Development of a multiple-hybrid population for genome-wide association studies: theoretical consideration and genetic mapping of flowering traits in maize

Hui Wang¹, Cheng Xu², Xiaogang Liu¹, Zifeng Guo³, Xiaojie Xu¹, Shanhong Wang¹, Chuanxiao Xie¹, Wen-Xue Li², Cheng Zou¹ & Yunbi Xu¹,²

Various types of populations have been used in genetics, genomics and crop improvement, including bi- and multi-parental populations and natural ones. The latter has been widely used in genome-wide association study (GWAS). However, inbred-based GWAS cannot be used to reveal the mechanisms involved in hybrid performance. We developed a novel maize population, multiple-hybrid population (MHP), consisting of 724 hybrids produced using 28 temperate and 23 tropical inbreds. The hybrids can be divided into three subpopulations, two diallels and NC (North Carolina Design) II. Significant genetic differences were identified among parents, hybrids and heterotic groups. A cluster analysis revealed heterotic groups existing in the parental lines and the results showed that MHPs are well suitable for GWAS in hybrid crops. MHP-based GWAS was performed using 55 K SNP array for flowering time traits, days to tassel, days to silk, days to anthesis and anthesis-silking interval. Two independent methods, PEPIS developed for hybrids and TASSEL software designed for inbred line populations, revealed highly consistent results with five overlapping chromosomal regions identified and used for discovery of candidate genes and quantitative trait nucleotides. Our results indicate that MHPs are powerful in GWAS for hybrid-related traits with great potential applications in the molecular breeding era.

Genetic mapping of important agronomic traits, followed by marker-assisted selection (MAS), provides a powerful tool for crop genetic improvement. Genes can be mapped through four basic approaches: linkage analysis using bi- or multi-parental populations, association or linkage disequilibrium (LD) analysis using natural populations, comparative analysis using mutated populations and near-isogenic (introgression) lines, and selective analysis using sub-populations based on selective sweeps.

Association mapping has been used to detect the underlying major genes in the gene pools and their introgression to improve traits in major crop breeding programs¹. It has been based on two basic methods, one using candidate gene-based markers to confirm the association² and the other using whole genome scan³, the latter being called genome wide association studies (GWAS). GWAS using single nucleotide polymorphism (SNP) marker loci has successfully identified genes and pathways for agronomic traits in many crops of economic importance, including rice⁴, maize⁵, wheat⁶, sorghum⁷ and barley⁸. This method generally consists of five stages: selection of diverse germplasm, estimation of the level of population structure, phenotypic evaluation, genotyping for candidate genes or whole genome genotyping, and statistical test for genotype-phenotype association⁹.

In contrast to linkage mapping, GWAS based linkage disequilibrium (LD) offers a potentially useful and robust approach for mapping causal genes with moderate or large effects¹⁰, which has several advantages: extensive genetic variations in a more representative genetic background, higher resolution, and utilization of historic phenotypic data on cultivars without the need to develop special mapping populations¹¹. The simple statistical model for GWAS is

¹Institute of Crop Science, National Key Facility of Crop Gene Resources and Genetic Improvement, Chinese Academy of Agricultural Sciences, Beijing 100081, China. ²International Maize and Wheat Improvement Center (CIMMYT), El Batán 56130, Texcoco, Mexico. Correspondence and requests for materials should be addressed to Y.X. (email: y.xu@cgiar.org)
focusing on single-SNP tests, and the test results frequently show high false positives owing to specific problems such as population structure, relatedness and polygenic background effects. Therefore, a variety of statistical analytical methods have been developed, such as, the mixed linear model (Q + K model), which is the most popular method that effectively eliminates false positives by incorporating population structure (Q) and relative kinship matrix (K)\(^{12}\), multi-trait mixed model (MTMM) for multiple traits\(^{13}\), multi-locus mixed-model (MLMM) based on multiple loci\(^{14}\), factored spectrally transformed linear mixed model (FaST-LMM) with the number and square of rank of the relationship among individuals\(^{15}\), settlement of MLM under progressively exclusive relationship (SUPER) using influential bin markers and a small set of markers to define the relationship among the individuals\(^{16}\), multi-trait set linear mixed-model (mtSet-LMM) between sets of variants and multiple traits\(^{17}\), and a random-SNP-effect MLM (RMLM) with a modified Bonferroni correction and a multi-locus model with less rigorous selection criteria from RMLM (MRMLM)\(^{18}\).

There are many types of populations that have been used in genetics, genomics and crop improvement\(^{19,20}\). These populations have been used individually, and in very few cases, in combination. The primary objective for this article was to review all available populations, and introduce the concept of multiple-hybrid population (MHP) as a new population type, which is more suitable for GWAS in hybrid crops using hybrid vigor. Using maize as an example, we developed an MHP from diallel and NC II mating designs. We will present the experimental design, parent classification, data analysis strategies and applications of the MHP in conventional quantitative genetics, GWAS, and breeding.

**Populations in Genetics, Genomics and Crop Improvement**

Populations used in genetics, genomics and crop improvement have their advantages and disadvantages. For the convenience of comparison, we will first discuss all available populations and then focus on the MHP in the next section.

**Biparental populations.** The populations derived from two parental lines are most widely used in genetics, genomics and breeding, with the following advantages and disadvantages:

- **F\(_2\):** containing sufficient genetic information, but failing to distinguish the genotypes just by phenotyping owing to the presence of heterozygotes, and thus the phenotyping is usually measured with its derived F\(_2\)\(_5\) families, progenies produced by bulked pollination of F\(_2\) individuals, and the F\(_2\) genotypes maintained by multiple tillers and ratoning procedures\(^{19,20}\).
- **BC\(_1\):** including two types of genotypes which could be distinguished theoretically just by phenotyping if the target gene is dominant, but difficult to maintain the materials for long time, although the genotypes may be maintained by following the procedures for F\(_2\) individuals as discussed above.
- **RIL (recombinant inbred line):** consisting of genetically stable genotypes, which can be maintained permanently without change of the population constitution, and suitable for distinguishing dominant and co-dominant status, but difficult to be bred for some species, especially for cross-pollinated plants\(^{21,22}\).
- **DH (doubled haploid):** including homogeneous and breeding-true genotypes and reflecting the gene segregation and recombination rate of F\(_1\) gametes, with different protocols available for the development of DH lines including microspore culture, chromosome elimination, and haploid inducers, depending on crop species\(^{19,20}\).

An immortalized F\(_1\) population can be developed through crosses among RIL or DH populations, which can be used permanently for QTL analysis\(^{20,23}\).

**Multiparental populations.** There are three major types of multi-parental populations, four-way crosses (4WC), NAM (nested association population) and MAGIC (multiparent advanced generation inter-cross). The 4WC were widely used in commercial animal and plant breeding, formed by a cross between two hybrids, F\(_1\) (P\(_1\) × P\(_2\)) and F\(_1\)\(_3\) (P\(_3\) × P\(_4\)), by which genes and QTL can be identified\(^{24}\) with epistatic QTL mapping developed using penalized maximum likelihood (PML)\(^{12}\).

NAM population is developed by crossing a reference genotype (line) with a panel of lines maximizing the genetic diversity and RIL families are then developed. Different sets of RIL populations share a common parental line, while RILs within each population are derived from two parents. The NAM population combines the advantages of both bi-parental segregating and natural populations, which can be used for integrated linkage and association mapping\(^{26}\). The first NAM population in plants was developed by using 200 RILs from each of the 25 crosses derived from 25 diverse inbred lines crossed with a reference inbred line B73, which resulted in nearly 5000 RILs in total\(^{27}\). This NAM population has been successfully applied to the study of flowering time\(^{28}\), resistance to southern corn leaf blight\(^{29}\), leaf architecture\(^{30}\) and carbon and nitrogen metabolism\(^{31}\). However, 48.7% of marker genotypes were from B73 among these RILs and allelic frequencies from common parent are much higher than those from other 2 parents with phenotype deviation due to the only one common parent.

The MAGIC population starts with multiple bi-parental intercrosses based on variety-specific founders selected, and then every two bi-parental F\(_1\)s are intermated to produce a double hybrid. The two double hybrid F\(_2\)s are intermated again to develop a hybrid with eight parents involved\(^{29}\). This process continues until all parental lines are included in a final hybrid, and then RILs can be developed by selfing at least five generations or DHs generated through an available protocol\(^{32}\). The first MAGIC population has been developed in *Arabidopsis thaliana* by using a set of 527 RILs descended from a heterogeneous stock of 19 intermated accessions\(^{32}\). MAGIC populations have been developed to significantly increase the effectiveness of whole genome scan in wheat\(^{33}\), rice\(^{34}\) and maize\(^{35}\).

