

Index selection on seed traits under direct, cytoplasmic and maternal effects in multiple environments

Wenying Zhang^{a, b}, Haiming Xu^a, Jun Zhu^{a, *}

^a Institute of Bioinformatics, College of Agriculture and Biotechnology, Zhejiang University, Hangzhou 310029, China

^b Institute of Plant Genetics and Breeding, College of Agriculture, Yangtze University, Jingzhou 434025, China

Received for publication 18 June 2008; revised 25 September 2008; accepted 18 November 2008

Abstract

Crop seeds are important sources of protein, oil, and carbohydrates for food, animal feeds, and industrial products. Recently, much attention has been paid to quality and functional properties of crop seeds. However, seed traits possess some distinct genetic characteristics in comparison with plant traits, which increase the difficulty of genetically improving these traits. In this study, diallel analysis for seed models with genotype by environment interaction (*GE*) effect was applied to estimate the variance-covariance components of seed traits. Mixed linear model approaches were used to estimate the genetic covariances between pair-wise seed and plant traits. The breeding values (BV) were divided into two categories for the seed models. The first category of BV was defined as the combination of direct additive, cytoplasmic, and maternal additive effects, which should be utilized for selecting stable cultivars over multi-environments. The three genetic effects, together with their *GE* interaction, were included in the second category of BV for selecting special lines to be grown in specific ecosystems. Accordingly, two types of selection indices for seed traits, i.e., general selection index and interaction selection index, were developed and constructed on the first and the second category BV, respectively. These proposed selection indices can be applied to solve the difficult task of simultaneously improving multiple seed traits in various environments. Data of crop seeds with regard to four seed traits and four yield traits based on the modified diallel crosses in Upland cotton (*Gossypium hirsutum* L.) were used as an example for demonstrating the proposed methodology.

Keywords: index selection; seed trait; mixed linear model approach; genotype by environment interaction

Introduction

Artificial selection is usually performed for multiple traits in order to select the best individual. In breeding practice, several strategies, such as tandem selection, independent culling, and index selection, can be applied to achieve this target. Theoretically, the selection index is never inferior to the other two methods (Young, 1961). Selection index theory was firstly developed by Smith (1936) in plants and Hazel (1943) in animals, which is

generally called Smith-Hazel index now. Usually, the traits examined in the selection indices are controlled by nuclear genome (Geidel et al., 2000; Jannink et al., 2000), while extranuclear effects were seldom considered in this process. Also, the breeding value in the Smith-Hazel index is simply considered as additive effect.

Crop seeds are important sources of oil, protein, and carbohydrate for food, animal feeds or industrial products. The value of crop seeds mainly depends on the content and quality of seed components. Diploid embryos and triploid endosperms are the major storage organs of seeds. The specialness of growth and development of crop seeds increase the difficulty of understanding the genetic complexity of

* Corresponding author. Tel: +86-571-8697 1731.
E-mail address: jzhu@zju.edu.cn

seed traits. Accordingly, the methods for improving multiple seed traits, such as index selection, were seldom addressed.

Maternal effect is an important component of extranuclear effect in mammals (Pakkasmaa et al., 2003) and plants (Lemontey et al., 2000). Mitochondria exist in both plant and animal cells, but chloroplasts only exist in plant cells. The genome of these organelles can greatly contribute to some special traits (Snowder et al., 2004). In plant seeds, cytoplasmic and maternal effects are the major sources of extranuclear effects (van Sanford and Matzinger, 1982). To evaluate the extranuclear genetic effects, Beavis et al. (1987) proposed a quantitative genetic model for traits influenced by cytoplasmic genes that were considered as the only source of maternal effects. Foolad and Jones (1992) presented a model considered all sources, i.e., embryo, endosperm, testa/perisperm and cytoplasm, of maternal contributions to quantitative seed-related traits. With more details, Zhu and Weir (1994a, 1994b) developed seed models for analyzing cytoplasmic and maternal effects in diploid seeds and triploid endosperms. Furthermore, Zhu (1996) extended these models by including genotype \times environment (GE) interaction effects.

van Vleck (1970) extended index selection on multiple traits by employing the genetic models with maternal effects reported by Willham (1963), which included only the direct and maternal genetic components but not other complicated genetic effects. However, the methods for estimating the genetic variance-covariance of indirect additive and maternal additive genetic effects were not illuminated by van Vleck. Moreover, cytoplasmic effects were not included in the study of van Vleck (1970).

Multi-environment selection is a pendent problem puzzling plant and animal breeders. Falconer (1952) suggested that genotypic performances in two environments could be treated as two genetically correlated characters. Following Falconer's theory, James (1961) developed a statistical method of index selection for animal traits in two environments. Later, van Sanford et al. (1993) performed index selection on genetic correlations of cultivar performance based on the genetic correlations between primary locations and the target environments. Atlin et al. (2000) studied the selection response of genotypic value in a large region and constituent sub-regions. Although selection indices were used in these studies, the essential issue was single-trait improvement in varied environments, which was not indeed exerting the predominance of index selection in the aspect of multi-trait improvement.

In this study, a novel method of index selection on seed traits is proposed based on the estimation of variable ge-

netic parameters. The method integrates the problem of seed trait selection and multiple environment selection into a whole procedure, which facilitates improvement of seed quality in various environments. Data of crop seeds with regard to four seed traits and four yield traits based on the modified diallel crosses in Upland cotton (*Gossypium hirsutum* L.) were used as an example for demonstrating the proposed methodology.

