RESEARCH ARTICLE

# QTL mapping for combining ability in different population-based NCII designs: a simulation study

LANZHI LI[1], CONGWEI SUN[1], YUAN CHEN[1], ZHIJUN DAI[1], ZHEN QU[2], XINGFEI ZHENG[2], SIBIN YU[3], TONGMIN MOU[3], CHENWU XU[4]* and ZHONGLI HU[2]*

[1]*Hunan Provincial Key Laboratory for Biology and Control of Plant Disease and Insect Pests, College of Bio-Safety Science and Technology, Hunan Agricultural University, 410128 Changsha, People's Republic of China*
[2]*State Key Laboratory of Hybrid Rice, College of Life Science, Wuhan University, 430072 Wuhan, People's Republic of China*
[3]*National Key Laboratory of Crop Genetic Improvement, Huazhong Agricultural University, 430070 Wuhan, People's Republic of China*
[4]*Division of Biostatistics and Quantitative Genetics, College of Agricultural Science, Yangzhou University, 225009 Yangzhou, People's Republic of China*

## Abstract

The NCII design (North Carolina mating design II) has been widely applied in studies of combining ability and heterosis. The objective of our research was to estimate how different base populations, sample sizes, testcross numbers and heritability influence QTL analyses of combining ability and heterosis. A series of Monte Carlo simulation experiments with QTL mapping were then conducted for the base population performance, testcross population phenotypic values and the general combining ability (GCA), specific combining ability (SCA) and Hmp (midparental heterosis) datasets. The results indicated that: (i) increasing the number of testers did not necessarily enhance the QTL detection power for GCA, but it was significantly related to the QTL effect. (ii) The QTLs identified in the base population may be different from those from GCA dataset. Similar phenomena can be seen from QTL detected in SCA and Hmp datasets. (iii) The QTL detection power for GCA ranked in the order of DH(RIL) based > $F_2$ based > BC based NCII design, when the heritability was low. The recombinant inbred lines (RILs) (or DHs) allows more recombination and offers higher mapping resolution than other populations. Further, their testcross progeny can be repeatedly generated and phenotyped. Thus, RIL based (or DH based) NCII design was highly recommend for combining ability QTL analysis. Our results expect to facilitate selecting elite parental lines with high combining ability and for geneticists to research the genetic basis of combining ability.

[Li L., Sun C., Chen Y., Dai Z., Qu Z., Zheng X., Yu S., Mou T., Xu C. and Hu Z. 2013 QTL mapping for combining ability in different population-based NCII designs by a simulation study. *J. Genet.* **92**, 529–543]

## Introduction

Breeding strategies for developing hybrids with high-yield potential and better grain quality require an expected level of heterosis and combining ability (Griffing 1956; Basbag *et al.* 2007; Riedelsheimer *et al.* 2012). To enhance the efficiency of hybrid breeding, many attempts have been made to select suitable parents and combinations (Verhoeven *et al.* 2006; Shukla and Pandey 2008; Riedelsheimer *et al.* 2012). An important approach for this purpose is combining ability analysis, which facilitate the selection of desirable parents and crosses for the exploitation of heterosis.

Combining ability has conventionally been estimated using partial–diallel designs (e.g., NCII design, factorial (FCT) designs, or circulant designs) with variance component analysis (Verhoeven *et al.* 2006). This approach is less labour intensive and can still yield considerable power for combining ability analysis (Kempthorne and Curnow 1961; Dhillon and Singh 1978). Under the NCII design, scientists have used the terms general combining ability (GCA) and specific combining ability (SCA) to designate the average performance of parent and hybrid combinations. However, whereas classic quantitative genetic studies attend to the

*For correspondence. E-mail: Chenwu Xu, qtls@yzu.edu.cn; Zhongli Hu, huzhongli@whu.edu.cn.

**Keywords.** combining ability; NCII design; QTL analysis; digenic; multiple alleles.

collective effects of all polygenes for a given trait, they cannot identify specific genomic regions that can be selected and introgressed (Verhoeven *et al.* 2006).

With the advent and development of molecular markers, it is now possible to dissect complex polygenic systems into individual Mendelian factors. Substantial progress has been made in quantitative trait loci (QTL) analysis based on a variety of genetic designs, including triple testcross (TTC) and North Carolina design III (NCIII) (Kusterer *et al.* 2007; Melchinger *et al.* 2007; Garcia *et al.* 2008; Li *et al.* 2008; Reif *et al.* 2009; He and Zhang 2011; He *et al.* 2012). These classic genetic designs focus on researching the genetic basis of heterosis. Only a few studies have used the NCII design for QTL analysis of combining ability (Qi *et al.* 2012; Qu *et al.* 2012). Combining the benefits of the NCII design with QTL mapping, the generalizable genetic architecture of complex traits could be estimated by QTLs detected by different allelic combinations among different parents. This approach focusses on the power of estimating QTL variance and gene action by traditional quantitative genetic approaches.

In our previous study (Qu *et al.* 2012), we developed a QTL-mapping method for dissecting combining ability and heterosis of agronomic traits using a backcross recombinant inbred lines (BCRIL)-based NCII design. Our results revealed that several QTLs associated with the combining ability of agronomic traits were similar to the QTLs associated with agronomic traits related to performance that have been detected in BCRIL (Qu *et al.* 2012). Qi *et al.* (2012) reported that the genetic basis of GCA of the traits is different from that of the yield-related traits per se.

To further uncover the genetic basis of combining ability and heterosis of quantitative traits using NCII design, two issues related to the detection of QTLs need to be addressed. First, it is unclear whether our method of mapping QTLs for combining ability could also be applied to other population-based NCII designs. Second, what are the factors that influence the detection efficiency of QTL mapping for combining ability (Qu *et al.* 2012). The objective of the present study was to estimate how different base populations, sample sizes, heritability and gene frequency influence QTL analyses of combining ability and heterosis. A series of Monte Carlo simulation (Fishman 1995; Berg 2004) experiments was carried out to confirm the proposed approach for BC, $F_2$, DH and RIL base populations.

## Methods

### *Genetic models for mapping QTLs in the NCII design with various base populations*

Under the NCII design, base populations ($F_2$, BC, DH or RIL) were derived from two inbred lines ($P_1$ and $P_2$) that differed significantly in the quantitative traits of interest and possessed abundant polymorphic molecular markers. A random sample of $n$ individuals from each base population was crossed with $k$ testers to produce $kn$ testcross families. All of the $kn$ families were planted with $x$ replicates. Molecular marker information was obtained from all the $n$ individuals from the base populations, whereas quantitative traits were measured for all the $knx$ testcross progeny. Phenotypic observations were denoted by $y_{tij}$, where $t\sim1, 2,\ldots k$, $i\sim1, 2,\ldots n$ and $j\sim1, 2,\ldots x$. Two models were considered: the additive-dominance model with two alleles or multiple alleles at each locus. In the absence of multiple alleles at QTL loci, $k = 2$, and the QTL genotype in the tester is a mixture of $QQ$ and $qq$, with the genotype frequency of $QQ$ and $qq$ taken as $q$ and $p$, respectively. In cases of multiple alleles at QTL loci in a NCII mating design, $k \geq 4$.

Assume that a quantitative trait is controlled by the QTL Q, with two alleles ($Q$ and $q$). This QTL is located near molecular marker M (with two alleles $M$ and $m$) on the same chromosome. The recombination fraction between molecular marker M and QTL Q is $r$. The genotypes of the two inbred lines ($P_1$ and $P_2$) and their $F_1$ progeny are $MMQQ$, $mmqq$ and $MmQq$, respectively. $F_1$ individuals produce four types of gametes, $MQ$, $mq$ $Mq$, and $mQ$ with frequencies $\frac{1}{2}(1-r)$, for the first two and $\frac{1}{2}r$ for the remaining.

Under the NCII design, the GCA and SCA effects derived from the base population (BC, $F_2$, DH or RIL) and its related TC populations could be theoretically applied to the QTL analysis. The formulae for this theory are deduced in tables 1, 2 and 4 for cases with two alleles at each locus and in tables 3 and 5 in cases with multiple alleles at each locus in an additive-dominance model for BC and $F_2$ based NCII design. The genetic effect symbols used in this study correspond to those established in our previous study (Qu *et al.* 2012).

### *BC-based NCII design*

The genotypes and genotype effects of markers and QTLs for combining ability and heterosis in the BC-based NCII design are listed in tables 1–3. According to the expected genetic values of the GCA and SCA effects, subpopulation subtraction of GCA (or SCA) effect between the two genotypes of molecular marker M ($MM$ and $Mm$) is equal to a fixed value multiplied by a coefficient $(1-2r)$ (tables 1 and 2). The Hmp (midparental heterosis) effect is equal to zero when the base population is $BC_1$ ($F_1 \times P_1$ $MMQQ$) and the genotype of the QTL in the testcross is $QQ$ (table 1). The Hmp effect is also zero when the base population is $BC_2$ ($F_1 \times P_1$ $mmqq$) and the genotype of the QTL in the testcross is $qq$ (table 2). In other cases, Hmp is equal to the dominance effect multiplied by the coefficient $(1-2r)$. When there are multiple alleles at each locus, subtraction of the mean Hmp, GCA or SCA value of two subpopulations is also equal to a fixed value multiplied by a coefficient $(1-2r)$ (table 3).

