Rice

Genome-Based Identification of Heterotic Patterns in Rice

Ulrike Beukert[1], Zuo Li[1], Guozheng Liu[1], Yusheng Zhao[1], Nadhigade Ramachandra[2], Vilson Mirdita[3], Fabiano Pita[4], Klaus Pillen[5] and Jochen Christoph Reif[1*]

Abstract

Background: Hybrid rice breeding facilitates to increase grain yield and yield stability. Long-term success of hybrid breeding depends on the recognition of high-yielding complementary heterotic patterns, which is lacking in crops like rice.

Result: The main goal of this study was to evaluate the potential and limits to use genomics for establishing heterotic patterns in rice. For this purpose, data of a commercial hybrid rice breeding program targeted to India was analyzed, including 1,960 phenotyped hybrids from three market segments and 262 genotyped parental lines. Our cross-validation study revealed that grain yield of all potential single-crosses can be accurately predicted. Based on the full matrix of hybrid performances, high-yielding heterotic patterns were identified. These heterotic patterns increased grain yield up to 9% compared to the currently employed groups. Heterotic groups of around 14 individuals reflect a good compromise between long-term and short-term selection response.

Conclusions: Our findings clearly underlined the benefits of a genome-based establishment of heterotic patterns in rice as a requirement for a sustainable long-term success of hybrid rice breeding.

Keywords: Heterotic Groups, Heterotic Pattern, Hybrid Rice, Genome-Wide Predictions

Background

Rice belongs to the three leading food crops providing the majority of calories to feed the world (International Rice Research Institute 2013). The demand for rice is raising steadily due to changes in customer priorities and population growth (Khush 2013). Thus, rice production has to be increased. Unfortunately, the potential of expanding the cultivated area is limited (Khush 2005) and, therefore, rice production has to be increased by enhancing yields per area (Khush 2013). Nevertheless, worldwide selection gain in rice is only 1.0% per year and much lower than the required increase in yield potential of 2.4% (Ray et al. 2013). Thus, there exits the strong need to develop and implement improved breeding tools (Phillips 2010; Zhu et al. 2016). A promising approach to boost yield per area consists in hybrid breeding, which has been successfully applied in rice improvement for some target regions such as China (Khush 2005; Khush 2013).

The advantages of hybrid versus line breeding are the exploitation of heterosis resulting in higher grain yield (Xu et al. 2014) and an enhanced yield stability (Longin et al. 2012). Furthermore, hybrid breeding simplifies to stack major dominant genes and results in substantial return on investments, which is required to refinance future breeding progress (Longin et al. 2012). Hybrid rice breeding programs were initiated in China in 1964 based on a cytoplasmic male sterility (CMS) system. The first hybrid rice variety was released in 1976 (Barclay 2010). In addition to CMS, photoperiod sensitive male sterility has been exploited to produce hybrid seeds on a large scale (Chen and Liu 2014; Li et al. 2007). Rice grain yield per area increased through hybrid breeding by approximately 40% during the past three decades (Zhu et al. 2016). As a consequence, hybrid varieties covered more than 50% of China's total rice growing area in 2003 (Cheng et al. 2007a). Moreover, the success of hybrid rice in China encouraged

* Correspondence: reif@ipk-gatersleben.de
[1]Department of Breeding Research, Leibniz Institute of Plant Genetics and Crop Plant Research (IPK) Gatersleben, Corrensstraße 3, 06466 Stadt Seeland, Germany
Full list of author information is available at the end of the article

embarking on a national program for the development of hybrid rice for India in 1989 (Viraktamath 2010).

The optimum exploitation of heterosis requires that the available germplasm is structured into genetically diverse heterotic groups. A heterotic group refers to a collection of genotypes resulting in similar hybrid performance when crossed with individuals belonging to a complementary and genetically distinct germplasm group. A specific combination of two heterotic groups leading to high-yielding hybrids is defined as heterotic pattern (Melchinger and Gumber 1998). Hybrid breeding based on the concept of heterotic patterns leads to more pronounced variance of the breeding values in contrast to the variance of the dominance deviations, which enhances recurrent selection gain (Reif et al. 2007).

Heterotic patterns have been established in the past either considering hybrid seed production traits such as seed yield of the female lines and high pollination capability of the male lines (Reif et al. 2005) or they have been developed empirically by testing combining ability among available germplasm (Fischer et al. 2010). The latter approach is afflicted by the large number of possible hybrid combinations among available elite inbred lines. To solve this challenge, Zhao et al. (2015) developed a three-step approach to search for heterotic patterns: (1) The performance of all possible single-cross combinations is determined through genome-wide or metabolite-based hybrid predictions. (2) The predicted hybrid performances are used to identify superior heterotic patterns based on a simulated annealing algorithm. (3) The optimum size of the heterotic pattern is determined balancing the expected short- and long-term selection.

The genetically distant subgroups of *indica* and *japonica* have been suggested as a promising heterotic pattern for hybrid rice breeding (Cheng et al. 2007a). Nevertheless, fertility barriers hamper so far the exploitation of this heterotic pattern (Ikekashi and Araki 1984; Liu et al. 1996). Several approaches have been used to reduce the sterility occurring when crossing *indica* and *japonica* (Guo et al. 2016; Ouyang et al. 2010), but none of them has been successfully implemented into hybrid rice breeding programs. The search for heterotic groups within the major germplasm pools is limited to studies based on tropical *indica* rice (Wang et al. 2014; Xie et al. 2013). Moreover, previous studies used molecular marker-based genetic distances as proxy of the heterosis or hybrid performance, which is unreliable for unrelated parental inbred lines (Melchinger 1999; Xu et al. 2014; Xu et al. 2016). Thus, other approaches exploiting for example genome-wide hybrid predictions are needed to recognize promising heterotic groups for hybrid rice breeding.

Our study is based on genomic and phenotypic data of 1,960 rice hybrids adapted to the Indian sub-continent, which have been evaluated for grain yield in two to four locations. The hybrids were grouped based on grain size, shape and appearance into three market segments: Long grain (LS) segment with grain type of long slender, length of larger than 6 mm, length/breadth ratio of larger than 3; Medium (MM) grain segment with grain type of medium slender, length of less than 6 mm, length/breadth ratio of 2.5 to 3.0 mm; Short (SS) grain segment with grain type of short slender, length of less than 6 mm, length/breadth ratio of larger than 3. Our main goal was to evaluate the potential and limits of genome-based establishment of heterotic patterns in rice. In particular, the objectives were to (1) investigate the accuracy for genome-wide prediction of hybrid performance in rice, (2) evaluate the benefits of heterotic patterns identified with a simulated annealing algorithm, and (3) assess the optimal size of the heterotic patterns balancing short- and long-term selection gain.