**Mating-design populations.** There are several populations that can be derived from specific mating designs, including:
- Diallel: multiple parents are crossed to produce all possible bi-parental F1s.
- North Carolina Design (NC) I: some individuals randomly selected as males from a biparental F1 population are crossed with other randomly selected as females.
- NC II: n parental lines are divided into two groups, one group as males and the other as females, to produce hybrids of all possible combinations.
- NC III: n individuals are selected from an F2 population to backcross its parents, P1 and P2.
- TTC (triple test crosses): an extension of NC III, i.e., n individuals (n > 20) are selected from an F2 population to backcross with both parental lines, P1 and P2, and their F1s.
- sTTC (simplified triple test-crosses): n cultivars or strains are selected from the germplasm pool to cross with two varieties or strains, P1 and P2, with the highest and lowest phenotypic values.

**Natural populations.** The natural populations are those consisting of many individuals that are fixed or can be maintained individually through selfing or outcrossing, including inbred lines, varieties, landraces and wild relatives, which have been widely used in association mapping. There are two types of natural populations. The natural populations composed of elite lines have lower diversity and slower LD decay with a longer distance, while those with a core collection of diverse germplasm have rich genetic diversity and more rapid LD decay with a shorter distance. Both types can be combined in applications. First, rough mapping for a particular trait can be performed using natural populations with lower diversity and a small amount of markers to cover the whole genome; then, using natural populations with rich diversity in the genomic region of interest, a particular gene can be fine mapped for gene cloning.

**Multiple-Hybrid Populations (MHP) for GWAS**

Populations described above, except for those derived from various mating designs, are most appropriate for crops using inbreds or varieties for commercialization but not suitable for hybrid crops that use hybrids for production, as genetic researches based on inbreds reveal neither what are hidden in hybrids nor what are associated with hybrid performance. Populations for hybrid crops and hybrid performance may come from two sources, one with test-crosses using all individuals from each of the populations as parents to testcross with one or several testers, and the other with a large number of multiple hybrids generated from such mating designs described in the previous section as a population, which can be simply called as multiple hybrid population (MHP). Considering all available mating designs that have been used in quantitative genetics, hybrids from a full diallel design with n parents would be one of the best MHPs, as the full set hybrids can be divided into three subsets, two diallels and one NC II, as needed. As such an attempt with a limited number of hybrids, a partial NC II was used for QTL mapping to test effects associated with traits, and the significant effects identified by empirical Bayesian were used to estimate genetic effects for the missing F1s with elite parents and hybrids predicted. The generated hybrids, along with the parental genotype and phenotype, can be used to predict the missing hybrids and the rest hybrids. The relative sizes of the three subsets can be optimized to best match different ecotypes of parental lines, and the number of hybrids can be optimized so that a best design can be achieved with the smallest sample sizes as possible for the largest “genetic gain” (Fig. 1). As an example, we developed an MHP with hybrids from Griffing IV diallel crosses and NC II design using temperate and tropical elite maize inbreds as parental lines.

The parents used for developing MHP can be a representative sample of the population to which inference is desired and a core collection from a gene bank, varieties or landraces representing the elite germplasm for a breeding program, or a set of inbred lines representing a synthetic outcrossing population. MHPs have several distinct characteristics that make them very unique compared to other types of populations and very useful in genetics and plant breeding.

**Suitability for combining ability and heterosis analysis.** The evaluation of hybrids produced by inbred lines is an important step towards the development of hybrid varieties in crops such as maize and this process theoretically should be done for all possible ways of hybridization (diallel crosses), where selection of favorable traits for each inbred can be determined. The diallel and NC II design analysis can provide elaborate information on the genetic identity of genotypes especially on dominance-recessiveness relationships and some genetic interactions. With diallel or NC designs, many studies have been conducted for combining ability and heterosis analysis. For example, 30 crosses developed in maize with six parents according to Griffing III method to identify hybrids expressing a high level of heterosis. The overall increase of GCA and SCA was found in NC II design between some pairs of combinations.

A complete diallel series of crosses involving eight parents was developed to evaluate GCA and SCA for corn grits. A partial NC II design for cotton and rapeseed using epistatic association mapping revealed additive and additive by dominance interaction effects for GCA, and dominance related effects for SCA. Mid-parent heterosis, dominance and dominance by dominance interaction effects affect heterosis more than over-dominance and complete dominance.

The MHP developed in our research will be used to estimate combining ability, heterosis, and genotype by environment interaction and details will be reported elsewhere.

**Easy for sharing through parental lines.** Compared to natural populations used for GWAS, which needs to share at least several hundreds, even thousands, of varieties, MHP just needs to maintain and share a much less number of samples. Sharing germplasm accessions that can be used as parental lines is more frequent for germplasm exchange. For n varieties, n(n−1)/2 hybrids can be produced. For example, 100 varieties can be used to generate...
99 × 100/2 = 4500 hybrids. It would be much more difficult to share 4500 samples compared with sharing 100 parental lines. In our study, the MHP, consists of 724 hybrids, generated through just crossing 51 maize inbreds with each other (Fig. 1), which is time- and labor-saving.

**Flexibility for testing across diverse environments.** Diverse germplasm including inbreds and hybrids can be further investigated to identify new sources of genetic variations for fundamental experiments and commercial breeding. MHPs provide an opportunity for testing across diverse environments because of their better adaptation compared to inbred populations. The most typical example is the large-scale yield trials and multi-location tests before commercialization. Testing of multiple hybrids across diverse environments has been also used in genetic study. For example, testcrosses between Tx714 and 346 diverse inbreds evaluated under well watered and non-irrigated trials showed high genetic variance, and for 10 quantitative trait variants for agronomic traits as revealed by 60,000 SNPs, three of them explained 5–10% of phenotypic variation in grain yield under both water conditions. In our research, hybrids from the MHP showed a high level of genetic and phenotypic variation across different environments (data not shown).

Through multi-environmental trials, adapted hybrids can be identified for specific environments. The difficulty of choosing appropriate varieties has contributed to restricted breeding progress for biotic and abiotic stress tolerance in highly variable target environments. Hybrids produced from diverse inbreds will show significantly different responses to diverse environments and thus the hybrids most adapted to a specific environment can be identified and used for further testing and breeding. In this regard, 91 hybrids made among seven high- and seven low-Fe-Zn
content lines were evaluated in six locations, and one low-Fe-Zn parental line showed a significant positive GCA effect and its hybrids emerged as a highly promising variety\(^5\). A total of 84 hybrids were investigated for their resistance to parasitic weed *Striga* under artificial infestation and free environments in two locations, and three inbreds with high GCA effect for yield and *Striga* resistance and four high yielding hybrids showed the potential for further use\(^6\). In our MHP, some elite hybrids, especially large-scale commercialized ones, showed stronger adaptation under various environmental conditions.

**Suitability for both hybrid and inbred crops.** Progeny selection is one of the most important stages in plant breeding. Crossing patterns and hybrids as those generated in MHP are commonly used in both hybrid and inbred breeding programs. For hybrid crops, the crosses are usually made among heterotic groups to achieve a high level of heterosis and better adaptation to specific locations and crop seasons. A total of 91 hybrids, evaluated along with their parental lines for combining ability and heterotic pattern, showed significant heterosis and better adaptation for hybrids development. MHP can be also used in self-pollinated species, although production of hybrids is much more difficult than outcrossing species. There are numerous available examples using diallel crosses in self-pollinated crops. For example, non-additive gene effect for most of yield components was revealed based on 28 rice hybrids from a partial diallel crossing of eight inbreds, five of which showed significant favorable SCA effect on yield\(^4\). Earlier stem elongation in floating rice can greatly improve the chance of floating rice to survive in the flood, and therefore a set of 6 × 6 half-diallel crosses with four floating and two non-floating rice varieties was developed with results indicating that the additive effect was higher than dominant effects and the dominant alleles were concentrated in the floating parental lines\(^49\). A partial diallel based on six *indica* and seven *japonica* rice genotypes was used to investigate the genetic variations of yield and cold tolerance, with significant heterosis and combining ability revealed for tested characters\(^50\).

**Savings in genotyping.** As only the parental lines need to be genotyped, from which hybrid genotypes can be inferred, using MHP can save great investment in genotyping, compared to genotyping the same number of inbreds. As we calculated in the precious section, genotypes for 4,500 hybrids can be inferred from 100 parental lines genotyped, the former being 45 folds larger. In *Brassica napus*, 205 SSR markers were used to examine the polymorphisms among 441 parents (298 sterile and 143 restorer lines), and the genotypes for the partial NC II hybrids could be then deduced from their parents\(^37\), with 8 main-effect and 37 interacted QTL identified for oil content, which could be used to predict 10 elite restorer lines, 10 elite sterile lines and 10 elite hybrids.

**A Maize Multiple Hybrid Population**

**An MHP from diallel and NC II designs.** A total of 51 maize inbred lines, representing a broad selection of breeding germplasm from temperate and tropical regions, were used for development of an MHP (Fig. 1; Table 1). These included 28 temperate and 23 tropical inbred lines as parents, by which three subpopulations of MHP were developed: (1) temperate diallel consisting of 325 crosses made in Griffing IV using 26 elite maize inbred lines, including 13 U.S. and 13 Chinese inbred lines that represent different heterotic groups; (2) temperate and tropical NC II consisting of 263 hybrids made between 13 temperate and 21 tropical inbred lines; (3) tropical diallel consisting of 136 crosses made in Griffing IV using 17 tropical inbred lines as parents, most of which come from CIMMYT.