### *DH and RIL-based NCII designs*

The genotypes and genotype effects of markers and QTLs for combining ability and heterosis in the DH (RIL)-based NCII design used here are listed in previous study (Qu *et al.* 2012).

**Table 1.** The genotype and genotype effect of molecular marker (*M*) and QTL for combining ability and heterosis with two alleles at each locus in backcross population ($F_1 \times P_1$ *MM*QQ).

| $F_1$ gamete | *MQ* | *Mq* | *mQ* | *mq* |
|---|---|---|---|---|
| Genotype in BC plant | *MMQQ* | *MMQq* | *MmQQ* | *MmQq* |
| Frequency | $\frac{1}{2}(1-r)$ | $\frac{1}{2}r$ | $\frac{1}{2}r$ | $\frac{1}{2}(1-r)$ |
| Genotype effect in BC plant | $m+a$ | $m+d$ | $m+a$ | $m+d$ |
| Genotype in test cross[a] | *QQ* | *QQ* : *Qq* (1:1) | *QQ* | *QQ* : *Qq* (1:1) |
| Genotype effect in test cross[a] | $m+a$ | $m+\frac{1}{2}(a+d)$ | $m+a$ | $m+\frac{1}{2}(a+d)$ |
| Genotype effect in Hmp[a] | 0 | 0 | 0 | 0 |
| Genotype effect in Sca[a] | $-qa-\left(1+\frac{1}{2}q\right)d$ | $qa-\frac{1}{2}qd$ | $-qa-\left(1+\frac{1}{2}q\right)d$ | $qa-\frac{1}{2}qd$ |
| Mean value of subpopulation of Sca[a] | $\overline{MM}_{Sca}=(2r-1)qa+\left(-1-\frac{1}{2}q+r\right)d$ | | $\overline{Mm}_{Sca}=(1-2r)qa+\left(-\frac{1}{2}q-r\right)d$ | |
| Subpopulation subtraction of Sca[a] | $\overline{MM}_{Sca}-\overline{Mm}_{Sca}=(1-2r)(-2qa-d)$ | | | |
| Genotype in test cross[b] | *Qq* | *QQ* : *Qq* (1:1) | *Qq* | *QQ* : *Qq* (1:1) |
| Genotype effect in test cross[b] | $m+d$ | $m+\frac{1}{2}(d-a)$ | $m+d$ | $m+\frac{1}{2}(d-a)$ |
| Genotype effect in Hmp[b] | $d$ | 0 | $d$ | 0 |
| Mean value of subpopulation of Hmp[b] | $\overline{MM}_{Hmp}=(1-r)d$ | | $\overline{Mm}_{Hmp}=rd$ | |
| Subpopulation subtraction of Hmp[b] | $\overline{MM}_{Hmp}-\overline{Mm}_{Hmp}=(1-2r)d$ | | | |
| Genotype effect in Sca[b] | $pa-\left(1+\frac{1}{2}p\right)d$ | $-pa-\frac{1}{2}pd$ | $pa-\left(1+\frac{1}{2}p\right)d$ | $-pa-\frac{1}{2}pd$ |
| Mean value of subpopulation of Sca[b] | $\overline{MM}_{Sca}=(1-2r)pa+\left(-1-\frac{1}{2}p+r\right)d$ | | $\overline{Mm}_{Sca}=(2r-1)pa+\left(-\frac{1}{2}p-r\right)d$ | |
| Subpopulation subtraction of Sca[b] | $\overline{MM}_{Sca}-\overline{Mm}_{Sca}=(1-2r)(2pa-d)$ | | | |
| Mean value of subpopulation of Gca | $\overline{MM}_{Gca}=\left(q-\frac{1}{2}\right)a+\frac{1}{2}pd-\left[\left(\frac{1}{2}-2p\right)a+\frac{1}{2}d\right]r$ | | $\overline{Mm}_{Gca}=pa-\frac{1}{2}qd+\left[\left(\frac{1}{2}-2p\right)a+\frac{1}{2}d\right]r$ | |
| Subpopulation subtraction of GCA | $\overline{MM}_{Gca}-\overline{Mm}_{Gca}=(1-2r)\left[\left(\frac{1}{2}-2p\right)a+\frac{1}{2}d\right]$ | | | |

*MM* and *mm* denote the two genotype of molecular marker M; *QQ*, *Qq* and *qq* denote the three genotype of QTL; *r*, recombinant probability between molecular marker M and QTL; *m*, overall mean value; *a* and *d*, additive and dominant effects, respectively; *p* and *q*, genotype frequency of QTL *QQ* and *qq* in tester, respectively ($p+q=1$).
[a]When the genotype of QTL in tester is *QQ*, and its genotype frequency is *q*.
[b]When the genotype of QTL in tester is *qq*, and its genotype frequency is *p*.

**Table 2.** The genotype and genotype effect of molecular marker (*M*) and QTL for combining ability and heterosis with two alleles at each locus in backcross population ($F_1 \times P_1$ mmqq).

| $F_1$ gamete | *MQ* | *Mq* | *mQ* | *mq* |
|---|---|---|---|---|
| Genotype in BC plant | *MmQq* | *Mmqq* | *mmQq* | *mmqq* |
| Frequency | $\frac{1}{2}(1-r)$ | $\frac{1}{2}r$ | $\frac{1}{2}r$ | $\frac{1}{2}(1-r)$ |
| Genotype effect in BC plant | $m+d$ | $m-a$ | $m+d$ | $m-a$ |
| Genotype in test cross[a] | *QQ* : *Qq* (1:1) | *Qq* | *QQ* : *Qq* (1:1) | *Qq* |
| Genotype effect in test cross[a] | $m+\frac{1}{2}(a+d)$ | $m+d$ | $m+\frac{1}{2}(a+d)$ | $m+d$ |
| Genotype effect in Hmp[a] | 0 | $d$ | 0 | $d$ |
| Mean value of subpopulation of Hmp[b] | $\overline{Mm}_{Hmp}=rd$ | | $\overline{mm}_{Hmp}=(1-r)d$ | |
| Subpopulation subtraction of Hmp[a] | $\overline{Mm}_{Hmp}-\overline{mm}_{Hmp}=-(1-2r)d$ | | | |
| Genotype effect in SCA[a] | $-\frac{1}{2}a-\frac{1}{2}pd$ | $\frac{3}{2}qd$ | $-\frac{1}{2}a-\frac{1}{2}pd$ | $\frac{3}{2}qd$ |
| Mean value of subpopulation of SCA[a] | $\overline{Mm}_{Sca}=(r-1)a+\left(-\frac{1}{2}p+\frac{3}{2}r-pr\right)d$ | | $\overline{mm}_{Sca}=\frac{1}{2}ra-\frac{1}{2}(r+2qr-3q)d$ | |
| Subpopulation subtraction of SCA[a] | $\overline{Mm}_{Sca}-\overline{mm}_{Sca}=(1-2r)\left(-\frac{1}{2}a-\frac{3}{2}d-pd\right)$ | | | |
| Genotype in test cross[b] | *Qq* : *qq* (1:1) | *qq* | *Qq* : *qq* (1:1) | *qq* |
| Genotype effect in test cross[b] | $m+\frac{1}{2}(d-a)$ | $m-a$ | $m+\frac{1}{2}(d-a)$ | $m-a$ |
| Genotype effect in Hmp[b] | 0 | 0 | 0 | 0 |
| Genotype effect in SCA[b] | $-\frac{1}{2}a+\frac{1}{2}pd$ | $-\frac{1}{2}pd$ | $-\frac{1}{2}a+\frac{1}{2}pd$ | $-\frac{1}{2}pd$ |
| Mean value of subpopulation of SCA[b] | $\overline{Mm}_{Sca}=\frac{1}{2}[(1-r)a-(1-2r)pd]$ | | $\overline{Mm}_{Sca}=-\frac{1}{2}[ra+(1-2r)pd]$ | |
| Subpopulation subtraction of SCA[b] | $\overline{Mm}_{Sca}-\overline{mm}_{Sca}=(1-2r)\left(-\frac{1}{2}a+pd\right)$ | | | |
| $(Gca)_j$ | $\frac{3}{4}a+\left(\frac{1}{4}-\frac{1}{2}p\right)d$ | $-\frac{1}{4}a+\left(\frac{1}{2}p-\frac{1}{4}\right)d$ | $\frac{3}{4}a+\left(\frac{1}{4}-\frac{1}{2}p\right)d$ | $-\frac{1}{4}a+\left(\frac{1}{2}p-\frac{1}{4}\right)d$ |
| Mean value of subpopulation of GCA | $\overline{Mm}_{Gca}=\left(\frac{3}{4}-r\right)a+\left[\frac{1}{4}-2p-\left(\frac{1}{2}-p\right)r\right]d$ | | $\overline{Mm}_{Gca}=\left(r-\frac{1}{4}\right)a+\left[\left(p-\frac{1}{2}\right)r+\frac{1}{2}p-\frac{1}{4}\right]d$ | |
| Subpopulation subtraction of GCA | $\overline{Mm}_{Gca}-\overline{mm}_{Gca}=(1-2r)\left[a+\left(\frac{1}{2}-p\right)d\right]$ | | | |

*MM* and *mm*, two genotype of molecular marker M; *QQ*, *Qq* and *qq*, three genotype of QTL; *r*, recombinant probability between molecular marker M and QTL; *m*, overall mean value; *a* and *d*, additive and dominant effect, respectively; *p* and *q*, genotype frequency of QTL *QQ* and *qq* in tester, respectively ($p+q=1$). $(Gca)_j$ ($j=1\sim4$) denotes the general combining ability of the *j*th genotype of BC plant.