Materials and methods

Genetic model

Zhu and Weir (1994a, 1994b) investigated the cytoplasmic and maternal effects for plant seeds containing diploid embryos or triploid endosperms. They proposed genetic models based on the Willham (1963) method by including seed direct genetic effect (G_o), cytoplasmic effect (C), and maternal nuclear genetic effect (G_m). Zhu (1996) further extended these seed models by including environmental effect (E) and genotype by environment interaction effect (GE). In general, the mean observation (y_{hijkl}) in the l -th block of the k -th genetic entry from inbred lines i and j in the h -th environment can be expressed as

$$y_{hijkl} = \mu + E_h + G_{ijk} + GE_{hijk} + B_{l(h)} + e_{hijkl} \quad (1)$$

where, μ and E_h refer to the population mean value and the environmental effect, respectively, fixed; G_{ijk} is the total genetic main effect, random; and GE_{hijk} is the total GE interaction effect, random; $B_{l(h)}$ and e_{hijkl} represent the block effect and the random error component, respectively, random. The blocks are set to eliminate the maternal common environmental effect by certain experimental designs, e.g., randomized block design.

The total main genetic effect G_{ijk} and the total GE interaction effect GE_{hijk} in formula (1) can be further partitioned following Zhu and Weir (1994a, 1994b) and Zhu (1996) under the assumptions: (1) no paternal effects; (2) no maternal interaction effects; (3) no epistatic effects; (4) constant inheritance of cytoplasmic genes through maternal lines; and (5) no covariance between genetic effects of seed traits and that of plant traits.

The phenotypic variance (V_p) for equation (1) can be partitioned as

$$\begin{aligned} V_p &= V_G + V_{GE} + V_e \\ &= (V_A + V_D + V_C + V_{Am} + V_{Dm}) \\ &\quad + (V_{AE} + V_{DE} + V_{CE} + V_{AmE} + V_{DmE}) + V_e \end{aligned}$$

where, V_G = variance of genetic main effects, V_{GE} = variance of GE interaction effects, V_e = variance of residual effects, V_A = variance of direct additive effects, V_D = variance of direct dominance effects, V_C = variance of cytoplasmic effects, V_{Am} = variance of maternal additive effects, V_{Dm} = variance of maternal dominance effects, V_{AE} = variance of additive by environment interaction effects, V_{CE} = variance of cytoplasm by environment interaction effects, V_{DE} = variance of dominance by environment interaction effects, V_{AmE} = variance of maternal additive by environment interaction effects, and V_{DmE} = variance of maternal dominance by environment interaction effects.

Phenotypic covariance (C_P) between two seed traits can be partitioned and defined as

$$\begin{aligned} C_P &= C_G + C_{GE} + C_e \\ &= (C_A + C_D + C_C + C_{Am} + C_{Dm}) \\ &\quad + (C_{AE} + C_{DE} + C_{CE} + C_{AmE} + C_{DmE}) + C_e \end{aligned}$$

The above genetic models can be employed to estimate the genetic parameters of seed traits. For agronomic traits (called plant traits), which are measured on maternal plants, the total genetic value $G + GE$ can be simply denoted as

$$G + GE = (Am + Dm + C) + (AmE + DmE + CE)$$

where, Am = additive effects, Dm = dominance effects, and C = cytoplasmic effects, respectively; AmE , DmE , and CE are their GE interaction effects, respectively.

The phenotypic variance and covariance of plant traits can be partitioned as follows

$$\begin{aligned} V_P &= V_G + V_{GE} \\ &= (V_C + V_{Am} + V_{Dm}) + (V_{CE} + V_{AmE} + V_{DmE}) + V_e \\ C_P &= C_G + C_{GE} \\ &= (C_C + C_{Am} + C_{Dm}) + (C_{CE} + C_{AmE} + C_{DmE}) + C_e \end{aligned}$$

Variances of seed trait were estimated by MINQUE (0/1) (Zhu and Weir, 1994a, 1994b), and covariances between seed traits and plant traits were estimated following Zhu (1993). These variances and covariances were utilized in the following calculation for selection index. The Jack-knife technique was applied by taking the sampling means of genetic entries for estimating the standard errors of estimated variances and covariances.

Proposed selection index approach

The Smith-Hazel index includes two indices, phenotypic index I and genotypic index H . They can be expressed as

$$I = \sum_{i=1}^m b_i x_i = b^T x,$$

and

$$H = \sum_{j=1}^n w_j g_j = w^T g$$

where, x is the vector of m known phenotypic values x_i , g is the vector of n known genetic values g_j , w is the vector of n known relative economic weights w_j , and b is the vector of m index coefficients to be estimated. The index coefficients were estimated by the method of least-square (LS), as

$$b = P^{-1}Cw$$

where, P is the matrix of phenotypic variance-covariance of x_i , C is the matrix containing all covariances between x_i and g_i . Here, the genetic value (g_j) can be considered as breeding value (B).

In construction of selection index on seed traits, two types of characters may be included in the information traits *via* seed traits and plant traits. For seed traits, the general breeding value of a seed trait can be written as

$$B_{G(S)} = A_{(S)} + C_{(S)} + Am_{(S)}$$

where, $A_{(S)}$, $C_{(S)}$, and $Am_{(S)}$ are the direct additive effect, cytoplasmic and maternal additive genetic effects of seed trait, respectively. The interaction breeding value of a seed trait can be written as

$$B_{GE(S)} = AE_{(S)} + CE_{(S)} + AmE_{(S)}$$

where, $AE_{(S)}$, $CE_{(S)}$, and $AmE_{(S)}$ are the GE interaction effects of the seed traits.

