

University of Nebraska - Lincoln

DigitalCommons@University of Nebraska - Lincoln

---

Agronomy & Horticulture -- Faculty Publications

Agronomy and Horticulture Department

---

2020

## Genome-wide identification and analysis of heterotic loci in three maize hybrids

Hongjun Liu

*Shandong Agricultural University*

Qin Wang

*Shanghai Normal University*

Mengjiao Chen

*Shanghai Normal University*

Yahui Ding

*Shandong Agricultural University*

Xuerong Yang

*Shandong Agricultural University*

Follow this and additional works at: <https://digitalcommons.unl.edu/agronomyfacpub>



Part of the [Agricultural Science Commons](#), [Agriculture Commons](#), [Agronomy and Crop Sciences Commons](#), [Botany Commons](#), [Horticulture Commons](#), [Other Plant Sciences Commons](#), and the [Plant Biology Commons](#)

---

Liu, Hongjun; Wang, Qin; Chen, Mengjiao; Ding, Yahui; Yang, Xuerong; Liu, Jie; Li, Xiaohan; Zhou, Congcong; Tian, Qilin; Lu, Yiqi; Fan, Danlin; Shi, Junpeng; Zhang, Lin; Kang, Congbin; Mingfei Sun, Mingfei; Li, Fangyuan; Wu, Yujian; Zhang, Yongzhong; Liu, Baoshen; Zhao, Xiang Yu; Feng, Qi; Yang, Jinliang; Han, Bin; Lai, Jinsheng; Zhang, Xian Sheng; and Huang, Xuehui, "Genome-wide identification and analysis of heterotic loci in three maize hybrids" (2020). *Agronomy & Horticulture -- Faculty Publications*. 1343. <https://digitalcommons.unl.edu/agronomyfacpub/1343>



This Article is brought to you for free and open access by the Agronomy and Horticulture Department at DigitalCommons@University of Nebraska - Lincoln. It has been accepted for inclusion in Agronomy & Horticulture -- Faculty Publications by an authorized administrator of DigitalCommons@University of Nebraska - Lincoln.

---

## Authors

Hongjun Liu, Qin Wang, Mengjiao Chen, Yahui Ding, Xuerong Yang, Jie Liu, Xiaohan Li, Congcong Zhou, Qilin Tian, Yiqi Lu, Danlin Fan, Junpeng Shi, Lin Zhang, Congbin Kang, Mingfei Mingfei Sun, Fangyuan Li, Yujian Wu, Yongzhong Zhang, Baoshen Liu, Xiang Yu Zhao, Qi Feng, Jinliang Yang, Bin Han, Jinsheng Lai, Xian Sheng Zhang, and Xuehui Huang

# Genome-wide identification and analysis of heterotic loci in three maize hybrids

Hongjun Liu<sup>1,†</sup>, Qin Wang<sup>2,†</sup>, Mengjiao Chen<sup>2</sup>, Yahui Ding<sup>1</sup>, Xuerong Yang<sup>1</sup>, Jie Liu<sup>2</sup> , Xiaohan Li<sup>1</sup>, Congcong Zhou<sup>3</sup>, Qilin Tian<sup>3</sup>, Yiqi Lu<sup>3</sup>, Danlin Fan<sup>3</sup>, Junpeng Shi<sup>4</sup>, Lin Zhang<sup>5</sup>, Congbin Kang<sup>1</sup>, Mingfei Sun<sup>1</sup>, Fangyuan Li<sup>1</sup>, Yujian Wu<sup>1</sup>, Yongzhong Zhang<sup>6</sup>, Baoshen Liu<sup>6</sup>, Xiang Yu Zhao<sup>1</sup>, Qi Feng<sup>3</sup>, Jinliang Yang<sup>7</sup>, Bin Han<sup>3</sup>, Jinsheng Lai<sup>4</sup>, Xian Sheng Zhang<sup>1,\*</sup> and Xuehui Huang<sup>2,\*</sup> 

<sup>1</sup>State Key Laboratory of Crop Biology, College of Life Sciences, Shandong Agricultural University, Tai'an, China

<sup>2</sup>Shanghai Key Laboratory of Plant Molecular Sciences, College of Life Sciences, Shanghai Normal University, Shanghai, China

<sup>3</sup>National Center for Gene Research, CAS Center for Excellence of Molecular Plant Sciences, Institute of Plant Physiology and Ecology, Shanghai Institutes for Biological Sciences, Chinese Academy of Sciences, Shanghai, China

<sup>4</sup>State Key Laboratory of Agrobiotechnology and National Maize Improvement Center, Department of Plant Genetics and Breeding, China Agricultural University, Beijing, China

<sup>5</sup>College of Agriculture, Northeast Agricultural University, Harbin, China

<sup>6</sup>State Key Laboratory of Crop Biology, College of Agronomy, Shandong Agricultural University, Tai'an, China

<sup>7</sup>Department of Agronomy and Horticulture, University of Nebraska-Lincoln, Lincoln, NE, USA

Received 3 April 2019;

revised 1 June 2019;

accepted 4 June 2019.

\*Correspondence (Tel 86 21 64322006;

fax 86 21 64322006; emails

zhangxs@sdau.edu.cn (XSZ);

xhhuang@shnu.edu.cn (XH))

†These authors contributed equally to this work.

## Summary

Heterosis, or hybrid vigour, is a predominant phenomenon in plant genetics, serving as the basis of crop hybrid breeding, but the causative loci and genes underlying heterosis remain unclear in many crops. Here, we present a large-scale genetic analysis using 5360 offsprings from three elite maize hybrids, which identifies 628 loci underlying 19 yield-related traits with relatively high mapping resolutions. Heterotic pattern investigations of the 628 loci show that numerous loci, mostly with complete–incomplete dominance (the major one) or overdominance effects (the secondary one) for heterozygous genotypes and nearly equal proportion of advantageous alleles from both parental lines, are the major causes of strong heterosis in these hybrids. Follow-up studies for 17 heterotic loci in an independent experiment using 2225 F<sub>2</sub> individuals suggest most heterotic effects are roughly stable between environments with a small variation. Candidate gene analysis for one major heterotic locus (*ub3*) in maize implies that there may exist some common genes contributing to crop heterosis. These results provide a community resource for genetics studies in maize and new implications for heterosis in plants.

**Keywords:** heterosis, maize, genomics, quantitative trait loci, molecular breeding.

## Introduction

Heterosis, or known as hybrid vigour, refers to the phenomenon where the F<sub>1</sub> hybrid has greater phenotypic performance than both inbred parents (Bruce, 1910; East, 1936; Jones, 1917; Shull, 1908). Heterosis forms the foundation of modern breeding (Birchler *et al.*, 2016). As the widespread of the phenomenon, heterosis has been discovered and used in production for many crop species, such as maize (*Zea mays*), rice (*Oryza sativa*) and canola (*Brassica napus*). During recent decades, the development of a number of elite hybrids in these crops had greatly contributed to the food security through increasing crop yield per unit area across the world (Schnable and Springer, 2013). To further exploit heterosis potentials in hybrids by a genomics-assisted breeding (GAB), the understanding of heterosis mechanisms, coupled with the discovery of the major genetic loci involved in yield heterosis performances in the major crops, will be of great interests.

The molecular basis of heterosis has been explored from different perspectives for many decades (Garcia *et al.*, 2008; Hua *et al.*, 2002; Ko *et al.*, 2016; Kusmec *et al.*, 2017; Lu *et al.*, 2003; Luo *et al.*, 2011; Ma *et al.*, 2018; Tang *et al.*, 2010; Wang *et al.*, 2016, 2019). Several major models for heterosis have been proposed, including dominance complementation and single-locus overdominance. In the model for dominance

complementation, beneficial and dominant alleles from one parental line compensate for deleterious and recessive alleles from the other. For the case of single-locus overdominance, the heterozygous genotype has a better performance than both homozygous genotypes (Hollick and Chandler, 1998). Recently, in a number of plant species (e.g. *Arabidopsis*, rice, maize, sorghum and tomato), genetic mapping has been used to identify the detailed gene loci contributing to hybrid performances and investigate their mechanisms (Swanson-Wagner *et al.*, 2006; Riedelsheimer *et al.*, 2012; Zhou *et al.*, 2012; Yao *et al.*, 2013; Dapp *et al.*, 2015; Birchler *et al.*, 2016; Li *et al.*, 2016; Yang *et al.*, 2017a,b; Liu *et al.*, 2019). For example, in tomato and rice, the orthologue gene of *Arabidopsis* flowering locus *SFT* for tomato (Krieger *et al.*, 2010) and *Hd3a* for rice (Huang *et al.*, 2016), respectively, has been found to show single-gene overdominance. In another case, repulsion linkage between two quantitative trait loci (QTLs) underlying plant height (*qHT7.1* and *Dw3*) with dominance effects caused the occurrence of heterosis in sorghum (Li *et al.*, 2015).