Results

Variance due to specific combining ability effects played a prominent role in hybrid rice populations

We observed a wide range of Best Linear Unbiased Estimations leading to significant ($P < 0.001$) genetic variances within all three market segments (Fig. 1). In total, 19, 4, and 10% of the hybrids significantly ($P < 0.05$) outperformed leading commercial varieties for the market

Fig. 1 Best linear unbiased estimations for grain yield performance of hybrids and checks of the market segments LS, MM, and SS
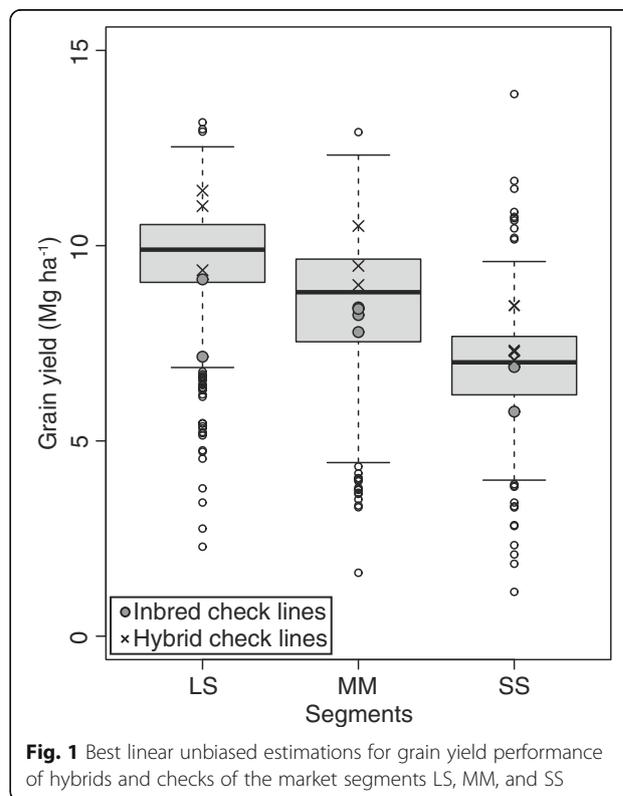
segments LS, MM, and SS, respectively. Genetic variance components involving general combining ability effects σ^2_{GCA} showed higher values for male than for female lines (Table 1). The variance of specific combining ability effects σ^2_{SCA} was for all segments significantly ($P < 0.05$) larger than zero and contributed on average to 42% of the total genetic variance. Heritability estimates ranged from 0.26 to 0.61.

Analyses of population structure revealed genetically distinct parental pools for market segments MM and SS but not for LS

The analyses of linkage disequilibrium (LD) revealed a fast decay of LD with increasing physical distance (Additional file 1: Figure S1). The sharp decay of LD underlines the potential of using the populations for high-resolution genome-wide association mapping. The population structure was examined by applying principal coordinate analyses. The analyses of the 262 genotyped lines revealed absence of a population structure among market segments (Additional file 1: Figure S2). Despite this, differences in quality requirements are large and parents are used only rarely across segments MM, SS, and LS. Thus, we analyzed in the following each market segment separately. In segment LS, male and female lines did not clustered separately (Fig. 2). This was further substantiated by the distributions of Rogers' distances with highest values between male lines as well as between male and female lines (Fig. 2b). In contrast, for segments MM and SS, principal coordinate analyses and distributions of pairwise Rogers' distances revealed genetically distinct parental pools (Fig. 2c, d, e and f).

Hybrid performance can be predicted with high accuracy

The predicted hybrid performance values of the full diallel matrix formed the basis to search for promising heterotic patterns. We used a chess-board like cross-validation strategy to evaluate the prediction ability of the hybrid performance (Additional file 1: Figure S3). The hybrids were split into an estimation set and the test groups T2, T1, and T0 with decreasing relationship to the estimation set. The accuracy observed in the T2 scenario is relevant for compiling the full hybrid performance matrix, because the parents of all predicted hybrids were evaluated in other single-cross combinations. The prediction abilities ranged for the T2 scenario from 0.33 to 0.58 (Fig. 3). The prediction accuracies were estimated by standardizing the prediction abilities with the square root of heritabilities and amounted to 0.72 averaged across the three market segments.

Detected heterotic patterns are stable across varying group sizes

We used the predicted performances of all potential single-cross hybrids in order to identify high-yielding heterotic patterns with population sizes ranging from two to 20 lines using a previously developed simulated annealing algorithm. In this search, we assumed that lines can be clustered irrespective to their restoration ability. For all three segments, selection and clustering of lines into heterotic groups were stable across the full range of examined population sizes (Additional file 1: Table S1). Interestingly, only few female lines were selected to establish heterotic patterns: In market segment MM, no female line was selected. In the market segments LS and SS, 38 and 40% of the females were selected, respectively. The selected heterotic patterns substantially outperformed the overall mean with a maximum difference ranging from 14% for market segment SS to 28% for market segment MM (Fig. 4a, b and c).

The performances of the identified heterotic patterns were compared with the average hybrid performance of crosses between male and female lines. In order to make a fair comparison, we focused on a standardized population size of 256 hybrids. The hybrids of the identified heterotic patterns surpassed the superior female and male single crosses by 2, 9, and 7% for market segments LS, MM, and SS, respectively (Fig. 5).

Short- and long-term success of the identified heterotic patterns

We used the theoretical selection limit as a parameter to assess the long-term success of the detected heterotic groups. The selection limit increased with growing heterotic group size and plateaued at around 14 individuals across the different market segments (Fig. 4a, b and c). We further assessed the long-term success of the heterotic patterns by estimating genetic

Table 1 Second degree statistics for hybrid grain yield (Mg ha^{-1}) experiments in market segments LS, MM, and SS performed in three, two, and four environments, respectively

Source	LS	MM	SS
$\sigma^2_{Genotype}$	0.99***	0.53***	0.81***
$\sigma^2_{GCA\ male}$	0.30**	0.27	0.24***
$\sigma^2_{GCA\ female}$	0.24**	0.13	0.11***
σ^2_{SCA}	0.45***	0.13*	0.46***
$\sigma^2_{Genotype\ x\ Location}$	1.48***	2.10***	0.90***
$\sigma^2_{GCA\ male\ x\ Location}$	0.88***	1.39***	0.33***
$\sigma^2_{GCA\ female\ x\ Location}$	0.11***	0.26***	0.12***
$\sigma^2_{SCA\ x\ Location}$	0.49***	0.45**	0.45***
σ^2_{error}	0.59	0.97	1.34
h^2	0.59	0.26	0.61