**Chinese temperate inbreds used as MHP parents.** Among 15 Chinese temperate inbreds used to develop the MHP (Table 2), Ye478, HZ4 (Huangzao 4), Dan340, Mo17, Tie7922 and Qi319 have been used as common tester lines for six Chinese heterotic groups, Reid, SPT, LRC, Lancaster, PA and PB, respectively\(^51\). These key inbreds have been playing a very important role in development of both inbreds and hybrids across maize regions in China. For example, Chang7-2 and LX9801 are among lines derived from HZ4 in summer maize region, Huang-Huai-Hai River Zone\(^52\). A total of 20 commercial hybrids have been developed by using Dan340-derived lines as parents, and 17 elite inbreds have been developed using Dan340’s sister lines\(^53\). As a common tester line for heterotic group SPT, HZ4 has been used to develop more than 40 hybrids and nearly 70 elite inbreds, including J853, LX9801 and Chang7-2\(^54\). H21, an inbred line with high drought tolerance derived from Pioneer hybrid P78599\(^55\), and two superior hybrids made directly were planted widely in Huang-Huai-Hai River Zone. Four elite hybrids widely used in production were developed by Nan21-3, which was developed by selfing the elite exotic hybrid\(^56\). Qi205, a quality protein maize (QPM), was a breakthrough progress in protein maize in China\(^57\). More than 50 varieties, including Ludan50, Ludan981 and Ludan963, were released with Qi319, highly resistant to southern corn rust, and its derived lines as parents\(^58\). Over twenty inbreds were derived from Si287 with two widely used hybrids, Jidan27 and Jidan46. Developed by seven continuous selflings from a U.S. hybrid 3382, Tie7922 was the parent for 30 hybrids and numerous inbreds, including TieC8605-2, Dan9046 and Liao2345\(^59\). Ye478 has been used to produce nearly 50 hybrids, such as Zhongdan8, Yedan12, Yedan13, and Yedan19. Zheng22, belonging to the heterotic group LRC, was used in Yuyu18, a hybrid widely used in production. Zheng58 is the female parent of ‘Zhengdan958’, which had been planted over 4 million ha per year in three consecutive years\(^60\). Dan398, developed from Pioneer hybrid P78599 from tropical zone, was used to make several widely planted hybrids, mainly Danke series\(^61\). Huang C was the male for one of the best hybrids, Nongda108, which was awarded the First-Class Prize for China National Science and Technology Progress in 2002\(^62\).

**Tropical inbreds used as MHP parents.** Tropical germplasm has abundant genetic variability and special resistance to diseases, pests and abiotic stresses. Introgression of favorable genes from tropical maize can broaden the genetic basis of temperate maize, improve abiotic and biotic stress tolerance, optimize heterotic patterns, and develop improved temperate-tropical hybrids. There are two primary ways for utilization of tropical germplasm in temperate maize breeding. One way is to conduct adaptive selection for tropical germplasm in transition regions, development
### Table 1. 51 maize inbred lines used for development of a multiple-hybrid population (MHP).

| No | Name | Pedigree | Origin | Ecotype |
|----|------|----------|--------|---------|
| P1 | NK764 | NK235 × B73 | USA | Temperate |
| P2 | AS6103 | (AS10631 × A62) × (B14AH × B14AH) | USA | Temperate |
| P3 | F42 | B73 mutation selection | USA | Temperate |
| P4 | FAPW | B14AH × B37Ht | USA | Temperate |
| P5 | LH51 | Mo17 Backcross 5 recovery | USA | Temperate |
| P6 | LK9801 | Ye502 × H2 | China | Temperate |
| P7 | NS701 | A632 × B37Ht | USA | Temperate |
| P8 | PHG83 | PHH14 × PH207 | USA | Temperate |
| P9 | PHN47 | Pioneer Hi-Beed proprietary inbred | USA | Temperate |
| P10 | Dan340 | Lu Jiu Kuan × C060 | China | Temperate |
| P11 | H21 | American hybrid P78599 | China | Temperate |
| P12 | HZ4 | Improved from landrace, Tang Si Ping Tou | China | Temperate |
| P13 | Nan21-3 | Yugoslavian Hybrid | China | Temperate |
| P14 | Q205 | (YeAl141 × ZhongX1017) × Population70 | China | Temperate |
| P15 | Ye78 | U8112 × Shen5003 | China | Temperate |
| P16 | B73 | BSSS C5 (Iowa Stiff Stalk Synthetic) | USA | Temperate |
| P17 | LH132 | (H93 × B73) × B73 | USA | Temperate |
| P18 | Mo17 | CI03 × 187-2 | USA | Temperate |
| P19 | PH4CY | PH7YO × PH1E | USA | Temperate |
| P20 | PH6WC | PH01N × PH09B | USA | Temperate |
| P21 | Chang7-2 | HZ4 × WeiChun | China | Temperate |
| P22 | Q319 | American Hybrid P78599 | China | Temperate |
| P23 | Si287 | 444 × 255 | China | Temperate |
| P24 | Tie7922 | American Hybrid P3832 | China | Temperate |
| P25 | Zheng22 | (Duqing × E28) × Lu JiuKuan | China | Temperate |
| P26 | Zheng58 | Variety plants from Ye478 | China | Temperate |
| P27 | Dan598 | (Dan340 × Danhuang11) × (Danhuang02 × Dan599) | China | Temperate |
| P28 | Huang C | (Huangxiao162 × ZJ300/02) × Tuxpeno | China | Temperate |
| P29 | TR0397 | [CML159/(CML159/[MSRXPOOL9]C1F2:205-1/(SU32)])-5-3-X-X-1-BB]F2-3x]-8-1-1-BB-1-B | CIMMYT | Tropical |
| P30 | CL04934A | CL-04934 (P4922H12-5-4xP23C2-11-1-2-2-2-B*10 | CIMMYT | Tropical |
| P31 | CML418 | G15C22H100*-2-2-6-1-2-BBBB | CIMMYT | Tropical |
| P32 | CML84 | G32C19E512-2-#1-B-#BBB-6-B | CIMMYT | Tropical |
| P33 | CML96 | P42CA623-1-2-B-#6-2-B | CIMMYT | Tropical |
| P34 | Chuan29 Female | Commercialized hybrid | China | Tropical |
| P35 | TR0395 | [(SML × SMQPM) × (MTL × SMQPM)]F1S6-1-25-BB-1-B | CIMMYT | Tropical |
| P36 | TR0403 | [CML389/CML176]-B-29-2-2-B*5 | CIMMYT | Tropical |
| P37 | TR0412 | [DTPWC8F31-4-2-1-6/CML444/[ZM521B-66-1-1-1-BB]-3-2-1-BB-B | CIMMYT | Tropical |
| P38 | TR0415 | [Eq320-92SEW2-77/[DMRESR-W]/EarlySel-#1-2-4-B/CML836]-B-11-3-2-2-#-B*4 | CIMMYT | Tropical |
| P39 | TR0423 | 052A621-BB-#16-2-1-B-B | CIMMYT | Tropical |
| P40 | 18-599 | American Hybrid P78599 | China | Tropical |
| P41 | 622016-ZCN-2 | (Zhongnoo No.2 × H99) × H99 × H99 | CIMMYT | Tropical |
| P42 | CML206 | [EV79920/EVPO44SRBC3]F1B378R-2-3SR-2-4-3-BB | CIMMYT | Tropical |
| P43 | CML312SR | [M3R/312]-117-2-2-1-B4-1-BB-B | CIMMYT | Tropical |
| P44 | CML312SRQ | CML312SRQ = [[CML-RCWQ83×CML312SR]×CML312SR]×CML312SR]-15-1-BB | CIMMYT | Tropical |
| P45 | CML330 | 89[SUWAN/422]/P4758/MP78518]-7-1-1-1-1-1-1-1-1-B-BB | CIMMYT | Tropical |
| P46 | CML504 | [COMPE2/P43SR/COMPE2]-#-20-1-1-B-1-BB-6-BB | CIMMYT | Tropical |
| P47 | CML85 | P34C5F21-2-1-2-2-# | CIMMYT | Tropical |
| P48 | DTM227B | DTPWC9-F104-5-4-1-1-BB-B | CIMMYT | Tropical |
| P49 | TR0501 | INTA-191-2-1-2-1-B-B | CIMMYT | Tropical |
| P50 | TR0582 | P501SRc0-F2-47-3-1-1-B-B-B | CIMMYT | Tropical |
| P51 | Jiao51 | Improved from landrace | China | Tropical |

of temperate-tropical composite populations and reciprocal recurrent selection for performance of temperate by tropical hybrids. Another method is to directly modify tropical maize that do not adapt in temperate conditions by selection for precocious, short flowering interval, and disease and lodging resistance. From the 1960s to early 1980s,
Table 2. Key Chinese inbreds used in a multiple-hybrid population (MHP)and their contribution to hybrid and inbred breeding programs. NCLB: Northern corn leaf blight; SCLB: Southern corn leaf blight; MDMV: Maize dwarf mosaic virus; MRDV: Maize rough dwarf virus; BYDV: Barley yellow dwarf virus; BLSB: Banded leaf sheath blight; ASI: anthesis-silking interval.

