Theor Appl Genet (2012) 125:235–246
DOI 10.1007/s00122-012-1828-4

ORIGINAL PAPER

Probability of success of breeding strategies for improving pro-vitamin A content in maize

Xuecai Zhang · Wolfgang H. Pfeiffer · Natalia Palacios-Rojas · Raman Babu · Howarth Bouis · Jiankang Wang

Received: 21 September 2011 / Accepted: 11 February 2012 / Published online: 27 March 2012

Abstract Biofortification for pro-vitamin A content (pVAC) of modern maize inbreds and hybrids is a feasible way to deal with vitamin A deficiency in rural areas in developing countries. The objective of this study was to evaluate the probability of success of breeding strategies when transferring the high pVAC present in donors to elite modern-adapted lines. For this purpose, a genetic model was built based on previous genetic studies, and different selection schemes including phenotypic selection (PS) and marker-assisted selection (MAS) were simulated and compared. MAS for simultaneously selecting all pVAC genes and a combined scheme for selecting two major pVAC genes by MAS followed by ultra performance liquid chromatography screening for the remaining genetic variation on pVAC were identified as being most effective and cost-efficient. The two schemes have 83.7 and 84.8% probabilities of achieving a predefined breeding target on pVAC and adaptation in one breeding cycle under the current breeding scale. When the breeding scale is increased by making 50% more crosses, the probability values could reach 94.8 and 95.1% for the two schemes. Under fixed resources, larger early generation populations with fewer crosses had similar breeding efficiency to smaller early generation populations with more crosses. Breeding on a larger scale was more efficient both genetically and economically. The approach presented in this study could be used as a general way in quantifying probability of success and comparing different breeding schemes in other breeding programs.

Communicated by M. Frisch.

X. Zhang · J. Wang (✉)
Institute of Crop Science, The National Key Facility for Crop Gene Resources and Genetic Improvement, and CIMMYT China Office, Chinese Academy of Agricultural Sciences, No. 12 Zhongguancun South Street, Beijing 100081, China
e-mail: wangjk@caas.net.cn; jkwang@cgiar.org

X. Zhang · N. Palacios-Rojas · R. Babu
International Maize and Wheat Improvement Center (CIMMYT), Apdo. Postal 6-641, 06600 Mexico, D.F., Mexico

W. H. Pfeiffer
HarvestPlus Challenge Programme and International Center for Tropical Agriculture (CIAT-HarvestPlus), A.A. 6713 Cali, Colombia

H. Bouis
HarvestPlus Challenge Program, International Food Policy Research Institute (IFPRI), 2033 K Street, NW Washington, DC 20006-1002, USA

Introduction

Vitamin A deficiency has been recognized by the World Health Organization as a major public health problem that afflicts more than 100 million people in the developing world, especially young children and women. HarvestPlus (http://www.harvestplus.org) is a Challenge Program of the Consultative Group on International Agricultural Research (CGIAR) that holds the view that micronutrient malnutrition can be reduced by enhancing the nutrient content of staple crops through modern breeding technologies. This approach is referred to as biofortification, a strategy that is more sustainable and cost-efficient compared with the traditional nutrition interventions such as supplementation, food fortification, and dietary diversification (Pfeiffer and McClafferty 2007; Bouis and Welch 2010).

Maize is a model cereal crop for developing strategies to solve global micronutrient deficiencies, showing promise for pro-vitamin A biofortification (Wurtzel 2010). It is a staple

food for more than one billion people in sub-Saharan Africa and Latin America, and exhibits a broad range of genetic variation for pVAC (Kurilich and Juvik 1999; Egesel et al. 2003; Chander et al. 2008b; Menkir et al. 2008). However, most modern elite-adapted maize inbreds and hybrids have low levels of pVAC (i.e., 1–2 μg/g), compared with some donor parents, whose pVAC can reach 10 μg/g (Pfeiffer and McClafferty 2007; Pixley et al. 2010). High performance liquid chromatography (HPLC) is currently used for testing pVAC in maize grain (Weber 1987); however, it is expensive and time-consuming, which makes it impossible to apply in the large-scale selection of early-generation breeding lines. Less expensive and high-throughput methods such as near infrared reflectance (Brenna and Bernardo 2004; Tallada et al. 2009) and ultra performance liquid chromatography (UPLC) are being developed for the large-scale screening of pVAC. The precision of these methods is usually lower compared with HPLC.

Progress in developing saturated genetic maps and quantitative trait loci (QTL) mapping methods (Bernardo 2002; Li et al. 2007a) has led to the intensive use of QTL mapping in various segregating populations to study the genetic control of carotenoids in maize grain (Wong et al. 2004; Chander et al. 2008a). Major genes with large genetic effects (i.e., *crtRB1* and *LcyE*), have been identified and confirmed across different populations (Harjes et al. 2008; Yan et al. 2010). Though biochemical regulation of carotenoid biosynthesis in maize endosperm has not yet been fully characterized, key genes encoding major structural enzymes that synthesize and regulate carotenoids have been isolated, characterized and cloned (Buckner et al. 1990; Li et al. 1996, 2007b; Sun et al. 1996; Tian and DellaPenna 2001; Bai et al. 2009). These genetic studies provide not only a better understanding of the genetic control of carotenoids in maize grain, but also the opportunity to use marker-assisted selection (MAS) to enhance expression of the trait through breeding (Bouis and Welch 2010).

Micronutrients have not been the major target traits of modern maize breeding. Most genetic materials rich in the favored micronutrients are usually poorly adapted and show poor agronomic performance, whereas most elite-adapted maize inbreds and hybrids have low micronutrient content. Our objective in this study was to simulate breeding strategies for transferring high pVAC from donor parental lines to elite-adapted lines. Specifically, we tried to answer questions relevant to pVAC improvement in maize breeding. For example, based on breeders' current knowledge of the genetic structure of pVAC and of parental lines, can a target concentration for pVAC in maize grain be achieved in just one breeding cycle? What are the most efficient breeding strategies for achieving this target? How would the probability of success change if the investment in maize breeding were increased?

Materials and methods

The hybrid breeding simulation tool of QuHybrid

As a genetic and breeding simulation tool based on the QU-GENE platform (Podlich and Cooper 1998), QuLine is capable of simulating most breeding methods for selecting fixed lines as cultivars (Wang et al. 2003, 2004). Breeding methods that can be simulated in QuLine include mass selection, pedigree breeding (including single seed descent), bulk population breeding, backcross breeding, topcross (or three-way cross) breeding, doubled haploid breeding, MAS, and combinations and modifications of these methods. QuHybrid was developed to simulate breeding programs for selecting hybrids as cultivars. While retaining most QuLine functionalities, QuHybrid is able to make testcross and hybrid performance predictions; it is thus possible to simulate the hybrid breeding program and optimize the hybrid breeding strategy. Both breeding simulation tools are freely available from http://www.uq.edu.au/lcafs/qugene/.