[a]When the genotype of QTL in tester is *QQ*, and its genotype frequency is *q*.

[b]When the genotype of QTL in tester is *qq*, and its genotype frequency is *p*.

**Table 3.** The genotype and genotype effect of molecular marker ($M$) and QTL for combining ability and heterosis with multiple alleles at each locus in backcross population.

| $F_1$ gamete | $MQ$ | $Mq$ | $mQ$ | $mq$ |
|---|---|---|---|---|
| Genotype in BC plant | $MMQQ$ | $MMQq$ | $MmQQ$ | $MmQq$ |
| Frequency | $\frac{1}{2}(1-r)$ | $\frac{1}{2}r$ | $\frac{1}{2}r$ | $\frac{1}{2}(1-r)$ |
| Genotype effect in BC plant | $m+a$ | $M+d$ | $m+a$ | $m+d$ |
| Genotype in test cross | $Q_iQ$ | $Q_iQ : Q_iq$ (1:1) | $Q_iQ$ | $Q_iQ : Q_iq$ (1:1) |
| Genotype effect in test cross | $m+g_i$ | $m+\frac{1}{2}(g_i+g_i')$ | $m+g_i$ | $m+\frac{1}{2}(g_i+g_i')$ |
| Genotype effect in Hmp | $g_i-\frac{1}{2}(a+a_i)$ | $\frac{1}{2}(g_i+g_i'-d+a_i)$ | $g_i-\frac{1}{2}(a+a_i)$ | $\frac{1}{2}(g_i+g_i'-d+a_i)$ |
| Mean value of subpopulation of Hmp | $\overline{MM}_{Hmp}=g_i-\frac{1}{2}(a+a_i)-\left[\frac{1}{2}(g_i-g_i'+d-a_i)-a\right]r$ | | $\overline{Mm}_{Hmp}=\frac{1}{2}(g_i+g_i'-d+a)+\left[\frac{1}{2}(g_i-g_i'-a_i+d)-a\right]r$ | |
| Subpopulation subtraction of Hmp | $\overline{MM}_{Hmp}-\overline{Mm}_{Hmp}=(1-2r)\left[\frac{1}{2}(g_i-g_i'-a+d)-a_i\right]$ | | | |
| Genotype effect in SCA | $\frac{1}{2}(G+G'-g_i-g_i')$ | $G-g_i$ | $\frac{1}{2}(G+G'-g_i-g_i')$ | $G-g_i$ |
| Mean value of subpopulation of SCA | $\overline{MM}_{Sca}=\frac{1}{2}(G-G'-g_i+g_i')+\frac{1}{2}r(G-G'-g_i+g_i')$ | | $\overline{Mm}_{Sca}=G-g_i-\frac{1}{2}r(G-G'-g_i+g_i')$ | |
| Subpopulation subtraction of SCA | $\overline{MM}_{Sca}-\overline{Mm}_{Sca}=-\frac{1}{2}(1-2r)(G-G'-g_i+g_i')$ | | | |
| $(Gca)_j$ | $-\frac{1}{2}(G+G')$ | $-G$ | $-\frac{1}{2}(G+G')$ | $-G$ |
| Mean value of subpopulation of GCA | $\overline{MM}_{Gca}=-\frac{1}{2}(G+G')+\frac{1}{2}(G'-G)r$ | | $\overline{Mm}_{Gca}=-G-\frac{1}{2}(3G+G')r$ | |
| Subpopulation subtraction of GCA | $\overline{MM}_{Gca}-\overline{Mm}_{Gca}=\frac{1}{2}(1-2r)(G-G')$ | | | |

$MM$ and $mm$, two different genotype of molecular marker M; $Q$ and $q$, two alleles of QTL; $Q_i (i=1\sim k)$, multiple alleles of QTL in tester; $r$, recombinant value between molecular marker M and QTL; $m$, overall mean value; $a$, additive effect in BC plant; $a_i (i=1\sim k)$, additive effect in $i$th tester; $g_i$ and $g_i'$ $(i=1\sim k)$, genotypic value of the homozygote and heterozygote of QTL, respectively. $G=\frac{1}{k}\sum_{i=1}^{k}g_i$, $G'=\frac{1}{k}\sum_{i=1}^{k}g_i'$. $(Gca)_i$ $(i=1\sim k)$, general combining ability of the $i$th tester; $(Gca)_j$ $(j=1\sim 4)$, general combining ability of the $j$th genotype of BC plant. The genotype of QTL in tester is $Q_iQ_i$, and its genotype frequency is $1/k$

**Table 4.** The genotype and genotype effect of molecular marker (*M*) and QTL for combining ability and heterosis with two alleles at each locus in $F_2$ population.

| Genotype of $F_2$ plant | *MMQQ* | *MMQq* | *MMqq* | *MmQQ* | *MmQq* | *Mmqq* | *mmQQ* | *mmQq* | *mmqq* |
|---|---|---|---|---|---|---|---|---|---|
| Frequency in $F_2$ plant | $\frac{1}{4}(1-r)^2$ | $\frac{1}{2}r(1-r)$ | $\frac{1}{4}r^2$ | $\frac{1}{2}r(1-r)$ | $\frac{1}{2}(1-2r+2r^2)$ | $\frac{1}{2}r(1-r)$ | $\frac{1}{4}r^2$ | $\frac{1}{2}r(1-r)$ | $\frac{1}{4}(1-r)^2$ |
| Genotype effect in $F_2$ | $m+a$ | $m+d$ | $m-a$ | $m+a$ | $m+d$ | $m-a$ | $m+a$ | $m+d$ | $m-a$ |
| Genotype of test cross[a] | *QQ* | *QQ* : *Qq*(1:1) | *Qq* | *QQ* | *QQ* : *Qq*(1:1) | *Qq* | *QQ* | *QQ* : *Qq*(1:1) | *Qq* |
| Genotype effect in test cross[a] | $m+a$ | $m+\frac{1}{2}(a+d)$ | $m+d$ | $m+a$ | $m+\frac{1}{2}(a+d)$ | $m+d$ | $m+a$ | $m+\frac{1}{2}(a+d)$ | $m+d$ |
| Genotype effect in *Hmp*[a] | 0 | 0 | $d$ | 0 | 0 | $d$ | 0 | 0 | $d$ |
| Mean value of subpopulation of Hmp[a] | | $\overline{MM}_{Hmp}=r^2d$ | | | $\overline{Mm}_{Hmp}=r(1-r)d$ | | | $\overline{mm}_{Hmp}=(1-r)^2d$ | |
| Subpopulation subtraction of Hmp[a] | | $\overline{\mathrm{MM}}_{Hmp}-\overline{\mathrm{Mm}}_{Hmp}=-(1-2r)rd$ | | | $\overline{\mathrm{Mm}}_{Hmp}-\overline{\mathrm{mm}}_{Hmp}=(1-2)(r-1)d$ | | | $\overline{\mathrm{MM}}_{Hmp}-\overline{\mathrm{mm}}_{Hmp}=-(1-2r)d$ | |
| Genotype effect in *Sca*[a] | $-qd$ | 0 | $qd$ | $-qd$ | 0 | $qd$ | $-qd$ | 0 | $qd$ |
| Mean value of subpopulation of Sca[a] | | $\overline{MM}_{Sca}=-(1-2r)qd$ | | | $\overline{Mm}_{Sca}=0$ | | | $\overline{mm}_{Sca}=(1-2r)qd$ | |
| Subpopulation subtraction of Sca[a] | | $\overline{MM}_{Sca}-\overline{Mm}_{Sca}=-(1-2r)qd$ | | | $\overline{Mm}_{Sca}-\overline{mm}_{Sca}=-(1-2r)qd$ | | | $\overline{MM}_{Sca}-\overline{mm}_{Sca}=2(1-2r)qd$ | |
| Genotype of test cross[b] | *Qq* | *Qq* : *qq* (1:1) | *qq* | *Qq* | *Qq* : *qq* (1:1) | *qq* | *Qq* | *Qq* : *qq* (1:1) | *qq* |
| Genotype effect in test cross[b] | $M+d$ | $m+\frac{1}{2}(d-a)$ | $m-a$ | $m+d$ | $m+\frac{1}{2}(d-a)$ | $m-a$ | $m+d$ | $m+\frac{1}{2}(d-a)$ | $m-a$ |
| Genotype effect in *Hmp*[b] | $d$ | 0 | 0 | $d$ | 0 | 0 | $d$ | 0 | 0 |
| Mean value of subpopulation of Hmp[b] | | $\overline{MM}_{Hmp}=(1-r)^2d$ | | | $\overline{Mm}_{Hmp}=r(1-r)d$ | | | $\overline{mm}_{Hmp}=r^2d$ | |
| Subpopulation subtraction of Hmp[b] | | $\overline{\mathrm{MM}}_{Hmp}-\overline{\mathrm{Mm}}_{Hmp}=(1-2r)(1-r)d$ | | | $\overline{\mathrm{Mm}}_{Hmp}-\overline{\mathrm{mm}}_{Hmp}=(1-2r)rd$ | | | $\overline{\mathrm{MM}}_{Hmp}-\overline{\mathrm{mm}}_{Hmp}=(1-2r)d$ | |
| Genotype effect in *Sca*[b] | $pd$ | 0 | $-pd$ | $pd$ | 0 | $-pd$ | $pd$ | 0 | -*pd* |
| Mean value of subpopulation of SCA[b] | | $\overline{MM}_{Sca}=(1-2r)pd$ | | | $\overline{Mm}_{Sca}=0$ | | | $\overline{mm}_{Sca}=-(1-2r)pd$ | |
| Subpopulation subtraction of SCA[b] | | $\overline{MM}_{Sca}-\overline{Mm}_{Sca}=(1-2r)qd$ | | | $\overline{Mm}_{Sca}-\overline{mm}_{Sca}=(1-2r)qd$ | | | $\overline{MM}_{Sca}-\overline{mm}_{Sca}=2(1-2r)qd$ | |