| Inbreds       | Characteristics                                      | Hybrids released                                      | Inbreds developed                          | Hybrids released with derivatives               |
|---------------|------------------------------------------------------|------------------------------------------------------|--------------------------------------------|------------------------------------------------|
| LX9801        | Compact plant type, high combining ability, strong resistance to NCLB, SCLB, MDMV and head smut | Ludan981, Ludan984                                   | LX03-2, Jing92                             | Jingke968, Ludan9066                           |
| Chang7-2      | Compact plant type, small tiller size, less branch number, abundant pollen and high GCA | Anyu5, Zhengan958, Jidan7, Jifeng96                  | Xun926, Xun92-8, K17, HD568                | Xundan20, Zhongdan909                          |
| Dan340        | Erect leaves above ear, flourishing roots, strong resistance to NCLB and head smut | Yedan13, Danyu15, Jidan159 and Jidan27               | Tie9010, Xi502, Dan232                    | Xiyu3, Yedan20, Tiedan12, Dan605               |
| H21           | Broad leaf blade, long internode, small ASI, strong resistance to NCLB, head smut and stalk rot | Jindan33, Xinkang5                                   |                                          |                                                |
| LX9801        | Compact plant type, high combining ability, strong resistance to NCLB and head smut | Ludan981, Xingchen97, M81                           |                                          | Neyongyu39, Zhongxing56, Ludan999             |
| HZ4           | High combining ability, strong disease resistance, compact plant type | Yandian14, Yedan2, Yedan4                           | Chang7-2, J444, K12, 897, J854, Tianyan4, J856, 196, J853 |                                                |
| Nan21-3       | Flourishing roots, stiff stalk, small ASI, strong resistance to NCLB and dwarf mosaic disease | Nansandanjiao, Nanqidanjiao, Nanqidanjiao, Nanqianjiao, Nanhuangdanjiao | 286-4, Bai286                           | Wandan11, Ke118, Wandan14                      |
| Qi205         | High combining ability, strong resistance, high seed yield and good adaptability | Chengdan17, Chengdan18, Zhongdan940                  | Tie925, ISA01                              | Ludan718, Xieshiang11                         |
| Qj319         | Compact-type plant, abundant pollen, strong resistance to lodging and southern corn rust | Ludan50, Ludan963                                   | LX9999, JN15, PCS8                        | Ludan606, Jingke308, Tiantai33                |
| Si287         | Slender leaves, abundant pollen, high seed-setting rate, and high combining ability | Jidan27, Jidan46                                    | KM36                                      | Jinongda115, Jinongda578                      |
| Tie7922       | Stiff stem, strong resistance to drought, lodging and waterlogging, high resistance to NCLB and SCLB, head smut and stalk rot | Tiedan8, Yayu2, Simi25                               | TieC605-2, Dan9046, Liao2345              | Tiedan10, Tiedan19, Danyu2, Danyu26, Liao2345 |
| Ye478         | High combining ability, compact-type, short and stout stem, high comprehensive agronomic traits | Zhongdan8, Yedan12, Yedan13, Yedan19                 | K22, Jing99, Yuqing858                    | Shandan21, Shan930, Jingu16, Yuqing1           |
| Zheng22       | Stocky stem, superior root system, strong resistance to drought, NCLB and SCLB, MRDV and BYDV | Yuyu18                                              | Shen151, Zheng22you                       | Shenyu17, Shenyu20, Wuke1                    |
| Zheng58       | High combining ability, compact plant type, strong resistance to NCLB, SCLB and lodging | Ludan9902                                           | M6, WK858                                 | Huannong18, Weike702                          |
| Dan598        | Good stems elasticity and strong resistance to BLSB and head smut | Danyu26, Danyu39, Danke2151, Dongdan60              | M5972, Y14, 765                           | Liangyu99, Huatian866, Danyu502               |
| Huang C       | Compact plant type, high lysine content, strong resistance to NCLB and SCLB | Nongda108, Liyu10                                   | 168, T14-3B                               | Changyu1, Jindan66, Zhongyu335               |
| Jiao51        | High combining ability, good grain quality, stiff stem and lodging resistance | Jiaosandanjiao, Andan136, Yudan11, Yuyu15           |                                          |                                                |
| 18-599        | High combining ability, high seed yield, good quality and good comprehensive agronomic traits | Shengyu9, Dong315, Chuxin11                          | Zhongyu335, Yuanyu18                     | YA0419                                       |
| Chuan 29 Female | High combining ability and heterosis, high ferric content | Chuanandan29                                         |                                          |                                                |

U.S. scientists made great efforts for utilization of tropical maize. With more than 20,000 varieties of tropical maize collected\cite{63}, they found that maize inbreds containing tropical germplasm were not only a useful source to expand the genetic diversity of temperate inbreds, but also competitive in crosses with temperate materials, producing high yielding hybrids\cite{64}. For example, temperate by tropical hybrids showed greater adaptation than temperate hybrids under heat stress environment\cite{65}. In China, some tropical germplasm, such as Tuxpeno, Suwan and Mohuang9, were introduced in 1980s, with several excellent inbreds developed, such as 8703, S37, Taixi113 and 8501. In our MHP 23 tropical inbred lines widely used in breeding programs of China and CIMMYT were included as parental lines, and our results also showed that the tropical lines had great potential for improvement of yield traits for temperate maize. Among three Chinese tropical inbred lines, Jiao51 was selected from landrace 'Jiaomaerhuangzao' in Guizhou, by artificial selection for adaptation and productivity lead to the differentiation between temperate and tropical maize. Improving the adaptability of temperate maize largely depends upon the use of tropical maize germplasm, which hosts rich sources of favorable genes and alleles\cite{66}, particularly for tolerance to abiotic and biotic stresses. A shortcut to the favorable alleles hosting in tropical maize is to develop temperate-tropical hybrids. Considering the significant differentiation between the two groups, the best way to test temperate-tropical crosses is by NC II, where a set of hybrids released with derivatives were used as parents to cross with a set of tropical inbreds.

Temperate by tropical hybrids in maize production. Maize exhibits an astounding capacity for environmental adaptation with wide distribution around the world, but genetic bottlenecks resulting from natural and artificial selection for adaptation and productivity lead to the differentiation between temperate and tropical maize. Improving the adaptability of temperate maize largely depends upon the use of tropical maize germplasm, which hosts rich sources of favorable genes and alleles\cite{67}, particularly for tolerance to abiotic and biotic stresses. A shortcut to the favorable alleles hosting in tropical maize is to develop temperate-tropical hybrids. Considering the significant differentiation between the two groups, the best way to test temperate-tropical crosses is by NC II, where a set of temperate inbreds are used as parents to cross with a set of tropical inbreds.

Moving maize from tropical to temperate means the selection for adaptation to changing day length and temperature conditions. Such photo-thermal responses have been investigated intensively. An improved population BS16...
Agronomic performance of parental lines and their hybrids. Agronomic traits for inbreds and hybrids included in our MHP were evaluated for three years (2013–2015) in two locations (Beijing and Henan) with randomized block design each with two replications. Phenotypic performance across environments was evaluated using the best linear unbiased predictions (BLUPs) estimated by SAS PROC MIXED procedure with genotype, environment and year as random effects (Table 3). Compared to temperate parental lines, tropical inbreds showed higher plant and ear heights, delayed growth and development with significant larger days to tassel, silk and anthesis. For yield traits, however, temperate inbreds were overall better than tropical lines with higher grain number, row number, hundred grain weight and grain weight per plant. Among three hybrid subpopulations, significant delayed flowering time were found in both NC II and tropical diallel hybrids, and greater flowering time interval, anthesis-silking interval, was shown in NC II hybrids. For yield traits, NC II crosses showed higher ear length, ear diameter and hundred grain weight, while temperate diallels had greater grain number.

In conclusion, tropical germplasm can be used to improve the yield potential for temperate lines, while temperate germplasm, such as those from the U.S., are important donors to broaden genetic basis for specific breeding programs. As high genetic diversity and great trait variation were observed among parental lines and hybrids, the MHP developed in this study can be used in GWAS for many important traits. With known pedigrees and genotypic information from high-density markers, the structure and relationship among parental lines and thus among the hybrids can be well estimated and thus controlled in GWAS.

Genotypic information generated for the MHP. The 51 parental lines used to develop the MHP have been genotyped by a newly developed maize 55K-SNP chip. The chip contains 30,133 SNPs selected from 600 K Affymetrix® Axiom® Maize Genotyping Array, which evenly distributed on the genome, 4,049 high polymorphic SNPs from widely used 56K-SNP chip MaizeSNP50, 9,395 SNPs from whole genome RNA-seq76, 4,067 SNPs that are tropical specific and generated by resequencing to fill the gaps in the B73 reference genome, and 132 SNPs from the tags for published transgenic events. Genotyping the parents with the improved 55K-SNP chip will benefit genetic and breeding studies because of the improved genome coverage, compared to other SNP chips available.