Genetic models used in simulation

We considered a genome of ten chromosomes, similar to the maize genome. Two traits were defined, one for pVAC in maize grain, and the other for field adaptation. Based on previous genetic studies (Wong et al. 2004; Chander et al. 2008a; Harjes et al. 2008; Yan et al. 2010), four additive genes, distributed on four chromosomes, were assumed to control pVAC (Table 1). The total additive effect of these genes was 7.5 μg/g and the mid-parent value of pVAC was 7.5 μg/g. Therefore, the maximum pVAC value was 15 μg/g when all favorable alleles were present in one individual, and the minimum value was 0. Each pVAC gene was completely linked with a co-dominant molecular marker, which allows selecting a gene through its linked marker. When there is no closely linked marker, the use of two flanking markers can also assure the selection of the target gene (Wang et al. 2007). The broad-sense heritability of pVAC at the individual plant level was set at 0.4 when pVAC was tested by UPLC, which is faster and less expensive than HPLC, but also less precise. Heritability was 0.8 when pVAC was tested by HPLC (Weber 1987), which is time-consuming, expensive, but more precise (Table 1).

Adaptation in this study was defined as an integrated index of various breeding target traits (other than pVAC), such as maturity, plant height, yield and yield components, and resistance or tolerance to biotic and abiotic stresses. Adaptation was used to evaluate whether the final selected lines had similar or improved agronomic traits, compared with the initial adapted lines. Heritability at the individual plant level was set at 0.2 for adaptation (Hallauer 2007).

Table 1 Four pVAC genes, their additive effects and explained phenotypic variances (PVE), and genotypes in the donor and adapted parents

Locus	Additive effect (μg/g)	PVE (%)		Donor parent[b]	Adapted parent[b]
		$H = 0.8$[a]	$H = 0.4$[a]		
pVAC1	2.28	0.28	0.14	*AA*	*aa*
pVAC2	2.28	0.28	0.14	*BB*	*bb*
pVAC3	1.72	0.16	0.08	*cc*	*CC*
pVAC4	1.22	0.08	0.04	*DD*	*dd*

[a] H represents the broad-sense heritability. $H = 0.4$ indicates that pVAC was tested by UPLC, and 0.8 indicates that pVAC was tested by HPLC

[b] *A–a*, *B–b*, *C–c*, and *D–d* are the two alleles at the four pVAC loci, respectively. Three favorable alleles, i.e., *A*, *B* and *D*, exist in the donor population, and one favorable allele, i.e., *C*, exists in the adapted population

It was assumed that adaptation was controlled by a total of 100 genes evenly distributed on the ten chromosomes; the additive effect of each gene was 0.5. The maximum value for adaptation was 100 when all favorable alleles were fixed in one genotype.

We generated three populations, representing adapted, donor, and tester lines, respectively. The adapted population consisted of 40 modern elite inbred lines, where the frequency of favorable adaptation alleles was 0.90. Of the four favorable alleles, the one at locus *pVAC3* was present in the adapted population (Table 1). Thus, the average genotypic values were 90 for adaptation and 3.44 μg/g for pVAC. There were ten inbred lines in the donor population, where the frequency of favorable adaptation alleles was 0.60. Favorable alleles at *pVAC1*, *pVAC2* and *pVAC4* were present in this population (Table 1). Thus, the average genotypic values were 60 for adaptation and 11.56 μg/g for pVAC. The tester population consisted of one inbred line, where the frequency of favorable adaptation alleles was assumed at 0.80, and all favorable alleles for pVAC were absent. Thus, the average genotypic values were 80 for adaptation and zero for pVAC. We understand that different testers may result in varied combining abilities of the final selected lines. The choice of suitable testers maximizing the genetic gain on hybrid performance was not the topic of this study. Also as the study is looking at pyramiding of additive QTL, the choice of tester is not relevant here. Therefore, we did not include results from other potential testers.

Breeding strategies used in simulation

The breeding target is to select inbred lines whose pVAC is ≥11.5 μg/g and whose adaption is ≥90 (or no worse than that of current inbred lines) after one breeding cycle. A breeding cycle begins with crossing and ends when the selected advanced lines are almost fixed. The pVAC breeding program being implemented at CIMMYT currently makes 40 crosses, and generates 600 BC_1F_2 individuals for each cross every year (Fig. 1; Table 2). Screening for adaptation (including disease resistance, maturity, yield, etc.) is conducted based on the field performance at all generations except BC_1F_3 and BC_1F_6, where selection is based on testcross performance. For testcross selection, each $BC_1F_{2:3}$ family is crossed with the tester line, and top families, i.e., those with the best testcross performance in the field, are selected. Each selected family is harvested in bulk to form BC_1F_4 for the next generation (Fig. 1). Testcross selection is performed once again in BC_1F_6 to choose the best families with superior combining ability. Final retained lines are tested in the BC_1F_7 generation by HPLC to confirm whether the target of pVAC ≥11.5 has been met after one cycle of breeding (Fig. 1). The procedure described above is similar to the single backcrossing breeding strategy used in CIMMYT's wheat breeding programs (Wang et al. 2009a).

After selecting for adaptation, pVAC could be selected in the BC_1F_1 and BC_1F_2 generations as well, using the self-pollinated seeds of individual plants (Fig. 1). Two phenotypic selection (PS) schemes were simulated by considering two methods for measuring pVAC, i.e., *UPLC* and *HPLC* (Table 3). Hypothetically, MAS can be conducted for the four pVAC genes (i.e. *MAS4* in Table 3), as follows. In the BC_1F_1 generation, individuals homozygous for *pVAC3*, and heterozygous for *pVAC1*, *pVAC2* and *pVAC4*, were selected. The total proportion for selecting pVAC in the BC_1F_1 generation was estimated at $0.5^4 = 0.0625$, as the selected proportion was 0.5 for each of the four selected loci assuming no segregation distortion. It should be noted that the favorable allele at *pVAC3* was fixed in the BC_1F_1 generation. In the BC_1F_2 generation, individuals homozygous for *pVAC1*, *pVAC2* and *pVAC4* were selected so as to fix the four pVAC alleles. The total proportion for selecting pVAC in the BC_1F_2 generation was estimated at $0.25^3 = 0.015625$, as the selected proportion was 0.25 for each of the three selected loci. Segregation distortion for pVAC markers may be observed in practical breeding population, but was not considered in this simulation study.

