Estimation of Genetic Parameters of Selfing Population Derived from an Equilibrium One I. Theory

1Twenty Liana, 2Nasrullah, 2Woerjono Mangoendijojo and 2Panjisakti Basunanda
1Assessment Institute for Agricultural Technology of Central Kalimantan (BPTP Kalimantan Tengah), Jl. G. Obos Km. 5, Palangka Raya, Kalimantan Tengah, 73112, Indonesia
2Faculty of Agriculture, Gadjah Mada University, Jl. Flora Bulaksumur, Yogyakarta, Indonesia

ABSTRACT
Evaluation of inbred lines to develop hybrid is a key in a hybrid program. Inbred lines are extracted from the base population. During the process, extraction of inbred lines requires time and it is therefore interesting that during the extraction process, useful information appears for breeders, whether it is valuable or not. In response to the described useful information that appears for breeders, we formulate the concepts to estimate the value of the base population by genetic parameters. Using a single multi-allelic locus model, in addition to mean and variance, an inbreeding population derived by selfing an equilibrium one has other parameters similar to those of an F2 from a line cross useful for a hybrid program. The theory and estimation procedure are discussed.

Key words: Genetic parameters, selfing population, mean, variance

INTRODUCTION
Commercial hybrid plant has potential to increase the world’s agricultural productivity. The first step in hybrid program is the development of inbred lines for a hybrids parents. Inbred lines are a population of identical or nearly identical plants that are usually developed by self-pollination (Sleper and Poehlman, 2006) and breeders have been developing a large number of inbred lines and evaluating their performance in crosses (Guo et al., 2013). During the early inbred line development from base population, each plant possesses different combinations of genes, resulting in various combinations of traits that respond differently. At this point, the inbreds sort themselves into unique patterns as the offspring plants segregate from the parent lines. During the phases of inbred line development, researchers select the seeds from the best plants in the best rows and plant those seeds for the next generation of testing. For most hybrid breeding programs, only a small proportion of crosses can be evaluated in the field. An accurate prediction of hybrid performance prior to and after some field testing is of crucial importance in maize breeding (Guo et al., 2013). Extracting inbred lines from the base population takes time, so it will be interesting if, during the extraction, there is useful information that appears for breeders on whether it is valuable or not. In order to get valuable results for the development of elite inbred, we must know genetic parameter from the inbred lines.

An equilibrium population consists of inbred lines and hybrids. The hybrids do not know their superiority, so they need to be extracted. The process of extraction of an equilibrium population will form an inbreeding population. One consequence of inbreeding is changes in the distribution of genetic variance (Lopez-Fanjul et al., 1989; Fernandez et al., 1995; Fowler and Whitlock, 1999; Whitlock and Fowler, 1999). Under the assumption of segregation generation starting from selfing heterozygous individuals in the base population can form the structure of the F2 genotypes (Falconer and Mackay, 1996; Kearsey and Pooni, 1996). If we look at the groups of heterozygous genotype in an equilibrium population, they are a form of group pairs of inbred lines which have the possibility of being good hybrids. One parameter that is often used to assess the inbreeding
population is variance component. Theory has shown that genetic parameters are present in an inbreeding population (Robertson, 1952; Mather and Jinks, 1982; Kearsey and Pooni, 1996; Hallauer et al., 2010).

With respect to a quantitative trait, a random mating population consists of a mixture of homozygous and non-homozygous individuals where the latter is at least heterozygous for one locus. When the population is in equilibrium, their frequency is the product of the frequency of the genes they possess. Information on dominance variance in the population and the average degree of dominance are the basis prior to starting a hybrid program: Inbred lines-homozygous individuals-are extracted, crossed in pairs and evaluated for heterosis in replicated experiments. Extraction is time consuming but there is no guarantee that elite in breds may be obtained in the evaluation phase. Using a single locus model with an arbitrary number of alleles, similar parameters relevant to a hybrid program are developed.

The parameters: Heterosis, additive and dominance variance refer to the mixture of F2S populations derived from all possible pairs of the in breds. Thus they are weighted averages.

Here we discuss the genetic parameters for selfing population through variance estimating of segregation generation in S1 (F = 1/2) and S2 (F = 3/4) families. In response to the described useful information that appears for breeders, we formulate the concepts to estimate the value of the base population by genetic parameters.