Applications of the MHP with diallel and NC II mating designs. The MHP proposed in this article can be used to estimate combining ability and heterosis for yield and yield components using hybrid phenotypes compared with their parents, which can be further analyzed in GWAS along with other traits such as hybrid phenotypes per se. Associated information can be used for genome-wise prediction of hybrid performance. The MHP and its derived secondary populations can be also used as populations in genetics, genomics and plant breeding. As basic information required for applications of the MHP, a cluster analysis was performed to classify the parental lines (Fig. 2a). A neighbor-joining tree was constructed based on Roger’s genetic distance, which clearly separated the parental lines into two major groups, temperate and tropical. The 28 temperate inbred lines were further divided into six heterotic groups, Lancaster, LRC, Reid, PA, SPT and PB, as revealed in previous studies.

Genetic analysis of combining ability and hybrid performance. Evaluation of combining ability and heterosis is the first step for breeding inbred lines towards the development of commercial hybrids. This process theoretically should go through for all parental lines and their possible hybrids. Both diallel and NC II designs can provide elaborate information on the genetic identity of genotypes, especially on dominance-recessive relationships and some genetic interactions. As an example, a complete diallel crosses involving eight parents was produced to
construct HapMaps in maize with three different versions of HapMap developed. RNA-seq and other sequencing combinations (haplotypes) can be developed to cover the whole genome. For example, haplotypes have been used to infer genetic diversity in various markers such as SNP, copy number variations (CNV) and structure variation (SV), and their alleles and genotypes have been identified. These high-density chips provide a quick approach to reveal alleles, genotypes, and a new 55K-SNP chip with improved genome coverage.

GWAS based on markers, alleles and haplotypes. Genotyping of parental lines and then inferring their hybrid genotypes can be done through various genotyping platforms including sequencing (both whole genome and simplified genome sequencing or genotyping by sequencing, GBS) and chip-based genotyping. After genotyping, various markers such as SNP, copy number variations (CNV) and structure variation (SV), and their alleles and combinations (haplotypes) can be developed to cover the whole genome. For example, haplotypes have been used to construct HapMaps in maize with three different versions of HapMap developed. RNA-seq and other sequencing technologies such as methylation-based sequencing have been used to identify expressed genes, epialleles and haplotypes associated with epigenetics.

Resequencing and GBS have been widely used in maize germplasm fingerprinting, genetic diversity analysis and GWAS.

Using numerous markers identified through various sequencing technologies, high density chips have been developed. In maize there are several high-density SNP chips available, including 56 K SNP chip, 600K-SNP chip and a new 55K-SNP chip with improved genome coverage. In human, high-density chips containing SV markers have been also developed. These high-density chips can be updated with functional- and gene-markers so that more gene-related information can be generated. High-density chips provide a quick approach to reveal alleles, genotypes, haplotypes for a large number of samples with fixed and comparable markers. They have been widely used in maize germplasm fingerprinting, genetic diversity analysis and GWAS.

The currently available studies are mostly limited in a single application of genotyping methods, each with specific disadvantages. For example, gene arrays can not determine accurate positions of the target genes in multiple cell type, GBS may generate large quantities of missing data, and whole-genome resequencing shows high coverage but poor pertinence versus targeted resequencing. Therefore, simultaneous use of multiple genotyping techniques might be imperative. In soybean, 1.4 million tag SNPs were used as a reference to impute a large set of SNPs with a panel of 301 soybean accessions through whole genome resequencing, GBS and SNP-array (SoySNP50K), and this imputation can be used to fill in missing genotypes and untyped loci with high accuracy. In maize, CRTRB1, LCYE and other key genes or genomic regions that govern rate-critical steps in the upstream pathway were identified for various carotenoids using GWAS with 380 CIMMYT inbred lines genotyped by 55 K chip and GBS.

By genotyping the MHP parental lines with the 55 K chip, the genotypes of hybrids can be deduced. In the current MHP, 724 hybrids have been generated by using 51 parental lines, which is much less than what can be generated in all possible ways (51 × (51-1)/2 = 1275). The SNP loci that are heterozygous in one of the two parental lines are scored as missing. SNP markers with missing data rate <20% were extracted to deduce hybrid genotypes. A total of 37,527 SNPs (with minor allele frequency >5% and missing data rate less than 25%) for 724 hybrids were used for structure analysis and GWAS. Polygenic component and main effect (additive and dominant effect) analysis were estimated by a web tool PEPIS (the pipeline for estimating EPISstatic genetic effects, http://bioinfo.noble.org/ PolyGenic_QTL/) (Fig. 3). For PEPIS, three files, including additive genotypic data, dominance data and phenotypic data, were uploaded. The genotype of each hybrid at each specific marker locus was coded as three genotypes, A for one homozygous genotype, B for the second and H for heterozygous genotype. The K matrix corresponding to the marker-generated kinship was calculated and used to estimate variance components with restricted maximum likelihood method (REML). The polygenic structure for the trait was examined with the given variance ratios, and genome scanning for main and epistatic effects was performed for each marker (main) effect and marker pair interaction (epistatic) effects. With the sub-pipelines in PEPIS, the sub-pipeline 1 was used to calculate kinship matrix, and sub-pipeline 2 for polygenic component analysis and genome scanning for main and epistatic effect QTL. As a comparison, Q + K model for classical GWAS analysis was also performed. Population structure (Q)

| Trait                        | Temperate inbreds | Tropical inbreds | Temperate diallel | Tropical diallel | NC II |
|------------------------------|-------------------|------------------|-------------------|------------------|-------|
| Plant height                 | 182.73±(20.88)    | 183.36±(14.94)   | 252.56±(19.85)    | 247.85±(11.37)   | 269.73±(14.87) |
| Ear height                   | 74.24±(14.78)     | 90.26±(15.98)    | 111.13±(14.60)    | 116.25±(9.20)    | 131.86±(13.78) |
| Days to tassel               | 56.41±(3.26)      | 65.59±(3.82)     | 51.89±(2.07)      | 58.50±(3.82)     | 56.81±(2.28)   |
| Days to silk                 | 61.39±(3.14)      | 70.21±(4.23)     | 55.89±(1.89)      | 62.30±(3.26)     | 60.86±(2.39)   |
| Days to anthesis             | 59.53±(3.22)      | 68.80±(4.03)     | 54.16±(1.91)      | 60.90±(3.03)     | 59.05±(2.36)   |
| Anthesis-silking interval    | 2.19±(0.33)       | 2.55±(0.28)      | 1.68±(0.33)       | 1.60±(0.53)      | 1.87±(0.35)    |
| Ear length                   | 13.66±(1.43)      | 12.45±(1.28)     | 17.33±(1.42)      | 15.65±(0.53)     | 17.96±(1.02)   |
| Ear diameter                 | 38.91±(2.84)      | 35.08±(2.19)     | 46.16±(2.34)      | 46.15±(1.66)     | 46.85±(2.35)   |
| Row number                   | 13.68±(1.43)      | 12.04±(0.89)     | 15.62±(1.20)      | 14.35±(1.02)     | 15.02±(1.38)   |
| Grain number per row         | 21.66±(2.42)      | 18.68±(1.77)     | 35.50±(3.03)      | 33.00±(1.45)     | 34.78±(2.70)   |
| Grain number                 | 300.76±(45.01)    | 237.41±(25.18)   | 554.37±(56.91)    | 474.25±(42.90)   | 522.45±(53.96) |
| Hundred grain weight         | 26.56±(3.43)      | 23.63±(3.04)     | 29.32±(3.26)      | 31.45±(1.85)     | 33.13±(3.08)   |
| Grain weight per plant       | 52.91±(10.23)     | 44.77±(6.38)     | 128.90±(18.96)    | 135.35±(3.91)    | 134.70±(8.64)  |

Table 3. Agronomic traits observed for parental lines and hybrids in a multiple-hybrid population (MHP). Different letters in the same row indicate significant difference by Duncan test at 1% probability, and standard errors are given in parenthesis.
was determined by fastSTRUCTURE\textsuperscript{96}, the number of subpopulation was corrected by neighbor-joining genetic distance\textsuperscript{77}, a cluster dendrogram was constructed using FigTree v1.4.2 software (Fig. 2b), and the phenotypic contribution of population structure was estimated with SAS PROC GLM procedure. The kinship matrix (K) and its contribution to phenotype were calculated by Tassel 5.0 software\textsuperscript{97}. The LD (r\textsuperscript{2}) was calculated to estimate the degree of LD between pairwise SNPs and the sliding window size was set as 500 with a step of 50 markers using Tassel 5.0\textsuperscript{(7)} (Fig. 2c). Four flowering traits, DTT (days to tassel), DTS (days to silk), DTA (days to anthesis) and ASI (anthesis-silking interval) were taken as examples for GWAS in this report (Fig. 4). Association analysis were conducted using the compressed mixed linear mixed model (CMLM) and the P3D method including 37,527 SNP markers, population structure (Q), kinship matrix (K) and phenotypic information with Tassel 5.0 software\textsuperscript{97}. The casual genes or QTN associated were identified with a significant threshold (-log\textsubscript{10} P ≥ 5.81) corrected by the Bonferroni test.