**Table 4** (*contd*)

| Genotype of $F_2$ plant | *MMQQ* | *MMQq* | *MMqq* | *MmQQ* | *MmQq* | *Mmqq* | *mmQQ* | *mmQq* | *mmqq* |
|---|---|---|---|---|---|---|---|---|---|
| $(Gca)_j$ | $\frac{1}{2}a+\left(q-\frac{1}{2}\right)d$ | 0 | $-\frac{1}{2}a+\left(p-\frac{1}{2}\right)d$ | $\frac{1}{2}a+\left(q-\frac{1}{2}\right)d$ | 0 | $-\frac{1}{2}a+\left(p-\frac{1}{2}\right)d$ | $\frac{1}{2}a+\left(q-\frac{1}{2}\right)d$ | 0 | $-\frac{1}{2}a+\left(p-\frac{1}{2}\right)d$ |
| Mean value of subpopulation of GCA | $\overline{MM}_{Gca}=(1-2r)\left[\frac{1}{2}a+\left(q-\frac{1}{2}\right)d\right]$ | | | $\overline{Mm}_{Gca}=0$ | | | $\overline{mm}Gca=-(1-2r)\left[\frac{1}{2}a+\left(q-\frac{1}{2}\right)d\right]$ | | |
| Subpopulation subtraction of GCA | $\overline{MM}_{Gca}-\overline{Mm}_{Gca}=(1-2r)\left[\frac{1}{2}a+\left(q-\frac{1}{2}\right)d\right]$ | | | $\overline{Mm}_{Gca}-\overline{mm}_{Gca}=(1-2)\left[\frac{1}{2}a+\left(q-\frac{1}{2}\right)d\right]$ | | | $\overline{MM}_{Gca}-\overline{mm}_{Gca}=(1-2r)\,[a+(q-p)d]$ | | |

*MM* and *mm*, two genotype of molecular marker M; *QQ*, *Qq* and *qq*, three genotype of QTL; *r*, recombinant probability between molecular marker M and QTL; *m*, overall mean value; *a* and *d*, additive effect and dominant effect, respectively; *p* and *q*, genotype frequency of QTL *QQ* and *qq* in tester, respectively ($p+q=1$). $(Gca)_j$ ($j=1\sim9$), general combining ability of the *j*th genotype of $F_2$ plant.

[a]When the genotype of QTL in tester is *QQ*, and its genotype frequency is *q*.

[b]When the genotype of QTL in tester is *qq*, and its genotype frequency is *p*.

$\overline{Mm}_{Sca}-\left(\overline{MM}_{Sca}+\overline{mm}_{Sca}\right)/2=0.$

$\overline{Mm}_{Gca}-\left(\overline{MM}_{Gca}+\overline{mm}_{Gca}\right)/2=0.$

### *$F_2$-based NCII design*

The genotypes and genotype effects of markers and QTLs for combining ability and heterosis in the $F_2$ population are listed in tables 4 and 5. Subtraction of the mean Hmp, GCA or SCA effects from any two subpopulations is equal to a fixed value multiplied by a coefficient $(1 - 2r)$.

### *Simulation experiment I*

The purpose of the simulation experiment was to evaluate the statistical power of QTL mapping for combining ability in different population-based NCII designs. The simulated genome consisted of four chromosomes (chromosomes 1–4), with 25 evenly spaced markers covering each chromosome at an average marker interval of 10.0 cM. We simulated three main-effect QTLs, each of which was located in the marker interval. The three main-effect QTLs were located at 192.0, 105.0 and 196.0 cM of chromosome 1, 2 and 4, respectively.

For the additive-dominance model with two alleles at each locus, the number of testcrosses (m) for each NCII family was set at 2. When the $F_2$ population was used as the base population, the genetic parameters were as follows: $\mu = 100.00$; $a_1 = 1.50$, $a_2 = 2.00$ and $a_3 = -1.00$ for $QTL_1$, $QTL_2$ and $QTL_3$, respectively; and $d_1 = 1.50$, $d_2 = -1.00$ and $d_3 = 2.00$ for $QTL_1$, $QTL_2$ and $QTL_3$, respectively. When the BC, DH or RIL population was set as the base population, the genetic parameters were as follows: $\mu = 100.00$; $a_1 = 1.20$, $a_2 = 1.50$ and $a_3 = -1.00$ for $QTL_1$, $QTL_2$ and $QTL_3$, respectively; and $d_1 = 1.50$, $d_2 = -1.00$ and $d_3 = 2.00$ for $QTL_1$, $QTL_2$ and $QTL_3$, respectively. When the QTL genotype of the tester is $QQ$, its genotype frequency is $q$. When the QTL genotype of the tester is $qq$, its genotype is $p$. In the simulations, $q = \frac{1}{2}$. Each treatment was replicated 100 times.

For the additive-dominance model with multiple alleles at each locus, the number of testcrosses (m) for each NCII family was set at 4. Certain genetic parameters, such as population mean values and additive and dominant effects of the three QTLs, were the same as those under the additive-dominance model with two alleles at each locus. The other genetic parameters were set as follows: for the first testcross population, $g_1 = -2.00$, $g_2 = 1.80$, and $g_3 = 3.00$ and $g'_1 = 1.00$, $g'_2 = -1.40$, and $g'_3 = 1.60$ for $QTL_1$, $QTL_2$ and $QTL_3$, respectively; for the second testcross population, $g_1 = 1.60$, $g_2 = 2.00$, and $g_3 = -1.00$ and $g'_1 = 1.40$, $g'_2 = -2.00$, and $g'_3 = 1.60$ for $QTL_1$, $QTL_2$ and $QTL_3$, respectively; for the third testcross population, $g_1 = 1.00$, $g_2 = 1.50$, and $g_3 = 2.00$ and $g'_1 = -1.20$, $g'_2 = 1.00$, and $g'_3 = -3.00$ for $QTL_1$, $QTL_2$ and $QTL_3$, respectively; for the fourth testcross population, $g_1 = -0.50$, $g_2 = -1.00$, and $g_3 = -2.00$ and $g'_1 = 1.00$, $g'_2 = 1.50$, and $g'_3 = 1.80$ for $QTL_1$, $QTL_2$ and $QTL_3$, respectively. $g_i$ and $g'_i$ $(i = 1 \sim k)$ denote the genotypic value of the homozygote and heterozygote of QTL, respectively. The broad heritability $(h_B^2)$ of each model was set at three levels: 20%, 50% and 80%, which represent low, middle and high heritability. The sample size $(n)$ or the number of individuals in the base population (BC, $F_2$, DH or RIL) was set at two levels: 200 and 400. Molecular marker information for TC populations was derived from the corresponding base population in the NCII design. Each treatment was replicated 100 times.