**METHODOLOGY**

**Development of the theory:** Consider a given locus in an equilibrium population $\pi_0$ where there is no guarantee that allele $A_i$ (i = 1, 2, ..., s) with frequency $p_i$. The genetic structure of $\pi_0$ is:

$$\pi_0 = \sum_{i,j} p_i p_j A_i A_j$$

The population mean is:

$$\mu_0 = \sum_{i,j} p_i p_j \mu_{ij}$$

where, $\mu_{ij}$ is genotypic value of $A_i A_j$. Modelling $\mu_{ij}$ statistically as $\mu_{ij} = \mu_0 + \alpha_i + \alpha_j + \delta_{ij}$ with $\alpha_i$ and $\alpha_j$ is additive effect of allele $A_i$ and $A_j$, and $\delta_{ij}$ is the dominance deviation or the interaction effect of allele $A_i$ and $A_j$. The genetic variance of the population is measured according to Kempthorne (1969), Steel and Torrie (1980) and Nyquist (1990):

$$\sigma_0^2 = \sigma_A^2 + \sigma_D^2$$

Where:

$$\sigma_A^2 = 2 \sum_i p_i \sigma_i^2$$

$$\sigma_D^2 = \sum_{i,j} p_i p_j \delta_{ij}^2$$

Inbreeding population derived from $\pi_0$ is $\pi_i = \sum p_i A_i A_i$ with mean $\mu_i = \sum p_i \mu_{ij}$ and genetic variance $\sigma_i^2 = \sum p_i (\mu_{ij} - \mu_i)^2$. However, if inbreeding is up to an inbreeding coefficient of $F$, the resulting inbreeding population is $\pi_F = (1-F)\pi_0 + F\pi_1$ and is depicted in Table 1. Note that $\pi_1$ is a mixture of $\pi_0$ and $\pi_1$ with frequency of $(1-F)$ and $F$, respectively. Therefore, the mean of $\pi_1$, denoted by $\mu_1$, is the expected value of the mean:

$$\mu_1 = (1-F)\mu_0 + F\mu_1$$

(1)

while the variance, denoted by is the sum of variance of the mean, $(1-F)(\mu_0 - \mu_1)^2 + F(\mu_1 - \mu_1)^2 = F(1-F)(\mu_0 - \mu_0)^2$ and the mean of the variance $(1-F)\sigma_0^2 + F\sigma_1^2$. Thus:

$$\sigma_1^2 = (1-F)\sigma_0^2 + F\sigma_1^2 + F(1-F)(\mu_1 - \mu_1)^2$$

(2)

An equation that according to Crow and Kimura (1970), Jain (1982) and Nyquist (1990) it was formulated for the first time by Wright (1951).

With selfing ($F = \frac{1}{2}$), the population structure, denoted by $\pi_{S0}$ or $\pi_{S1}$ is:

$$\pi_{Si} = \sum_{i,j} p_i p_j (\frac{1}{4}A_i A_i + \frac{1}{2}A_i A_j + \frac{1}{4}A_j A_j)$$

(3)

The population mean is:

$$\mu_{Si} = \sum_{i,j} p_i p_j (\frac{1}{4}\mu_{ii} + \frac{1}{2}\mu_{ij} + \frac{1}{4}\mu_{jj})$$

$$= \mu_0 - \frac{1}{2} \sum_{i,j} p_i p_j H_{ij}$$

(4)

Where:

$$H_{ij} = \mu_{ij} - \frac{1}{2}(\mu_{ii} + \mu_{jj})$$

$$\mu_{Si} = \mu_0 - \frac{1}{2} H^*$$

Where:

$$H^* = \sum_{i,j} p_i p_j H_{ij}$$

$H^*$ as weighted intra varietal heterosis or weighted average heterosis in Hardy-Weinberg population.