Probability level: *** < 0.001, ** < 0.01, * < 0.05, not assigned not significant

Fig. 2 Principal coordinate analyses of parental lines (**a**) for market segments LS, (**c**) MM, and (**e**) SS with relating distribution of Rogers' distances within and between parental pools (**b**) for market segments LS, (**d**) MM, and (**f**) SS
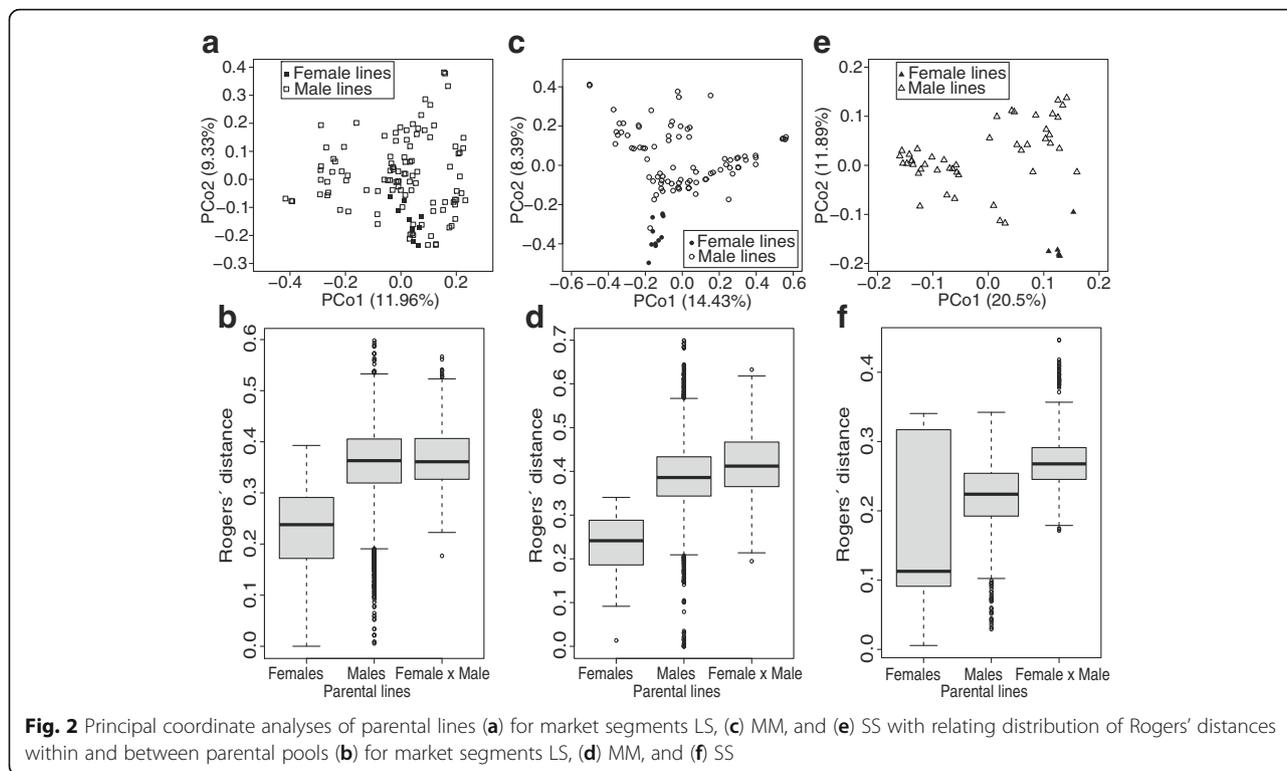
representativeness, which also plateaued at around 14 genotypes. In contrast, the short-term success assessed as the population mean was highest at small population sizes and decreased nearly linearly with increasing size of the heterotic patterns.

Fig. 3 Prediction ability of grain yield performance for different subgroups of market segments LS, MM, and SS

Discussion

More than 50% of the total cultivated rice area in China is grown with hybrid varieties, which was enabled by substantial investments in rice research and breeding (Cheng et al. 2007b). Encouraged by the success of hybrid rice in enhancing the rice production and productivity in China, India initiated a national program for the development and large scale adaption of hybrid rice in 1989 (Viraktamath 2010). Plant breeders in the government and private sectors have launched hybrid varieties for different states in India. These hybrids have a 10 to 44% higher yield compared to popular high-yielding line varieties (Wanjari et al. 2006). Thus, a good beginning has been made by ushering in to an era of hybrid rice breeding and production in India (Spielman et al. 2013). Long-term success of hybrid breeding depends on the establishment of complementary heterotic groups (Reif et al. 2007; Zhao et al. 2015). Nevertheless, heterotic groups in rice are not clearly defined (Xie et al. 2012). This encouraged us to assess the potential and limits of a recently developed three-step approach to search for high-yielding heterotic patterns using data of a commercial hybrid rice breeding program. The hybrid rice breeding program is centered on a CMS hybrid seed production strategy. Lines have been clustered based on their restoration ability into female and male groups. The molecular analyses revealed for two of the examined rice
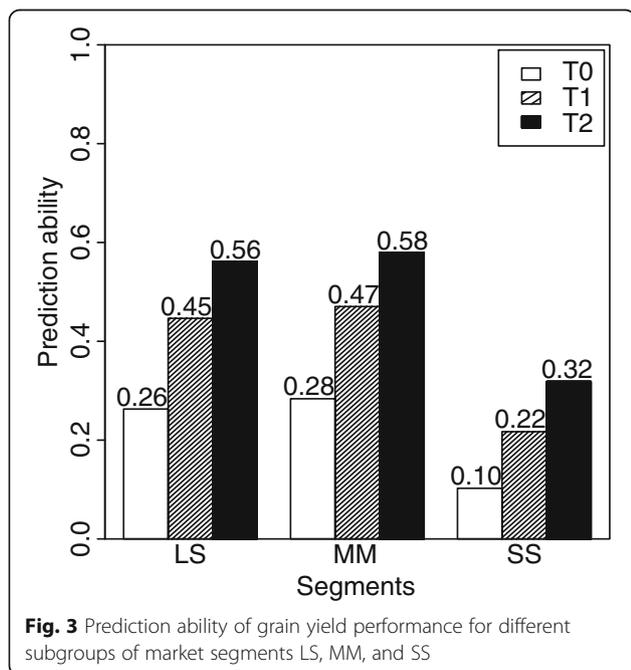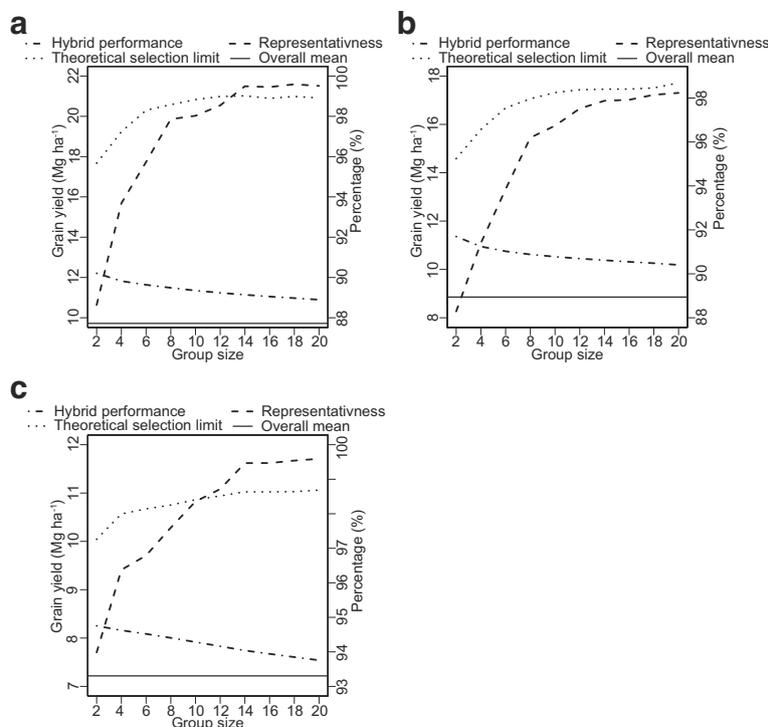
Fig. 4 Short- (Hybrid performance) and long-term success (Representativeness, Theoretical selection limit) in dependence of heterotic group size for market segments (**a**) LS, (**b**) MM, and (**c**) SS