For plant traits, the general breeding value ($B_{G(P)}$) available for multiple environments is

$$B_{G(P)} = C_{(P)} + Am_{(P)}$$

The GE interaction breeding value that can be only appreciable for specific environments is

$$B_{GE(P)} = CE_{(P)} + AmE_{(P)}$$

The total breeding value of a plant trait is

$$B_{(P)} = B_{G(P)} + B_{GE(P)}$$

The covariance between the phenotypic value of seed trait i ($x_{i(S)}$) and the general breeding value of seed trait j ($B_{G(S)j}$) is

$$C(x_{i(S)}, B_{G(S)j}) = C(G_{i(S)}, B_{G(S)j}) + C(\varepsilon_{i(S)}, B_{G(S)j})$$

where, $x_{i(S)}$ and $G_{i(S)}$ are the phenotypic and total genetic main effects of the seed trait, respectively; $\varepsilon_{i(S)}$ is the residual error. It is usually assumed that there is no covariance between $\varepsilon_{i(S)}$ and $B_{G(S)j}$, yielding

$$\begin{aligned} C(x_{i(S)}, B_{G(S)j}) &= C(G_{i(S)}, B_{G(S)j}) \\ &= C((A_{i(S)} + D_{i(S)} + C_{i(S)} + Am_{i(S)} \\ &\quad + Dm_{i(S)}), (A_{j(S)} + C_{j(S)} + Am_{j(S)})) \\ &= C(A_{i(S)}, A_{j(S)}) + C(A_{i(S)}, C_{j(S)}) \\ &\quad + C(A_{i(S)}, Am_{i(S)}) + C(D_{i(S)}, C_{j(S)}) \end{aligned}$$

$$\begin{aligned}
&+ C(C_{i(S)}, A_{j(S)}) + C(C_{i(S)}, C_{j(S)}) \\
&+ C(C_{i(S)}, Am_{j(S)}) \\
&+ C(Am_{i(S)}, A_{j(S)}) + C(Am_{i(S)}, C_{j(S)}) \\
&+ C(Am_{i(S)}, Am_{j(S)}) + C(Dm_{i(S)}, C_{j(S)})
\end{aligned}$$

For simplicity, the covariance between the direct additive genetic effect of seed trait i and maternal effect of seed trait j and the covariance between cytoplasmic effect and the maternal genetic effects or direct genetic effects are ignored. Then, covariance between the phenotypic values of seed trait i and the general breeding value of seed trait j can be written as

$$\begin{aligned}
C(x_{i(S)}, B_{Gj(S)}) &= C(A_{i(S)}, A_{j(S)}) + C(C_{i(S)}, C_{j(S)}) \\
&+ C(Am_{j(S)}, Am_{j(S)})
\end{aligned}$$

Also, the covariance between the phenotypic values of seed trait i and the GE interaction breeding value of seed trait j can be expressed as

$$\begin{aligned}
C(x_{i(S)}, B_{GEj(S)}) &= C(AE_{i(S)}, AE_{j(S)}) + C(CE_{i(S)}, CE_{j(S)}) \\
&+ C(AmE_{i(S)}, AmE_{j(S)})
\end{aligned}$$

Similarly, the covariance between the phenotypic value of plant trait i ($x_{i(P)}$) and the genotypic value of seed trait j ($B_{j(S)}$) can be written as

$$C(x_{i(P)}, B_{j(S)}) = C(C_{i(P)}, C_{j(S)}) + C(Am_{i(P)}, Am_{j(S)})$$

where, $C_{i(P)}$ and $Am_{i(P)}$ are the cytoplasmic and additive genetic effect of plant trait i , respectively; $C_{j(S)}$ and $Am_{j(S)}$ are the cytoplasmic and maternal additive genetic effect of seed trait j , respectively. Then, the covariance between the phenotypic value of plant trait i ($x_{i(P)}$) and the general breeding value of plant trait j ($B_{Gj(P)}$) is

$$C(x_{i(P)}, B_{Gj(P)}) = C(C_{i(P)}, C_{j(P)}) + C(Am_{i(P)}, Am_{j(P)})$$

And, the covariance between the phenotypic value of plant traits i ($x_{i(P)}$) and the GE interaction breeding value of plant trait j ($B_{GEj(P)}$) can be written as

$$\begin{aligned}
C(x_{i(P)}, B_{GEj(P)}) &= C(CE_{i(P)}, CE_{j(P)}) + \\
&C(AmE_{i(P)}, AmE_{j(P)})
\end{aligned}$$

By adopting the above methods and the estimated results regarding variance and covariance components, two types of indices, general selection index (I_G) and interaction selection index (I_I) can be constructed: (a) general index, constructed on the basis of estimation related to general breeding value can be employed in multiple environments, and (b) interaction index, constructed on the basis of information involving both general breeding value and GE interaction breeding value can be confined to specific situation.

The predicted response (R) of the index selection can be calculated following Lin (1978) as

$$R = i \cdot ((b^T Pb)/(w^T Gw))^{1/2}$$

where, i is the selection intensity, G is the genetic variance-covariance matrix of breeding value.

The response for single trait j is

$$R_j = i \cdot ((b^T c_j)/(b^T P b) V_{(B_j)})^{1/2}$$

where, c_j is the j -th column of covariance matrix C .