When MAS is applied for all four pVAC genes, i.e. *MAS4*, no PS on pVAC is needed since the four genes explain all the genetic variation for pVAC. For comparison, MAS for fewer genes was also considered (Table 3). For example, MAS was conducted only for the largest-effect genes, and PS was conducted for selecting other genes. *MAS1 + UPLC* and *MAS1 + HPLC* were the two schemes in which the largest-effect pVAC gene was selected by MAS, and the other three were selected by the UPLC and HPLC phenotyping test, respectively (Table 3). Similarly, *MAS2 + UPLC* and *MAS2 + HPLC* were the two schemes where the two largest-

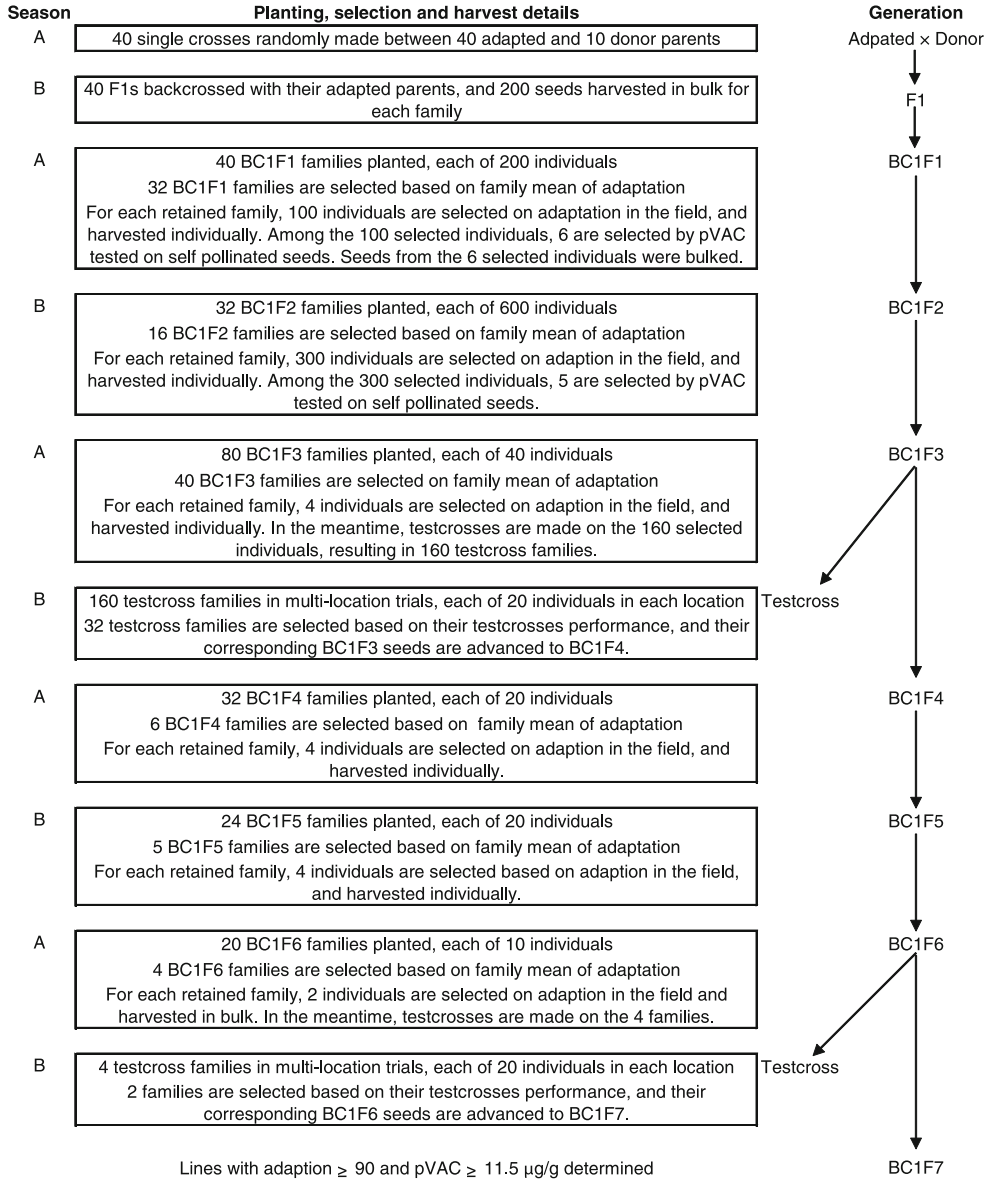

Fig. 1 Flowchart of the breeding strategy currently used at CIMMYT for improving pVAC. Forty crosses are made, 600 BC_1F_2 individuals from each cross are grown, and pVAC is tested by UPLC in early generations, and by HPLC in late generations

effect pVAC genes were selected by MAS, and *MAS3 + UPLC* and *MAS3 + HPLC* were the two schemes where the three largest-effect pVAC genes were selected by MAS. Varied proportions of PS on pVAC were used to make sure each selection scheme had similar selection intensity.

Breeding scales and the evaluation of genetic gain and cost

Two factors were considered in defining the breeding scale: (1) number of crosses, and (2) BC_1F_2 population size of each cross. There were four levels for the number of crosses (i.e., 40, 60, 80 and 100), and five levels for BC_1F_2 population size (i.e., 400, 600, 800, 1,000 and 1,200), representing various operational scales or investments. In total, we simulated 20 breeding scales and 9 selection schemes (Table 4). QuHybrid was used to simulate the breeding procedure. One thousand runs were repeated for each strategy, from which the probability of success was calculated. By success we mean that at least one target line was selected at the end of one breeding cycle. In the cost–benefit analysis, we assume that it costs $2 to grow one maize plant, to make one single cross, backcross or self-pollination; $5 to acquire one marker data point; and $20 to test pVAC by UPLC and $50 to test pVAC by HPLC in one sample (Pfeiffer and McClafferty 2007). The calculation of cost in each breeding generation for selection scheme *HPLC* is given in detail in the last column in Table 2. Total cost of each scheme was given in Table 4. As expected, the use of HPLC in selecting

Table 2 Flowchart of the breeding strategy used at CIMMYT for improving pVAC, where 40 crosses are made, 600 BC_1F_2 individuals from each cross are grown, and pVAC is tested by HPLC

Generation	Seed propagation[a]	Generation advance[b]	Families	Family size	Among family selection[c]	Within family selection[c]		Cost ($)[d]
						Adaptation	pVAC	
Parents	Selfing	Bulk	50	10	1.0	1.0	1.0	1,080
F1	Hand pollination	Bulk	40	10	1.0	1.0	1.0	880
BC1F1	Backcrossing	Bulk	40	200	0.8	0.5	0.0625	182,400
BC1F2	Selfing	Pedigree	32	600	0.5	0.5	0.015625	288,000
BC1F3	Selfing	Pedigree	80	40	0.5	0.1	1.0	6,720
BC1F3T	Testcross	Bulk	160	20	0.2	1.0	1.0	6,400
BC1F4	Selfing	Pedigree	32	20	0.2	0.2	1.0	1,328
BC1F5	Selfing	Pedigree	24	20	0.2	0.2	1.0	1,000
BC1F6	Selfing	Bulk	20	10	0.2	0.2	1.0	408
BC1F6T	Testcross	Bulk	4	20	0.5	1.0	1.0	160
BC1F7	Selfing	Bulk	2	10				140

[a] Seed propagation specifies how the seed is formed for the next generation. Available methods in QuHybrid are testcross, clone, selfing, random mating, backcross, topcross, doubled haploids, and noselfing (random mating but selfing pollination avoided)

[b] There are two options in QuHybrid for generation advanced method, one is pedigree and the other is bulk. When pedigree is used, each selected individual will form a family in the next generation. When bulk is used, all selected individuals in a family will be harvested in bulk and form one family in the next generation