market segments that male and female pools were genetically divergent (Fig. 2). Two major restorer genes *Rf3* and *Rf4* are known for the underlying CMS-WA cytoplasm (Zhang et al. 1997; Zhang et al. 2002). Thus, a conversion of female into male lines or vice versa can be realized but is sometimes hampered by modifier genes influencing the penetrance of the restorer genes. Despite this, we assumed in our study that the

hybridization system is not restricting the grouping of lines.

Hybrids significantly surpassed yield performance of commercial line varieties

The yield advantage of rice hybrids over released line varieties, which is often denoted as commercial heterosis, was reported to be in the range of 10 to 20% (Cheng et al. 2007a; Huang et al. 2016; Wang et al. 2014; Xie et al. 2013). We estimated the magnitude of the commercial heterosis by contrasting grain yield of the phenotyped hybrids with the performance of the highest-yielding line variety included in the trials. The commercial heterosis amounted to 25, 34, and 37% for the segments LS, MM, and SS, respectively (Fig. 1). The slightly higher values compared to the previously reported ones (Cheng et al. 2007a; Huang et al. 2016; Wang et al. 2014; Xie et al. 2013) can be explained by the moderate phenotyping intensity and the augmented design used in our study with unreplicated evaluation of hybrids but replicated evaluation of line varieties. The different allocation of resources for checks and hybrids was caused by restrictions in hybrid seed production. Summarizing, our findings underlined the yield advantage of hybrids over inbred lines and underpinned the potential to boost grain yield through hybrid rice breeding.
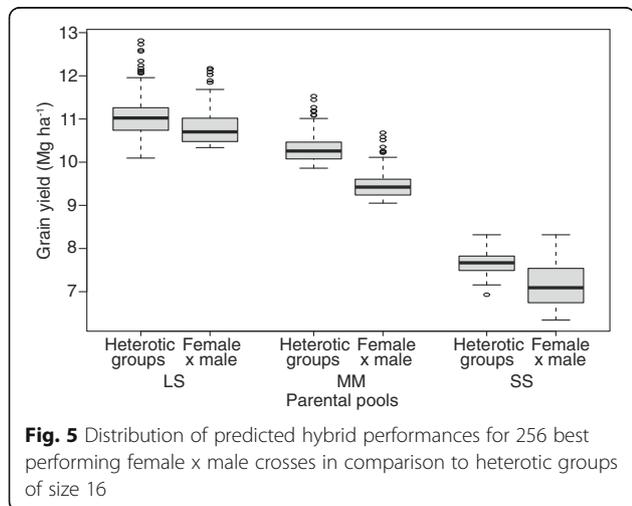
Fig. 5 Distribution of predicted hybrid performances for 256 best performing female x male crosses in comparison to heterotic groups of size 16

Pronounced variance of specific combining ability effects

A predominance of σ^2_{GCA} over σ^2_{SCA} leads to an accurate prediction of the hybrid performance based on GCA effects (Zhao et al. 2013). The experimental resources, which need to be installed to estimate GCA effects, are a linear function of the number of parents. This is in contrast to the resources required to estimate SCA effects, which are a quadratic function of the number of parents.

The incomplete factorial mating designs for the market segments LS, SS, and MM differed in the number of non-phenotyped hybrids (Additional file 1: Figure S4). Nevertheless, this should not bias the estimates of the variance components as we can assume missing at random (Little and Rubin 2002). A favorable ratio of σ^2_{GCA} versus σ^2_{SCA} is expected in crops with genetically divergent heterotic groups such as in hybrid maize breeding (Reif et al. 2007). In contrary to this, we observed that σ^2_{SCA} amounts to 25 to 57% of the total genetic variance in the hybrid populations for the three market segments (Table 1). The σ^2_{SCA} is higher compared to previous estimates in other selfing species such as wheat (Zhao et al. 2015) or barley (Philipp et al. 2016) and can be explained only partially by a lack of genetic divergence between the parental lines (Fig. 2). Another reason for the low estimates of σ^2_{GCA} is the reduced diversity in the pool of female lines, which is most likely due to the bottleneck in selecting suitable maintainer lines. The relevance of σ^2_{SCA} indicates that it is challenging to identify superior hybrid combinations based on GCA effects. Moreover, only slow recurrent selection gain is expected when interpreting σ^2_{GCA} as an estimate of the variance of the breeding values.

High accuracy observed for genome-wide hybrid prediction

We implemented ridge regression best linear unbiased prediction (RR-BLUP) of the hybrid performances considering additive and dominance effects (Zhao et al. 2014). The ability of prediction was evaluated applying a chess-board like cross-validation scenario (Additional file 1: Figure S3). The prediction ability averaged across the three market segments increased with increasing relatedness between the estimation and test sets from 0.21 for the T0 scenario to 0.49 for the T2 scenario (Fig. 3). These findings confirmed the high relevance of relatedness as the driving force of the prediction ability (Desta and Ortiz 2014; Zhao et al. 2015).