Worked example

To illustrate the methods proposed in the present study, we analyzed the data on Upland cotton (*Gossypium hirsutum* L.), and constructed selection indices from various aspects. The experimental germplasm comprised of five varieties or strains, 1) Zhongmiansuo 7 (ZMS 7), 2) HG-12, 3) PD0111, 4) PD0458, and 5) GL-5. Genotype 1) is from China and others are from the USA. In 2003, the five parents were crossed in a diallel with reciprocal crosses. In 2004 and 2005, a modified diallel mating system consisting of parents, F_1 , F_2 , BC_1 , BC_2 , RBC_1 , and RBC_2 generations was conducted. Backcrossing population BC_1 or BC_2 were made using F_1 as the female and P_1 or P_2 as the male; reciprocal backcross population RBC_1 or RBC_2 were made using P_1 or P_2 as the female and F_1 as the male. The experiment was conducted in a randomized complete block design (RCBD) with three replications at the experimental station of Zhejiang University, Hangzhou, China.

Four seed traits, oil content (OC, %), protein content (PC, %), oil index (OID = oil weight per 100 kernels, g), and protein index (PID = protein weight per 100 kernels, g) were measured on seed kernel sampled randomly at each plot. In addition, four yield traits, lint yield (LY, g/plant), boll number (BN), boll weight (BW, g), and lint percentage (LP, %) were also investigated.

Results

For the construction of selection indices, four types of information traits were taken into account: 1) two seed traits, OC and PC, as objective traits, and another two seed traits, OID and PID, as the secondary traits; 2) two seed traits, OC and PC, together with LY, as objective traits, no secondary traits; 3) two seed traits, OC and PC, as objective traits, and three yield component traits, BN, BW, and LP as the secondary traits; 4) OC and PC as objective traits but not acting as information traits, and three yield component traits, BN, BW, and LP as the secondary traits. Note that in the former three cases, objective traits were

part of the information traits.

The selection fraction is fixed as $\alpha = 0.05$, with $i = 2.063$ as the selection intensity. For simplicity, the economic weights of all the information traits are set to 1 here. The data were analyzed on an IBM computer by programs written in the C language. The whole analysis procedure was divided into four parts.

The first part of analysis comprised of the estimation of genetic variance components of the four seed traits (Table 1).

Table 1
Estimates of variance components for four seed traits in Upland cotton

| Parameter | Oil % | Protein % | Oil index | Protein index |
|-----------|--------|-----------|-----------|---------------|
| V_A | 0.427* | 0.569* | 0.019* | 0.026* |
| V_D | 0.111 | 0.225 | 0.000 | 0.000 |
| V_C | 0.000 | 0.941* | 0.000 | 0.008 |
| V_{Am} | 1.278* | 0.152* | 0.036* | 0.085* |
| V_{Dm} | 0.000 | 0.064* | 0.000 | 0.000 |
| V_{AE} | 0.000 | 0.457* | 0.000 | 0.001 |
| V_{DE} | 0.000 | 0.000 | 0.007* | 0.002* |
| V_{CE} | 0.439* | 0.819* | 0.028 | 0.052 |
| V_{AmE} | 0.000 | 0.000 | 0.000 | 0.000 |
| V_{DmE} | 1.098* | 0.042** | 0.055* | 0.030* |
| V_e | 1.170 | 0.460 | 0.051 | 0.063 |
| V_P | 4.523 | 3.729 | 0.196 | 0.267 |

* and ** are significant at the level of $P < 0.05$ and $P < 0.01$, respectively.

Table 2
Phenotypic and genetic variance of seed traits and yield traits in Upland cotton

| Parameter | Oil % | Protein % | Oil index | Protein index | Boll no. | Boll weight | Lint % | Lint yield |
|--------------------|-------|-----------|-----------|---------------|----------|-------------|--------|------------|
| V_{BG} | 1.706 | 1.662 | 0.055 | 0.119 | 3.514 | 0.052 | 0.587 | 82.558 |
| $V_{BG} + V_{BGE}$ | 2.145 | 2.938 | 0.083 | 0.171 | 4.486 | 0.171 | 1.094 | 101.851 |
| V_P | 4.524 | 3.728 | 0.196 | 0.266 | 7.258 | 0.290 | 2.030 | 180.579 |

V_{BG} is the variance of breeding value for objective trait; V_{BGE} is the variance of genotype \times environment interaction breeding value for objective trait; V_P is the phenotypic variance of information trait.

Table 3
Phenotypic and genetic covariance between seed traits and between yield traits in Upland cotton

| Parameter | Oil % & protein % | Oil % & oil index | Oil % & protein index | Protein % & oil index | Protein % & protein index | Oil index & protein index | Boll no. & boll weight | Boll no. & lint % | Boll no. & lint yield | Boll weight & lint % | Boll weight & lint yield | Lint % & lint yield |
|------------------------|-------------------|-------------------|-----------------------|-----------------------|---------------------------|---------------------------|------------------------|-------------------|-----------------------|----------------------|--------------------------|---------------------|
| C_{B_e} | -0.970 | -0.269 | -1.580 | 2.613 | 3.285 | 0.581 | -0.143 | 3.686 | 19.049 | 0.945 | -0.115 | 12.906 |
| $C_{B_e} + C_{B_{GE}}$ | -1.583 | -0.909 | -0.023 | 7.204 | 8.669 | 0.610 | -1.770 | 56.571 | 89.410 | 29.878 | 30.978 | 224.513 |
| C_P | -5.429 | 0.351 | -3.117 | 1.727 | 3.591 | 1.121 | -2.721 | 28.193 | 68.575 | 16.626 | 15.889 | 106.263 |

C_{B_e} is the covariance between the phenotypic values in the phenotypic index I and the general genotypic values in the aggregate genotype H in the selection index; $C_{B_{GE}}$ is the covariance between the phenotypic values in the phenotypic index I and the genotype \times environment interaction values in the aggregate genotype H ; C_P is the phenotypic covariance between the information traits.