### *Simulation experiment II*

In order to investigate how the different gene frequency in testers influence QTL mapping for combining ability and heterosis. Under the additive-dominance model with two alleles at each locus, $q = \frac{1}{2}, \frac{1}{3}, \frac{1}{4}, \frac{1}{5}, \frac{2}{5}, \frac{1}{6}, \frac{1}{8}, \frac{3}{8}$ (when $p = \frac{1}{2}, \frac{2}{3}, \frac{3}{4}, \frac{4}{5}, \frac{3}{5}, \frac{5}{6}, \frac{7}{8}, \frac{5}{8}$) for testers in RIL-based NCII design were set. Each treatment was replicated 30 times. With these gene frequencies of $p$ and $q$, all conditions of gene frequency were taken into accounted when the denominator of the gene frequency is 2 to 6 and 8. For example, when the denominator is 6, gene frequency of $\frac{1}{6}, \frac{2}{6}\left(\frac{1}{3}\right), \frac{3}{6}\left(\frac{1}{2}\right), \frac{4}{6}\left(\frac{2}{3}\right), \frac{5}{6}$ were set in our model.

### *QTL mapping*

The QTL analysis was performed separately for the phenotype values of the base populations and the GCA, SCA and Hmp datasets derived from each base population and its related TC populations. Hmp, SCA and GCA effects were estimated with the following equations: $(Hmp)_{ij} = y_{ij} - \left(y_i + y_j\right)/2$; $(Sca)_{ij} = y_{ij} - \bar{y}_{..} - (Gca)_i - (Gca)_j$; $(Gca)_i = \bar{y}_{i.} - \bar{y}_{..}$, where $(Hmp)_{ij}$, the midparental heterosis value of the TC hybrid from the parental lines' base population (BC, DH, RIL or $F_2$), line i and tester j; $y_{ij}$, the phenotypic value of the TC hybrid between the parental lines' base population (BC, DH, RIL or $F_2$), line i and tester j; $y_i$, the phenotypic value of base population line i; $y_j$, the phenotypic value of tester j; $(Sca)_{ij}$, SCA effect; $\bar{y}_{..}$, overall mean; $(Gca)_i$, GCA effect of base population line i; $(Gca)_j$, GCA effect of tester j; and $\bar{y}_i$, mean performance of the three hybrids between base population line $i$ and the $m$ testers.

Analyses of the main-effect QTL (M-QTL) were conducted for each dataset (including GCA, SCA, Hmp, base population and TC populations) by network interval mapping (NWIM) using QTL network 2.0 (Yang *et al.* 2008). A logarithm of the odds (LOD) score of 3.0 was selected as the threshold for the presence of a M-QTL, based on the total map distance and the average distance between markers. QTLs that were detected in different populations or for different traits were considered common if their estimated map positions were within a distance of 20 cM (Groh *et al.* 1998). This is a common approach used in comparative mapping (Frascaroli *et al.* 2007; Li *et al.* 2008, 2010). For each simulated QTL, the estimate for the QTL parameter was the average of the corresponding estimates from the counted samples. The number of QTLs detected in every 100 simulated datasets represents the detection power of the QTL.

**Table 5.** The genotype and genotype effect of molecular marker (*M*) and QTL for combining ability and heterosis with multiple alleles at each locus in the $F_2$ population.

| Genotype of $F_2$ plant | *MMQQ* | *MMQq* | *MMqq* | *MmQQ* | *MmQq* | *Mmqq* | *mmQQ* | *mmQq* | *mmqq* |
|---|---|---|---|---|---|---|---|---|---|
| Frequency in $F_2$ plant | $\frac{1}{4}(1-r)^2$ | $\frac{1}{2}r(1-r)$ | $\frac{1}{4}r^2$ | $\frac{1}{2}r(1-r)$ | $\frac{1}{2}\left(1-2r+2r^2\right)$ | $\frac{1}{2}r(1-r)$ | $\frac{1}{4}r^2$ | $\frac{1}{2}r(1-r)$ | $\frac{1}{4}(1-r)^2$ |
| Genotype effect in $F_2$ | $m+a$ | $m+d$ | $m-a$ | $m+a$ | $m+d$ | $m-a$ | $m+a$ | $m+d$ | $m-a$ |
| Genotype of testcross | $Q_iQ$ | $Q_iQ:Q_iq(1{:}1)$ | $Q_iq$ | $Q_iQ$ | $Q_iQ:Q_iq(1{:}1)$ | $Q_iq$ | $Q_iQ$ | $Q_iQ:Q_iq(1{:}1)$ | $Q_iq$ |
| Genotype effect in testcross | $m+g_i$ | $m+\frac{1}{2}(g_i+g_i')$ | $m+g_i'$ | $m+g_i$ | $m+\frac{1}{2}(g_i+g_i')$ | $m+g_i'$ | $m+g_i$ | $m+\frac{1}{2}(g_i+g_i')$ | $m+g_i'$ |
| Genotype effect in *Hmp* | $g_i-\frac{1}{2}(a+a_i)$ | $\frac{1}{2}(g_i+g_i'-d-a_i)$ | $g_i'-\frac{1}{2}(-a+a_i)$ | $g_i-\frac{1}{2}(a+a_i)$ | $\frac{1}{2}(g_i+g_i'-d-a_i)$ | $g_i'-\frac{1}{2}(-a+a_i)$ | $g_i-\frac{1}{2}(a+a_i)$ | $\frac{1}{2}(g_i+g_i'-d-a_i)$ | $g_i'-\frac{1}{2}(-a+a_i)$ |
| Mean value of subpopulation of Hmp | $\overline{MM}_{Hmp}=g_i-\frac{1}{2}(a+a')+r(g_i'-g_i+a-d)+r^2(g_i+g_i'-a')$ | | | $\overline{Mm}_{Hmp}=(r-r^2)d+\frac{1}{2}(g_i+g_i'-d-a')$ | | | $\overline{mm}_{Hmp}=g_i'-\frac{1}{2}(a'-a)+r(g_i+g_i'-d-a')+r^2(-2g_i-a+d+a')$ | | |
| Subpopulation subtraction of Hmp | $\overline{\mathrm{MM}}_{Hmp}-\overline{\mathrm{MM}}_{Hmp}=(1-2r)\times[(1-r)A+(2r-1)B-rC]$ | | | $\overline{\mathrm{MM}}_{Hmp}-\overline{\mathrm{mm}}_{Hmp}=(1-2r)(A-C)$ | | | $\overline{\mathrm{Mm}}_{Hmp}-\overline{\mathrm{mm}}_{Hmp}=(1-2r)\times[rA+(1-2r)B+(r-1)C]$ | | |
| Genotype effect in *Sca* | $\frac{1}{2}(g_i+g_i'-G+G')$ | 0 | $-\frac{1}{2}(g_i+g_{i\prime}-G+G')$ | $\frac{1}{2}(g_i+g_i'-G+G')$ | 0 | $-\frac{1}{2}(g_i+g_i'-G+G')$ | $\frac{1}{2}(g_i+g_i'-G+G')$ | 0 | $-\frac{1}{2}(g_i+g_i'-G+G')$ |
| Mean value of subpopulation of SCA | $\overline{MM}_{Sca}=\frac{1}{2}(1-2r)(g_i-g_i'-G+G')$ | | | $\overline{Mm}_{Sca}=0$ | | | $\overline{mm}_{Sca}=-\frac{1}{2}(1-2r)(g_i-g_i'-G+G')$ | | |
| Subpopulation subtraction of SCA | $\overline{MM}_{Sca}-\overline{Mm}_{Sca}=-\frac{1}{2}(1-2r)(g_i-g_i'-G+G')$ | | | $\overline{MM}_{Sca}-\overline{mm}_{Sca}=(1-2r)(g_i-g_i'-G+G')$ | | | $\overline{Mm}_{Sca}-\overline{mm}_{Sca}=\frac{1}{2}(1-2r)(g_i-g_i'-G+G')$ | | |

**Table 5** (*contd*)

| Genotype of $F_2$ plant | *MMQQ* | *MMQq* | *MMqq* | *MmQQ* | *MmQq* | *Mmqq* | *mmQQ* | *mmQq* | *mmqq* |
|---|---|---|---|---|---|---|---|---|---|
| $(Gca)_j$ | $\frac{1}{2}(G-G')$ | 0 | $\frac{1}{2}(G'-G)$ | $\frac{1}{2}(G-G')$ | 0 | $\frac{1}{2}(G'-G)$ | $\frac{1}{2}(G-G')$ | 0 | $\frac{1}{2}(G'-G)$ |
| Mean value of subpopulation of GCA | $\overline{MM}_{Gca}=\frac{1}{2}(1-2r)(G-G')$ | | | $\overline{Mm}_{Gca}=0$ | | | $\overline{mm}_{Gca}=-\frac{1}{2}(1-2r)(G-G')$ | | |
| Subpopulation subtraction of GCA | $\overline{MM}_{Gca}-\overline{Mm}_{Gca}=\frac{1}{2}(1-2r)(G-G')$ | | | $\overline{MM}_{Gca}-\overline{mm}_{Gca}=(1-2r)(G-G')$ | | | $\overline{Mm}_{Gca}-\overline{mm}_{Gca}=\frac{1}{2}(1-2r)(G-G')$ | | |

*MM* and *mm*, two different genotype of molecular marker M; *Q* and *q*, two alleles of QTL; $Q_i(i=1\sim k)$, multiple alleles of QTL in tester; *r*, recombinant value between molecular marker M and QTL; *m*, overall mean value; *a*, additive effect in $F_2$ plant; *d*, dorminance effect in $F_2$ plant; $a_i(i=1\sim k)$, additive effect in *i*th tester; $g_i(i=1\sim k)$, genotypic value of the homozygote in tester line i, and $g_i'$ $(i=1\sim k)$, genotypic value of the heterozygote of QTL which derived from $F_2$ plant × tester line *i*, respectively. $G=\frac{1}{k}\sum_{i=1}^{k}g_i$, $G'=\frac{1}{k}\sum_{i=1}^{k}g_i'$·$(Gca)_j$ $(j=1\sim 9)$, general combining ability of the *j*th genotype of $F_2$ plant. The genotype of QTL in tester is $Q_iQ_i$,, and its genotype frequency is $1/k$.