The prediction accuracy was estimated by standardizing the prediction ability with the square root of heritability estimates to compare our findings across market segments and also with other studies. Prediction accuracies were comparable for segment LS and MM but lower for segment SS (Fig. 3). This difference can be explained by a smaller population size (Additional file 1: Table S2) in combination with a more unbalanced factorial mating design (Additional file 1: Figure S4) for segment SS than LS and MM. Zhao et al. (2015) have shown using a resampling strategy based on 1,604 wheat hybrids that both the number of parents used in the factorial mating design and the balanced representation of parental lines used to generate hybrids are important factors driving the prediction accuracy.

Previous studies observed that the prediction accuracies were smaller for self-pollinating than for outcrossing crops (Philipp et al. 2016; Massman et al. 2012; Technow et al. 2014). Moreover, prediction accuracies were higher for crops with clearly defined heterotic groups versus crops with absence of heterotic groups (Philipp et al. 2016). In line with these expectations, we observed prediction accuracies, which were slightly smaller than reported previously for maize hybrids (Massman et al. 2012) and similar to those reported for wheat (Zhao et al. 2015) and rice hybrids (Xu et al. 2014; Xu et al. 2016). Summarizing, genome-wide prediction is a useful and appropriate method to predict hybrid performances in rice.

The identified heterotic patterns outyielded the currently used hybrid populations

The search for promising heterotic patterns, which maximizes the grain yield in the hybrid populations, was performed using a simulated annealing algorithm based on the predicted hybrid performance matrices. The detected heterotic patterns outyielded the population means by 14% for SS to 28% for LS and MM (Fig. 4). The lower advantage of SS compared with LS and MM can be explained by the limited genetic diversity of the SS segment (Fig. 2). Moreover, the detected heterotic patterns resulted in a yield improvement of 7 to 35% compared to hybrids between the existing male and female pools (Fig. 5). Interestingly, male lines were selected more frequently for the heterotic patterns than female lines. This finding is in line with those reported by Huang et al. (2016), who observed also a higher yield potential of male parents either pointing to more beneficial alleles carried by male than female lines or to a bottleneck in selecting stable maintainer lines.

Compromise between short- and long-term success was accomplished with heterotic groups each including 14 lines

The efficiency and success of heterotic patterns can be assessed by short- and long-term response to selection (Zhao et al. 2015). Short-term success is maximized by a high selection intensity choosing only the very best performing parental lines. In contrast, long-term success increases with increasing diversity and heterotic group

size. Heterotic group sizes should reflect a compromise between short- and long-term successes.

Short-term selection response was assessed based on the hybrid performance of the selected heterotic pattern. We ignored a combination of hybrid performance and expected selection gain, i.e., the usefulness criterion, which was previously used (Zhao et al. 2015) because the usefulness criterion is mainly influenced by the hybrid performance. The assessment of long-term selection gain was based on the parameters of theoretical selection limit and genetic representativeness as suggested previously (Zhao et al. 2015). The genetic representativeness is an index estimating the proportion of specific genomes within the full population represented by the identified heterotic groups (Druet et al. 2014). The theoretical selection limit denotes the maximum hybrid performance realized by reciprocal recurrent selection based on the detected heterotic groups.

We observed that the parameters of long-term response plateaued at approximately 14 genotypes per heterotic group (Fig. 4). This group size was also related to an acceptable level of short-term selection gain; thus, representing a good compromise between short- and long-term selection response. Zhao et al. (2015) reported a similar heterotic group size of 16 individuals per group achieving a suitable balance between short-term and long-term selection gain for European hybrid wheat breeding. Summarizing, a beneficial hybrid breeding in rice can be realized with a heterotic group size of 14 parents, which is connected to a suitable level of short-term as well as long-term selection gain.

Conclusions

Hybrid breeding and the effective utilization of heterosis crucially depends on the identification of heterotic patterns. In this study, we have evaluated a three-step approach to detect heterotic patterns in rice using data of a commercial hybrid breeding program. Our findings revealed that hybrid rice breeding based on the identified heterotic patterns holds the potential to boost grain yield and represents an important step for the long-term success of hybrid rice breeding.

Methods
Plant material
The plant material included rice hybrids from the breeding program of Bayer Crop Science, adapted to the Indian sub-continent. The hybrids were grouped based on grain size, shape and appearance into three market segments: Long grain (LS) segment with grain type of long slender, length of larger than 6 mm, length/breadth ratio of larger than 3; Medium (MM) grain segment with grain type of medium slender, length of less than 6 mm, length/breadth ratio of 2.5 to 3.0 mm; Short (SS) grain

segment with grain type of short slender, length of less than 6 mm, length/breadth ratio of larger than 3. All parental lines were derived from the indica rice pool with some introgressions from the japonica group. The lines were grouped into restorer (male) and maintainer (female) lines, because hybrids are produced based on a cytoplasmic male sterility (CMS) system. Most lines were used specifically for a certain market segment and only 17 lines were used as parents across different market segments (Additional file 1: Figure S5). The total number of unique lines amounted to 270, from which 262 were used for genotyping.

For the hybrid production in segment LS, 109 males and 13 females were used, segment MM consisted of 95 male and 16 female parents and in segment SS crosses were made between 47 male and 10 female lines (Additional file 1: Figure S4). The lower number of female than male lines is because it is easier to breed restorer than maintainer lines. The parental lines were crossed using an unbalanced factorial mating design. Phenotypic data comprised of 625, 935, and 400 hybrids for the LS, MM, and SS segment, respectively.

Field trials
The phenotypic data is based on grain yield trials conducted in the year 2014. The hybrids were evaluated in three, two, and four locations for LS, MM, and SS, respectively. The locations were Faizabad (FZB, 26.78°N; 82.13°E; loam, 97 m a.s.l.), Hyderabad (HYD, 17.44°N; 78.11°E, black loam, 569 m a.s.l.), Dhantori (DHN, 28.2° N; 75.45°E, loam, 257 m a.s.l.), Mysore (MYS, 12.5°N; 76.67°E, loam, 709 m a.s.l.), and Raipur (RPR, 21.19°N; 81.56°E, loam, 298 m a.s.l.). In all locations, genotypes were sown in nursery beds and 30 days aged seedlings were manually transplanted in the puddled main field conditions.