Maize has been a model plant species for genetic studies of heterosis over a century, largely owing to its outcrossing habits, the availability of diverse germplasm and the importance of maize yield in global agricultural production (Riedelsheimer *et al.*, 2012). Nowadays, the majority of maize varieties grown

throughout the world are hybrids crossed between two inbred lines with contrasting pedigree relationships. In the United States, the maize germplasm was classified into two major heterotic groups – Stiff Stalk (SS) and Non-Stiff Stalk (NSS). Inbred parents selected from these two heterotic groups contribute most of the genetic gains in hybrid production. Normally, these maize hybrids are usually much taller and have larger ears than the inbred parents (Figure 1a). Here, using 5360 offsprings from three elite hybrids in maize, we performed a large-scale genetic mapping and heterosis analysis, aiming to identify important loci contributing to heterosis advantage in elite maize hybrids and provide a useful data resource for heterosis studies in plants.

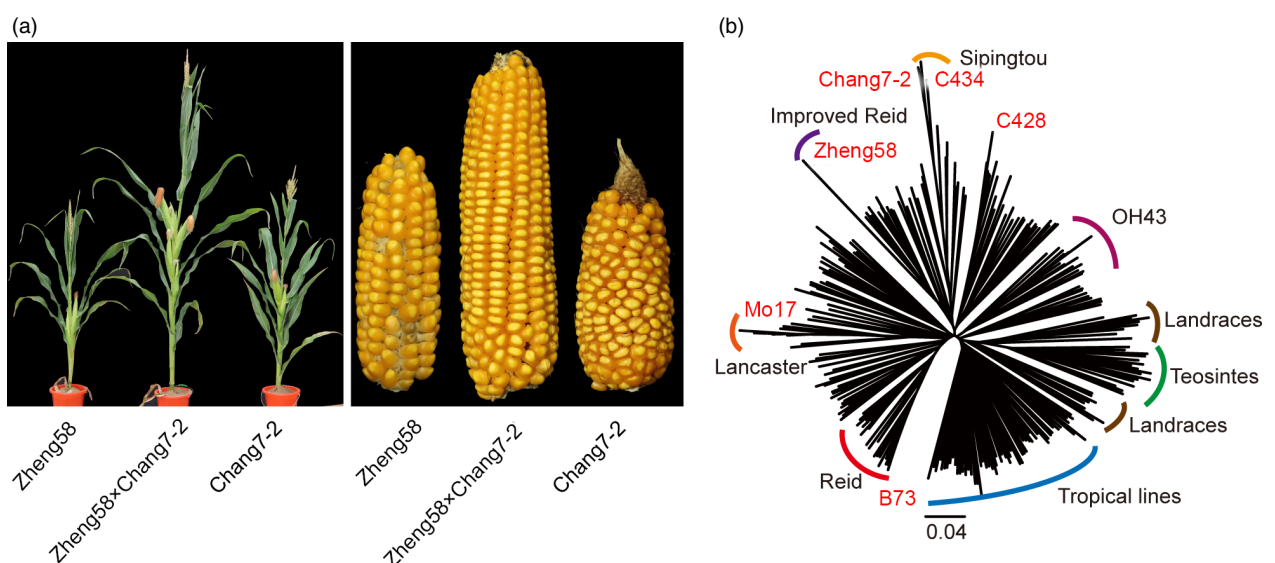
## Results

### Population construction and genome sequencing

Three representative hybrid crosses in maize, Zheng58 × Chang7-2, B73 × Mo17 and C428 × C434, were selected in this study. The cultivar Zhengdan958, which is the F<sub>1</sub> plant from the cross between two inbred parental lines Chang7-2 and Zheng58 (Figure 1a), displays a high yield performance and is the most widely grown hybrid in China (account for 13% of the maize planting areas in China since 2004). The hybrid cross between Mo17 and B73 is a classical combination in heterosis studies, especially in the United States. Another hybrid Siyu-2 from the C428 × C434 cross is a newly developed maize cultivar in China, and the parental line C434 is derived from Chang7-2 in breeding. The six inbred parental lines were sequenced using Illumina platform with more than 30 × coverage for each. The resequencing reads of six parental lines were combined with published data of 281 diverse maize lines (including teosintes, landraces and modern cultivars from different periods of breeding history in the United States and China; Chia *et al.*, 2012; Jiao *et al.*, 2012; Bukowski *et al.*, 2018) to construct a neighbour-joining tree using 2.67 million whole-genome SNPs (minor allele frequency > 0.05). As expected, teosintes were grouped into a

single clade which was surrounded by the landraces and tropical lines. All the six parental lines were identified as temperate lines from five genetically distinct clades, including those of well-known heterotic groups of Reid (B73), Lancaster (Mo17), Improved Reid (Zheng58) and Sipingtuo (Chang7-2 and C434, Figure 1b). We further used the sequencing data to identify high-quality SNPs for each hybrid cross. There were totally 6 807 279 SNPs (between Chang7-2 and Zheng58), 5 285 206 SNPs (between Mo17 and B73) and 4 094 055 SNPs (between C434 and C428) across the maize genome (AGPv4). These SNPs between parents covered the majority of genomic regions, thus facilitating follow-up genotyping in the F<sub>2</sub> populations (Figure S1a-c). Interestingly, there were two very large *identical by state* (IBS) segments on chromosome 2 (located within the 53–183 Mb interval) and chromosome 7 (located within the 8–128 Mb interval) between C434 and C428 covering over half of the chromosomes (Figure S1c), indicating that the parental lines C428 and C434 had a shared common ancestor in the recent breeding.

We then generated three sets of large F<sub>2</sub> populations from the F<sub>1</sub> hybrid combinations, with 2567 F<sub>2</sub> individuals, 1505 F<sub>2</sub> individuals and 1288 F<sub>2</sub> individuals for Zheng58 (Female) × Chang7-2 (Male), B73 (Female) × Mo17 (Male) and C428 (Male) × C434 (Female), respectively. The genomes of all the 5360 lines were sequenced individually using Illumina platform, each with ~0.2 × coverage, resulting in a total of ~2 Tb sequence data. The raw reads of each F<sub>2</sub> individuals were aligned to the maize reference genome for genotype calling. Using the software SEG-Map, high-density genetic maps were generated subsequently for each population (Figure 2; Huang *et al.*, 2009). In the F<sub>2</sub> population of Zheng58 × Chang7-2, the total length of the genetic map is 1588.8 cM. For each F<sub>2</sub> individuals in Zheng58 × Chang7-2, there were generally 2–4 recombination events on each chromosome (Figure 2a). We observed that the recombination varied extremely along chromosomes – all chromosomes had very low recombination rates around peri-centromeric regions, and high



**Figure 1** Heterosis advantage of maize hybrids and whole-genome variation of their parental lines. (a) Phenotype of the maize hybrid Zhengdan958 and its parental inbred lines. The plant height and ears of inbred lines Zheng58 (left), Chang7-2 (right) and the F<sub>1</sub> (middle) are shown. (b) The phylogenetic positions of six parental lines (Zheng58, Chang7-2, B73, Mo17, C428 and C434) in the neighbour-joining tree of 287 maize inbred lines, in which the known information on the clades of maize germplasm is indicated.

recombination rates around peri-telomere regions (Figure 2b, the plots for Zhengdan958 population), which is consistent with the recombination pattern evaluated from maize haplotype map (Bukowski *et al.*, 2018; Chia *et al.*, 2012; Rodgers-Melnick *et al.*, 2015). The exceptions occurred on chromosomes 1 and 4 – we found that besides the peri-telomere region, there was another region with extremely low recombination rates (located within the 224–245 Mb interval of chromosome 1 and the 201–228 Mb interval of chromosome 4), probably suggesting the existence of some other factors to restrict the recombination (e.g. large genomic inversions between parental lines).