$A=g_i-\frac{1}{2}(a+a_i)$; $B=\frac{1}{2}(g_i+g_i'-d-a_i)$; $C=g_i'-\frac{1}{2}(-a+a_i)$.

$\overline{Mm}_{Hmp}-(\overline{MM}_{Hmp}+\overline{mm}_{Hmp})/2=-\frac{1}{2}(6r^2-6r+1)d-\frac{1}{2}(g_i+g_i'-a_i)$.

$\overline{Mm}_{Sca}-(\overline{MM}_{Sca}+\overline{mm}_{Sca})/2=0$.

$\overline{Mm}_{Gca}-(\overline{MM}_{Gca}+\overline{mm}_{Gca})/2=0$.

## Results and discussion

Tables 1–5 in electronic supplementary material summarize the QTL mapping results from eight datasets (base population, two TC populations, one GCA dataset, two SCA datasets and two Hmp datasets) under the additive-dominance model with two alleles ($q = 1/2$) at each locus in BC-based, $F_2$-based, DH-based, RIL (selfing)-based and RIL (sib-mating)-based NCII designs, respectively. Tables 6–10 in electronic material summarize the QTL mapping results from 14 datasets (base population, four TC populations, GCA dataset, four SCA datasets and four Hmp datasets) under the additive-dominance model with four alleles at each locus in different population based NCII designs, respectively.

The results showed that many QTLs were accurately detected in set intervals in chromosomes 1, 2 and 4 for every simulated situation (tables 1–10 in electronic supplementary material). In particular, the QTLs successfully detected in the GCA and SCA datasets further confirm that QTL mapping for combining ability is feasible under different population-based NCII designs. Quantitative genetic studies focussing on parental GCA and SCA for a particular complex trait were popular precisely because they allowed an estimation of the genetic architecture of the phenotype, with inference extending back to the reference population (Griffing 1956; Verhoeven *et al.* 2006).

Similar to He *et al.* (2001) and Su *et al.* (2010), our results indicate that increasing the broad heritability and sample size (base population sample size in the present study) can enhance the QTL detection power (tables 1–10 in electronic supplementary material). He *et al.* (2001) also demonstrated that the efficiency of the QTL detection power increased sharply when the sample size was raised from 100 to 400. Similar results were obtained in the present study (data not shown).

Increasing the number of testers did not necessarily enhance the QTL detection power for GCA, but it was significantly related to the QTL effect. For example, in the BC-based NCII design with two testcross populations, three set QTLs for GCA were detected at a near-significant level in every simulation (table 1 in electronic supplementary material). Conversely, with four testcross populations, the QTL set for GCA on chromosome 4 was hardly detectable in our simulations with a sample size of 200 (table 6 in electronic supplementary material). This result may be interpreted by supposing that the genetic effect of the QTL on GCA was $G-G'$ ($G = \frac{1}{k}\sum_{i=1}^{k} g_i$, $G' = \frac{1}{k}\sum_{i=1}^{k} g_i' \cdot g_i$ and $g_i' (i = 1 \sim k)$ denote the genotypic values of the homozygous and heterozygous QTLs, respectively) multiplied by a coefficient. As a consequence, if $G$ and $G'$ have the same magnitude, the QTL will not be identified in the GCA dataset, no matter how many ($k$) testers are included in the NCII design. Our results may, to some extent, reconcile previous disagreements in the literature regarding optimal QTL mapping designs. In particular, Rao and Li (2000) recommended the use of a small number of large families (two to five families, each with 250–1000 progeny), whereas Xu (1998) and Xie *et al.* (1998), both suggested a larger number of intermediate-sized families. Wu and Jannink (2004) found the highest QTL detection power for strategies employing relatively few parents (5–10). Verhoeven *et al.* (2006) demonstrated that the trade-off between the number of families and family size differed between experimental designs. Zhu and He (2001) pointed that the GCA effect of sweetcorn can be tested accurately by one selected tester line and at least three testers are needed when not selecting. Three distinct factors may contribute to this disagreement. First, different statistical methods were used. In Wu and Jannink (2004) study, the authors discarded potential information related to QTL effects that contribute to differences among families such that the higher the number of families, the more among-family information was lost. Second, different mating designs were used. Wu and Jannink (2004) used a circulant diallel design, whereas a half-diallel mating design without selfing was used by Verhoeven *et al.* (2006) and in the present work. The last and perhaps most important factor is the use of different statistical methods. Previous studies (Griffing 1956; Xu 1998; Zhu and He 2001; Wu and Jannink 2004) explored the variance components of the QTL and polygenes linked to a given trait. However, we detected the position and QTL effects of each QTL locus for the traits.

Our results showed that, for general combining ability, the genetic effect and position of gca-QTLs (QTLs detected in the GCA dataset) were greatly influenced by the testcross population although increasing the number of testers did not necessarily enhance the QTL detection power for GCA. gca-QTLs may indeed be different for different testers, even with the same base population (e.g., the gca-QTL detected in table 4 in electronic supplementary material is different from that in table 9 in electronic supplementary material). Therefore, the more testcross populations included in the NCII design, the more meaningful the gca-QTL detection results will be.

Although the number of testers has little impact on the detection of GCA and SCA QTLs, the QTL effect does influence QTL detection power. When the recombination rate $r$ is fixed, the absolute value of QTL effect becomes larger and hence, the QTL can be detected more easily. This relationship is well illustrated by the following four examples.

(i) With two alleles at each locus, when $p = q = \frac{1}{2}$, the QTL detection power in the GCA dataset is approximately equal to or lower than that calculated for the base population phenotype value dataset. For example, in the DH-based NCII design, when broad heritability was high (50% or 80% in the present study), the number of times the three QTLs were detected in both DH population and DH-GCA dataset was close to 100 in 100 simulations. However, when the sample size was relatively small (200 in the present study)

and broad heritability was low (20% in the present study), the three set QTLs were detected in the DH population just 76, 89 and 39 times, respectively, out of 100 simulations and only, 55, 76 and 21 times in the DH-GCA dataset (table 3 in electronic supplementary material). This finding is consistent with the derived formulae in table 2 in electronic supplementary material in our previous study (Qu *et al.* 2012). A subpopulation subtraction between the two genotypes of molecular marker M (*MM* and *mm*) is equal to:

$$\overline{MM}_{DH} - \overline{mm}_{DH} = 2\,(1-2r)\,a \qquad (1)$$

$$\overline{MM}_{Gca} - \overline{mm}_{Gca} = 2\,(1-2r)\left[\frac{1}{2}a + \left(q-\frac{1}{2}\right)d\right] \qquad (2)$$

In formula (2), $\left[\frac{1}{2}a + \left(q-\frac{1}{2}\right)d\right]$ is GCA effect. When $q=\frac{1}{2}$, the QTL effect in formula (1) is almost twice as strong as that in formula (2). Similar phenomena were observed in the BC, RIL and $F_2$-based NCII designs. However, it should be noted that when $q \neq \frac{1}{2}$ under the additive-dominance model, the gene actions of QTLs investigated in the GCA dataset would differ from the corresponding activities in the base population because of the dominance effect (d).

The QTL mapping results from datasets (base population, testcross population, GCA, SCA and Hmp datasets) with different gene frequency of testers in RIL-based NCII design are listed in tables 11–17 in electronic supplementary material. From figure 1, in RIL-based NCII design with low broad heritability (20%), we can see that, no matter how different the gene frequency in testers, the larger the QTL effect (absolute value) the higher the QTL detection power. When the QTL effect (absolute value) is large, the number of times the three QTLs were detected in GCA dataset was close to 30 in 30 simulations. This means that seldom false positive and false negative QTLs are detected when QTL effect (absolute value) is large. However, when the broad heritability is high (80%), the number of times the three QTLs were detected in GCA dataset was close to 30 in 30 simulations, no matter the QTL effect is large or small.