Population structure was assessed by fast STRUCTURE for K values ranging from 1 to 10, and the population was divided into six subgroups with the correction of NJ genetic distance (Fig. 2b). The phenotypic contribution of population structure for DTT, DTS, DTA and ASI were 0.56, 0.58, 0.58 and 0.03, respectively, and the kinship contribution for phenotype were 0.99, 0.99, 0.98 and 0.52, respectively. The distributions of r\textsuperscript{2} over the whole genome were presented in Fig. 2c. The r\textsuperscript{2} values declined sharply as the distance increased and the average r\textsuperscript{2} was estimated at ~100 kb, when the cut-off value for r\textsuperscript{2} was set to 0.2. The polygenic variance component ratios for the flowering traits are shown in Fig. 3. Additive genetic variance accounted for 79.5%, 76.8% and 78.5% of trait variance for DTT, DTS and DTA, respectively, whereas additive by additive genetic variance accounted for 50.2% variance for ASI, dominant effects accounted secondarily for 16.3%, 18.6% and 16.4%, and 20.7%, respectively. Our result is consistent with the previous study for the additive effects for the flowering traits with unimportant epistasis\textsuperscript{28}. In PEPIS, polygenic dominant effects accounted secondarily for 16.3%, 18.6%, 16.4% and 20.7%, respectively. Our result is consistent with the previous study for the additive effects for the flowering traits with unimportant epistasis\textsuperscript{28}. In PEPIS, polygenic model for additive, dominant, additive by additive, additive by dominant, dominant by additive and dominant by dominant adopted to hybrids and main-effect distribution including additive and dominant are shown in Fig. 4a, and it was confirmed by the results for additive model in Tassel (Fig. 4b). Therefore, we can run GWAS with hybrid phenotypes using the Tassel software for additive effect analysis for the traits that are largely controlled by additive genetic variance. Given the additive effect accounted for the most for the flowering traits and the purpose for this paper is GWAS with maize MHP, dominant effects analysis will be discussed elsewhere.

A large number of significant markers were identified for flowering traits and significant signals around known loci were used to identify candidate genes and QTN co-located with SNPs based on the B73 reference genome (http://www.maizegdb.org/gbrowse). Comparing GWAS analyses from PEPIS and Tassel, we found consistent results for flowering traits with five overlapping peak regions significant for pleiotropy allelic effects for four flowering traits on chromosomes 2, 3, 6 and 8 (Fig. 4). A peak was mapped closed to GRMZM2G065276 on bin2.06, a homolog of FCA on Arabidopsis chromosome 4\textsuperscript{98}. Arabidopsis FCA combined with another gene FPA encoded RNA binding protein, downregulated transcription of FLC, and promoted flowering\textsuperscript{99}. Similarly, two QTN were detected on bin3.07, which were significantly associated with DTS (as qdsilk1 and qdsilk8 mapped in a previous research\textsuperscript{100}). The related loci mdh3 (malate dehydrogenase3) showing four alleles in Kalmia latifolia\textsuperscript{101} regulated maize pollen tube growth\textsuperscript{102} and encoded a cytosolic malate dehydrogenase for glyoxylate cycle activity in Arabidopsis\textsuperscript{103}. The peak on bin6.01 contained GRMZM2G004959 which had a homologous gene in Arabidopsis (AT3G61230.1) expressing in pollen and tube growth\textsuperscript{104}. Peak SNPs on the short arm of chromosome 8 (bin 8.01) revealed a phosphatidyethanolamine-binding protein (ZCN9) gene GRMZM2G021614, the homologous gene in rice LOC_Os05g50890.1 OsGH3.5 coding JA-amino acid synthetase1 modulating light and JA signal in the photomorphogenesis\textsuperscript{105} and flower opening and anther dehiscence\textsuperscript{106}.

Table 4. Agronomic traits for the hybrids from parental inbreds in different heterotic groups. Different letters in the same row indicate significant difference by Duncan test at 1% probability, and standard errors are given in parenthesis.
For ungenerated hybrids (1275–724 = 551), their performance can be predicted using genomic best linear unbiased prediction (GBLUP)\textsuperscript{108} or empirical Bayesian approach\textsuperscript{37}, and in general Bayes and GBLUP possess high accuracy. With an optimized scheme, we should be able to predict performance for much more hybrids with much less hybrids tested. For example, we may be able to predict a full set of diallel hybrids based on the results from NC II mating design using the same set of the parental lines. A statistical method has been developed and applied to such prediction in rice\textsuperscript{108}, where a set of 278 selected IMF2 hybrids, developed from a RIL population between Zhenshan 97 and Minghui 63\textsuperscript{22}, was used as a training sample to predict the rest 21,667 hybrids, with 16% yield increase in the top 100 selection compared with the average of all hybrids. The results revealed GBLUP as the best method for prediction compared with LASSO and SSVS. The GBLUP involves three stages: construction of respective kinship matrix for training sample and all hybrids, parameter estimation via cross-validation and prediction for missing and rest hybrids\textsuperscript{108}. The empirical Bayesian approach was developed and used for a partial NC II mating design in \textit{Brassica napus} including 284 \textit{F}_1 hybrids, and 143 restorers and 298 sterile lines were used to predict the phenotypes for missing hybrids, with QTL effects estimated by empirical Bayesian. GCA and SCA would be estimated and elite parents and hybrids could be predicted\textsuperscript{37}. In our following research, the GBLUP approach\textsuperscript{108} will be used for analysis of the MHP and reported elsewhere. A recent research showed that predictability of yield was nearly twofold with metabolomics data compared with genomic prediction, despite the latter would be the most efficient method for high-heritability traits\textsuperscript{109}, and hybrid prediction by metabolomics data may be a novel pathway in crop breeding programs that should be explored more in the future.

**Breeding using MHPs.** Breeding hybrid crops is an important tool that may remarkably promote yield enhancement from 30 to 400% resulting of the heterosis\textsuperscript{110}, and hybrid crops were used to investigate heterosis for parental inbreds with maximum combining ability. The populations created should have a higher possibility of acquiring diverse and accurate phenotype performance than natural populations\textsuperscript{111}. Diallel and NC mating designs have been widely used in plant breeding programs to maximize selection response and the opportunity for managing coancestry for breeding populations. A prominent feature of diallel and NC mating designs is that both designs provide breeding populations with known pedigrees, which can be used to estimate specific parental genetic effects for backward selection\textsuperscript{112}. Estimation of combining ability and heterosis can be used to guide our future breeding activities using the hybrids and their parental lines and derivatives. GCA and SCA, two essential factors in developing breeding strategies, are two main genetic parameters that can be obtained from diallel and NC analysis. GCA reflects the stand or fall of the inbred lines per se while SCA is the ability to develop elite hybrids with specific partners. A line × tester population from a collection of 302 diverse inbreds with a common tester B73 was used to predict heterosis by inbred performance and genetic distance between parents for traits, and the result suggested that heterosis could be explained by trait-dependence\textsuperscript{113}. An unbalanced NC II for 400 hybrids and 79 inbreds was used for joint analysis of hybrids and parental lines to predict the performance of untested hybrids\textsuperscript{114}. A set of NC
II crosses between 285 Dent and two Flint inbreds genotyped with 56,110 SNPs and 130 metabolites was used to predict combining ability, allowing a reliable screening of large collections of diverse inbred lines for the potential to create superior hybrids. A set of 136 hybrids and 17 parental lines was used to estimate combining ability and heterosis for stress and non-stress environments, and SCA was significantly correlated with genetic distance. The MHP used in this research contains three subpopulations from temperate, tropical and temperate by tropical hybrids, which can be used for breeding different ecotypes and for gene pyramiding and recombination to combine the merits from different maize germplasm sources.

Secondary populations derived from the MHP. Many secondary populations can be generated from the parental lines and hybrids of MHP by selfing the hybrids or crossing among parental lines and hybrids. The secondary populations include not only F$_2$, BC and RILs, but also MAGIC and NAM. MHP-based researches will provide important information on genetics and genomics such as genes, alleles, haplotypes and population structure about their parental lines and associated hybrids, which can be used to guide further researches on the secondary populations.

As discussed in the previous section, the MHP developed in this study can be shared through distribution of the 51 parental lines, which is much simpler than sharing all possible hybrids (1275) that can be generated from the 51 parents. By sharing the parental lines, a flexible number of hybrids can be generated and tested for agronomic traits of interest to specific users. Genotypic information will be generated and accumulated through genotyping parental lines using different genotyping platforms conducted by collaborators worldwide, which can be shared through a website. As a start point of sharing, we are happy to distribute the 51 parental lines, along with their genotypic information available so far, to those who are interested in using the MHP in their genetics and breeding researches. As a return, we would appreciate all collaborators for sharing their phenotype information generated for specific traits in specific environments or locations and genotypic information generated with their own genotyping platforms.