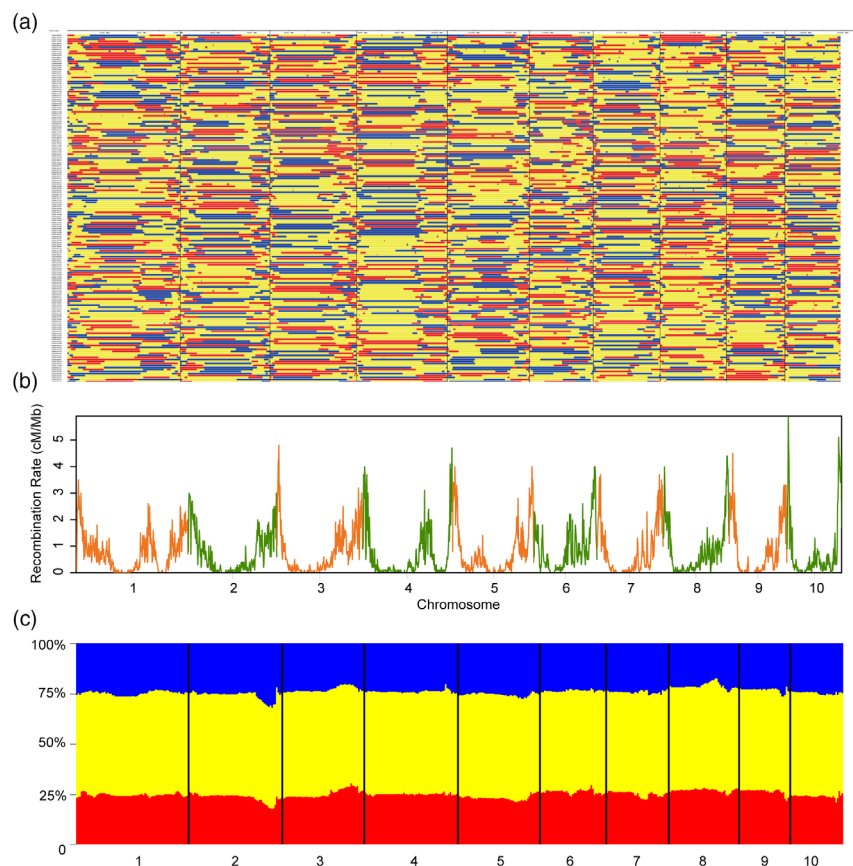
For QTLs mapping, a total of 19 agronomic traits were evaluated for  $F_2$  individual, inbred parents and the hybrids in Shandong Agricultural University Research Farm (eastern China, N36°01', E117°0') in 2017. The traits included flowering time, plant height, leaf length/width, kernel yield per ear and other yield component traits (Figure S2). For most traits, the  $F_2$  population exhibited normal distribution (Figure S3). Moreover, the parental lines were planted and phenotyped for these traits, which displayed a high amount of heterosis (Table S1), and a strong correlation of the degree of heterosis between populations could be observed (Figure S4). Furthermore, we performed genetic dissections using a composite interval-mapping method for each population, resulting in 256, 214 and 158 QTLs identified (with LOD cut-off  $>3.5$ ) for the populations of Zheng58  $\times$  Chang7-2, B73  $\times$  Mo17 and C428  $\times$  C434, respectively (Figure 3a-c). As displayed in Figure 3, most traits in maize hybrids were controlled by numerous QTLs with modest or small effects, in contrast to those in rice hybrids where there were only a few large-effect QTLs underlying one trait (Huang *et al.*, 2016).

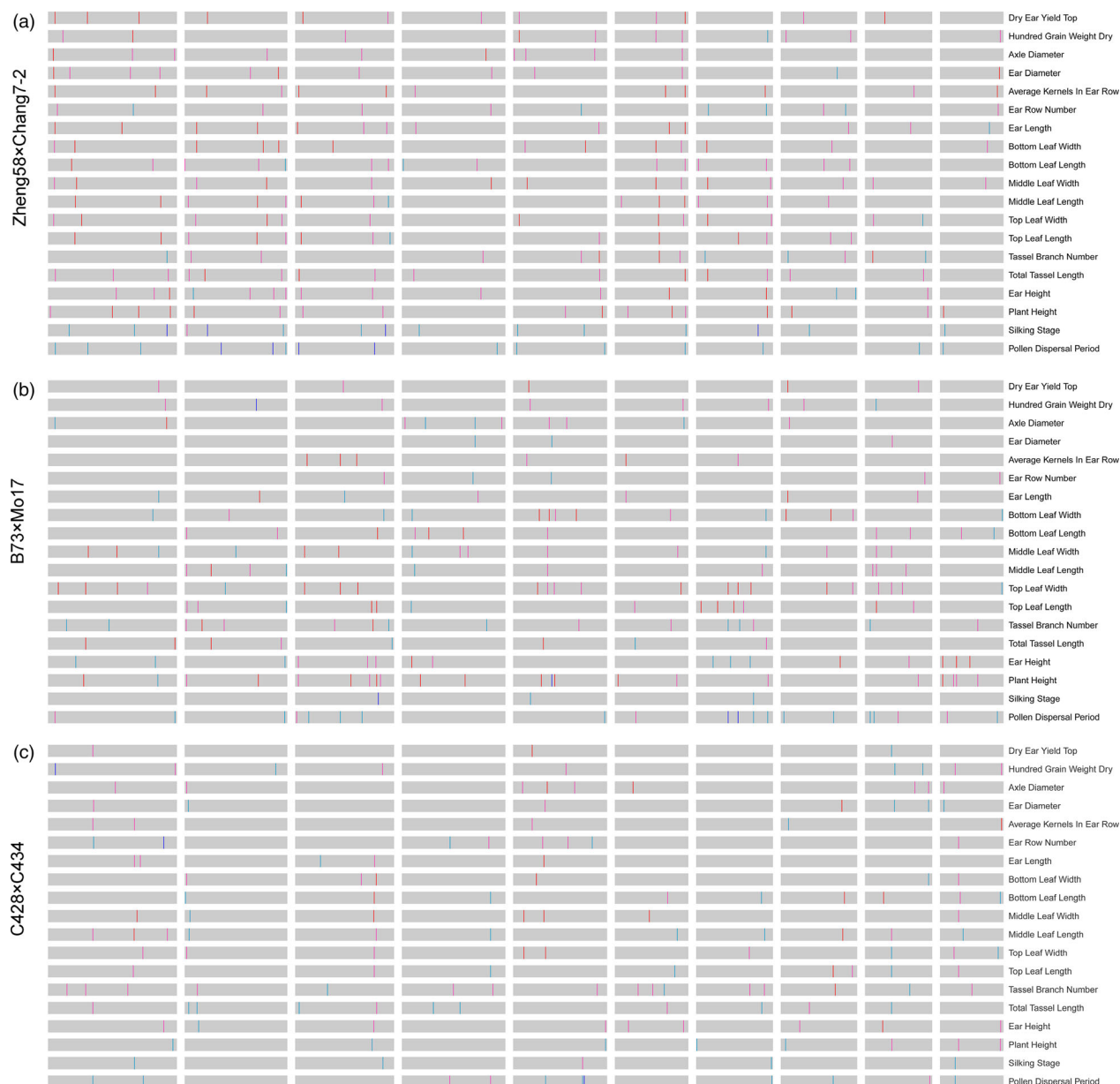
Moreover, we noticed that many QTLs underlying multiple traits were located within the same genomic regions, because some traits were partially related (e.g. the QTL around the 159–169 Mb interval on chromosome 6 underlying several leaf size-related traits in Zheng58  $\times$  Chang7-2). Additionally, there were totally 76 overlapping QTLs between populations underlying the same traits – for example the QTL on chromosome 7 for leaf length and the QTL on chromosome 5 for flowering time (Figure S5).

### Global analysis of heterotic effects

We make a further inquiry that how the QTLs lead to the phenomenon of strong better-parent heterosis in these maize hybrids. Therefore, for each QTL the genetic effect of the heterozygous genotype (MF, where M and F indicate the alleles from male and female parental lines, respectively) was compared with those of homozygous genotypes (MM and FF). When putting together the estimates of  $d/a$  for all the QTLs underlying 19 traits, the overall pattern was that – complete–incomplete dominance acted as the principle one, followed with overdominance (Figure 4a–c). It should be noticed that the number of overdominant QTLs was probably overestimated, although we have used a large population size with a high genotype resolution. One of the reasons is that there must be some pseudo-overdominance cases (very close linkage of two advantageous alleles with repulsion phases in two inbred parents) which our genetic mapping cannot well distinguish with true overdominance cases. Another reason is that, for minor-effect QTLs, there were more statistical errors when evaluating their phenotypic effects, leading to the false positives in overdominance cases. Hence, as expected by the hypothesis (statistical

**Figure 2** High-resolution genotyping of  $F_2$  lines by whole-genome resequencing and the distribution of recombination rates in the hybrid Zhengdan958. (a) The recombination bin map of the  $F_2$  population ( $n = 2567$ ) from a cross between Zheng58 and Chang7-2, in which the horizontal axis indicates the genomic regions and the vertical axis indicates the  $F_2$  lines. Zheng58/Zheng58 homozygous type is shown in red, Zheng58/Chang7-2 heterozygous type is shown in yellow, and Chang7-2/Chang7-2 homozygous type is shown in blue. (b) Plots of the recombination rates of the Zhengdan958  $F_2$  population throughout the maize chromosomes. (c) Statistic of separation ratio of the three genotypes. Zheng58/Zheng58 homozygous type is shown in red, Zheng58/Chang7-2 heterozygous type is shown in yellow, and Chang7-2/Chang7-2 homozygous type is shown in blue.





