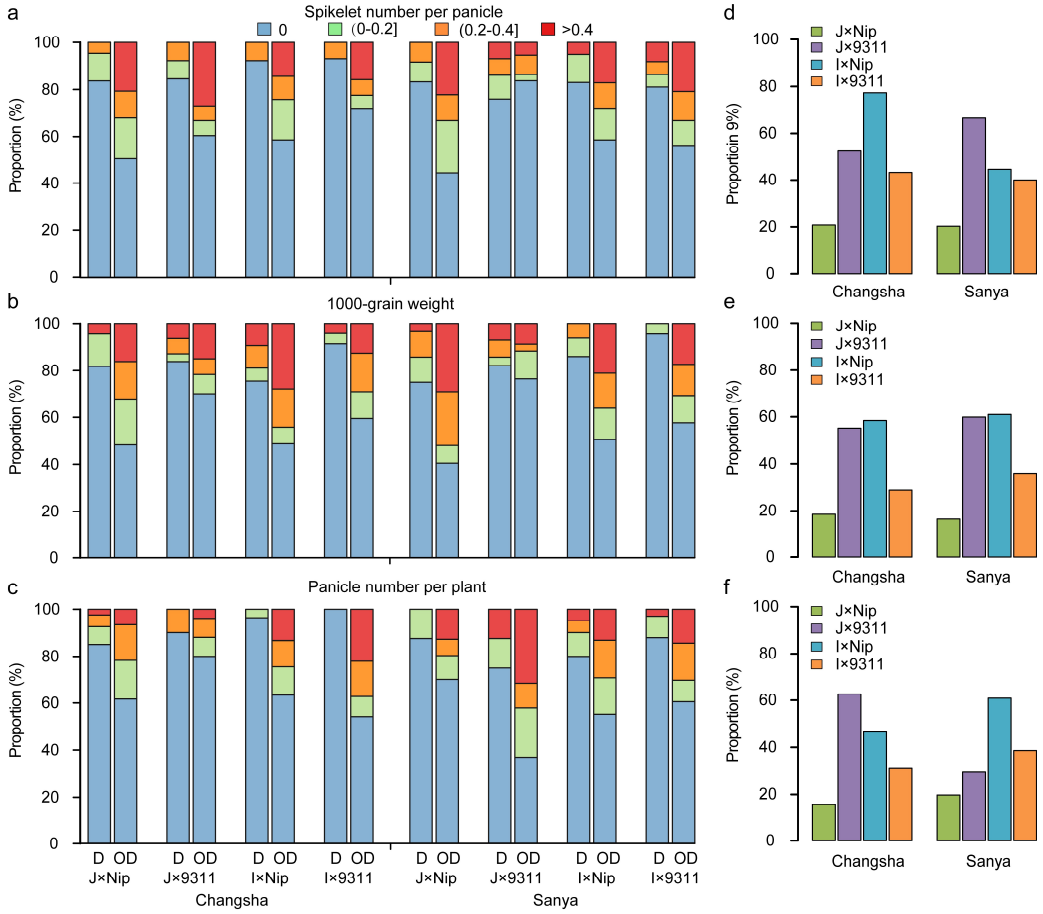

Figure 2: Distribution of repulsive dominant alleles in different types of QTLs and hybrid combinations. (a)-(c), the proportion of different magnitudes of repulsive degree in dominant (D) and over-dominant (OD) QTLs, the magnitudes with 0-0.2, 0.2-0.4 and >0.4 were presented as blue, light green, orange and red, respectively. (d)-(f), the proportion of combinations with repulsive dominant alleles averaged by QTLs containing repulsive dominant alleles, for traits SPP, KGW and PNP.