---

**Table 1:** Composition of population, frequency, mean and variance when the coefficient of inbreeding is $F$

<table>
<thead>
<tr>
<th>Population</th>
<th>Frequency</th>
<th>Mean</th>
<th>Variance</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\pi_0$</td>
<td>1-F</td>
<td>$\mu_0$</td>
<td>$\sigma_0^2$</td>
</tr>
<tr>
<td>$\pi_1$</td>
<td>F</td>
<td>$\mu_1$</td>
<td>$\sigma_1^2$</td>
</tr>
</tbody>
</table>

and:

$$\sigma_1^2 = \sum_{i,j} p_i p_j \delta_{ij}^2$$
Note that the expression in parentheses in the second term in Eq. 3 is an F2 genotypic array derived from \( A_i A_i \times A_j A_j \) cross \((i \neq j)\). Denote the variance of this F2 by \( \sigma^2_{e_{ij}} = \sigma^2_{A_i} + \sigma^2_{D_{ij}} \). Then variance within the S1 family is that with a common S0 parent:

\[
\sigma^2_{S_{ij}|S_0} = \sum_{i\neq j} p_i p_j (\sigma^2_{A_i} + \sigma^2_{D_{ij}}) = \sigma^2_{A_1} + \sigma^2_{D_0}. \tag{5}
\]

Further selfing gives an S2 population \((F = \frac{1}{4})\), the structure of which is:

\[
\pi_i = \Sigma p^2_i A_i A_i + \Sigma p_i p_j [\frac{1}{4} A_i A_i + \frac{1}{2} (\frac{1}{4} A_i A_i + \frac{1}{2} A_j A_j) + \frac{1}{4} A_j A_j].
\]

Note that in the second term, the expression in parentheses in the second term of Eq. 3 is an F2 and those in brackets is an F3 derived from \( A_i A_i \times A_j A_j \) cross. Being related to the difference of \( \frac{1}{2} \) in Fisher notation with a subscript (i.e., \( \sigma^2_{A_j} \)).

Denote this by then:

\[
\pi_i = \Sigma p^2_i A_i A_i + \Sigma p_i p_j [\frac{1}{4} A_i A_i + \frac{1}{2} (\frac{1}{4} A_i A_i + \frac{1}{2} A_j A_j) + \frac{1}{4} A_j A_j].
\]

Within the S2 family variance is the variance within S2 families having a common S0 parent:

\[
\sigma^2_{S_{ij}|S_0} = \frac{1}{2} (\sigma^2_{A_1} + \sigma^2_{D_0}). \tag{7}
\]

while within the S2 family variance having a common S1 parent is a half of F2 variance. Denoting this by \( \sigma^2_{S_{ij}|S_1} \), then:

\[
\sigma^2_{S_{ij}|S_1} = \frac{1}{2} \sigma^2_{A_1} + \frac{1}{2} \sigma^2_{D_1}. \tag{8}
\]

then using Eq. 7, 8 estimate of variance of F1 mean can be obtained:

\[
\sigma^2_{S_{ij}|S_0} - \sigma^2_{S_{ij}|S_1} = \sigma^2_{A_1} + \frac{1}{2} \sigma^2_{D_0}. \tag{9}
\]

**Estimation procedure:** From theory development, some parameters are of interest: \( \mu_0, \mu_1, H^*, \sigma^2_{e_0}, \sigma^2_{A_1}, \sigma^2_{D_0}, \sigma^2_{A_0}, \) and \( H^* \).

For estimating these parameters, from an equilibrium population, self a random sample of \( n \) individuals to generate \( S_1 \) families. After saving some seeds from each ear for a replicated experiment, plant the remaining seeds in ear-to-rom. In the i-th \((i = 1, 2, ..., n_i)\) row, self a random sample of \( n_i \) individuals to generate \( S_2 \) families. Evaluate these:

\[
N = \sum_{i=1}^{n} n_i
\]

\( S_2 \) families together with \( n S_1 \) families.

For simplicity of discussion, let the evaluation be conducted in a randomized complete block with \( r \) blocks as replicates. Analysis is done as shown in Table 2.