[c] For among family selection, the number for each generation is the proportion of families to be selected. For within family selection, the two numbers for each generation are proportions of individuals to be selected on field adaptation and on pVAC

[d] We assumed the cost was $2 to grow one maize plant, to make one single cross, backcross or self-pollination, $5 to acquire one marker data point, and $20 to test pVAC by UPLC and $50 by HPLC in one sample

Table 3 Genotypes to be selected by markers and proportions in phenotypic selection (PS) on pVAC in two generations

Selection scheme[a]	BC_1F_1		BC_1F_2	
	Genotypes to be selected	Proportion to be selected in PS	Genotypes to be selected	Proportion to be selected in PS
PS	N.A.	0.0625	N.A.	0.015625
MAS4	*Aa Bb CC Dd*	N.A.	*AA BB DD*	N.A.
MAS1 + PS	*Aa*	0.125	*AA*	0.0625
MAS2 + PS	*Aa Bb*	0.25	*AA BB*	0.25
MAS3 + PS	*Aa Bb Dd*	0.5	*AA BB DD*	N.A.

[a] *PS* represents only phenotypic selection was conducted for pVAC, which can be based on either UPLC or HPLC. *MAS4* represents all four pVAC genes were selected by markers and no phenotypic selection on pVAC was conducted. *MAS1 + PS* represents that *pVAC1* was selected by its associated marker and the remaining variation by phenotypic selection. *MAS2 + PS* represents that *pVAC1* and *pVAC2* were selected by markers and the remaining variation by phenotypic selection. *MAS3 + PS* represents that *pVAC1*, *pVAC2* and *pVAC3* were selected by markers and the remaining variation by phenotypic selection

pVAC is most expensive. In comparison, UPLC and MAS greatly reduce the cost to run the breeding program.

Results

Comparison of selection schemes by probability of success

A selection scheme is successful if at least one maize inbred line whose pVAC is $\geq$11.5 μg/g and adaptation is $\geq$90 is selected after one breeding cycle. Probability of success values for the nine selection schemes and 20 breeding scales are presented in Table 5. Great differences can be observed for the nine schemes on a fixed breeding scale, indicating the importance of choosing appropriate selection schemes in breeding. Due to the high precision of HPLC in pVAC testing, PS based on HPLC has much higher probability of success than PS based on UPLC, indicating the importance of high phenotyping precision in selection. When the four pVAC genes have molecular markers that

Table 4 Cost (in K$) of each breeding scheme in simulation

Number of crosses	Segregating population size	Selection scheme								
		UPLC	*HPLC*	*MAS4*	*MAS1 + UPLC*	*MAS2 + UPLC*	*MAS3 + UPLC*	*MAS1 + HPLC*	*MAS2 + HPLC*	*MAS3 + HPLC*
40[a]	400	194.1	386.1	178.1	146.1	150.1	170.1	218.1	180.1	182.1
	600[a]	248.6	488.6	224.6	184.6	190.6	216.6	268.6	223.6	228.6
	800	300.0	588.0	268.0	220.0	228.0	260.0	316.0	264.0	272.0
	1,000	354.3	690.3	314.3	258.3	268.3	306.3	366.3	307.3	318.3
	1,200	405.9	789.9	357.9	293.9	305.9	349.9	413.9	347.9	361.9
60	400	290.8	578.8	266.8	218.8	224.8	254.8	326.8	269.8	272.8
	600	372.5	732.5	336.5	276.5	285.5	324.5	402.5	335.0	342.5
	800	449.6	881.6	401.6	329.6	341.6	389.6	473.6	395.6	407.6
	1,000	531.1	1035.1	471.1	387.1	402.1	459.1	549.1	460.6	477.1
	1,200	608.0	1184.0	536.0	440.0	458.0	524.0	620.0	521.0	542.0
80	400	387.3	771.3	355.3	291.3	299.3	339.3	435.3	359.3	363.3
	600	496.3	976.3	448.3	368.3	380.3	432.3	536.3	446.3	456.3
	800	598.7	1174.7	534.7	438.7	454.7	518.7	630.7	526.7	542.7
	1,000	707.7	1379.7	627.7	515.7	535.7	611.7	731.7	613.7	635.7
	1,200	810.2	1578.2	714.2	586.2	610.2	698.2	826.2	694.2	722.2
100	400	483.9	963.9	443.9	363.9	373.9	423.9	543.9	448.9	453.9
	600	620.1	1220.1	560.1	460.1	475.1	540.1	670.1	557.6	570.1
	800	748.2	1468.2	668.2	548.2	568.2	648.2	788.2	658.2	678.2
	1,000	884.7	1724.7	784.7	644.7	669.7	764.7	914.7	767.2	794.7
	1,200	1012.7	1972.7	892.7	732.7	762.7	872.7	1032.7	867.7	902.7

[a] Underlined values represent the number of crosses and segregating population size currently used in CIMMYT

are completely linked, MAS for all pVAC genes shows the highest probability of success among all schemes. There are other combined MAS/PS schemes between *PS + UPLC* and *MAS4*, but the difference between *MAS2 + PS* and *MAS4* or *MAS3 + PS* and *MAS4* is minor.

Effect of an increased breeding scale

For the current breeding scale (number of crosses = 40 and BC_1F_2 population size = 600), probability of success is 43.7% when PS for pVAC is based on UPLC, and 73.8% when PS is based on HPLC (Table 5). When the four pVAC genes have molecular markers that are completely linked, MAS for all pVAC genes shows a probability of success of 83.7%. Other combined MAS/PS schemes have probability of success between 60.0 and 84.8%. Thus, there is still a risk that the breeding target will be missed using the current breeding scale.

If the number of crosses stays at 40, but the size of the BC1F2 generation doubles, i.e., BC_1F_2 population size = 1,200, PS based on HPLC will have a greater than 90% probability of achieving the breeding target, and *MAS4* will have a probability above to 95% (Table 5). The risk of missing the breeding target is greatly reduced by the increase in breeding population size. A comparable increase in breeding scale can also be achieved by doubling the number of crosses. When the number of crosses is doubled, i.e., number of crosses of 80, but the size of BC1F2 stays at 600, each selection scheme shows similar probability of success as that of the breeding scale number of crosses of 40 and BC_1F_2 population size of 1,200 (Table 5). Thus, the risk can also be reduced by increasing the number of crosses as well.

In Table 5, similar probability of success can be observed for the same breeding scale. For example, number of crosses = 40 and BC_1F_2 population size = 1,200, number of crosses = 60 and BC_1F_2 population size = 800, and number of crosses = 80 and BC_1F_2 population size = 600 have similar probability of success no matter which scheme is used. Therefore, increasing the number of crosses or the increasing the segregating population size can increase the probability of success significantly.