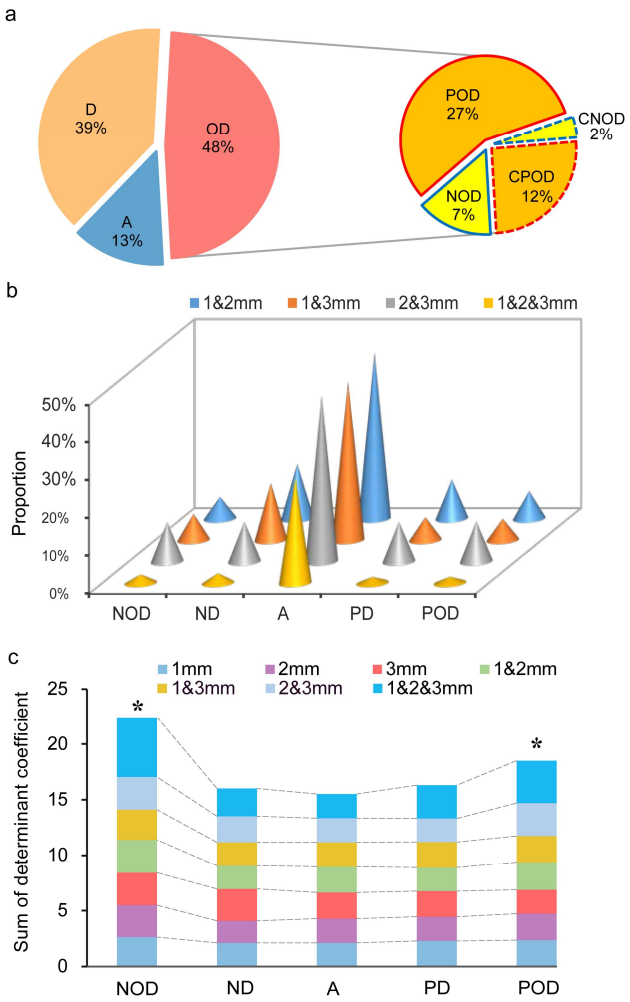

Figure 3: Non-additive is more sensitive to background changes than additive at the transcriptional level. (a) The expression patterns in rice young panicles among 9311, Peiai 64s (PA64S) and their hybrid Liangyoupei 9 (LYP9); A, D, OD, POD, NOD, CPOD and CNOD represent the expression patterns, additive, dominant and over-dominant, positive over-dominant, negative over-dominant, positive over-dominant with complementation and negative over-dominant with complementation, respectively. (b) Consistency for different expression patterns among different tissues (including 1 mm, 2 mm and 3 mm young panicles) in the combination of LYP9. 1&2mm means the same expression pattern in 1 mm and 2 mm panicles, and similar for other symbols of 2&3mm and 1&2&3mm. ND and PD mean negative and positive dominant effect, respectively. NOD and POD mean negative and positive over-dominant effect, respectively. (c) Sum of determination coefficient between transcription factors and their targets genes with different expression patterns in rice young panicles among 9311, Peiai 64s (PA64S) and their hybrid Liangyoupei 9 (LYP9).