All of the four seed traits exhibited significant direct additive and maternal additive variances. Very high and significant cytoplasmic variance for PC was observed, which accounts for 25.23% of the total phenotypic variation. In contrast, PID exhibited a small (2.99%) and insignificant cytoplasmic variance. Cytoplasmic variance was not found for OC and OID. On the other hand, OC and PC exhibited significant variances of cytoplasmic interaction with environment, which was not found in OID and PID. The phenotypic variances and covariances, which are the prerequisite for construction of matrix P , are presented in Table 2. The phenotypic and genetic covariances to be used in matrix C are listed in Table 3, which can be further used in the construction of selection index.

The second part of analysis is based on the estimation of phenotypic and genetic variances (Table 2) and phenotypic and genetic covariances (Table 3) for lint yield and its three component traits. Estimation of phenotypic and genetic covariance between seed traits and yield traits are consequently conducted (Table 4). These estimated results are the basic requirements for construction of matrix P , G , and C in the present study for selection indices.

In the third step, four types of combination sets of information traits were considered to estimate the general selection index to be used in various environments. In the first case, OC and PC were considered to be objective traits, while OID and PID were set to be the secondary

Table 4
Phenotypic and genetic covariance between seed traits and yield traits in Upland cotton

| Trait | Parameter | Boll no. | Boll weight | Lint % | Lint yield |
|---------------|------------------------|----------|-------------|--------|------------|
| Oil % | C_{B_g} | 2.330 | -0.327 | 0.209 | 6.521 |
| | $C_{B_g} + C_{B_{ge}}$ | 1.621 | -0.225 | 0.151 | 10.589 |
| | C_P | 1.965 | -0.149 | 0.156 | 12.317 |
| Protein % | C_{B_g} | -0.990 | 0.447 | -0.657 | -14.940 |
| | $C_{B_g} + C_{B_{ge}}$ | -0.192 | 0.678 | -0.470 | -17.543 |
| | C_P | -4.298 | 0.606 | -0.628 | -18.441 |
| Oil index | C_{B_g} | -0.387 | 0.058 | 0.004 | -1.717 |
| | $C_{B_g} + C_{B_{ge}}$ | -0.673 | 0.088 | -0.095 | -1.170 |
| | C_P | -0.574 | 0.099 | -0.095 | -0.731 |
| Protein index | C_{B_g} | -1.004 | 0.133 | -0.097 | -3.875 |
| | $C_{B_g} + C_{B_{ge}}$ | -1.266 | 0.176 | -0.172 | -3.749 |
| | C_P | -1.197 | 0.178 | -0.191 | -3.473 |

C_{B_g} is the covariance between the phenotypic values in the phenotypic index I and the general genotypic values in the aggregate genotype H in the selection index; $C_{B_{ge}}$ is the covariance between the phenotypic values in the phenotypic index I and the genotype \times environment interaction values in the aggregate genotype H in the selection index; C_P is the phenotypic covariance between the information traits.

traits. Because all the four traits were seed traits, it was easier to estimate their phenotypic and genetic variances and covariances. The general selection index was $I_{G(i)} = 0.384x_{OC} + 0.432x_{PC} + 1.274x_{OID} - 0.288x_{PID}$. The predicted response was 3.028, and the single responses for OC and PC were 0.313 and 2.489, respectively (Table 5). High yield and good quality are the basic requirements for plant and animal breeding. Thus, in the second case, LY was considered as an objective trait. The genetic correlation between OC and LY was positive (Table 4), whereas the genetic correlation between PC and LY was negative, which was consistent with the results of Turner et al. (1976). As a result, the single response for OC and LY was positive, namely, 1.579 and 1.184, respectively. The single response for PC was negative (-1.840). The index was $I_{G(ii)} = 2.244x_{OC} + 1.454x_{PC} + 0.406x_{LY}$. In the third case, three yield component traits, BN, BW, and LP served as the secondary traits to provide genetic information for selection for OC and PC. The predicted response was 1.492. The index in this case was $I_{G(iii)} = 3.322x_{OC} + 1.666x_{PC} - 2.388x_{BN} + 4.018x_{BW} + 0.297x_{LP}$. In the fourth case, only three yield component traits acted as information traits. The total predicted response was 0.409, which was smaller than that in the previous three cases (Table 5). The index was $I_{G(iv)} = 0.060x_{BN} - 0.131x_{BW} + 0.019x_{LP}$.