Cost and benefit analysis

Different selection schemes have different breeding costs, from which the cost of selecting one target line can also be calculated. The cost of selecting one target line is defined in this study as the ratio of the breeding cost in one cycle to the number of target inbred lines selected at the end of the

Table 5 Probability of success (%) when number of crosses was equal to 40, 60, 80, and 100, and BC_1F_2 population size was equal to 400, 600, 800, 1,000, and 1,200

Number of crosses	Segregating population size	Selection scheme								
		UPLC	*HPLC*	*MAS4*	*MAS1 + UPLC*	*MAS2 + UPLC*	*MAS3 + UPLC*	*MAS1 + HPLC*	*MAS2 + HPLC*	*MAS3 + HPLC*
40[a]	400	28.3	50.5	62.8	38.7	54.5	56.6	56.4	56.8	60.4
	600[a]	43.7	73.8	83.7	60.0	81.6	82.5	76.5	84.8	82.4
	800	56.2	86.3	93.1	70.2	91.0	89.8	90.4	92.9	91.7
	1,000	64.2	88.5	96.6	74.1	92.6	93.0	91.9	94.0	93.4
	1,200	66.9	94.3	96.5	80.1	95.7	96.3	96.1	97.4	96.7
60	400	48.3	76.2	85.4	56.6	80.3	81.7	80.9	84.8	82.1
	600	58.0	87.3	94.8	72.3	91.9	92.8	91.5	95.1	92.2
	800	68.2	93.7	97.3	83.1	96.3	95.3	97.6	97.2	95.2
	1,000	79.1	97.1	99.0	88.8	98.2	98.2	98.6	99.4	98.3
	1,200	82.5	98.3	99.9	91.2	99.4	99.6	98.8	99.6	99.2
80	400	56.3	87.4	93.2	67.8	89.7	89.7	89.0	93.1	90.3
	600	68.7	94.3	97.1	78.5	96.2	95.4	94.5	97.5	95.9
	800	74.3	97.1	99.4	86.4	98.2	98.3	97.8	99.1	98.9
	1,000	83.8	99.0	99.8	92.1	98.7	99.3	99.2	99.6	99.5
	1,200	86.9	99.0	99.9	93.1	99.7	100.0	99.7	100.0	99.7
100	400	58.0	89.1	97.7	72.9	93.2	94.0	92.0	95.0	94.3
	600	76.2	97.4	99.4	87.6	98.4	97.8	97.9	99.2	98.9
	800	83.9	99.1	99.9	93.2	99.1	99.7	99.6	99.7	99.7
	1,000	92.6	99.9	100.0	96.0	100.0	99.9	99.5	100.0	99.9
	1,200	92.4	99.9	100.0	97.3	100.0	100.0	99.7	100.0	100.0

[a] Underlined values represent the number of crosses and segregating population size currently used in CIMMYT

breeding cycle; this is a suitable index for measuring the cost-efficiency of a selection scheme (Fig. 2). For a certain breeding scale, the selection scheme with higher probability of success is generally more cost-efficient as well (Fig. 2). The cost of selecting one target line for *MAS4* is similar to the cost of selecting one target line for *MAS2* and one for *MAS3* (*MAS4* not shown in Fig. 2), regardless of whether UPLC or HPLC is used in PS for pVAC. Breeding with larger scales not only resulted in higher probability of success (Table 5) but also was more cost-efficient. A clear trend is evident, i.e., that a large breeding scale decreases the cost of selecting one target line, regardless whether HPLC (Fig. 2a) or UPLC (Fig. 2b) is used in PS for pVAC.

Scheme *MAS2 + UPLC* was identified as being most cost-efficient (Fig. 2). When the number of crosses of 60 and selection scheme *MAS2 + HPLC* were applied, the cost of selecting one target line decreased from $215k to $138k, as BC_1F_2 population size increased from 400 to 1,200. In comparison, when selection scheme *MAS2 + UPLC* was applied, the cost decreased from $190 k to $130 k, as BC_1F_2 population size increased from 400 to 1,200 (Fig. 3). Though more efficient genetically, PS based on HPLC is less cost-efficient compared with PS based on UPLC (Fig. 3).

Robustness of the genetic and population models

Computer simulation greatly facilitates the comparison of a large number of possible crossing and selection scenarios in breeding. However, simulation results may always depend on the genetic models used. From previous genetic studies, it is generally accepted that pVAC has a fairly simple genetic architecture controlled by a few additive genes. Adaptation was considered in this study as a combined trait that may be controlled by many genes having complicated genetic actions. To validate the simulation results in this study, we built one other adaption model that included both major and minor genes. Four major genes were considered, each having an effect of five on adaptation, which is ten times the effect in the previous model. Each major gene was linked with a pVAC gene. For this genetic model, the probabilities of successful results for all selection schemes are shown in Table 6. These results are not exactly the same as the results in Table 5, but major observations remain unchanged. To give a few examples, PS based on HPLC resulted in higher probability of success than PS based on UPLC; scheme *MAS4* was most efficient based on the number of selected target lines; combined MAS for two or three pVAC genes and PS were as efficient as *MAS4*; doubling

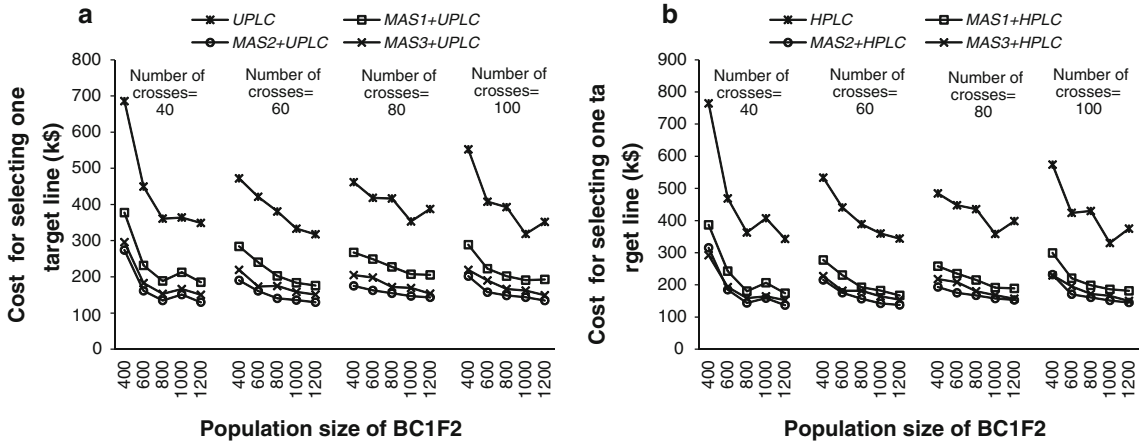

Fig. 2 Cost of selecting one target line for each selection scheme. The number of crosses was equal to 40, 60, 80, and 100, and the BC_1F_2 population size was equal to 400, 600, 800, 1,000, and 1,200. **a** UPLC was used to test the value of pVAC in phenotypic selection, **b** HPLC was used to test the value of pVAC in phenotypic selection. Results of scheme *MAS4* are almost identical to those of *MAS3 + PS*, and therefore are not shown