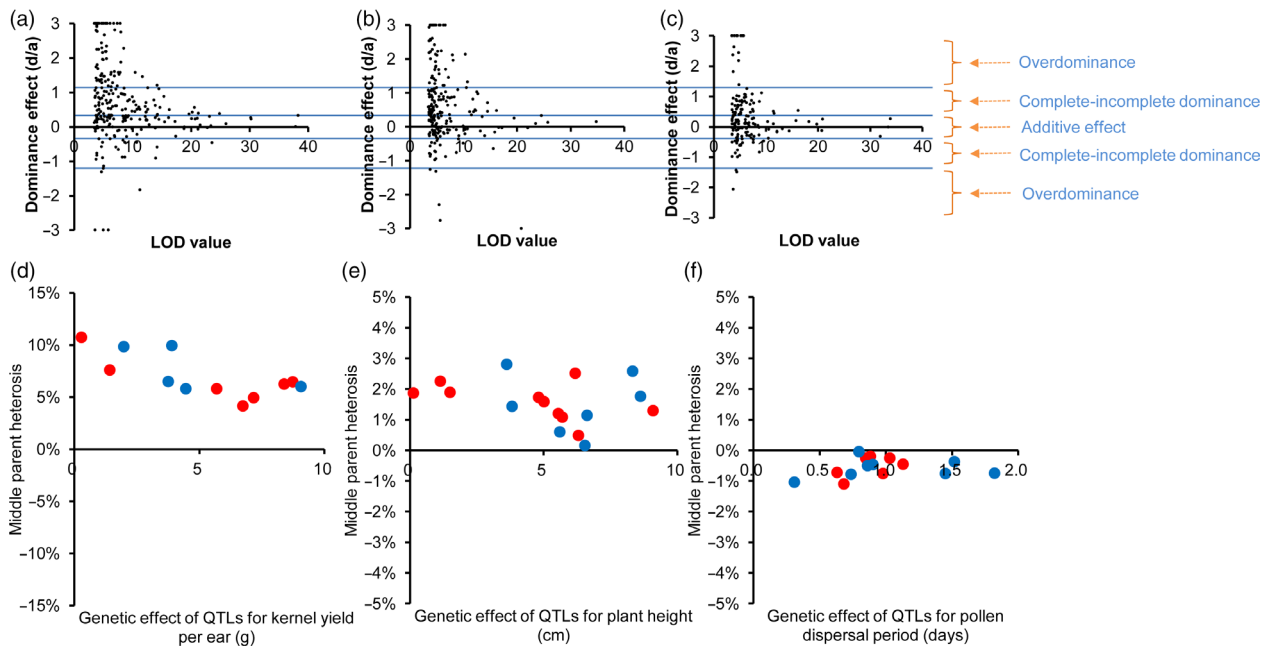
**Figure 3** Genome-wide QTL mapping for yield-related traits in three  $F_2$  populations. (a) Plots of 256 QTLs ( $LOD > 3.5$ ) underlying 19 agronomic traits in Zheng58  $\times$  Chang7-2  $F_2$  population. The genomic location of each QTL peak was indicated in the maize genome by black vertical lines. (b) Plots of 214 QTLs ( $LOD > 3.5$ ) underlying 19 yield-related traits in B73  $\times$  Mo17  $F_2$  population. (c) Plot of 158 QTLs ( $LOD > 3.5$ ) underlying 19 yield-related traits in C428  $\times$  C434  $F_2$  population. The heterotic effects of the QTLs are indicated by deep red ( $d/a > 1$ ), light red ( $0 < d/a \leq 1$ ), light blue ( $-1 < d/a \leq 0$ ) and dark blue ( $d/a \leq -1$ ), respectively.

errors in minor-effect QTLs leading to overestimated overdominance cases), we found that the number of overdominant QTLs decreases rapidly with increasing of the LOD value (Figure 4a-c). Taken these factors into considerations, we proposed that complete-incomplete dominance should be the predominant pattern in maize heterosis, consistent with previous reports in maize (Yang *et al.*, 2017a,b) and similar with the pattern in hybrid rice (Huang *et al.*, 2016).

Considering that the population from Zheng58  $\times$  Chang7-2 has a large sample size ( $n = 2567$ ), a strong yield heterosis, and an unbiased segregation distortion (25%, 50% and 25% for genotypes MM, MF and FF, respectively) across the whole maize genome (see Figure 2c), genetic dissections of yield-related QTLs

and evaluation of heterosis effects should be more powerful and precise in this population. Hence, 2567  $F_2$  individuals from Zheng58  $\times$  Chang7-2 were used for in-depth genetic analyses. The hybrid Zhengdan958 was compared with their parental lines to evaluate the degree of heterosis. The  $F_1$  line showed strong heterosis for kernel yield per ear (trait-based better-parent heterosis BPH index = 259%), modest heterosis for plant height (trait-based BPH index = 52%) and negative heterosis for flowering time (trait-based BPH index = -13%). We further investigated why the level of heterosis in  $F_1$  (ZhengDan958) varied for different agronomic traits using the QTL data. In the  $F_2$  population, 15, 17 and 12 QTLs were identified for the three representative traits including flowering time, plant height and





**Figure 4** Evaluation of dominance effects for the QTLs. (a-c) Plot of the dominance effects ( $d/a$  values) and their LOD values for the QTLs in the population Zheng58  $\times$  Chang7-2 (a), B73  $\times$  Mo17 (b) and C428  $\times$  C434 (c). The QTLs with  $d/a$  values of  $>3$  or  $<-3$  are plotted to be  $= 3$  or  $= -3$  for display purposes. (d-e) Plot of middle parent heterosis index and their allele effects for the QTLs underlying ear yield (d), plant height (e), and flowering time (f) in the population Zhengdan958. The alleles with better yield performance, more plant height or late flowering time that were contributed from Zheng58 and Chang7-2 are coloured by red and blue, respectively.

kernel yield per ear, respectively. The degree of locus-based MPH index was strong (ranging from 4% to 10%, for each individual QTL) for QTLs underlying kernel yield, modest for QTLs underlying plant height (from 0% to 3%) and slightly negative (from  $-1\%$  to 0%) for QTLs underlying flowering time (Figure 4d–f), which was consistent with the overall heterosis of each trait by comparing the  $F_1$  and its parental lines. This consistency indicates the level of heterosis of one trait depended on the heterotic effects of individual QTLs underlying the trait. We further addressed the question – what about the proportion of advantageous alleles from either parental line in ZhengDan958. In a previous work for hybrid rice, we had found that most of the advantageous alleles (the allele leading to more grain-yield in rice) come from male parents (Huang *et al.*, 2016). However, the situation in hybrid maize is completely different, which could be reflected from ‘the contributor sources’ (or called, ‘directions’) of QTLs for yield, height and flowering (Figure 4d–f, dots in blue and red). For total 256 QTLs identified in the Zheng58  $\times$  Chang7-2 population, the proportions of advantageous alleles from Chang7-2 and Zheng58 were 49.6% and 50.4%, respectively.

#### Replication of heterotic effects and candidate gene analysis

The yield-related QTLs and their heterotic effects may be not fully stable in different environmental locations due to some environmental factors or technical noises (or called random errors). In order to evaluate the degree of these implications, we carried out an independent experiment using another set of 2225  $F_2$  individuals from ZhengDan958 (Zheng58  $\times$  Chang7-2) population which planted in Northeast Agricultural University Research Farm (N45°46', E126°52'), Heilongjiang, 2018. In this replication experiment, we did not re-genotyped the whole maize genomes but selected 17 major QTLs identified in Shandong (eastern

China) in 2017, which underlay kernel yield per ear and yield component traits (Table 1). Molecular markers around the selected QTLs were designed, and all the  $F_2$  individuals in the population of Heilongjiang were genotyped at these loci. Further genetic analysis demonstrated that 15 of the 17 QTLs showed significant effects in Heilongjiang (Table 1). Moreover, the heterotic effects of the QTLs were partially consistent (Figure 5,  $r = 0.6$  for the correlation of the log of  $d/a + 1$  values between two locations), but at several heterotic loci the  $d/a$  values had some variations between Shandong and Heilongjiang, for example from the incomplete dominance effect in Shandong to the overdominance effect in Heilongjiang, possibly due to genotype by environmental interactions.