Conclusions
Compared with bi- and multi-parental populations and natural ones, MHPs are more suitable for hybrid crops. From various mating designs, a large number of hybrids can be generated to form MHPs, and a set of diallel hybrids from a relatively large number of parental lines would be the best. Each set of diallel hybrids can be divided into three subsets of hybrids with distinct genetic properties, two diallels each representing one ecotype and one NC II representing between-ecotype hybrids. A partial set of hybrids can be generated, selectively or randomly, from the full set, and then used to predict the rest ungenerated hybrids. By sharing parental lines along with their genotypic information, a large number of hybrids can be generated and tested by worldwide collaborators.

In addition to conventional quantitative genetics on combining ability, heterosis and hybrid performance, MHPs can be widely used in modern genetics, genomics and breeding. By high-density genotyping, MHPs can be used in GWAS for phenotypic data collected under diverse environments, including traits of agronomic importance, and heterosis and combining ability as well, which can be based on markers, alleles and haplotypes. MHPs per se and their derived secondary populations can be used in breeding for both inbred lines and hybrids. Taking 724 hybrids derived from 51 parental lines and four flowering traits as an example, we compared two independent GWAS methods, PEPIS software developed for hybrids and TASSEL software designed for inbred line populations. The two

---

**Figure 3.** Polygenic component ratios of the flowering traits for maize multiple-hybrid population (MHP). Additive effect accounted for the most of variance for DTT (days to tassel), DTS (days to silk) and DTA (days to anthesis), while additive by additive effect accounted for 50.2% of the variance for ASI (anthesis-silking interval).
methods revealed highly consistent results with five overlapping chromosomal regions identified and used for discovery of candidate genes and QTN. Our results indicate that MHPs are powerful in GWAS for hybrid-related traits and will be widely used in the molecular breeding era.

References
1. Desta, Z. A. & Ortiz, R. Genomic selection: genome-wide prediction in plant improvement. *Trends Plant Sci.* 19, 592–601 (2014).
2. Thornsberry, J. M. *et al.* Dwarf8 polymorphisms associate with variation in flowering time. *Nat. Genet.* 28, 286–289 (2001).
3. Hansen, M. *et al.* Linkage disequilibrium mapping of the bolting gene in sea beet using AFLP markers. *Genet. Res.* 77, 61–66 (2001).
4. Yano, K. *et al.* Genome-wide association study using whole-genome sequencing rapidly identifies new genes influencing agronomic traits in rice. *Nat. Genet.* 48, 927–934 (2016).
5. Li, Y. *et al.* Identification of genetic variants associated with maize flowering time using an extremely large multi-genetic background population. *Plant J.* 86, 391–402 (2016).
6. Muqaddasi, Q. H. *et al.* Genome-wide association mapping of anther extrusion in hexaploid spring wheat. *PLoS One* 11, e0155494 (2016).
7. Upadhyaya, H. D. *et al.* Association mapping of germinability and seedling vigor in sorghum under controlled low-temperature conditions. *Genome* 59, 137–145 (2015).
8. Maurer, A., Darba, V. & Pillen, K. Genomic dissection of plant development and its impact on thousand grain weight in barley through nested association mapping. *J. Exp. Bot.* 67, 2507–2518 (2016).
9. Erozh, E. S., Yu, J. & Buckler, E. S. Applications of linkage disequilibrium and association mapping in maize *In Molecular genetic approaches to maize improvement* (ed. Kriz, A. L. & Larkins, B. A.) 173–195 (Springer Berlin Heidelberg, 2009).
10. Yu, J. & Buckler, E. S. Genetic association mapping and genome organization of maize. *Curr. Opin. Biotechnol.* 17, 155–160 (2006).
11. Flint-Garcia, S. A., Thornsberry, J. M. & Buckler, E. S. Structure of linkage disequilibrium in plants. *Annu. Rev. Plant Biol.* 54, 357–374 (2003).
12. Yu, J. *et al.* A unified mixed-model method for association mapping that accounts for multiple levels of relatedness. *Nat. Genet.* 38, 203–208 (2006).
13. Korte, A. *et al.* A mixed-model approach for genome-wide association studies of correlated traits in structured populations. *Nat. Genet.* 44, 1066–1071 (2012).
14. Segura, V. *et al.* An efficient multi-locus mixed-model approach for genome-wide association studies in structured populations. *Nat. Genet.* 44, 825–830 (2012).
15. Lippert, C. *et al.* FaST linear mixed models for genome-wide association studies. *Nat. Methods* 8, 833–835 (2011).
16. Wang, Q. *et al.* A SUPER powerful method for genome wide association study. *PLoS One* 9, e107684 (2014).
17. Casale, F. P. *et al.* Efficient set tests for the genetic analysis of correlated traits. *Nat. Methods* 12, 755–758 (2015).
18. Wang, S. B. *et al.* Improving power and accuracy of genome-wide association studies via a multi-locus mixed linear model methodology. *Sci. Rep.* 6, 19444 (2016).
19. Xu, Y. Populations in genetics and breeding. *In Molecular plant breeding* (CABI, UK). 4, 113–145 (2010).
20. Zou, C., Wang, P. & Xu, Y. Bulked sample analysis in genetics, genomics and crop improvement. *Plant Biotechnol. J.* 14, 1941–1955 (2016).
21. Burr, B. *et al.* Gene mapping with recombinant inbreds in maize. *Genetics* 118, 519–526 (1988).
22. Xu, Y. Global view of QTL: rice as a model. *In Quantitative genetics, genomics, and plant breeding* (ed. Kang, M. S.) 109–134 (CABI UK, 2002).
23. Hua, J. et al. Single-locus heterotic effects and dominance by dominance interactions can adequately explain the genetic basis of heterosis in an elite rice hybrid. *Proc. Natl. Acad. Sci. USA* **100**, 2574–2579 (2003).
24. Rao, S. & Xu, S. Mapping quantitative trait loci for ordered categorical traits in four-way crosses. *Hereditas* **81**, 214–224 (1998).
25. He, X. H. et al. Mapping of epistatic quantitative trait loci in four-way crosses. *Theor. Appl. Genet.* **122**, 33–48 (2011).
26. Yu, J. et al. Genetic design and statistical power of nested association mapping in maize. *Genetics* **178**, 339–351 (2008).
27. McMullen, M. D. et al. Genetic properties of the maize nested association mapping population. *Science* **325**, 737–740 (2009).
28. Buckler, E. S. et al. The genomic architecture of maize flowering time. *Science* **325**, 714–718 (2009).
29. Kump, K. L. et al. Genome-wide association study of quantitative resistance to southern leaf blight in the maize nested association mapping population. *Nat. Genet.* **43**, 163–168 (2011).
30. Tian, F. et al. Genome-wide association study of leaf architecture in the maize nested association mapping population. *Nat. Genet.* **43**, 159–162 (2011).
31. Zhang, N. et al. Genome-wide association of carbon and nitrogen metabolism in the maize nested association mapping population. *Plant Physiol.* **168**, 575–583 (2015).
32. Kover, P. X. et al. A multiparent advanced generation inter-cross to fine-map quantitative traits in *Arabidopsis thaliana*. *PLoS Genet.* **5**, e1000551 (2009).
33. Huang, B. E. et al. A multiparent advanced generation inter-cross population for genetic analysis in wheat. *Plant Biotechnol J.* **10**, 826–839 (2012).
34. Bandillo, N. et al. Multi-parent advanced generation inter-cross (MAGIC) populations in rice: progress and potential for genetics research and breeding. *Rice* **6**, 11 (2013).
35. Dell’Acqua, M. et al. Genetic properties of the MAGIC maize population: a new platform for high definition QTL mapping in Zea mays. *Genome Biol.* **16**, 167 (2015).
36. Rafalski, A. & Morgante, M. Corn and humans: recombination and linkage disequilibrium in two genomes of similar size.
37. Dell’Acqua, M. et al. Analysis of photoperiod sensitivity within a collection of tropical maize populations.
38. Dell’Acqua, M. et al. Using tropical maize germplasm for temperate breeding.
39. Dell’Acqua, M. et al. The breeding and application of maize hybrid Zhengdan 958.
40. Dell’Acqua, M. et al. Genetic variation in fifty-three maize inbred lines in relation to drought tolerance at seeding stage.
41. Dell’Acqua, M. et al. Mapping quantitative trait loci for ordered categorical traits in four-way crosses.
42. Dell’Acqua, M. et al. Determination of combining abilities and heterotic patterns of fourteen medium to late maturing Iranian maize inbred lines using diallel mating design.
43. Dell’Acqua, M. et al. Heterosis studies in diallel crosses of maize.
44. Dell’Acqua, M. et al. The breeding and use of Nan21-3 higher combining ability corn line.
45. Dell’Acqua, M. et al. Genetic design and statistical power of nested association mapping in Zea mays. *Genome Biol.* **16**, 167 (2015).
115. Riedelsheimer, C. 
116. Betrán, F. J., Ribaut, J. M., Beck, D. & Gonzalez de León, D. Genetic diversity, specific combining ability, and heterosis in tropical maize
113. Flint-Garcia, S. A. 
114. Schrag, T. A. 
108. Xu, S., Zhu, D. & Zhang, Q. Predicting hybrid performance in rice using genomic best linear unbiased prediction. 
105. Danilevskaya, O. N. 
106. Riemann, M., Riemann, M. & Takano, M. Rice 
104. Ye, J. & Xu, M. Actin bundler PLIM2s are involved in the regulation of pollen development and tube growth in 
102. Frova, C. & Sari-Gorla, M. Quantitative trait loci (QTLs) for pollen thermotolerance detected in maize. 
103. Pracharoenwattana, I., Cornah, J. E. & Smith S. M. Arabidopsis peroxisomal malate dehydrogenase functions in 
101. Levri, M. A. A measure of the various modes of inbreeding in 