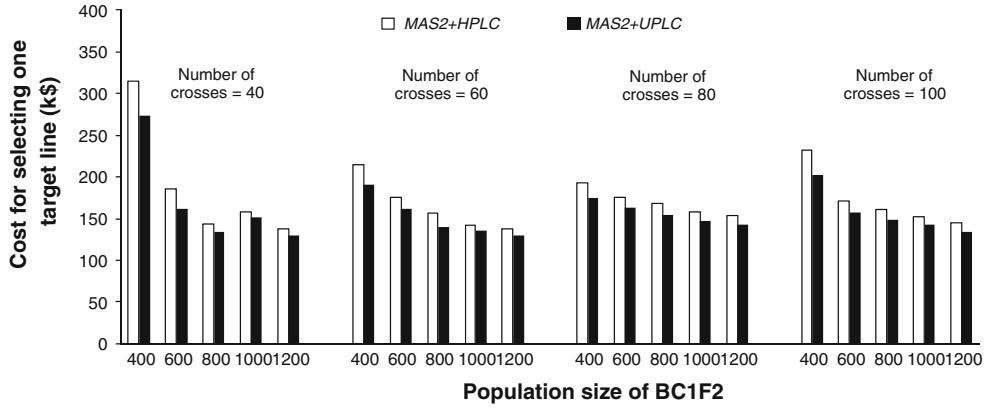

Fig. 3 Cost of selecting one target line when the two largest pVAC genes were selected by their associated markers and the remaining genetic variation on pVAC was selected by HPLC and UPLC. The number of crosses was equal to 40, 60, 80, and 100, and the BC_1F_2 population size was equal to 400, 600, 800, 1,000, and 1,200

the current breeding scale can assure a probability of success close to 90%; and the increase in number of crosses had similar consequence as the increase in BC_1F_2 population size under a fixed breeding scale (Table 6).

In the present study, adaptation in donor parents was assumed to be 60. However, donors with favorable pVAC may have lower levels of adaptation if they come from more ancient maize germplasm. Additional simulations were conducted assuming the adaptation of donor parents was only 40 (Table 7). As expected, probability of success for all selection schemes in Table 7 were lower than probability of success in Table 5, indicating that it is more difficult to achieve the breeding target when donors are less adapted. In this case, if the breeding scale is increased to number of crosses = 60 and BC_1F_2 population size = 1,200, or to number of crosses = 80 and BC_1F_2 population size = 800, the selection scheme can still be identified to make sure that the breeding target will be achieved with a probability of around 90% (Table 7). When the donors have an even lower adaptation, one may want to consider the two times backcrossing strategy (Wang et al. 2009a). Another option may be to adopt two cycles of the improvement strategy without significantly increasing the breeding scale. In the first cycle, the breeding target is to achieve the high level of pVAC (i.e., pVAC $\geq$ 11.5 μg/g) and improved adaptation (i.e., adaptation > 60). In the second cycle, the breeding target is to achieve pVAC $\geq$ 11.5 μg/g and adaptation $\geq$90.

Discussion

Computer simulation makes it possible to evaluate the efficiency of breeding strategies both genetically and economically. Under the current investment, available genetic materials and existing breeding method, how likely can a

Table 6 Probability of success (%) in a genetic model including large-effect adaptation genes, and linkage between major adaptation genes and pVAC genes

Number of crosses	Segregating population size	Selection scheme								
		UPLC	*HPLC*	*MAS4*	*MAS1 + UPLC*	*MAS2 + UPLC*	*MAS3 + UPLC*	*MAS1 + HPLC*	*MAS2 + HPLC*	*MAS3 + HPLC*
40[a]	400	7.9	13.3	52.1	27.7	55.8	59.5	44.6	56.5	60.5
	600[a]	13.9	22.8	52.7	41.0	65.0	65.2	62.6	66.3	67.5
	800	19.8	29.7	69.6	55.3	80.1	76.4	76.5	82.8	80.5
	1,000	18.1	29.7	79.6	59.0	86.1	86.2	83.0	88.8	86.5
	1,200	24.1	35.1	87.3	65.7	90.1	89.5	86.1	91.4	88.9
60	400	13.7	17.0	55.4	37.8	63.9	66.9	59.8	68.3	68.8
	600	17.2	22.8	74.8	55.3	82.5	84.5	79.9	87.4	85.8
	800	19.7	28.5	87.7	64.9	89.8	93.6	85.0	92.1	94.3
	1,000	25.4	35.3	92.4	68.5	93.5	92.8	90.1	94.4	93.1
	1,200	27.4	36.7	94.2	78.3	96.4	96.7	94.8	97.1	96.7
80	400	18.4	26.6	69.7	51.2	81.1	85.2	74.6	82.2	84.6
	600	22.8	33.3	88.6	62.7	90.3	91.5	84.9	90.8	92.2
	800	28.8	40.4	92.6	71.5	94.7	95.6	92.1	94.7	96.6
	1,000	34.7	46.4	95.0	78.5	97.0	97.4	96.4	97.5	97.2
	1,200	35.7	45.6	98.2	85.3	97.9	98.6	97.5	98.8	98.6
100	400	17.8	26.9	83.2	60.1	88.8	91.1	82.5	91.3	93.3
	600	27.2	39.5	92.2	71.6	94.4	94.3	90.3	96.1	95.6
	800	32.0	44.3	96.0	79.8	97.3	97.8	95.7	97.6	98.0
	1,000	43.3	56.3	98.7	86.2	99.2	99.3	97.7	98.4	99.4
	1,200	42.3	57.3	99.5	90.4	99.7	99.7	99.1	100.0	99.8

[a] Underlined values represent the number of crosses and segregating population size currently used in CIMMYT

breeding target be achieved? How much could the probability of success be increased if we increase the breeding effort and investment? These questions can be properly investigated by simulation. Though breeding scheme used for the micronutrient breeding activities in CIMMYT was applied as an example, other breeding methods can be simulated and then optimized by applied simulation tools QuLine and QuHybrid as well. Modified pedigree and bulk breeding strategy simulated in this study has been widely used in CIMMYT, not only by maize breeders but also by wheat breeders. This methodology is more cost-efficient than the traditional pedigree breeding and has proved to be very successful (Wang et al. 2003, 2009a). By applying bulk, we may not know which F_2, F_3 or F_4 individual derives which final fixed line, but parental lines deriving each fixed line are still known, which provides the most important information for the next cycle of breeding.

In previous simulation studies, we normally used genetic gain to compare different breeding methods/scenarios (Wang et al. 2003, 2004, 2007, 2009a, b). In this paper, we used probability of success as a criterion for the comparison of breeding methods. Results of this study showed that probability of success of current breeding scale was as large as 70% (Table 5), and can be much smaller when donor parents are less adapted (Table 7). Increasing the breeding scale can make sure there is a greater than 90% probability of achieving the maize breeding target on pVAC in one breeding cycle. In CIMMYT, two generations of maize can be grown in 1 year (Fig. 1). F_1 hybrids normally need 7–8 generations (depending on genetic diversity between the two parental lines) of self-pollination and selection (include line per se and testcross) so as to derive pure inbred lines which can be used to make new hybrids and/or to derive new inbreds. Thus, it takes 3.5–4 years to complete one breeding cycle, and to miss the target in one breeding cycle means the delivery may be delayed 3.5–4 years. Therefore, increasing the probability of success by increasing the breeding scale may become worthwhile when the breeding target has to be achieved with high probability in one breeding cycle.