We performed detailed analyses for one heterotic locus *KY4q19* which contained one candidate gene (the ortholog of *OssPL14* in maize) and showed variations in heterotic effects between the two locations Shandong and Heilongjiang. In the hybrid rice of *indica-japonica* crosses, *OssPL14* (also known as *IPA1*) was found to be one of the key genes contributing to heterosis advantage, displaying a strong overdominance effect (Huang *et al.*, 2016; Jiao *et al.*, 2010; Miura *et al.*, 2010). One of its orthologs in maize, *ub3* (Gene ID: *Zm00001d052890*), has been reported to be the causative gene of the QTL underlying plant architecture in maize (Chuck *et al.*, 2014; Liu *et al.*, 2015). This QTL (named as *KY4q19* here, containing the candidate gene *ub3*) was also identified in our populations, with large effects on regulating tassel branch number and kernel yield per ear (Figure 6a). For kernel yield per ear, the *KY4q19* heterozygote increased  $\sim 13.5\%$  more yield advantage than the homozygous male genotype *KY4q19(M)* (Figure 6b), acting through the way of nearly complete dominance in Shandong ( $d/a = 0.97$ ). In the experiment of Heilongjiang, a strong overdominance effect was observed for the *KY4q19*

**Table 1** Replications of 17 heterotic loci using another ZhengDan958 F<sub>2</sub> population in Heilongjiang, China

Trait	Chr.	Peak position (Mb)	Filed experiment in Shandong, China				Filed experiment in Heilongjiang, China				
			LOD	<i>d/a</i>	Heterotic effects	Parent with advantageous allele	<i>P</i> -value (ANOVA)	<i>d/a</i>	Heterotic effects	Parent with advantageous allele	
Hundred grain weight	1	36	12.4	0.21	Additive effect	Zheng58	<2e-16	0.74	Complete–incomplete dominance	Zheng58	
Kernel yield per ear	1	95	3.9	2.25	Overdominance	Chang7-2	5.5E-09	1.82	Overdominance	Chang7-2	
Plant height	1	218	8	2.75	Overdominance	Chang7-2	1.9E-07	1.32	Overdominance	Chang7-2	
Kernel yield per ear	2	55	7.4	3.26	Overdominance	Chang7-2	7.1E-05	3.02	Overdominance	Chang7-2	
Ear diameter	2	158	8.7	0.25	Complete–incomplete dominance	Chang7-2	<2e-16	0.51	Complete–incomplete dominance	Chang7-2	
Ear diameter	2	225	3.8	1.13	Complete–incomplete dominance	Chang7-2	5.0E-03	0.99	Complete–incomplete dominance	Chang7-2	
Kernel yield per ear	3	19	6.8	46.16*	Overdominance	Zheng58	1.1E-05	3.11	Overdominance	Zheng58	
Bottom leaf length	3	184	6.5	0.33	Complete–incomplete dominance	Zheng58	8.2E-12	0.92	Complete–incomplete dominance	Zheng58	
Kernel yield per ear	4	191	6.7	0.97	Complete–incomplete dominance	Zheng58	1.1E-08	4.04	Overdominance	Zheng58	
Bottom leaf width	5	29	6	0.96	Complete–incomplete dominance	Zheng58	1.1E-03	0.65	Complete–incomplete dominance	Zheng58	
Plant height	5	214	3.7	49.92*	Overdominance	Zheng58	1.8E-01	–0.37	NA†	NA†	
Bottom leaf length	6	101	9.5	0.79	Complete–incomplete dominance	Zheng58	1.1E-04	1.48	Overdominance	Zheng58	
Kernel yield per ear	6	169	7.4	6.36	Overdominance	Chang7-2	6.5E-02	5.60	NA†	NA†	
Bottom leaf length	7	169	38.3	0.33	Complete–incomplete dominance	Zheng58	<2e-16	0.30	Complete–incomplete dominance	Zheng58	
Kernel yield per ear	8	55	10	0.87	Complete–incomplete dominance	Chang7-2	4.1E-03	1.23	Complete–incomplete dominance	Zheng58	
Ear height	8	134	18.8	–0.06	Additive effect	Chang7-2	2.0E-04	0.05	Additive effect	Chang7-2	
Average kernels in ear row	10	138	4.2	1.36	Overdominance	Zheng58	1.0E-06	4.95	Overdominance	Zheng58	

\*For the two loci, there was a great phenotypic distinction of heterozygous genotypes (MF) to homozygous genotypes (MM and FF), but very small differences between two homozygous genotypes (i.e. between MM and FF), which resulted in outrageously high *d/a* ratios.

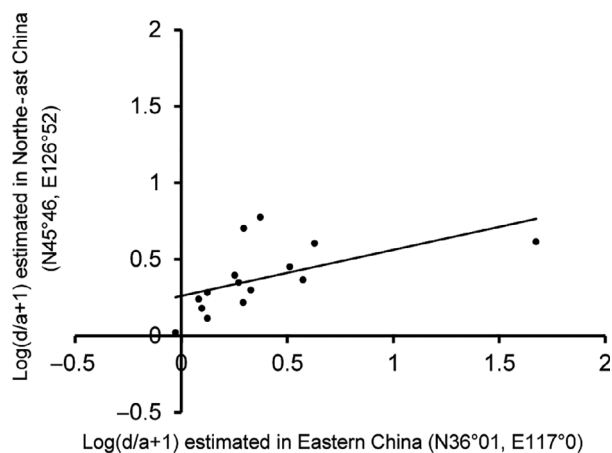
†For the two QTLs with less significant effects ( $P > 0.05$ ) in Heilongjiang, the heterotic effects and the parent with advantageous allele in Heilongjiang cannot be well evaluated.

heterozygote ( $d/a = 4.04$ ). Sequence comparisons between parents showed that there were no differences in coding regions of *ub3*, but many variants in the promoter regions (Figure 6c). Hence, the heterotic effect of *ub3* in yield probably resulted from an optimal expression level in its heterozygous state, which interestingly was consistent with the situation in rice (Huang *et al.*, 2016; Liu *et al.*, 2015).

## Discussion and implication

In order to dissect the genetic basis of heterosis, numerous studies demonstrated that the control of heterosis is complex in plants (Birchler, 2016; Huang *et al.*, 2016; Kusmec *et al.*, 2017; Li *et al.*, 2016; Luo *et al.*, 2011; Riedelheimer *et al.*, 2012; Wang *et al.*, 2016; Yang *et al.*, 2017a,b; Yao *et al.*, 2013; Zhou *et al.*,





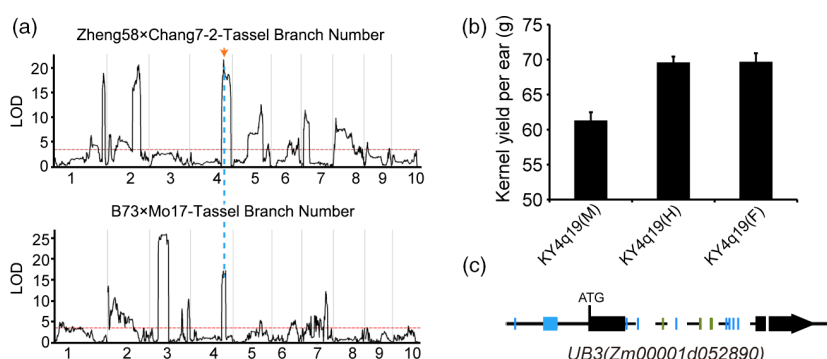
**Figure 5** The plot of heterotic effects estimated in eastern China (N36°01, E117°0) (x-axis) and north-east China (N45°46, E126°52) (y-axis) for 15 QTLs identified in maize Zhengdan958 hybrid. In the plots, the values of  $>\log(2)$  represent overdominance effects,  $0-\log(2)$  represent incomplete dominance effects and  $<0$  represent incomplete recessive effect. The Pearson correlation coefficient is indicated.

2012). Due to the complex maize genome and the complex genetic architecture of heterosis traits, genome-wide high-resolution identification of heterotic loci and comprehensive evaluation of their genetic effects in heterosis have been difficult in maize hybrids. In this study, we constructed three high-density genotype maps and identified 628 loci underlying 19 yield-related traits overall 5360  $F_2$  individuals, which provide a comprehensive list of the heterotic loci in maize hybrids and show that complete-incomplete dominance (the major one) and overdominance effects (the secondary one) for heterozygous genotypes exhibited an important role in heterosis. Of note, the random  $F_2$  populations represent a normal distribution (Figure 2c) which avoids the segregation distortion effect. The results provide a committee resource for breeders and seed industry to improve the potential yield by marker assisted selection and genomic prediction breeding (Frisch *et al.*, 2010; Schrag *et al.*, 2007, 2009, 2010, 2018; Windhausen *et al.*, 2012).