The experimental designs followed augmented designs including trials and blocks within replicated checks, but unreplicated entries. Restrictions in hybrid seed production and restricted budget were the main reason for choosing an augmented design with replicated checks, but unreplicated hybrids. The design of a field experiment is graphically illustrated for location HYD and segment MM (Additional file 1: Figure S6). The field experiments in the different locations were structured into trials. Every trial was further split into blocks including entries and five to seven checks. Plot size ranged from 3.6 m^2 to 4.2 m^2 and transplanting density was 33 seedlings m^{-2} using one seedling per hill.

Genomic marker data
In total 262 parental lines were genotyped with a 6 K SNP array based on an Illumina Infinium assay (Yu et al. 2014).

After excluding markers with minor allele frequency below 5%, 5,221 high-quality SNP markers were left for further analyses. For 612, 701, and 242 hybrids genotypic information was available for the LS, MM, and SS segment, respectively.

Statistical analyses of field trials

We analyzed every market segment independently. The variance components were estimated fitting the following model:

$$y_{jkmn} = \mu + m_j + g_j + l_k + t_{km} + b_{kmn} + (gl)_{jk} + \varepsilon_{jkmn},$$

where y_{jkmn} is the grain yield of jth genotype in n^{th} block, m^{th} trial and k^{th} location, μ is the overall mean, m_i is an effect for maturity of the observed genotype; g_j refers to the effect of the jth genotype, l_k to the effects of the kth location, t_{km} to the effect of the mth trial at the kth location, b_{kmn} to the effect of the nth block at the mth trial at the kth location, $(gl)_{jk}$ to the interaction effect between the jth genotype and the kth location and ε_{jkmn} to the residual. All effects except m_i were modeled as random. The estimated variance components were used to calculate the broad-sense heritability as:

$$h^2 = \frac{\sigma^2_{Genotype}}{\sigma^2_{Phenotype}}$$

$$= \frac{\sigma^2_{Genotype}}{\sigma^2_{Genotype} + \frac{\sigma^2_{GxL}}{No.\ of\ location} + \frac{\sigma^2_{error}}{No.\ of\ location\ \times\ No.\ of\ replicates}}$$

The genotypic variance was further decomposed into the variance due to general (GCA) and specific combining ability effects (SCA). Significance of variance components were tested using a likelihood-ratio test. We used the above outlined model and assumed fixed genotype effects to obtain the best linear unbiased estimations (BLUEs) of every genotype (Additional file 2). All described statistical approaches were conducted using R (The R Core Team 2016) and linear mixed models were implemented with the package ASReml-R (Gilmour et al. 2009).

Analyses of population structure

The genomic data were used to estimate the allele frequency of every marker. Linkage disequilibrium (LD) was assessed by the LD measure r^2 (Weir 1996). Rogers' distances (Rogers 1972) among all pairs of parental lines were calculated. The matrix of Rogers' distances was used to perform principal coordinate analyses (Gower 1966).

Prediction of hybrid performance

The genome-wide prediction models make use of all hybrids for which genotypic and phenotypic information

were available, i.e., 612, 701, and 242 hybrids for the LS, MM, and SS segment, respectively. For genome-wide prediction, we implemented ridge regression best linear unbiased prediction (RR-BLUP), modeling both additive and dominance effects as:

$$Y = 1_n\mu + Z_A a + Z_D d + e.$$

Y refers to BLUEs of the hybrids for grain yield. Vector 1_n includes only ones and its element number is equal to the number of n hybrids; μ refers to the overall mean. Design matrices Z_A and Z_D have a dimension of $n \times m$, where m is the number of markers and were defined using the F_∞ metric (Zhao et al. 2013). The vector e includes residual values for the single cross combinations. Estimations or predictions of a, d, and μ were done by implementing Henderson's mixed model equation (Henderson 1984). Normal distribution and constant variance for the additive and dominance effects were assumed (Zhao et al. 2013).

The accuracy of the prediction of grain yield was evaluated using cross-validations, which entails the division of the total population into estimation and test populations. Marker effects were estimated in the estimation population and used to predict the performance of hybrids in the test population. Since relatedness strongly influences prediction accuracy (Habier et al. 2007), a cross-validation strategy was used considering three test populations with varying degree of relatedness to the estimation population (Additional file 1: Figure S3): Test set T2 most closely related to the estimation set included only hybrids derived from the same parents as the hybrids that had been evaluated, while the less related test set T1 included hybrids sharing one parent with the hybrids in the estimation set and the least related test set T0 included only hybrids having no parents in common with the estimation set. Prediction ability was estimated as Pearson's correlation coefficient between the observed and the predicted hybrid performance.

Identification of heterotic groups

The search of heterotic patterns maximizing the hybrid performance was performed based on the predicted hybrid performances using a simulated annealing algorithm. The implementation of the simulated annealing algorithm is described in detail by Zhao et al. (2015). We assumed the same group size for two matching heterotic groups and performed the search for increasing group sizes from two to 20.

Assessing the short- and long-term success of heterotic groups

Beside the hybrid performance as a short-term success parameter, theoretical selection limit (Zhao et al. 2015)

and genetic representativeness describing long-term success were determined to evaluate the performance of the identified heterotic groups. The genetic representativeness gives the proportion of specific genomes within the full population represented by the identified heterotic groups (Druet et al. 2014). Representativeness of two specific heterotic groups depends from the additive relationship within heterotic groups beside additive relationship between the full population and heterotic groups and the genome proportion of heterotic groups. Additive relationship was calculated as one minus the Rogers' distance values of parental pairs (Melchinger et al. 1991). The final representativeness for two matching heterotic groups was the result of the mean value of all parental lines included in the selected heterotic patterns (Zhao et al. 2015). Theoretical selection limit is closely linked to the concept of maximal long-term selection response (Walsh and Lynch 2014) and represents the maximum hybrid performance, which can be reached by reciprocal recurrent selection based on the identified heterotic groups performing an infinite number of selection cycles. We assumed absence of migration, mutation, and epistasis for estimating of theoretical selection limit based on the predicted marker effects (Zhao et al. 2015).