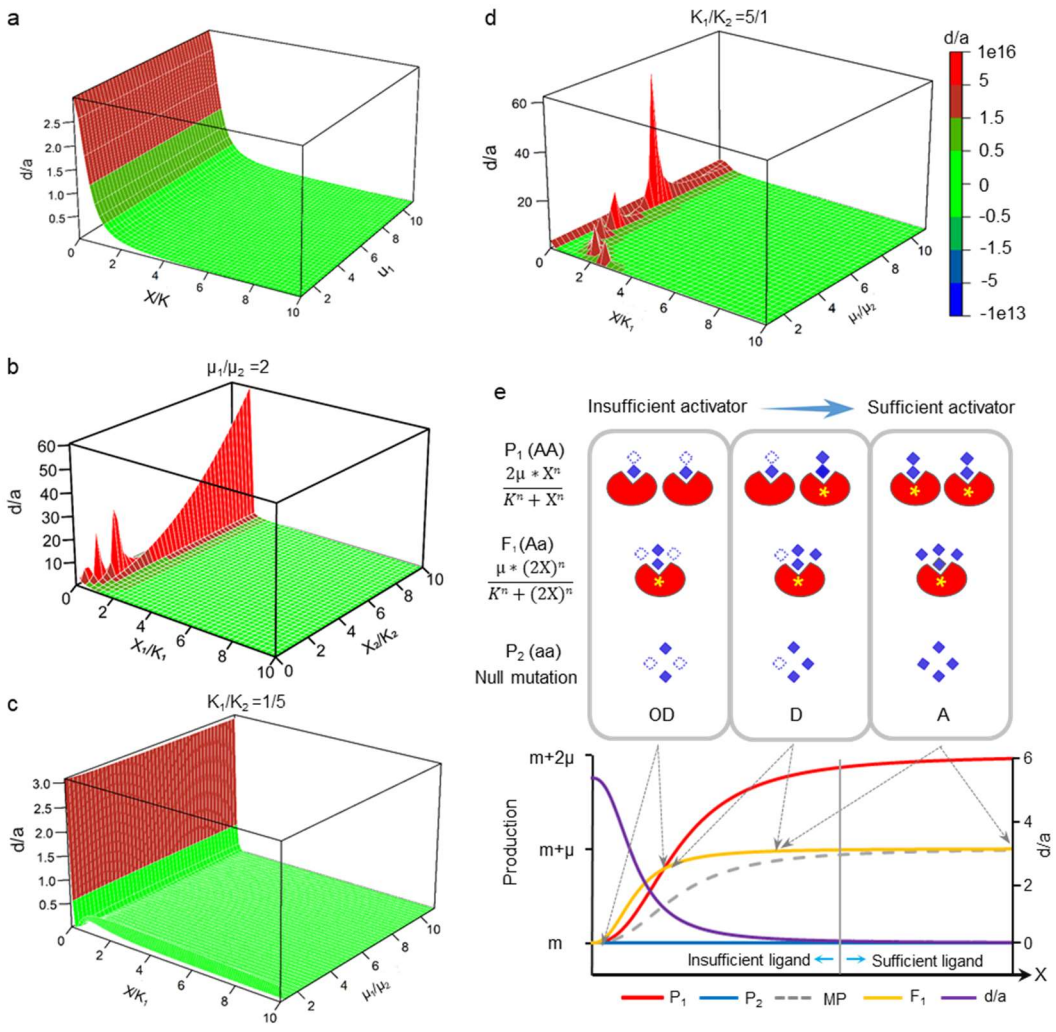

Figure 4: The diagram of Hill reactions illustrates the model of homo-insufficiency under insufficient background (HoIIB). (a) The simulated dominant degree of the target site under the activator background with different sufficiencies (X/K) and different μ_1 with null allele and one functional allele under one regulator background. (b) The simulated dominant degree of the target site with same homologous backgrounds, but the two alleles in F_1 are regulated by different factors in the background for positive regulation. (c) The simulated dominant degree of the target site with the same positive regulators or responsors as the background when allele 1 showing larger maximum function and higher affinity and allele 2 showing smaller maximum function and lower affinity ($\mu_1 > \mu_2$ and $K_1 / K_2=1/5$). (d) The simulated dominant degree of the target site with the same positive regulators or responsors as the background when allele 1 showing larger maximum function but lower affinity and allele 2 showing smaller maximum function but higher affinity ($\mu_1 > \mu_2$ and $K_1 / K_2=5/1$). (e) The red notched ellipse represents the target factor, and that with yellow star * indicates the target factor whom is activated by its