Breeding scale can be increased by increasing the number of crosses and the segregating population size. The issue on the number of crosses versus segregating population size has been investigated by various authors, and conflicting results have been obtained. Yonezawa and Yamagata (1978) and Weber (1979) concluded that number

Table 7 Probability of success (%) when donors are less adapted, i.e. the mean adaptation in the donor population was 40

Number of crosses	Segregating population size	Selection scheme								
		UPLC	*HPLC*	*MAS4*	*MAS1 + UPLC*	*MAS2 + UPLC*	*MAS3 + UPLC*	*MAS1 + HPLC*	*MAS2 + HPLC*	*MAS3 + HPLC*
40[a]	400	16.7	30.8	33.6	18.9	33.4	30.8	30.9	35.6	31.4
	600[a]	26.6	50.4	50.2	36.5	57.4	47.7	54.2	58.5	49.4
	800	37.4	65.4	66.7	47.6	69.9	62.2	69.9	71.9	64.8
	1,000	37.2	68.4	75.3	50.9	71.6	69.3	70.9	75.6	74.3
	1,200	48.3	76.6	77.1	55.8	82.7	74.9	77.7	79.6	76.4
60	400	25.8	51.4	50.5	37.4	56.9	49.1	54.3	59.6	49.9
	600	38.2	68.2	74.8	46.8	74.6	68.9	71.3	75.4	69.6
	800	46.2	78.6	81.7	57.8	81.0	74.4	79.5	82.5	75.0
	1,000	53.9	85.2	86.2	64.2	84.9	83.6	87.1	86.4	84.5
	1,200	59.2	88.6	91.9	71.8	90.4	89.4	91.9	92.4	89.2
80	400	36.0	65.2	67.1	50.2	74.6	61.6	68.1	69.8	62.4
	600	46.2	78.6	77.9	59.2	79.7	76.0	80.1	82.4	75.8
	800	55.5	84.0	90.1	67.5	88.2	84.8	87.7	89.9	83.1
	1,000	65.8	91.9	91.6	72.6	93.3	92.2	91.7	95.0	91.6
	1,200	67.4	92.6	95.5	80.0	96.3	95.4	95.5	97.4	94.7
100	400	36.9	65.8	76.4	47.4	69.9	74.6	70.8	74.2	73.9
	600	55.0	85.5	86.2	68.9	85.7	83.9	86.3	89.4	84.9
	800	63.8	90.2	91.6	76.7	94.2	91.4	92.3	94.7	92.7
	1,000	77.6	97.1	97.4	82.0	96.8	95.9	95.4	97.6	96.4
	1,200	76.3	97.6	97.8	86.8	98.6	97.5	98.1	98.9	97.9

[a] Underlined values represent the number of crosses and segregating population size currently used in CIMMYT

of crosses rather than segregating population size should be maximized to reduce the risk of excluding superior genotypes. In other words, segregating population size should be equal to one, and number of crosses should be maximized. In contrast, Huehn (1996) found that when a finite amount of resources (2,000 recombinant inbreds) was available, the selection response was largest when an intermediate level of number of crosses was used, i.e., number of crosses = 50–100. In this study, we found that, when the number of crosses is greater than 40, different combinations of number of crosses and segregating population size had minor effect on breeding efficiency under a fixed resource. For example in Table 5, when number of crosses = 60 and BC_1F_2 population size = 800, probability of success was 68.2, 83.1, 96.3, 95.3, and 97.3% for *UPLC*, *MAS1 + UPLC*, *MAS2 + UPLC*, *MAS3 + UPLC*, and *MAS4*, respectively. In comparison, when number of crosses = 80 and segregating population size = 600, probability of success was 68.7, 78.5, 96.2, 95.4, and 97.1% for above-mentioned selection schemes. The difference between different combinations of number of crosses and segregating population size was minor. However, we have noted that fewer crosses with larger population size may result in greater variance in the final selected breeding population.

Two most effective and cost-efficient selection schemes were identified: (1) simultaneous MAS for all pVAC genes (which is purely hypothetical because more than four genes are known to influence pVAC, and because we currently have markers for only two genes in the pVAC pathway), and (2) a combination of marker-assisted phenotypic selection, where two major pVAC genes were selected by markers and the remaining genetic variation was then selected by UPLC screening (a strategy that can be implemented at present). Breeding programs with larger scales were more efficient in terms of genetic gain and economic cost. Two pVAC test methods were considered in this study. One method is precise but expensive and time-consuming, which cannot be applied on high-throughput analysis. The other method is faster and less expensive, but less precise, which allows the high-throughput analysis. However, we will not exclude the possibility that the precision of UPLC may be much improved as the advance of new techniques. If HPLC and UPLC have same precision (represented by the correlation between true pVAC and observed pVAC), we use the one that is less expensive to reduce the breeding cost.

As anticipated, the use of gene and marker associations identified from genetic studies can improve breeding

efficiency, and enhance genetic gain. In this study, *MAS2 + UPLC* was identified as being an efficient selection scheme, in which diagnostic markers were used to screen the two largest pVAC genes, followed by phenotypic selection on pVAC tested by UPLC. Phenotypic selection thus performed has the potential to select the other two pVAC genes, and perhaps any other unidentified pVAC genes as well (Wang et al. 2009b). It is also useful when markers are not completely linked with the targeted genes (Wang et al. 2007).

Marker screening has been implemented in CIMMYT's pVAC breeding program for selecting favorable alleles at loci *LcyE* and *crtRB1* from a number of donor inbred lines. Promising advanced lines may have up to 15 μg/g pro-vitamin A carotenoids in the grain. New maize hybrids, using lines containing the favorable *crtRB1* allele, are now being tested in the field (Pixley et al. 2010). Based on preliminary field data, the new hybrids promise to achieve pro-vitamin A concentrations higher than 10 μg/g. The cost of acquiring one marker data point was estimated at $5 based on CIMMYT's current genotyping system. We understand that the cost may be lower in commercial molecular marker laboratories. In addition, the use of gel-free electrophoresis and a seed DNA-based genotyping system could also reduce the genotyping cost (Gao et al. 2008). When more advanced and cheaper genotyping methods become available, MAS should become more cost-efficient.

We assume pVAC were controlled by four additive genes, based on previous genetic studies. However, QU-GENE can handle more complicated models, such as epistasis and genotype by environment interactions. When more genes and more complicated effects are involved, it can be expected that the relative gain of MAS compared with PS will be reduced, unless the trait heritability is extremely low. So the conclusion from pVAC cannot be extended to other traits, whose genetic models are not similar to the model of pVAC. However, the approach presented in this study could be used as a general way for quantifying probability of success and comparing different breeding schemes in other breeding programs.