Maize was domesticated from teosinte (*Zea mays ssp. parviglumis*) approximately 9000 years ago and has become the most

productive and highest value commodity crop across the world. One of the main reasons for its rapid popularity in agriculture production globally is the effective use of heterosis. The large-scale genetic and genomic analyses in maize hybrids provided a number of important heterotic loci that may be quite useful in hybrid breeding and future functional studies and showed that the strong better-parent heterosis in maize profited from numerous advantageous alleles interspersing among parental lines. In maize hybrid, there were many QTLs with modest or small effects (probably owing to the presence of multiple paralogs and extensive variation in noncoding regulation regions), and the proportions of their advantageous alleles are much the same between male and female parents (Figure 4). The pattern is distinct from self-fertilizing plant species (e.g. rice), in which advantageous alleles from multiple large-effect QTLs had the chance for pyramid during long-term breeding of inbred lines (Huang *et al.*, 2016). Hence, the mating system (complete outcrossing) and the genome constitution (more paralogs and more intergenic regions) may partially explain the causes of strong heterosis in maize. Moreover, the development of high-throughput, high-density genotyping methods not only provide a resource to discover the genetic basis of heterosis and germplasm diversity, but also benefit to exploiting heterosis-related SNPs and enriching rare advantageous alleles, which could help breeder to improve the hybrid agronomic performance in maize breeding populations.

Another objective of this work is to seek for the common genetic patterns and causative genes contributing to heterosis in plants. From the view of the genome-wide pattern, many heterotic loci with incomplete dominance effects appeared to be the major cause in both rice and maize, with a few overdominance loci acting as the secondary factors. For the key gene contributors, clues have been detected for some important heterosis-related candidates in hybrid maize or hybrid rice which were orthologous to those well-known heterosis genes in other plant species (e.g. *OsSPL14* in rice and *ub3* in maize; *Hd3a* in rice and *SFT* in tomato), probably indicating convergent selection during hybrid breeding of a common set of core nodes related to plant architecture and flowering time pathways. Fine tuning of the function of these important genes related to plant architecture and flowering time probably lead to the occurrence of heterosis in plants.



**Figure 6** Candidate gene analysis of a major QTL *KY4q19*. (a) QTL mapping of tassel branch number in Zheng58  $\times$  Chang7-2 and B73  $\times$  Mo17  $F_2$  populations. The threshold value ( $\text{LOD} = 3.5$ ) is presented by a red horizontal dashed line, and the location of the QTL peak is presented by a blue vertical dashed line. (b) The performances of kernel yield per ear for three genotypes of the *KY4q19* locus in the Zhengdan958 population in Shandong. (c) Sequence differences (coloured in blue and green for SNPs and indels, respectively) between parents Zheng58 and Chang7-2 for the candidate gene *UB3* in the *KY4q19* locus.

Although both previous studies and our study showed complete–incomplete dominance could explain a large proportion of heterosis advantage, epistasis is considered as one of the important causes for heterosis in hybrid plants. The complexification of epistatic is the remaining challenge in heterotic studies. The challenge includes the following: (i) to search for potential epistasis interactions for  $n$  loci,  $n^2/2$  locus–locus interactions need be tested; and (ii) for each locus–locus pairs, there are at least 9 (i.e.  $3 \times 3$ ) allelic combinations. The current analysis algorithm for the epistasis interactions and the limited population size prevent us from conducting an effective epistasis analysis. With the development of well-designed genetic populations and efficient methods for epistasis analysis, we believe more genetic findings on epistasis interactions will help a better understanding of heterosis in maize and other crops.

## Experimental procedures

### Whole-genome sequencing

Three hybrid combinations in maize (Zheng58  $\times$  Chang7-2, B73  $\times$  Mo17 and C428  $\times$  C434) were selected to construct  $F_2$  populations, generating totally 5360 lines for whole-genome sequencing and phenotyping. For each  $F_2$  individual in the hybrid maize populations, the genomic DNA was extracted from fresh leaf tissues using N96 DNAsure Plant Kit (TianGen). Sequencing libraries were constructed using TruePrep Tagment Enzyme (Tn5 transposase) with an insert size of 300–500 bp (Picelli *et al.*, 2014). DNA samples of each  $F_2$  individual were bar-coded in the sequencing libraries and amplified using TruePrep Amplify Enzyme. The indexed DNA samples of  $\sim 384$   $F_2$  individuals with different barcodes were mixed together with a nearly equal molar concentration. Each mixture was loaded into one lane of the Illumina HiSeq4000 system.

### Genotype calling of the populations

Overall, six parental lines (Chang7-2, Zheng58, Mo17, B73, C434 and C428) and 5360  $F_2$  individuals in the three populations were sequenced, generating 150 bp paired-end reads containing totally  $\sim 2.4$  terabase sequences. Each parental line had approximately  $30\times$  genome coverage, and each  $F_2$  individual in the hybrid maize populations had approximately  $0.2\times$  genome coverage. The sequence reads of the six parental lines were aligned against the maize reference genome (AGPv4, downloaded from the Gramene database, [ftp://ftp.gramene.org/pub/gramene/release53/data/fasta/zea\\_mays/dna/](ftp://ftp.gramene.org/pub/gramene/release53/data/fasta/zea_mays/dna/)) with the BWA package (version 0.7.1) using default parameters, and PCR duplicates were removed by the 'MarkDuplicates' module in the Picard tools (version 1.119) (Li and Durbin, 2009). The raw reads were also re-aligned for the highly polymorphic regions using the 'IndelRealigner' function in the software package GenomeAnalysisTK (version 3.4.0; DePristo *et al.*, 2011). The sequence variants between parental lines were called using 'UnifiedGenotyper' in the GenomeAnalysisTK. The paired-end reads of each  $F_2$  individual were mapped onto the maize reference genome sequence using the software package SMALT (version 0.5.7). Only the uniquely mapped reads were used for subsequent SNP calling. Genotype calling of each  $F_2$  individual was carried out based on the SNP alleles between parents, using the pipeline SEG-Map (Huang *et al.*, 2009). For each  $F_2$  population, the genotypic data of each  $F_2$  individual were combined together for bin map constructions.

### Construction of neighbour-joining tree

Resequencing reads from 281 published maize inbred lines (102 teosintes, landraces and modern inbred lines from maize HapmapII and 179 temperate maize breeding lines from maize HapmapIII project) were incorporated with data of six parental lines in this study, and SNPs were called by using GATK pipeline. IBS distance matrix of 287 lines was calculated by Plink (v1.07) with parameters `-maf 0.05 -map3 -noweb -cluster -distance matrix -out`. Then, the neighbour-joining tree was constructed by phylip (v3.696) with default parameters.

### Planting and phenotyping of the $F_2$ individuals

The six inbred lines were used for population construction. In brief, sixty kernels from a single ear of Chang7-2, Zheng58, Mo17, B73, C434 and C428 were planted in Shandong Agricultural University Research Farm (eastern China, N36°01', E117°0'), Tai'an, Shandong province, China, in 2016. The  $F_1$  seeds from crosses of Zheng58  $\times$  Chang7-2, B73  $\times$  Mo17 and C428  $\times$  C434 were generated in this season. Subsequently, the  $F_1$  seeds of three combinations were planted for selfing in Shandong Agricultural University Research Nursery (South China, N18°24', E109°01'), Sanya, Hainan province, China, in 2016 winter. For each combination, the  $F_2$  population were planted in Shandong Agricultural University Research Farm (eastern China, N36°01', E117°0'), Tai'an, Shandong province, China, in 2017. All the  $F_2$  individuals were grown in the consecutive farmland with well-distributed soil status in the field. The ears of all the  $F_2$  individuals were open pollinated for evaluating yield traits. Detail description of phenotyping works for 19 agronomy traits can be found in Figure S2.