activators (as the background of the target factor), which are represented by the blue diamonds. Here, we assume that one allele of genotype A can be bond and activated by at least two units of activators, but the allele of genotype a is loss-function and can not be bond by the activator, and the activators can be randomly attached by each of two alleles in homozygote AA. The dotted blank diamonds are the required units of activators to activate all two alleles of AA in parent 1 (P_1). The target factor will show overdominance (OD), where the production of heterozygote (orange line, F_1) is higher than that of P_1 (red line), when the quantity of activator is too insufficient to activate even one allele of P_1 but can activate the allele A in F_1 in most cases; and the target factor will show (partial-) dominance (D), where the production of heterozygote is higher than the middle-parent (grey dotted line), when the quantity of activator is relatively insufficient to activate all two alleles of P_1 but can activate the allele A in F_1 in most cases; and when the quantity of activator is sufficient to activate all two alleles of P_1, the target factor will show additive effect, where the production of F_1 is similar to or equal to middle-parent (almost overlap between dotted grey line and orange line); so the dominance degree (d/a) of target factor (purple line) will decrease along with the increase of activator (i.e. from insufficient to sufficient). The parameter used here is, μ=1, n=2 and K = 1.

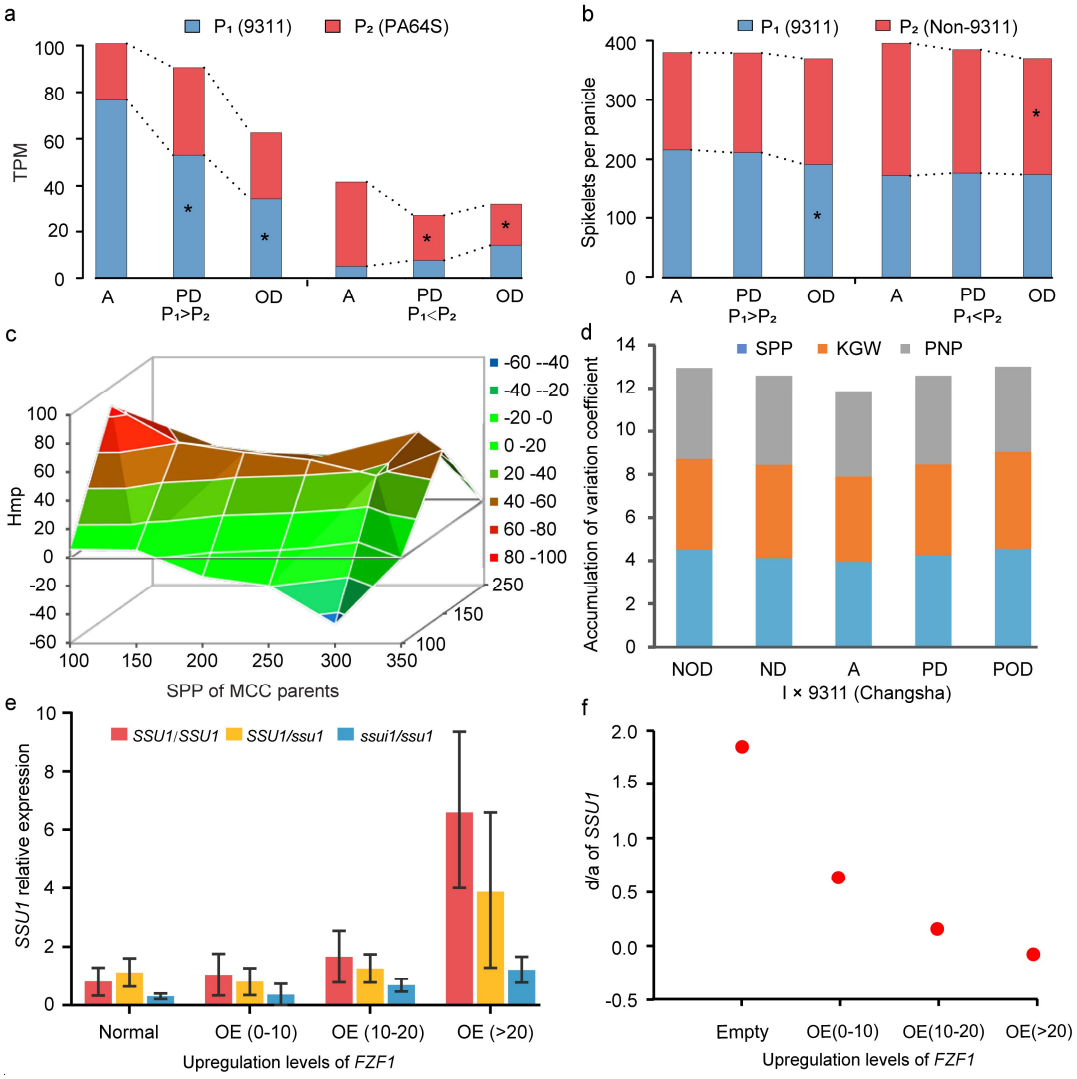

Figure 5: The background effect on additive, dominant and over-dominant effect at levels of transcription, QTL and trait and the result of validation experiment conducted in *Saccharomyces cerevisiae*. (a) The expression of genes with different expression patterns in 3 mm young panicles of two parents; the star means significant difference from additive effect. (b) The effects of 9311 genotype (P_1) and non-9311 genotype (P_2) for QTLs of spikelet number per plant (SPP) with different genetic effect types for combination of *japonica* and 9311 in Changsha; $P_1 > P_2$ means that 9311 genotype (P_1) has higher effect than non-9311 genotype in QTL, and vice versa for $P_1 < P_2$. (c) The middle-parent heterosis of SPP for