Table 5
Index coefficients and predicted responses of the index selection in different cases

| Index | Case | Parameter | Oil % | Protein % | Lint yield | Predicted response |
|-------|------|-----------|--------|-----------|------------|--------------------|
| I_G | i | R_j | 0.313 | 2.489 | - | - |
| | | R | - | - | - | 3.028 |
| | ii | R_j | 1.579 | -1.840 | 1.184 | - |
| | | R | - | - | - | 1.275 |
| | iii | R_j | -5.057 | 6.506 | - | - |
| | | R | - | - | - | 1.492 |
| | iv | R_j | 1.067 | -1.705 | - | - |
| | | R | - | - | - | 0.409 |
| I_I | i | R_j | -2.525 | 5.394 | - | - |
| | | R | - | - | - | 4.007 |
| | ii | R_j | 1.484 | -1.524 | 1.564 | - |
| | | R | - | - | - | 1.619 |
| | iii | R_j | -0.289 | 3.344 | - | - |
| | | R | - | - | - | 3.834 |
| | iv | R_j | 4.314 | -6.598 | - | - |
| | | R | - | - | - | 0.044 |

I_G is the general index for general environments, I_I is the interaction index for specific environment, b is the index coefficient in the selection index, R is the predicted response of aggregate genotypes H , R_j is the predicted response in the j -th objective trait due to selection on the phenotypic index I . Four types of information traits are taken into account in the construction of selection indices: (i) two seed traits, oil content and Protein content as objective traits, and two other seed traits, oil index and protein index as secondary traits; (ii) two seed traits, oil content and protein content, together with lint yield, as objective traits, no secondary traits; (iii) two seed traits, oil content and protein content as objective traits, and three yield component traits, boll number, boll weight, and lint percentage as secondary traits; and (iv) oil content and protein content as objective traits but not acting as information traits, and three yield component traits, boll number, boll weight, and lint percentage as the secondary traits.

In the fourth step, the interaction selection indices were calculated in a similar way and are listed in Table 5.

Discussion

Evidences from molecular biology (Gutierrez-Marcos et al., 2004) and statistical genetics (Zhang et al., 2004) have proved that genes or quantitative trait loci (QTLs) in female parents can influence some of the developmental processes on the embryos and endosperms. van Vleck (1970) treated direct and maternal additive effects as the main components of aggregate genetic value in animal index selection but cytoplasmic effect was ignored. Previ-

ous reports (Rajcan et al., 2002; Allen, 2005) have indicated that plant cytoplasmic genomes played an important role in the inheritance of seed quality characteristics. For seed traits in plant, direct additive effect, cytoplasmic effect, and maternal additive effect can be transmitted to the progenies and used by the breeders and growers to achieve the improved quality of seeds. Consequently, the concept of breeding value was expanded from the part of additive effect for the agronomic traits in the past to direct additive, cytoplasmic, and maternal additive effects for the seed traits in the present study. This approach can be easily applied in animal breeding experiment following Zhu and Weir (1994b).

Seed quality traits are quantitatively inherited traits. It is difficult, to some degree, to estimate genetic components for seed traits because they are controlled and influenced not only by the seed genome but also by the maternal plant genome. The genetic models proposed by Zhu and Weir (1994a, 1994b) facilitate effective estimation of the direct, maternal, and cytoplasmic effects. These approaches were successfully employed in some crops (Shi et al., 2003; Lin et al., 2005). In applied breeding experiments, modified diallel crosses consisting of parents (a set of completely inbred lines), F_1 s, reciprocal F_1 s, and BC_1 are essential for genetic analysis of animals and outcrossing plants (Zhu and Weir, 1994a). For self-pollinating species, parents, F_1 s and F_2 s are suitable because the seeds of F_1 plants belong to the F_2 generation. However, in breeding programs, selection is usually continued in subsequent generations of F_2 . In this situation, it is impractical for breeders to adopt the modified diallel to estimate the genetic parameters for seed traits. For instance, Gebre-Mariam and Larter (1996) studied genetic response to index selection for grain yield, kernel weight and percentage of protein in four wheat crosses in F_3 and F_4 populations in two different environments. The method of analysis of variance (ANOVA) was used to estimate the variance and covariance components. Total genetic effect was considered as “breeding value” in the calculation of index coefficients. Apparently, estimation of genetic parameters is not consistent with the principle of selection index. Such situations may be a compromise between theoretical rules and practical application.

The method of minimum norm quadratic unbiased estimation (MINQUE) and restricted maximum likelihood (REML) are generally used to estimate variances and covariances in statistical genetic analysis. However, MINQUE and REML require that pairwise traits to be estimated should have identical statistical model and equal design matrices. In addition, the REML method is appro-

priate for estimating asymptotic sampling variance or covariance components (Rao and Kleffe, 1988). Seed traits and plant traits have unequal design matrices. When both seed traits and plant traits are considered in the same selection index, the unbiased estimates of variance and covariance components can be obtained following Zhu (1993).

The relative economic weight of a trait was generally defined as net increase in profit of the production system for each unit of trait improvement. Accordingly, under most situations, economic factors, like price and profit, are directly or indirectly taken as criteria to determine the relative economic weights. Also, growth model was applied to derive the economic weights (Hermesch et al., 2003). Here, for simplicity, equal relative economic weights for all the traits were set to demonstrate the application of the proposed methodology for index selection under direct, cytoplasmic, and maternal effects. In the breeding practice, breeders can derive the relative economic weights according to different methods (Wricke and Weber, 1986).

The construction of selection index, in which trait combination is reasonable to be adopted, is a puzzling problem. Sivanadian and Smith (1997) indicated that only the traits with high relative products (wh^2) of the standard economic value (w) and heritability (h^2) could evoke large extra response. Adding a trait with lower wh^2 will result in less extra response. In the current example, index excluding objective traits (case *iv*) resulted in less response than index including objective traits (case *iii*) (Table 5). Thus, it can be seen that the objective traits are essential in the selection index. In addition, the aspects such as increases in measurement costs should be included in determination of economic weights and the selection programs. On the other hand, all the traits used in the selection index must be available at the same time. If there are some traits that can only be assessed during the harvest period, the selection procedures are correspondingly carried out at a later stage. This limits the flexibility of the selection procedures, to some degree.