### Quantitative genetic analyses

In each  $F_2$  population, QTL analysis of 19 agronomic traits was carried out using the composite interval-mapping (CIM) method with QTL Cartographer (version 2.5) with a step size of 2 cM, the cross type of SF<sub>2</sub> and a window size of 10 cM. The calculations of LOD values were based on likelihood ratio tests allowing both additive and dominance effects (Silva Lda *et al.*, 2012). QTLs with LOD value higher than 3.5 were used for follow-up heterosis analyses. After the QTLs were identified, genetic effects were assessed through comparisons of both homozygous genotypes and the heterozygous genotypes to search for advantageous alleles. The average phenotypic measurements of heterozygous genotypes and homozygous genotypes were further calculated for the estimation of the dominance effect index ( $d/a$ ) and the MPH index for each QTL. According to the estimates of dominance effects (short for  $d$ ) with those of additive effects (short for  $a$ ), each yield QTL underlying a certain trait could be divided into one of the five cases: overdominance ( $d/a > 1.25$ , in which the MF genotype displays more yield than both MM and FF), complete–incomplete dominance ( $0.25 < d/a \leq 1.25$ ), additive effect ( $-0.25 < d/a \leq 0.25$ ), incomplete recessive ( $-1.25 < d/a \leq -0.25$ ) and underdominance ( $d/a \leq -1.25$ , in which the MF genotype displays less yield than both MM and FF). In the replication experiment, another set of 2225  $F_2$  individuals from ZhengDan958 population were planted in Northeast Agricultural University Research Farm (N45°46', E126°52'), Heilongjiang, in 2018 for phenotyping. The  $F_2$  individuals were all genotyped at

17 target loci using a ligase detection reaction method. Statistic tests were performed through the analysis of variance (ANOVA), and the heterotic effects (the estimates of *d/a* values) were calculated as that in Shandong.

### Candidate gene analysis

The protein sequences of *ub3* were compared with its homologs in maize and rice using BLAST, ClustalW and MEGA5. PCR primers were designed for *ub3* for Sanger-based sequencing of the parental lines in AB3730. The PCR primers include *ub3* + 1106R: 5'CTGA GGGCCAACTGATCATGCCA3', *ub3* + 2211F: 5'CTGGCCTAGCC TAGTAGGAAGAGA3', *ub3* + 2394R: 5'GGTCAAAGCCAGTGGG AATGCT3', *ub3* + 3782R: 5'CTGGATGCTCTAGCCTCTAGGCA3', *ub3*-114F: 5'GTCAGCCAGCCAGCCCAAGA3', *ub3* + 1001F: 5'GTGCTCATTGCTTAGTCTGCCTCA3', *ub3* + 69R: 5'GCCGGCG TCCTCGAAGTAGAT3', *ub3*-3046F: 5'GGCGTGTITAGCATTAG CTCC3', *ub3*p2-3F: 5'CACTAGAACCAACGGTATACATGAC3', *ub3*p2-3R: 5'ATACCTGAGCCGACGATCTCAT3', *ub3*p3-3F: 5'TTGTAGACGCCAGGTAAGTCTGCTGT3', *ub3*p3-3R: 5'GTCAT GTATACCGTTGGTTCTAGTG3'.

The first eight primers were designed in this study, and the last four primers (*ub3*p2-3F/R and *ub3*p3-3F/R) were according to a previous study (Liu *et al.*, 2015). For exon and intron regions, the *ub3*-114F/*ub3* + 3782R pair was used for long fragment amplifications, while *ub3*-114F, *ub3* + 1106R, *ub3* + 1001F, *ub3* + 2394R, *ub3* + 2211F and *ub3* + 3782R primers were used for sequencing. For promoter regions, the *ub3*-3046F/*ub3* + 69R pair was used for long fragment amplifications, while the remaining primers (*ub3*-3046F, *ub3* + 69R, *ub3*p2-3F/R and *ub3*p3-3F/R) were used for sequencing.

### Accession numbers

The DNA sequencing data of the six parental lines are deposited in the European Nucleotide Archive under accession numbers PRJEB30082 (<http://www.ebi.ac.uk/ena/data/view/PRJEB30082>).

### Acknowledgements

This work was supported by State Key Laboratory of Crop Biology Foundation, the National Natural Science Foundation of China (31825015 to X.H., 31771799 to H.L. and 31801027 to Q.W.), Shanghai Municipal Education Commission (2017-01-07-00-02-E00039 to X.H.), the National Plant Transgenic Program (2016ZX08003-003 to X.Z) and Program of Shanghai Academic Research Leader (18XD1402900 to X.H.).

### Competing interests

The authors declare no competing financial interests.

### Author contributions

H.L., X.H. and X.S.Z. designed the research; Y.D., L.Z., X.L., C.K., M.S., F.L., Y.W., Y.Z., B.L. and X.Y.Z. contributed to create the genetic population and phenotyping for the whole related traits. Q.W., X.Y., J.L., M.C., C.Z., Q.T., Y.L., D.F., J.S. and Q.F. performed the genome sequencing and expression analysis. X.H., H.L., J.Y. and M.C. performed the data analysis and genetics analysis. X.H., H.L., B.H., J.Y., J.L. and X.S.Z. discussed the data and wrote the paper.

### References

- Birchler, J.A. (2016) Plant science: hybrid vigour characterized. *Nature*, **537**, 620–621.
- Birchler, J.A., Johnson, A.F. and Veitia, R.A. (2016) Kinetics genetics: incorporating the concept of genomic balance into an understanding of quantitative traits. *Plant Sci.* **245**, 128–134.
- Bruce, A.B. (1910) The mendelian theory of heredity and the augmentation of vigor. *Science*, **32**, 627–628.
- Bukowski, R., Guo, X., Lu, Y., Zou, C., He, B., Rong, Z., Wang, B. *et al.* (2018) Construction of the third-generation *Zea mays* haplotype map. *Gigascience*, **7**, 1–12.
- Chia, J.M., Song, C., Bradbury, P.J., Costich, D., de Leon, N., Doebley, J., Elshire, R.J. *et al.* (2012) Maize HapMap2 identifies extant variation from a genome in flux. *Nat. Genet.* **44**, 803–807.
- Chuck, G.S., Brown, P.J., Meeley, R. and Hake, S. (2014) Maize SBP-box transcription factors unbranched2 and unbranched3 affect yield traits by regulating the rate of lateral primordia initiation. *Proc. Natl. Acad. Sci. USA*, **111**, 18775–18780.
- Dapp, M., Reinders, J., Bediee, A., Balsera, C., Bucher, E., Theiler, G., Granier, C. *et al.* (2015) Heterosis and inbreeding depression of epigenetic Arabidopsis hybrids. *Nat. Plants*, **1**, 15092.
- DePristo, M.A., Banks, E., Poplin, R., Garimella, K.V., Maguire, J.R., Hartl, C., Philippakis, A.A. *et al.* (2011) A framework for variation discovery and genotyping using next-generation DNA sequencing data. *Nat. Genet.* **43**, 491.
- East, E.M. (1936) Heterosis. *Genetics*, **21**, 375–397.
- Frisch, M., Thiemann, A., Fu, J., Schrag, T.A., Scholten, S. and Melchinger, A.E. (2010) Transcriptome-based distance measures for grouping of germplasm and prediction of hybrid performance in maize. *Theor. Appl. Genet.* **120**, 441–450.
- Garcia, A.A., Wang, S., Melchinger, A.E. and Zeng, Z.B. (2008) Quantitative trait loci mapping and the genetic basis of heterosis in maize and rice. *Genetics*, **180**, 1707–1724.
- Hollick, J.B. and Chandler, V.L. (1998) Epigenetic allelic states of a maize transcriptional regulatory locus exhibit overdominant gene action. *Genetics*, **150**, 891–897.
- Hua, J.P., Xing, Y.Z., Xu, C.G., Sun, X.L., Yu, S.B. and Zhang, Q. (2002) Genetic dissection of an elite rice hybrid revealed that heterozygotes are not always advantageous for performance. *Genetics*, **162**, 1885–1895.
- Huang, X., Feng, Q., Qian, Q., Zhao, Q., Wang, L., Wang, A., Guan, J. *et al.* (2009) High-throughput genotyping by whole-genome resequencing. *Genome Res.* **19**, 1068–1076.
- Huang, X., Yang, S., Gong, J., Zhao, Q., Feng, Q., Zhan, Q., Zhao, Y. *et al.* (2016) Genomic architecture of heterosis for yield traits in rice. *Nature*, **537**, 629–633.
- Jiao, Y., Wang, Y., Xue, D., Wang, J., Yan, M., Liu, G., Dong, G. *et al.* (2010) Regulation of OsSPL14 by OsmiR156 defines ideal plant architecture in rice. *Nat. Genet.* **42**, 541–544.
- Jiao, Y., Zhao, H., Ren, L., Song, W., Zeng, B., Guo, J., Wang, B. *et al.* (2012) Genome-wide genetic changes during modern breeding of maize. *Nat. Genet.* **44**, 812–815.
- Jones, D.F. (1917) Dominance of linked factors as a means of accounting for heterosis. *Genetics*, **2**, 609.
- Ko, D.K., Rohozinski, D., Song, Q., Taylor, S.H., Juenger, T.E., Harmon, F.G. and Chen, Z.J. (2016) Temporal shift of circadian-mediated gene expression and carbon fixation contributes to biomass heterosis in maize hybrids. *PLoS Genet.* **12**, e1006197.
- Krieger, U., Lippman, Z.B. and Zamir, D. (2010) The flowering gene single flower truss drives heterosis for yield in tomato. *Nat. Genet.* **42**, 459–463.
- Kusmec, A., Srinivasan, S., Nettleton, D. and Schnable, P.S. (2017) Distinct genetic architectures for phenotype means and plasticities in *Zea mays*. *Nat. Plants*, **3**, 715–723.
- Li, H. and Durbin, R. (2009) Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics*, **25**, 1754–1760.
- Li, X., Li, X., Fridman, E., Tesso, T.T. and Yu, J. (2015) Dissecting repulsion linkage in the dwarfing gene *Dw3* region for sorghum plant height provides insights into heterosis. *Proc. Natl. Acad. Sci. USA*, **112**, 11823–11828.