combinations MCC parents and testers with different SPP. (d) The accumulation of average variation coefficient estimated in each QTL identified in *Indica* ×9311 combination for different types of genetic effects. (e) The relative expression of gene *SSU1* in different *SSU1* genotypes under different expression levels of its transcription factor (*FZF1*) in *Saccharomyces cerevisiae* BY4743; here,

SSU1/SSU1, ssu1/ssu1and SSU1/ssu1 represent the homologous genotype of wild type, the homologous genotype of mutant, and their heterozygous genotype, respectively; OE (0-10) means the strain with upregulated *FZF1* by 0-10 folds, and similar for OE (10-20) and OE (>20), and Empty means the strain with empty vector free of *FZF1*. (f) The dramatically decreased dominance degree of *SSU1* along with the increase of upregulation levels of its transcription factor *FZF1* in *Saccharomyces cerevisiae* BY4743.

Supplementary Files

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

- SupplementaryTable1TheListofricevarietiesusedinthisstudy.xlsx
- SupplementaryTable2Thephenotypicdatausedinthecurrentstudy.xlsx
- SupplementaryTable3ThetwowayANOVAofyieldtraitsinjaponicaparentsandtheirF1populations.xlsx
- SupplementaryTable4ThetwowayANOVAofyieldtraitsinindicaparentsandtheirF1populations.xlsx
- SupplementaryTable5SummaryoftheQTLsidentifiedbyGWASinparentpopulations.xlsx
- SupplementaryTable6SummaryoftheQTLsidentifiedbyGWASinF1populations.xlsx
- SupplementaryTable7SummaryoftheQTLsidentifiedbyGWASusingthemiddleparentheterosisvalues.xlsx
- SupplementaryTable8SummaryoftheQTLssharedamongthoseidentifiedusingmiddleparentheterosisvaluesF1phenotypesandparentphenotypes.xlsx
- SupplementaryTable9TheexpressionpatternofricegenesincombinationofPA64sand9311.xlsx
- SupplementaryTable10TheprimersusedinSSU1mutantconstructionandRTPCR.xlsx
- SupplementaryTable11ThecandidategeneswithinF1QTLsandHmpQTLsofSPPrelatedtraits.xlsx
- SupplementaryTable12ThecandidategeneswithinPQTLsF1QTLsandHmpQTLsofGWPPNPandKGW.xlsx
- SupplementaryTable13Theoverlappedcandidategeneswithinthe QTLswithdifferentgeneticeffecttypesbetweentesterssubspeciesandenvrionments.xlsx
- SupplementaryTable14TheGOenrichmentofcandiategeneswithinnonadditiveQTLsthatcanbeidentifiedrepeatedly.xlsx
- SupplementaryNoteandFiguresuse.pdf