Acknowledgments The authors wish to thank Dr. Kevin Pixley for his constructive comments and suggestions to a draft of the manuscript. This research was supported by the HarvestPlus Challenge Program of CGIAR. Development of the simulation tool QuHybrid was funded by the Generation Challenge Program (GCP) of CGIAR.

References

Bai L, Kim E, DellaPenna D, Brutnell TP (2009) Novel lycopene epsilon cyclase activities in maize revealed through perturbation of carotenoid biosynthesis. Plant J 59:588–599

Bernardo R (2002) Breeding for quantitative traits in plants. Stemma Press, Woodbury

Bouis HE, Welch RM (2010) Biofortification—a sustainable agricultural strategy for reducing micronutrient malnutrition in the global South. Crop Sci 50:S20–S32

Brenna OV, Bernardo N (2004) Application of near-infrared spectroscopy (NIRS) to the evaluation of carotenoids in maize. J Agric Food Chem 52:5577–5582

Buckner B, Kelson TL, Robertson DS (1990) Cloning of the *y1* locus of maize, a gene involved in the biosynthesis of carotenoids. Plant Cell 2:867–876

Chander S, Guo Y, Yang X, Zhang J, Lu X, Yan J, Song T, Rocheford T, Li J (2008a) Using molecular markers to identify two major loci controlling carotenoid contents in maize grain. Theor Appl Genet 116:223–233

Chander S, Meng Y, Zhang Y, Yan J, Li J (2008b) Comparison of nutritional trait variability in selected eighty-seven inbreds from Chinese maize (*Zea mays L.*) germplasm. J Agric Food Chem 56:6506–6511

Egesel CO, Wong JC, Lambert RJ, Rocheford TR (2003) Combining ability of maize inbreds for carotenoids and tocopherols. Crop Sci 43:818–823

Gao S, Martinez C, Skinner DJ, Krivanek AF, Crouch JH, Xu Y (2008) Development of a seed DNA-based genotyping system for marker-assisted selection in maize. Mol Breed 22:477–494

Hallauer AR (2007) History, contribution, and future of quantitative genetics in plant breeding: lessons from maize. Crop Sci 47:S5–S19

Harjes CE, Rocheford TR, Bai L, Brutnell TP, Kandianis CB, Sowinski SG, Stapleton AE, Vallabhaneni R, Williams M, Wurtzel ET, Yan J, Buckler ES (2008) Natural genetic variation in lycopene epsilon cyclase tapped for maize biofortification. Science 319:330–333

Huehn M (1996) Optimum number of crosses and progeny per cross in breeding self-fertilizing crops. I. General approach and first numerical results. Euphytica 91:365–374

Kurilich AC, Juvik JA (1999) Quantification of carotenoid and tocopherol antioxidants in *Zea mays*. J Agric Food Chem 47:1948–1955

Li Z, Matthews PD, Burr B, Wurtzel ET (1996) Cloning and characterization of a maize cDNA encoding phytoene desaturase, an enzyme of the carotenoid biosynthetic pathway. Plant Mol Biol 30:269–279

Li H, Ye G, Wang J (2007a) A modified algorithm for the improvement of composite interval mapping. Genetics 175:361–374

Li F, Murillo C, Wurtzel ET (2007b) Maize *Y9* encodes a product essential for 15-*cis* zeta-carotene isomerization. Plant Physiol 144:1181–1189

Menkir A, Liu W, White WS, Dixon BM, Rocheford T (2008) Carotenoid diversity in tropical-adapted yellow maize inbred lines. Food Chem 109:521–529

Pfeiffer WH, McClafferty B (2007) HarvestPlus: breeding crops for better nutrition. Crop Sci 47:S88–S105

Pixley K, Babu R, Yan J, Palacios-Rojas N (2010) Challenges, progress, & state of breeding for pro-vitamin A. First Global Conference on Biofortification: From Discovery to Delivery, 9–11 Nov. 2010. Washington, D.C.

Podlich DW, Cooper M (1998) QU-GENE: a platform for quantitative analysis of genetic models. Bioinformatics 14:632–653

Sun ZR, Gantt E, Cunningham FX (1996) Cloning and functional analysis of the β-carotene hydroxylase of *Arabidopsis thaliana*. J Biol Chem 271:24349–24352

Tallada JG, Palacios-Rojas N, Armstrong PR (2009) Prediction of maize seed attributes using a rapid single kernel near infrared instrument. J Cereal Sci 50:381–387

Tian L, DellaPenna D (2001) Characterization of a second carotenoid β-hydroxylase gene from *Arabidopsis* and its relationship to the LUT1 locus. Plant Mol Biol 47:379–388

Wang J, van Ginkel M, Podlich D, Ye G, Trethowan R, Pfeiffer W, DeLacy IH, Cooper M, Rajaram S (2003) Comparison of two breeding strategies by computer simulation. Crop Sci 43:1764–1773

Wang J, van Ginkel M, Trethowan R, Ye G, DeLacy I, Podlich D, Cooper M (2004) Simulating the effects of dominance and epistasis on selection response in the CIMMYT Wheat Breeding Program using QuCim. Crop Sci 44:2006–2018

Wang J, Chapman SC, Bonnett DB, Rebetzke GJ, Crouch J (2007) Application of population genetic theory and simulation models to efficiently pyramid multiple genes via marker-assisted selection. Crop Sci 47:580–588

Wang J, Singh RP, Braun HJ, Pfeiffer WH (2009a) Investigating the efficiency of the single backcrossing breeding strategy through computer simulation. Theor Appl Genet 118:683–694

Wang J, Chapman SC, Bonnett DG, Rebetzke GJ (2009b) Simultaneous selection of major and minor genes: use of QTL to increase selection efficiency of coleoptile length of wheat (*Triticum aestivum* L.). Theor Appl Genet 119:65–74

Weber WE (1979) Number and size of cross progenies from a constant total number of plants manageable in a breeding program. Euphytica 28:453–456

Weber EJ (1987) Carotenoids and tocols of corn grain determined by HPLC. J Am Oil Chem Soc 64:1129–1134

Wong JC, Lambert RJ, Wurtzel ET, Rocheford TR (2004) QTL and candidate genes phytoene synthase and ζ-carotene desaturase associated with the accumulation of carotenoids in maize. Theor Appl Genet 108:349–359

Wurtzel E (2010) Maize: a model cereal crop for developing strategies to solve global vitamin A deficiency. First Global Conference on Biofortification: From Discovery to Delivery, 9–11 Nov. 2010. Washington, D.C.

Yan J, Kandianis CB, Harjes CE, Bai L, Kim EH, Yang X, Skinner DJ, Fu Z, Mitchell S, Li Q, Salas-Fernandez MG, Zaharieva M, Babu R, Fu Y, Palacios N, Li J, DellaPenna D, Brutnell T, Buckler ES, Warburton ML, Rocheford T (2010) Rare genetic variation at *Zea mays crtRB1* increases β-carotene in maize grain. Nat Genet 42:322–327

Yonezawa K, Yamagata H (1978) On the number and size of cross combinations in a breeding programme of self-fertilizing crops. Euphytica 27:113–116