(ii) When the QTL effect is close to zero, the QTL is barely detected in the simulation. With multiple alleles at each locus, in BC-based NCII design, the subpopulation subtraction between the two genotypes of molecular marker M (*MM* and *Mm*) (table 3) is equal to:

$$\overline{MM}_{BC} - \overline{Mm}_{BC} = (1-2r)\,(a-d) \qquad (3)$$

$$\overline{MM}_{Gca} - \overline{Mm}_{Gca} = \frac{1}{2}\,(1-2r)\,(G-G') \qquad (4)$$

$G = \frac{1}{k}\sum_{i=1}^{k} g_i$, $G' = \frac{1}{k}\sum_{i=1}^{k} g_i' \cdot g_i$ and $g_i'$ $(i = 1 \sim k)$ denote the genotypic values of the homozygous and heterozygous QTLs, respectively. In our simulations, $(a-d)$ is equal to 0, 3 and −3, and $(G-G')$ is equal to −0.525, 1.30 and 0 for $QTL_1$, $QTL_2$ and $QTL_3$, respectively. Under these conditions, for the QTL effect is zero, the first QTL set on chromosome 1 detected in the BC population and the third QTL set on chromosome 4 in BC-GCA are hardly detected (table 1 in electronic supplementary material). In addition, the order of detection power of the three QTLs is $QTL_2 > QTL_1 > QTL_3$ in the BC-GCA data. The larger the absolute value of the QTL, the more easily the QTL can be detected, especially when broad heritability is low (tables 6–10 in electronic supplementary material).

(iii) In the $F_2$-based NCII design with two alleles at each locus, the proportion of the three QTLs detected in the Hmp dataset is less than that detected in the SCA dataset (table 7 in electronic supplementary material).

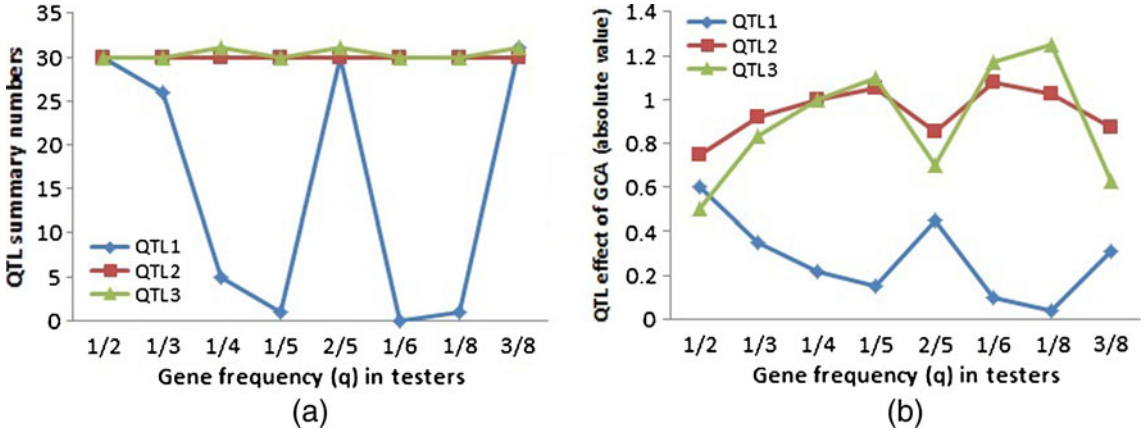


**Figure 1.** QTL mapping for GCA in RIL-based NCII design summary result when broad heritability is 20%.

For Hmp and SCA, subpopulation subtraction among three genotypes of molecular marker M (*MM*, *Mm* and *mm*) (table 4) is equal to:

$$\overline{\mathrm{MM}}_{Hmp} - \overline{\mathrm{Mm}}_{Hmp} = -(1-2r)\,rd \qquad (5)$$

$$\overline{\mathrm{MM}}_{Hmp} - \overline{\mathrm{mm}}_{Hmp} = (1-2r)\,(r-1)\,d \qquad (6)$$

$$\overline{\mathrm{MM}}_{Hmp} - \overline{\mathrm{mm}}_{Hmp} = -(1-2r)\,d \qquad (7)$$

$$\overline{\mathrm{MM}}_{Sca} - \overline{Mm}_{Sca} = -(1-2r)\,qd \qquad (8)$$

$$\overline{Mm}_{Sca} - \overline{mm}_{Sca} = -(1-2r)\,qd \qquad (9)$$

$$\overline{MM}_{Sca} - \overline{mm}_{Sca} = 2(1-2r)\,qd \qquad (10)$$

The recombination rate $r$ is set as 0.02, 0.05 and 0.06 for $QTL_1$, $QTL_2$ and $QTL_3$, respectively. Because $r$ is much lower than $q(q = 1/2)$, the QTL effect in the Hmp dataset is less than in the SCA dataset. Similar phenomena can be observed in the DH and RIL-based NCII designs. Interestingly, extremely few QTLs were detected in the HmpQQ dataset under the BC-based NCII design (table 6 in electronic supplementary material). This finding may be explained by the fact that when the $F_1$ progeny are backcrossed to the larger value parent $B_1$ (*MMQQ*), the subpopulation subtraction for HmpQQ between the two genotypes of molecular marker M (*MM* and *Mm*) is equal to 0 (table 1).

(iv) With multiple alleles at each locus, the proportion of the three QTLs detected in the Hmp dataset was not necessarily less than that found in the SCA dataset (tables 6–10). This result is consistent with the formulae for subpopulation subtractions deduced in tables 2 and 4. For example, in the BC-based NCII design, the following two formulae are not necessarily equal to or higher than one other.

$$\overline{MM}_{Hmp} - \overline{Mm}_{Hmp} = (1-2r)\left[\frac{1}{2}\left(g_i - g_i' - a + d\right) - a_i\right] \qquad (11)$$

$$\overline{MM}_{Sca} - \overline{Mm}_{Sca} = -\frac{1}{2}(1-2r)\left(G - G' - g_i + g_i'\right) \qquad (12)$$

It is interesting to note that the QTL detection power for GCA does not differ among the different population-based NCII designs when the sample size is large and the broad heritability is high (detection power close to 100). However, the QTL detection powers of the various designs differ when the broad heritability is low (tables 6–10 in electronic supplementary material). With multiple alleles at each locus, the subpopulation subtractions were defined as follows: $\overline{MM}_{Gca} - \overline{Mm}_{Gca} = \frac{1}{2}(1-2r)\left(G - G'\right)$ in the BC-based NCII design; $\overline{MM}_{Gca} - \overline{Mm}_{Gca} = \overline{Mm}_{Gca} - \overline{mm}_{Gca} = \frac{1}{2}(1-2r)\left(G - G'\right)$ and $\overline{MM}_{Gca} - \overline{mm}_{Gca} = (1-2r)\left(G - G'\right)$ in the $F_2$-based NCII design; and $\overline{MM}_{Gca} - \overline{Mm}_{Gca} = (1-2r)\left(G - G'\right)$ in the DH-based NCII design. For a fixed recombination rate $r$, the order of the QTL detection power for GCA should be DH > $F_2$ > BC. When the sample size was 200 and broad heritability was 20%, the detection power for the first QTL ranked as follows: DH (27) > $F_2$ (20) > BC (13), whereas the detection power for the second QTL was in the order of DH (100) > $F_2$ (99) > BC (58).

It should be noted that the recombination rate $r$ also impacts the QTL detection power. In the RIL-based NCII design, the expectations for every dataset were similar to those in the DH-based NCII design, except for $r$, which was replaced by $2r_m'/\left(1+2r_m'\right)$ or $4r_m''/\left(1+6r_m''\right)$. $r_m'$ and $r_m''$ were recombinant values for two RI populations (selfing population and sib-mating population, respectively) (Hu *et al.* 1995). For a fixed QTL position, $2r_m'/\left(1+2r_m'\right)$ and $4r_m''/\left(1+6r_m''\right)$ were less than $r$, so the subpopulation subtraction between *MM* and *mm* in the RIL-based NCII design is less than that determined for the DH-based NCII design. This difference is reflected by the fact that the number of QTLs detected in the RIL-based NCII design was clearly less than that of the DH-based NCII design. This pattern was particularly notable for conditions of low heritability (tables 3–5 and 8–10 in electronic supplementary material).

Among the different population-based NCII designs, we highly recommend the RIL (or DH) based NCII design for the following reasons. First, RIL (or DH) is a permanent segregating population, so it can be shared between research groups, and testcross progeny can be repeatedly generated and phenotyped as needed (Frascaroli *et al.* 2007, 2009; Melchinger *et al.* 2007; Li *et al.* 2008; He and Zhang 2011). Next, RIL (DH) offers a higher mapping resolution for allowing more recombinational segregation of tightly linked QTLs than other populations (Xiao *et al.* 1995; Li *et al.* 2008; Frascaroli *et al.* 2009; He and Zhang 2011). Further, the use of homozygous RIL (or DH) provides the highest statistical power for QTL detection (Li *et al.* 2008; Frascaroli *et al.* 2009; Schön *et al.* 2010). Finally, as combining ability was initially studied in homozygous inbred lines, further exploration of the RIL (or DH)-based NCII design is warranted.