**RESULTS AND DISCUSSION**

Inbreeding population may result from any inbreeding process. If this is so, only \( \mu_0, \mu_1, H^*, \sigma^2_{e_0} \) and \( \sigma^2_{e_1} \) can be estimated. Inbreeding through selfing will produce an F2 population if the selfed individuals are heterozygous. This unique characteristic is exploited to get parameters similar to the ones in line cross. In line cross, the parameters refer to a single F2 while ours refer to a mixture of F2 populations, each weighted by its frequency; hence, the average weighted parameters.

\[
H^* \text{ in Eq. 4 is exactly } d \text{ in Fisher notation with a subscript to remind that it refers to F2 population derived from } A_i A_i \times A_j A_j \text{ cross. Being related to the difference of } \mu_0 \text{ and } \mu_1, H^* \text{ measures inbreeding depression, considered the reverse of heterosis, disregarding the number of loci involved.}
\]

Estimation procedure discussed above is a general one in the sense that it applies also when \( S_2 \) families being evaluated may be generated from selfing \( S_1 \) ears not included in the replicated experiment as when the seed number is a constraint. However, when they are, additional information, i.e., \( \text{Cov}(S_i, S_j) = \sigma^2_{e_0} + \frac{1}{2} \sigma^2_{D_0} \) (Mather and Jinks, 1982) they can be

**Table 2: Analysis of variance for \( S_1 \) and \( S_2 \) families**

<table>
<thead>
<tr>
<th>Source of variance</th>
<th>DF</th>
<th>MS</th>
<th>E(MS)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Replication</td>
<td>r-1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Entry</td>
<td>N+n-1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Among ( S_1 ) families</td>
<td>n-1</td>
<td>M_1</td>
<td>( \sigma^2_{e_0} + \sigma^2_{A_1} )</td>
</tr>
<tr>
<td>Among ( S_2 ) families</td>
<td>N-1</td>
<td>M_4</td>
<td>( \sigma^2_{e_0} + \sigma^2_{D_0} )</td>
</tr>
<tr>
<td>Among ( S_2)S_0</td>
<td>n-1</td>
<td>M_3</td>
<td>( \sigma^2_{e_0} + \sigma^2_{D_0} )</td>
</tr>
<tr>
<td>Within ( S_2)S_0</td>
<td>N-n</td>
<td>M_3</td>
<td>( \sigma^2_{e_0} + \sigma^2_{D_0} )</td>
</tr>
<tr>
<td>Plot-to-plot ( S_2 ) error</td>
<td>(r-1)(N+n-1)</td>
<td>M_2</td>
<td>( \sigma^2_{e_2} )</td>
</tr>
<tr>
<td>Plot-to-plot ( S_2 ) error</td>
<td>(r-1)(N-1)</td>
<td>M_1</td>
<td>( \sigma^2_{e_2} )</td>
</tr>
</tbody>
</table>
used together with Eq. 8 to get $\sigma^2_i$ and $\sigma^2_{ip}$. The latter is better in the sense that the estimation procedure is a direct one; no need to estimate $\sigma^2_i$. Better still, all information is used and the estimate is obtained using the least square method.

Note that:

$$\sigma^2_i = \frac{1}{2} \sum_{i,j} p_i p_j (\mu_i - \mu_j)^2 = 2 \sum_{i,j} p_i p_j a_{ij}^2 = 2 \sigma^2_A,$$

using a notation of Fisher. It shows a better way to obtain $\sigma^2_i$. As:

$$\sigma^2_{ip} = \sum_{i,j} p_i p_j d_{ij}^2$$

their similarity to those of $F_2$ in a line cross see standard textbooks on Quantitative Genetics such as Falconer and Mackay (1996) and Hallauer and Filho (1981) is apparent. Thus they may be extended to a multi-locus model if there is no epistasis. Model extension with epistasis being included has also been available (Nasrullah et al., 1995; Kempthorne, 1969). The only difference is that they are weighted by their frequencies as their coefficient, instead of 1 as in a line cross.

A final point is to be mentioned. Though homozygous individuals in $p_n$ are inbreds regardless of the number of loci, non-homozygous ones are a mixture $F_1$ hybrids in a single locus model but a mixture of $F_1$ hybrids for one locus or more in a multi-locus model.

CONCLUSION

Genetic parameters can be constructed by inbreeding. Inbreeding through selfing will produce an $F_2$ population if the selfed individuals are heterozygous. The selfing families as expression in parentheses derived from $A_i A_i \times A_i A_i$ cross $(i,j)$, produce some parameters are of interest: $\mu_i, \mu_j, \sigma^2_i, \sigma^2_j, c_{ij}, a_{ij}, c_{ip}$, $\sigma^2_{ip}$ and $*$.

The genetic parameters will then allow assessment for proper or improper assessment of hybrid programs before extraction of inbred lines are done too much.

REFERENCES