It has been well established that there are negative correlations between seed oil and protein concentration (Panthee et al., 2005) and between yield and grain protein content (Chung et al., 2003). These inverse relationships between economically important traits increase the difficulty of improving all the traits simultaneously. Index selection is an appropriate method to realize this goal. In breeding practice, measurements of the mixture sample from an experimental plot are taken as phenotypic values, which

are applied to estimate the genetic parameters, such as variances and covariances. When the index coefficients are obtained, the performance of some individuals, which have been screened preliminarily, is applied to calculate corresponding indices. According to the magnitude of the indices and the determined selection fraction, important individuals can be screened out.

Genotype by environment interaction is an important component of genetic architecture of quantitative traits. Experiments conducted in multiple environments are beneficial to detect the genetic background of the complex traits. In the present study, two types of selection indices have been proposed, i.e., general index and interaction index, which have different backgrounds and various application potentials. General selection index is constructed on the basis of general breeding value and can be used to select promising lines, which should perform well in various environments. Interaction index is constructed on the ground of the conception of *GE* interaction breeding value and can be used to obtain superior cultivars, which can be cultivated in specific environment.

In succession of Smith-Hazel index, a series of extensions, such as index with Kempthorne-Nordskog restriction, index with Tallis restriction, Pesek-Baker index, Tai-Itoh-Yamada index, retrospective index, were put forth (Lin, 1978). These types of indices can be used under different conditions. For simplicity, only the Smith-Hazel index was applied in the working example. Other types of indices can be successfully implemented in this framework.

In traditional breeding programs, breeders take the phenotypic information of individuals for selection instead of genotypes because of their non-availability. One of the shortcomings of this procedure is the low efficiency. The successful development of molecular markers provides the possibility of direct selection upon genotypes through the strategy known as marker-assisted selection (MAS), which is especially effective for selection in early generations (Xu and Crouch, 2008). To maximize the rate of improvement quantitative traits, Lande and Thompson (1990) described a selection index combining molecular and phenotypic information, which only focused on one-trait improvement. Later, in common bean (*Phaseolus vulgaris* L.), Tar'an et al. (2003) developed a procedure integrated QTL-based index involving in several traits and genetic distance to identify lines that retain important QTLs and have desired genetic background. More recently, a model was illustrated to formulate selection index methods together with MAS covers both single trait selection and

multi-trait selection (Dekkers, 2007). However, all of those studies regarded additive effect as breeding value and did not place the ranges on multiple environments. The method proposed in this study can be extended to the situation of incorporation of index selection theory and QTL by environment interaction (QEI) in selection on seed traits.

Acknowledgements

This work was supported by the National Basic Research Program of China (No. 2006CB101708) and National Science and Technology Supporting Item of China (No. 2006BAD10A00). We gratefully acknowledge Professor Manfred Hühn and Dr. Andreas Müller in Plant Breeding Institute, Christian-Albrechts-University Kiel, Germany for critical reading of the manuscript.

References

- Allen, J.O. (2005). Effect of teosinte cytoplasmic genomes on maize phenotype. *Genetics* **169**: 863–880.
- Atlin, G.N., Baker, R.J., McRae, K.B., and Lu, X. (2000). Selection response in subdivided target regions. *Crop Sci.* **40**: 7–13.
- Beavis, W.D., Pollak, E., and Frey, K.J. (1987). A theoretical model for quantitatively inherited traits influenced by nuclear-cytoplasmic interactions. *Theor. Appl. Genet.* **74**: 571–578.
- Chung, J., Babka, H.L., Graef, G.L., Staswick, P.E., Lee, D.J., and Cregan, P.B. (2003). The seed protein, oil, and yield QTL on soybean linkage group I. *Crop Sci.* **43**: 1053–1067.
- Dekkers, J.C.M. (2007). Prediction of response to marker-assisted and genomic selection using selection index theory. *J. Anim. Breed. Genet.* **124**: 331–341.
- Falconer, D.S. (1952). The problem of environment and selection. *Amer. Nat.* **86**: 293–298.
- Foolad, M.R., and Jones, R.A. (1992). Models to estimate maternally controlled genetic variation in quantitative seed traits. *Theor. Appl. Genet.* **83**: 360–366.
- Gebre-Mariam, H., and Larter, E.N. (1996). Genetic response to index for grain yield, kernel weight and per cent protein in four wheat crosses. *Plant Breed.* **115**: 459–464.
- Geidel, H., Weber, W.E., Mechelke, W., and Haufe, W. (2000). Selection for sugar yield in sugar beet, *Beta vulgaris*, using different selection indices. *Plant Breed.* **119**: 188–190.
- Gutierrez-Marcos, J.F., Costa, L.M., Biderre-Petit, C., Khbaya, B., O'Sullivan, D.M., Wormald, M., Perez, P., and Dickinson, H.G. (2004). *Maternally expressed gene1* is a novel maize endosperm transfer cell-specific gene with a maternal parent-of-origin pattern of expression. *Plant Cell* **16**: 1288–1301.
- Hazel, L.N. (1943). The genetic basis for constructing selection indexes. *Genetics* **28**: 476–490.
- Hermesch, S., Kanis, E., and Eissen, J.J. (2003). Economic weights for feed intake in the growing pig derived from a growth model and an economic model. *J. Anim. Sci.* **81**: 895–903.