This research was supported by National Key Basic Research Program of China (2014CB138206), National Key Research and Development Program of China (2016YFD0101803), National Science Foundation of China (Grant #31271736), NSFC-CGIAR International Collaborative Program (31361140364), Beijing Science and Technology project (D14110005014003), Science and Technology Partnership Program, Ministry of Science and Technology (KY201402017), Agricultural Science and Technology Innovation Program (ASTIP) of CAAS, and CGIAR Research

Acknowledgements

Yu, Y. et al. Genetic diversity and structure of the core collection for maize inbred lines in China. Maydica 52, 181–194 (2007).

Lu, Y. et al. Molecular characterization of global maize breeding germplasm based on genome-wide single nucleotide polymorphisms. Theor. Appl. Genet. 120, 93–115 (2009).

Gore, M. A. et al. A first-generation haplotype map of maize. Science 326, 1115–1117 (2009).

Chia, J. M. et al. Maize HapMap2 identifies extant variation from a genome in flux. Nat. Genet. 44, 803–809 (2012).

Bukowski, R. et al. Construction of the third generation Zea mays haplotype map. bioRxiv 20969 (2015).

Meissner, A. et al. Reduced representation bisulphite sequencing for comparative high-resolution DNA methylation analysis. Nucleic Acids Res. 33, 5868–5877 (2005).

Mortazavi, A. et al. Mapping and quantifying mammalian transcriptomes by RNA-Seq. Nat. Methods 5, 621–628 (2008).

Lai, J. et al. Genome-wide patterns of genetic variation among elite maize inbred lines. Nat. Genet. 42, 1027–1030 (2010).

Eldridge, R. J. et al. A robust, simple genotyping-by-sequencing (GBS) approach for high diversity species. PLoS One 6, e19379 (2011).

Jiao, Y. et al. Genome-wide genetic changes during modern breeding of maize. Nat. Genet. 44, 812–815 (2012).

Ganal, M. W. et al. A large maize (Zea mays L.) SNP genotyping array: development and germplasm genotyping, and genetic mapping to compare with the B73 reference genome. PLoS One 6, e28334 (2011).

Unterseer, S. et al. A powerful tool for genome analysis in maize: development and evaluation of the high density 600 k SNP genotyping array. BMC Genomics 15, 823 (2014).

Komura, D. et al. Genome-wide detection of human copy number variations using high-density DNA oligonucleotide arrays. Genome Res. 16, 1575–1584 (2006).

Riedelsheimer, C. et al. Genome-wide association mapping of leaf metabolic profiles for dissecting complex traits in maize. Proc. Natl. Acad. Sci. USA 109, 8872–8877 (2012a).

Torkamaneh, D. & Belzile, F. Scanning and filling: ultra-dense SNP genotyping combining genotyping-by-sequencing, SNP array and whole-genome resequencing data. PLoS One 10, e0131533 (2015).

Suwannapha, W. B. et al. Genome-wide association analysis reveals new targets for carotenoid biofortification in maize. Theor. Appl. Genet. 128, 851–864 (2015).

Zhang, W. et al. PEPES: A pipeline for estimating epistatic effects in quantitative trait locus mapping and genome-wide association studies. PLoS Comput. Biol. 12, e1004925 (2016).

Rai, A., Stothers, M. & Pritchard, J. K. Faststructure: Variational inference of population structure in large SNP data sets. Genetics 197, 573–589 (2014).

Bradbury, P. J. et al. TASSEL: software for association mapping of complex traits in diverse samples. Bioinformatics 23, 2633–2635 (2007).

Siewiakowski, S., Liu, E., Magusin, A. & Dean, C. Cold-induced silencing by long antisense transcripts of an Arabidopsis Polycomb target. Nature 462, 799–802 (2009).

Liu, F. et al. The Arabidopsis RNA-binding protein FCA requires a lysine-specific demethylase 1 homolog to downregulate FLC. Mol. Cell 3, 398–407 (2007).

Ahler, B. S. B., Edwards M. D. & Stuber, C. W. Isozymatic identification of quantitative trait loci in crosses of elite maize inbreds. Crop Sci. 31, 267–274 (1991).

Levri, M. A. A measure of the various modes of inbreeding in 

Frova, C. & Sari-Gorla, M. Quantitative trait loci (QTLs) for pollen thermodtolerance detected in maize. Mol. Gen. Genet. 245, 424–430 (1994).

Pracharoenwattana, I., Cornah, J. E. & Smith S. M. Arabidopsis peroxisomal malate dehydrogenase functions in β-oxidation but not in the glyoxylate cycle. Plant J. 50, 381–390 (2007).

Ye, J. & Xu, M. Actin bundler PLIM2s are involved in the regulation of pollen development and tube growth in Arabidopsis. J. Plant Physiol. 169, 516–522 (2012).

Danilevskaya, O. N. et al. A genomic and expression comendium of the expanded PEP gene family from maize. Plant Physiol. 146, 250–264 (2008).

Riemann, M., Riemann, M. & Takano, M. Rice JASMONATE RESISTANT1 is involved in phytochrome and jasmonate signalling. Plant Cell Environ. 31, 783–792 (2008).

Xiao, Y. et al. OsJAR1 is required for JA-regulated floret opening and anther dehiscence in rice. Plant Mol. Biol. 86, 19–33 (2014).

Xu, S., Zhu, D. & Zhang, Q. Predicting hybrid performance in rice using genomic best linear unbiased prediction. Proc. Natl. Acad. Sci. USA 111, 12456–12461 (2014).

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

Xue, W., Guo, P. & Srivastava, H. K. Genetic component and combining ability analyses in relation to heterosis for yield and associated traits using three diverse rice-growing ecosystems. Field Crops Res. 88, 91–102 (2004).

Lippman, Z. B. & Zamir, D. Heterosis: revisiting the magic. Trends Genet. 23, 60–66 (2007).

Wu, H. W. & Matheson, A. C. Analyses of half-diallel mating designs with missing crosses: Theory and SAS program for testing and estimating GCA and SCA variance components. Silvae Genet. 50, 265–270 (2001).

Flint-Garcia, S. A. et al. Heterosis is prevalent for multiple traits in diverse maize germplasm. PLoS One 4, e7433 (2009).

Scharf, T. A. et al. Prediction of hybrid performance in maize using molecular markers and joint analyses of hybrids and parental inbreds. Theor. Appl. Genet. 120, 451–461 (2009).

Riedelsheimer, C. et al. Genomic and metabolic prediction of complex heterotic traits in hybrid maize. Nat. Genet. 44, 217–220 (2012b).

Betrán, E. J., Ribaut, J. M., Beck, D. & Gonzalez de León, D. Genetic diversity, specific combining ability, and heterosis in tropical maize under stress and nonstress environments. Crop Sci. 43, 797–806 (2003).
Program MAIZE. The authors would like to thank Dr. Yuan-Ming Zhang from Huazhong Agricultural University and two anonymous reviewers for their invaluable suggestions and comments, and Drs. Wenchao Zhang and Patrick X. Zhao from The Samuel Roberts Noble Foundation and Dr. Shizhong Xu from University of California, Riverside for their comments and assistance with PEPIS software.

**Author Contributions**
Conceived and designed the experiments: Y. Xu. Performed the experiments: H. Wang, X. Liu, C. Xie, Z. Guo and X. Xu. Analyzed the data: H. Wang, C. Xu and X. Liu. Contributed materials/analysis tools: Y. Xu, C. Xie, W. Li, C. Zou and S. Wang. Wrote the paper: Y. Xu and H. Wang.

**Additional Information**
**Competing financial interests:** The authors declare no competing financial interests.

**How to cite this article:** Wang, H. et al. Development of a multiple-hybrid population for genome-wide association studies: theoretical consideration and genetic mapping of flowering traits in maize. *Sci. Rep.* 7, 40239; doi: 10.1038/srep40239 (2017).

**Publisher’s note:** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

This work is licensed under a Creative Commons Attribution 4.0 International License. The images or other third party material in this article are included in the article’s Creative Commons license, unless indicated otherwise in the credit line; if the material is not included under the Creative Commons license, users will need to obtain permission from the license holder to reproduce the material. To view a copy of this license, visit http://creativecommons.org/licenses/by/4.0/

© The Author(s) 2016