Additional files

Additional file 1: for Genome-based identification of heterotic. **Figure S1.** Decay of linkage disequilibrium measured using r^2 with physical map distance. **Figure S2.** Principal coordinate analysis of all parental lines for market segments LS, MM, and SS. **Figure S3.** Applied cross validation scenario exemplifying the selected fractions for market segment MM. **Figure S4.** Unbalanced factorial crossing designs for market segment (A) LS, (B) MM, and (C) SS. **Figure S5.** Venn diagram of overlapping parental lines of the hybrid evaluation trials in market segments LS, MM, and SS. **Figure S6.** Experimental design and distribution of checks for hybrid experiment of market segment MM at location HYD. Orange and black lines represent the size of trials and blocks, respectively. White plots were not phenotyped. **Table S1.** Composition of heterotic groups selected with the simulated annealing algorithm as well as the average hybrid performance between both heterotic groups (Inter) for the market segments LS, MM, and SS. **Table S2.** Composition of estimation and test populations for market segment LS, MM, and SS. (DOCX 1811 kb)

Additional file 2: Best linear unbiased estimations for hybrid grain yield experiments in market segments LS, MM, and SS. (XLSX 76 kb)

Authors' contributions

JCR, KP, VM, and YZ conceived the design of this study. NG and FP coordinated the SNP genotyping and the experiments including the phenotypic trait measurements of the plant materials. UB, ZL, and GL conducted the analyses. UB and JCR wrote the manuscript. All authors have read and approved the final manuscript.

Competing interests

The authors declare that they have no competing interests.

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Author details

[1]Department of Breeding Research, Leibniz Institute of Plant Genetics and Crop Plant Research (IPK) Gatersleben, Corrensstraße 3, 06466 Stadt Seeland, Germany. [2]Bayer Bioscience, 500081 Hyderabad, India. [3]European Wheat Breeding Center, Bayer Crop Science, Am Schwabenplan 8, 06466 Stadt Seeland, Germany. [4]Biometrics and Breeding Research US, Bayer Crop Science, 407 Davis Dr., Morrisville, NC, USA. [5]K. Pillen, Chair of Plant Breeding, Martin-Luther-University Halle-Wittenberg, Betty-Heimann-Str. 3, 06120 Halle/Saale, Germany.

References

Barclay A (2010) Hybridizing the world. Rice Today 9(4):32–5

Chen L, Liu Y-G (2014) Male Sterility and Fertility Restoration in Crops. Annu Rev Plant Biol 65:579–606. doi:10.1146/annurev-arplant-050213-040119

Cheng S-H, Zhuang J-Y, Fan Y-Y, Du J-H, Cao L-Y (2007a) Progress in Research and Development on Hybrid Rice: A Super-domesticate in China. Ann Bot 100:959–66. doi:10.1093/aob/mcm121

Cheng S-H, Cao L-Y, Zhuang J-Y, Chen S-G, Zhan X-D, Fan Y-Y, Zhu D-F, Min S-K (2007b) Super Hybrid Rice Breeding in China: Achievements and Prospects. J Integr Plant Biol 49:805–10. doi:10.1111/j.1672-9072.2007.00514.x

Desta ZA, Ortiz R (2014) Genomic selection: genome-wide prediction in plant improvement. Trends Plant Sci 19:592–601. doi:10.1016/j.tplants.2014.05.006

Druet T, Macleod IM, Hayes BJ (2014) Toward genomic prediction from whole-genome sequence data: impact of sequencing design on genotype imputation and accuracy of predictions. Heredity 112:39–47. doi:10.1038/hdy.2013.13

Fischer S, Maurer HP, Würschum T, Möhring J, Piepho H-P, Schön CC, Thiemt E-M, Dhillon BS, Weissmann EA, Melchinger AE, Reif JC (2010) Development of Heterotic Groups in Triticale. Crop Sci 50:584–90. doi:10.2135/cropsci2009.04.0225

Gilmour AR, Gogel BJ, Cullis BR, Thompson R (2009) ASReml User Guide Release 3.0. VSN International Ltd, Hemel Hempstead, HP1 1ES, UK. https://www.vsni.co.uk/downloads/asreml/release3/UserGuide.pdf. Accessed 25 May 2016.

Gower JC (1966) Some distance properties of latent root and vector methods used in multivariate analysis. Biometrika 53:325–38

Guo J, Xu X, Li W, Zhu W, Zhu H, Liu Z, Luan X, Dai Z, Liu G, Zhang Z, Zeng R, Guang T, Fu X, Wang S, Zhang G (2016) Overcoming inter-subspecific hybrid sterility in rice by developing indica-compatible japonica lines. Sci Rep 6:26878. doi:10.1038/srep26878

Habier D, Fernando RL, Dekkers JCM (2007) The impact of genetic relationship information on genome-assisted breeding fs. Genetics 177:2389–97. doi:10.1534/genetics.107.081190

Henderson CR (1984) Applications of linear models in animal breeding, 3rd edn. University of Guelph, Ontario

Huang X, Yang S, Gong J, Zhao Q, Feng Q, Zhan Q, Zhao Y, Li W, Cheng B, Xia J, Chen N, Huang T, Zhang L, Fan D, Chen J, Zhou C, Lu Y, Weng Q, Han B (2016) Genomic architecture of heterosis for yield traits in rice. Nature 537:629–33. doi:10.1038/nature19760

Ikekashi H, Araki H (1984) Varietal Screening of Compatibility Types Revealed in F1 Fertility of Distant Crosses in Rice. Jpn J Breed 34:304–13

International Rice Research Institute (2013) Rice as human food. In: Rice Almanac, 4th edn. Los Banos, Philippines, pp 10–4, http://books.irri.org/9789712203008_content.pdf. Accessed 25 April 2016

Khush GS (2005) What it will take to Feed 5.0 Billion Rice consumers in 2030. Plant Mol Biol 59:1–6. doi:10.1007/s11103-005-2159-5

Khush GS (2013) Strategies for increasing the yield potential of cereals: case of rice as an example. Plant Breed 132:433–6. doi:10.1111/pbr.1991

Li S, Yang D, Zhu Y (2007) Characterization and Use of Male Sterility in Hybrid Rice Breeding. J Integr Plant Biol 49:791–804. doi:10.1111/j.1672-9072.2007.00513.x

Little RJA, Rubin DB (2002) Statistical analysis with incomplete data, 2nd edn. John Wiley & Sons, New York

Liu KD, Zhou ZQ, Xu CG, Zhang Q, Saghai Maroof MA (1996) An analysis of hybrid sterility in rice using a diallel cross of 21 parents involving indica, japonica and wide compatibility varieties. Euphytica 90:275–80

Longin CFH, Mühleisen J, Maurer HP, Zhang H, Gowda M, Reif JC (2012) Hybrid breeding in autogamous cereals. Theor Appl Genet 125:1087–96. doi:10.1007/s00122-012-1967-7

Massman JM, Gordillo A, Lorenzana RE, Bernardo R (2012) Genomewide predictions from maize single-cross data. Theor Appl Genet 126:13–22. doi:10.1007/s00122-012-1955-y