The drawbacks of our method are that: (i) because the QTL effects investigated for GCA and SCA are mixtures of additive and dominant effects as well as the genetic effects of multiple alleles, the QTL effect documented in this study is augmented effect, and (ii) for genetic basis of combining ability is very complicated and hardly to be explicated when take epistasis into account, only main-effect of QTL for combining ability was researched in present study. Many studies showed that epistasis play important effect on heterosis

(Li *et al.* 2008; Reif *et al.* 2009). The influence of epistasis on combining ability is in need for further study.

## Conclusion

We demonstrate that by combining an NCII design and network interval mapping (NWIM) with QTLnetwork ver. 2.0, we can dissect the genetic architecture of the combining ability of a complex trait. The combination of the NCII design with QTL mapping permits the identification of QTL loci for traits as well as assessment of the generalizability of the findings. If GCA and SCA loci detected, without crossbreeding all inbred lines in field, breeders can predict the combining ability of new combinations through identifying the molecular marker of new parental lines. The method documented in this study should be more accurate in combining ability prediction than by conventional method (e.g. analysis of variance analysis). Eventually, it will facilitate breeders in selecting elite parental lines with high combining ability.

#### Acknowledgements

This work was financially supported by the 973 program (no. 2011CB100102), the National Natural Science Foundation of China (no. 31000666), the China Postdoctoral Science foundation (no. 2012M511722) and Hunan Province Postdoctoral Science foundation (no. 2012RS4039).

## References

Berg B. A. 2004 *Markov Chain Monte Carlo simulations and their statistical analysis (with web-based Fortran code)*. World Scientific, Hackensack, USA.

Basbag S., Ekinci R. and Gencer O. 2007 Combining ability and heterosis for earliness characters in line6tester population of *Gossypium hirsutum* L. *Hereditas* **144**, 185–190.

Dhillon B. S. and Singh J. 1978 Evaluation of circulant partial diallel crosses in maize. *Theor. Appl. Genet.* **52**, 29–37.

Fishman G. S. 1995 *Monte Carlo: concepts, algorithms, and applications*. Springer, New York, USA.

Frascaroli E., Cane M. A., Landi P., Pea G. G., Villa M., Morgante M. and Pè M. E. 2007 Classical genetic and quantitative trait loci analyses of heterosis in a maize hybrid between two elite inbred lines. *Genetics* **176**, 625-644.

Frascaroli E., Canè M. A., Pè M. E., Pea G., Morgante M. and Landi P. 2009 QTL detection in maize testcross progenies as affected by related and unrelated testers. *Theor. Appl. Genet.* **118**, 993–1004.

Garcia A. A. F., Wang S., Melchinger A. E. and Zeng Z. B. 2008 Quantitative trait loci mapping and the genetic basis of heterosis in maize and rice. *Genetics* **180**, 1707–1724.

Griffing B. 1956 Concept of general and specific combining ability in relation to diallel crossing system. *Aust. J. Biol. Sci.* **9**, 463–493.

Groh S., Gonzáles-de-León D., Khairallah M. M., Jiang C., Bergvinson D., Bohn M. *et al.* 1998 QTL mapping in tropical maize: III. Genomic regions for resistance to Diatraea spp. and associated traits in two RIL populations. *Crop Sci.* **38**, 1062–1072.

He X. H., Hu Z. L. and Zhang Y. M. 2012 Genome-wide mapping of QTL associated with heterosis in the RIL-based NCIII design. *Chin. Sci. Bull.* **57**, 2655–2665.

He X. H., Xu C. W., Kuai J. M., Gu S. L. and Li T. 2001 Principal factors affecting the power of detection and accuracy of QTL mapping. *Acta Agron. Sin.* **27**, 469–475.

He X. H. and Zhang Y. M. 2011 A complete solution for dissecting pure main and epistatic effects of QTL in triple testcross design. *PLoS One* **6**, e24575.

Hu Z. L., Xie C., McDaniel G. R., Kuhler D. L. and Zhang X. 1995 A correlation method for detecting and estimating linkage between a marker locus and a quantitative trait locus using inbred lines. *Theor. Appl. Genet.* **90**, 1074–1078.

Kempthorne O. and Curnow R. N. 1961 The partial diallel cross. *Biometrics* **17**, 229–250.

Kusterer B., Piepho H. P., Utz H. F., Schön C. C., Muminovic J., Meyer R. C. *et al.* 2007 Heterosis for biomass-related traits in Arabidopsis investigated by quantitative trait loci analysis of the triple testcross design with recombinant inbred lines. *Genetics* **177**, 1839–1850.

Li L. Z., Lu K. Y., Chen Z. M., Mou T. M., Hu Z. L. and Li X. Q. 2010 Gene actions at loci underlying several quantitative traits in two elite rice hybrids. *Mol. Genet. Genomics* **284**, 383–397.

Li L. Z., Lu K. Y., Mou T. M., Hu Z. L. and Li X. Q. 2008 Dominance, overdominance and epistasis condition the heterosis in two heterotic rice hybrids. *Genetics* **180**, 1725–1742.

Melchinger A. E., Piepho H. P., Utz H. F., Muminovic J., Wegenast T., Törjék O. *et al.* 2007 Genetic basis of heterosis for growth-related traits in Arabidopsis investigated by testcross progenies of near-Isogenic lines reveals a significant role of epistasis. *Genetics* **177**, 1827–1837.

Qi H., Huang J., Zheng Q., Huang Y., Shao R., Zhu L. *et al.* 2012 Identification of combining ability loci for five yield-related traits in maize using a set of testcrosses with introgression lines. *Theor. Appl. Genet.* **126**, 369–377.

Qu Z., Li L. Z., Luo J. Y., Wang P., Yu S. B., Mou T. M. *et al.* 2012 QTL Mapping of combining ability and heterosis of agronomic traits in rice backcross recombinant inbred lines and hybrid crosses. *PLoS One* **7**, e28463.

Rao S. and Li X. 2000 Strategies for genetic mapping of categorical traits. *Genetica* **109**, 183–197.

Reif J. C., Kusterer B., Piepho H. P., Meyer R. C., Altmann T., Schön C. C. and Melchinger A. E. 2009 Unraveling epistasis with triple testcross progenies of near-isogenic lines. *Genetics* **181**, 247–251.

Riedelsheimer C., Czedik-Eysenberg A., Grieder C., Lisec J., Technow F., Sulpice R. *et al.* 2012 Genomic and metabolic prediction of complex heterotic traits in hybrid maize. *Nat. Genet.* **44**, 217–220.

Schön C. C., Dhillon B. S., Utz H. F. and Melchinger A E. 2010. High congruency of QTL positions for heterosis of grain yield in three crosses of maize. *Theor. Appl. Genet.* **120**, 321–332.

Shukla S. K. and Pandey M. P. 2008 Combining ability and heterosis over environments for yield and yield components in two-line hybrids involving thermosensitive genic male sterile lines in rice *Oryza sativa* L. *Plant Breed.* **127**, 28–32.

Su C. F., Zhao T. J. and Ga J. Y. 2010 Simulation comparisons of effectiveness among QTL mapping procedures of different statistical genetic models. *Acta Agron. Sin.* **36**, 1100–1107.

Verhoeven K. J. F., Jannink J. L. and McIntyre L. M. 2006 Using mating designs to uncover QTL and the genetic architecture of complex traits. *Heredity* **96**, 139–149.

Wu X. L. and Jannink J. L. 2004 Optimal sampling of a population to determine QTL location, variance, and allelic number. *Theor. Appl. Genet.* **108**, 1434–1442.

Xiao J., Li J., Yuan L. and Tanksley S. D. 1995 Dominance is the major genetic basis of heterosis in rice as revealed by QTL analysis using molecular markers. *Genetics* **140**, 745–754.

Xie C. Q., Gessler D. D. G. and Xu S. Z. 1998 Combining different line crosses for mapping quantitative trait loci using the identical by descent-based variance component method. *Genetics* **149**, 1139–1146.

Xu S. 1998 Mapping quantitative trait loci using multiple families of line crosses. *Genetics* **148**, 517–524.

Yang J., Hu C. C., Yu R. D., Xia Z., Ye X. Z. and Zhu J. 2008 QTLNetwork, mapping and visualizing genetic architecture of complex traits in experimental populations. *Bioinformatics* **24**, 721–723.

Zhu Y. P. and He F. M. 2001 Studies on inbred lines of sweet corn as testers and its number. *Southwest Chin. J. Agric. Sci.* **14**, 25–28.

Received 8 May 2013, in revised form 22 July 2013; accepted 6 August 2013
Published on the Web: 13 December 2013