- James, J.W.** (1961). Selection in two environments. *Heredity* **16**: 145–152.
- Jannink, J.L., Orf, J.H., Jordan, N.R., and Shaw, R.G.** (2000). Index selection for weed suppressive ability in soybean. *Crop Sci.* **40**: 1087–1094.
- Lande, R., and Thompson, R.** (1990). Efficiency of marker-assisted selection in the improvement of quantitative traits. *Genetics* **124**: 743–756.
- Lemontey, C., Mousset-Declas, C., Munier-Jolain, N., and Boutin, J.P.** (2000). Maternal genotype influences pea seed size by controlling both mitotic activity during early embryogenesis and final endoreduplication level/cotyledon cell size in mature seed. *J. Exp. Bot.* **51**: 167–175.
- Lin, C.Y.** (1978). Index selection for genetic improvement of quantitative characters. *Theor. Appl. Genet.* **52**: 49–56.
- Lin, J.R., Shi, C.H., Wu, M.G., and Wu, J.G.** (2005). Analysis of genetic effects for cooking quality traits of japonica rice across environments. *Plant Sci.* **68**: 1501–1506.
- Pakkasmaa, S., Merila, J., and O'Hara, R.B.** (2003). Genetic and maternal effect influences on viability of common frog tadpoles under different environmental conditions. *Heredity* **91**: 117–124.
- Panthee, D.R., Pantalone, V.R., West, D.R., Saxton, A.M., and Sams, C.E.** (2005). Quantitative trait loci for seed protein and oil concentration, and seed size in soybean. *Crop Sci.* **45**: 2015–2022.
- Rajcan, I., Kasha, K.J., Kott, L.S., and Beversdorf, W.D.** (2002). Evaluation of cytoplasmic effects on agronomic and seed quality traits in two doubled haploid populations of *Brassica napus* L. *Euphytica* **123**: 401–409.
- Rao, C.R., and Kleffe, J.** (1988). Estimation of Variance Components and Applications. (Amsterdam, North-Holland).
- Shi, C.H., Zhang, H.Z., Wu, J.G., Li, C.T., and Ren, Y.L.** (2003). Genetic and genotype \times environment interaction effects analysis for erucic acid content in rapeseed (*Brassica napus* L.). *Euphytica* **130**: 249–254.
- Sivanadian, B., and Smith, C.** (1997). The effect of adding further traits in index selection. *J. Anim. Sci.* **75**: 2016–2023.
- Smith, H.F.** (1936). A discriminant function for plant selection. *Ann. Eugen.* **7**: 240–250.
- Snowder, G.D., Hanford, K.J., and van Vleck, L.D.** (2004). Comparison of models including cytoplasmic effects for traits of Rambouillet sheep. *Livest. Prod. Sci.* **90**: 159–166.
- Tar'an, B., Michaels, T.E., and Pauls, K.P.** (2003). Marker-assisted selection for complex trait in common bean (*Phaseolus vulgaris* L.) using QTL-based index. *Euphytica* **130**: 423–432.
- Turner, J.H., Ramey, H.H., and Worley, S.** (1976). Relationship of yield, seed quality and fiber properties in Upland cotton. *Crop Sci.* **16**: 578–580.
- van Sanford, D.A., and Matzinger, D.F.** (1982). Direct and maternal variances of seedling characters in tobacco. *Crop Sci.* **22**: 1213–1218.
- van Sanford, D.A., Pfeiffer, T.W., and Cornelius, P.L.** (1993). Selection index based on genetic correlation among environments. *Crop Sci.* **33**: 1244–1248.
- van Vleck, L.D.** (1970). Index selection for direct and maternal genetic components of economic traits. *Biometrics* **26**: 477–488.
- Willham, R.L.** (1963). The covariance between relatives for characters composed of components contributed by related individuals. *Biometrics* **19**: 18–27.
- Wricke, G., and Weber, W.E.** (1986). Quantitative Genetics and Selection in Plant Breeding. (Berlin Germany: Walter de Gruyter Press).
- Xu, Y.B., and Crouch, J.H.** (2008). Marker-assisted selection in plant breeding: From publications to practice. *Crop Sci.* **48**: 391–407.
- Young, S.S.Y.** (1961). A further examination of the relative efficiency of three methods of selection for genetic gains under less restricted conditions. *Genet. Res.* **2**: 106–121.
- Zhang, L., Yang, M.C.K., Wang, X.L., Larkins, B.A., Gallo-Meagher, M., and Wu, R.L.** (2004). A model for estimating joint maternal-offspring effects on seed development in autogamous plants. *Physiol. Genomics* **19**: 262–269.
- Zhu, J.** (1993). Mixed model approaches for estimating covariances between two traits with unequal design matrices. *J. Biomath.* **8**: 24–30.
- Zhu, J.** (1996). Analytic methods for seed models with genotype \times environment interactions. *Acta Genet. Sin.* **23**: 11–22 (in Chinese with an English abstract).
- Zhu, J., and Weir, B.S.** (1994a). Analysis of cytoplasmic and maternal effects: I. A genetic model for diploid plant seeds and animals. *Theor. Appl. Genet.* **89**: 153–159.
- Zhu, J., and Weir, B.S.** (1994b). Analysis of cytoplasmic and maternal effects: II. Genetic models for triploid endosperms. *Theor. Appl. Genet.* **89**: 160–166.