Melchinger AE, Messmer MM, Lee M, Woodman WL, Lamkey KR (1991) Diversity and Relationships among U.S. Maize Inbreds Revealed by Restriction Fragment Length Polymorphisms. Crop Sci 31:669–78

Melchinger AE (1999) Genetic Diversity and Heterosis. In: Coors JG, Pandey S (eds) The Genetics and Exploitation of Heterosis in Crops, 1st edn. American Society of Agronomy, Crop Science of America, Madison, pp 99–118

Melchinger AE, Gumber RK (1998) Overview of Heterosis and Heterotic Groups in Agronomic Crops. In: Larnkey KR, Staub JE (eds) Concepts and Breeding of Heterosis in Crop Plants, 1st edn. Crop Science Society of America, Madison, pp 29–44

Ouyang Y, Liu Y-G, Zhang Q (2010) Hybrid sterility in plant: stories from rice. Curr Opin Plant Biol 13:186–92. doi:10.1016/j.pbi.2010.01.002

Philipp N, Liu G, Zhao Y, He S, Spiller M, Stiewe G, Pillen K, Reif JC, Li Z (2016) Genomic Prediction of Barley Hybrid Performance. Plant Genome 9:1–8. doi:10.3835/plantgenome2016.02.0016

Phillips RL (2010) Mobilizing Science to Break Yield Barriers. Crop Sci 50:99–108. doi:10.2135/cropsci2009.09.052

Ray DK, Mueller ND, West PC, Foley JA (2013) Yield trends are insufficient to double global crop production by 2050. PLoS One 8(6):e66428. doi:10.1371/journal.pone.0066428

Reif JC, Gumpert F-M, Fischer S, Melchinger AE (2007) Impact of Interpopulation Divergence on Additive and Dominance Variance in Hybrid Populations. Genetics 176:1931–4. doi:10.1534/genetics.107.074146

Reif JC, Hallauer AR, Melchinger AE (2005) Heterosis and heterotic patterns in maize. Maydica 50:215–23

Rogers JS (1972) Measures of genetic similarity and genetic distance. In: Wheeler MR (ed) Studies in Genetics VII, Univ, Texas Publ. 7213. University of Texas, Austin, pp 145–53

Spielman DJ, Kolady DE, Ward PS (2013) The prospects for hybrid rice in India. Food Sec. doi:10.1007/s12571-013-0291-7

Technow F, Schrag TA, Schipprack W, Bauer E, Simianer H, Melchinger AE (2014) Genome Properties and Prospects of Genomic Prediction of Hybrid Performance in a Breeding Program of Maize. Genetics 197:1343–55. doi:10.1534/genetics.114.165860

The R Core Team (2016) R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, https://cran.r-project.org/doc/manuals/fullrefman.pdf. Accessed 23 May 2016

Viraktamath BC (2010) Hybrid Rice in India - Current Status and Future Prospects. Directorate of Rice Research, Rajendranagar, Hyderabad, http://www.rkmp.co.in/sites/default/files/ris/research-themes/Hybrid%20rice%20in%20India.pdf. Accessed 25 Nov 2016

Walsh B, Lynch M (2014) Long-Term Response: Finite Population Size and Mutation. In: Evolution and Selection of Quantitative Traits., pp 181–202, http://nitro.biosci.arizona.edu/zbook/NewVolume_2/pdf/Chapter26.pdf. Accessed 23 Nov 2015

Wang K, Qiu F, Larazo W, Dela Paz MA, Xie F (2014) Heterotic groups of tropical indica rice germplasm. Theor Appl Genet 128:421–30. doi:10.1007/s00122-014-2441-5

Wanjari RH, Mandal KG, Ghosh PK, Adhikari T, Rao NH (2006) Rice in India: Present Status and Strategies to Boost Its Production Through Hybrids. J Sustain Agric 28:19–39. doi:10.1300/J064v28n01_04

Weir BS (1996) Genetic Data Analysis: Methods for Discrete Population Genetic Data, 2nd edn. Sinauer Associates, Sunderland

Xie F, Guo L, Ren G, Hu P, Wang F, Xu J, Li X, Qiu F, Dela Paz MA (2012) Genetic diversity and structure of indica rice varieties from two heterotic pools of southern China and IRRI. Plant Genet Res 10:186–93. doi:10.1017/S147926211200024X

Xie F, He Z, Esguerra MQ, Qiu F, Ramanathan V (2013) Determination of heterotic groups for tropical Indica hybrid rice germplasm. Theor Appl Genet 127:407–17. doi:10.1007/s00122-013-2227-1

Xu S, Xu Y, Gong L, Zhang Q (2016) Metabolomic prediction of yield in hybrid rice. Plant J. doi:10.1111/tpj.13242

Xu S, Zhu D, Zhang Q (2014) Predicting hybrid performance in rice using genomic best linear unbiased prediction. Proc Natl Acad Sci 111:12456–61. doi:10.1073/pnas.1413750111

Yu H, Xie W, Li J, Zhou F, Zhang Q (2014) A whole-genome SNP array (RICE6K) for genomic breeding in rice. Plant Biotechnol J 12:28–37. doi:10.1111/pbi.12113

Zhang G, Lu Y, Bharaj TS, Virmani SS, Huang N (1997) Mapping of the Rf-3 nuclear fertility-restoring gene for WA cytoplasmic male sterility in rice using RAPD and RFLP markers. Theor Appl Genet 94:27–33

Zhang QY, Liu YG, Zhang GQ, Mei MT (2002) Molecular mapping of the fertility restorer gene Rf-4 for WA cytoplasmic male sterility in rice. Acta Genet Sin 29:1001

Zhao Y, Li Z, Liu G, Jiang Y, Maurer HP, Würschum T, Mock H-P, Matros A, Ebmeyer E, Schachschneider R, Kazman E, Schacht J, Gowda M, Longin CFH, Reif JC (2015) Genome-based establishment of a high-yielding heterotic pattern for hybrid wheat breeding. Proc Natl Acad Sci 112:15624–9. doi:10.1073/pnas.1514547112

Zhao Y, Mette MF, Reif JC (2014) Genomic selection in hybrid breeding. Plant Breed 134:1–10. doi:10.1111/pbr.12231

Zhao Y, Zeng J, Fernando R, Reif JC (2013) Genomic Prediction of Hybrid Wheat Performance. Crop Sci 53:802–10. doi:10.2135/cropsci2012.08.0463

Zhu G, Peng S, Huang J, Cui K, Nie L, Wang F (2016) Genetic Improvements in Rice Yield and Concomitant Increases in Radiation- and Nitrogen-Use Efficiency in Middle Reaches of Yangtze River. Sci Report 6:21049. doi:10.1038/srep21049