- Li, D., Huang, Z., Song, S., Xin, Y., Mao, D., Lv, Q., Zhou, M. et al. (2016) Integrated analysis of phenome, genome, and transcriptome of hybrid rice uncovered multiple heterosis-related loci for yield increase. *Proc. Natl. Acad. Sci. USA*, **113**, E6026–E6035.
- Liu, L., Du, Y., Shen, X., Li, M., Sun, W., Huang, J., Liu, Z. et al. (2015) KRN4 controls quantitative variation in maize kernel row number. *PLoS Genet.* **11**, e1005670.
- Liu, J., Li, M., Zhang, Q., Wei, X. and Huang, X. (2019) Exploring the molecular basis of heterosis for plant breeding. *J. Integr. Plant Biol.* <https://doi.org/10.1111/jipb.12804>.
- Lu, H., Romero-Severson, J. and Bernardo, R. (2003) Genetic basis of heterosis explored by simple sequence repeat markers in a random-mated maize population. *Theor. Appl. Genet.* **107**, 494–502.
- Luo, X., Wu, S., Tian, F., Xin, X., Zha, X., Dong, X., Fu, Y. et al. (2011) Identification of heterotic loci associated with yield-related traits in Chinese common wild rice (*Oryza rufipogon* Griff.). *Plant Sci.* **181**, 14–22.
- Ma, J., Zhang, D., Cao, Y., Wang, L., Li, J., Lubberstedt, T., Wang, T. et al. (2018) Heterosis-related genes under different planting densities in maize. *J. Exp. Bot.* **69**, 5077–5087.
- Miura, K., Ikeda, M., Matsubara, A., Song, X.J., Ito, M., Asano, K., Matsuoka, M. et al. (2010) OsSPL14 promotes panicle branching and higher grain productivity in rice. *Nat. Genet.* **42**, 545–549.
- Picelli, S., Bjorklund, A.K., Reinius, B., Sagasser, S., Winberg, G. and Sandberg, R. (2014) Tn5 transposase and tagmentation procedures for massively scaled sequencing projects. *Genome Res.* **24**, 2033–2040.
- Riedelshimer, C., Czedik-Eysenberg, A., Grieder, C., Lisek, J., Technow, F., Sulpice, R., Altmann, T. et al. (2012) Genomic and metabolic prediction of complex heterotic traits in hybrid maize. *Nat. Genet.* **44**, 217–220.
- Rodgers-Melnick, E., Bradbury, P.J., Elshire, R.J., Glaubitz, J.C., Acharya, C.B., Mitchell, S.E., Li, C. et al. (2015) Recombination in diverse maize is stable, predictable, and associated with genetic load. *Proc. Natl. Acad. Sci. USA*, **112**, 3823–3828.
- Schnable, P.S. and Springer, N.M. (2013) Progress toward understanding heterosis in crop plants. *Annu. Rev. Plant Biol.* **64**, 71–88.
- Schrag, T.A., Maurer, H.P., Melchinger, A.E., Piepho, H.P., Peleman, J. and Frisch, M. (2007) Prediction of single-cross hybrid performance in maize using haplotype blocks associated with QTL for grain yield. *Theor. Appl. Genet.* **114**, 1345–1355.
- Schrag, T.A., Mohring, J., Maurer, H.P., Dhillon, B.S., Melchinger, A.E., Piepho, H.P., Sorensen, A.P. et al. (2009) Molecular marker-based prediction of hybrid performance in maize using unbalanced data from multiple experiments with factorial crosses. *Theor. Appl. Genet.* **118**, 741–751.
- Schrag, T.A., Mohring, J., Melchinger, A.E., Kusterer, B., Dhillon, B.S., Piepho, H.P. and Frisch, M. (2010) Prediction of hybrid performance in maize using molecular markers and joint analyses of hybrids and parental inbreds. *Theor. Appl. Genet.* **120**, 451–461.
- Schrag, T.A., Westhues, M., Schipprack, W., Seifert, F., Thiemann, A., Scholten, S. and Melchinger, A.E. (2018) Beyond genomic prediction: combining different types of omics data can improve prediction of hybrid performance in maize. *Genetics*, **208**, 1373–1385.
- Shull, G.H. (1908) The composition of a field of maize. *J. Hered.* **os-4**, 296–301.
- Silva Lda, C., Wang, S. and Zeng, Z.B. (2012) Composite interval mapping and multiple interval mapping: procedures and guidelines for using Windows QTL Cartographer. *Methods Mol. Biol.* **871**, 75–119.
- Swanson-Wagner, R.A., Jia, Y., DeCook, R., Borsuk, L.A., Nettleton, D. and Schnable, P.S. (2006) All possible modes of gene action are observed in a global comparison of gene expression in a maize F1 hybrid and its inbred parents. *Proc. Natl. Acad. Sci. USA*, **103**, 6805–6810.
- Tang, J., Yan, J., Ma, X., Teng, W., Wu, W., Dai, J., Dhillon, B.S. et al. (2010) Dissection of the genetic basis of heterosis in an elite maize hybrid by QTL mapping in an immortalized F<sub>2</sub> population. *Theor. Appl. Genet.* **120**, 333–340.
- Wang, H., Zhang, X., Yang, H., Liu, X., Li, H., Yuan, L., Li, W. et al. (2016) Identification of heterotic loci associated with grain yield and its components using two CSSL test populations in maize. *Sci. Rep.* **6**, 38205.
- Wang, C., Liu, Q., Shen, Y., Hua, Y., Wang, J., Lin, J., Wu, M. et al. (2019) Clonal seeds from hybrid rice by simultaneous genome engineering of meiosis and fertilization genes. *Nat. Biotechnol.* **37**, 283–286.
- Windhausen, V.S., Atlin, G.N., Hickey, J.M., Crossa, J., Jannink, J.L., Sorrells, M.E., Raman, B. et al. (2012) Effectiveness of genomic prediction of maize hybrid performance in different breeding populations and environments. *G3: Genes-Genomes-Genet.* **2**, 1427–1436.
- Yang, J., Mezouk, S., Baumgarten, A., Buckler, E.S., Guill, K.E., McMullen, M.D., Mumm, R.H. et al. (2017a) Incomplete dominance of deleterious alleles contributes substantially to trait variation and heterosis in maize. *PLoS Genet.* **13**, e1007019.
- Yang, M., Wang, X., Ren, D., Huang, H., Xu, M., He, G. and Deng, X.W. (2017b) Genomic architecture of biomass heterosis in Arabidopsis. *Proc. Natl. Acad. Sci. USA*, **114**, 8101–8106.
- Yao, H., Dogra Gray, A., Auger, D.L. and Birchler, J.A. (2013) Genomic dosage effects on heterosis in triploid maize. *Proc. Natl. Acad. Sci. USA*, **110**, 2665–2669.
- Zhou, G., Chen, Y., Yao, W., Zhang, C., Xie, W., Hua, J., Xing, Y. et al. (2012) Genetic composition of yield heterosis in an elite rice hybrid. *Proc. Natl. Acad. Sci. USA*, **109**, 15847–15852.

## Supporting information

Additional supporting information may be found online in the Supporting Information section at the end of the article.

**Figure S1** Genome-wide distribution of the genetic variation throughout the maize genome for Zheng58 × Chang7-2, B73 × Mo17 and C428 × C434, respectively.

**Figure S2** Detail descriptions of phenotyping works for 19 agronomical traits.

**Figure S3** Distribution of each phenotype across Zhengdan958 F<sub>2</sub> populations, parental lines, and F<sub>1</sub> hybrid lines.

**Figure S4** The plots and the correlation coefficient of MPH and BPH of the 19 traits between F<sub>2</sub> populations.

**Figure S5** Overlapping QTLs between populations.

**Table S1** Phenotype data of parental inbreds and F<sub>1</sub> hybrids for all 19 traits in three hybrids.