Journal of MSMSSEA
Vol.1, ICoMS-1 (2006), pp. 73-87

# THE APPLICATION OF STATISTICS IN MARKER ASSISTED SELECTION

ASEP SAEFUDDIN[1) 2)] AND FARIT MOCHAMAD AFENDI[2)]
[1)] Vice Rector of Planning, Development, and Cooperation,
Bogor Agricultural University
[2)] Deparment of Statistics, Bogor Agricultural University

## 1. INTRODUCTION

For centuries, human being had utilized many Gods' creation to fulfill their needs. The history has recorded that to fulfill the nutrition need, human had utilized many plants and animals through food gathering until raising them. However, the increase in human population has forced them to optimize the crop, milk or meat production through breeding program which basically is a process of genetical assembly to obtain new plant or animal having better performance. One of issues in this process is the selection of parents as source of genetic to assembly. The selection usually performed by choosing individual that showing good performance in the trait of interest such as has high production in meat, milk or crop. However, the nature of certain trait making the selection is difficult to be performed such as resistance for certain disease or carcass quality. The selection could become more complicated due to the effect of environment on the trait of interest.

On the other hand, the development in molecular biology has lead us to understand that genes, in the form of DNA, determine the expression of the traits. DNA (deoxyribonucleic acid) is a molecule that is shaped like a double helix and made up of pairs of nucleotides. DNA is packaged into chromosomes which are located within the nucleus of all cells. Every cell in the body contains all of the chromosomes that collectively make up the genome of that organism. DNA codes for amino acids which are linked together to make proteins. A gene is a stretch of DNA that specifies all of the amino acids that make up a single protein. Proteins are the building blocks of life. There are thousands of proteins in the body (encoded by thousands of genes). The interaction and structure of proteins determines the visible characteristics or phenotype of an organism, while the genotype refers to the genetic makeup. By knowing genes that code the proteins which affecting traits, we could determine the phenotype of the trait before they appear.

In the next section, the role of DNA in breeding program through the concept of Marker Assisted Selection will be discussed. Furthermore, the determination of whether certain gene is truly affecting the trait could be

performed by means of hypothesis testing which is the role of the statistics. Further explanation on the role of statistics in this issue will be presented in Section III. Finally, the prospect and preparation of Marker Assisted Selection is discussed in Section IV.

## 2. MARKER ASSISTED SELECTION

The idea behind marker assisted selection (MAS) is genes hunting, i.e. selecting genes having significant effects on the trait of interest. Some traits are controlled by single genes (e.g. hair colour) but most traits of economic importance are quantitative traits that most likely are controlled by a fairly large number of genes. However, some of these genes might have a larger effect. Such genes can be called major genes located at certain regions in chromosome or loci. These loci are then called Quantitative Trait Loci or QTL. Although the term QTL strictly applies to genes of any effect, in practice it refers only to major genes, as only these will be large enough to be detected and mapped on the genome. Following the pattern of inheritance at such QTL might assist in selection.

Recently, scientists have started to identify regions of DNA that influence traits. They have used the techniques of molecular biology and quantitative genetics to find differences in the DNA sequence in these regions. Tests have been developed to identify these subtle sequence differences and so identify whether an individual is carrying a segment of DNA that is positively or negatively associated with the trait of interest. These different forms of a genetic marker are known as DNA-marker. MAS then could be defined as the process of using the results of DNA-marker tests to assist in the selection of individuals to become the parents in the next generation of a genetic improvement program. That is, instead of using only a traditional selection program which based on the phenotype of the trait to increase the proportion of favorable alleles for the genes that affect a certain trait, specific DNA tests are used to assist in the selection of those favorable alleles. Genotyping allows for the accurate detection of specific DNA variations that have been associated with measurable effects on complex traits. It is important to remember that markers for complex traits are associated with only those genes that are located in close proximity to the marker and do not identify favorable alleles for all the other genes that are associated with the trait. Selecting an individual that carries favorable alleles of a marker, which is the allele that is associated with a positive impact on the trait of interest, can result in an improvement in the observed phenotype for that trait.

Potential benefits from marker-assisted selection are greatest for traits that:

- have low heritability (i.e. traits where an individual's measured value is a poor predictor of breeding value due to the large environmental influences on the observed value).
- are difficult or expensive to measure (e.g. disease resistance).
- cannot be measured until after the individual has already contributed to the next generation (e.g. reproduction or longevity).
- are currently not selected for because they are not routinely measured (e.g. tenderness).
- are genetically correlated with a trait that you do not want to increase (e.g. a marker that is associated with increased marbling but that is not also associated with those genes that increase backfat thickness).

The use of molecular marker has been applied in assisting the selection in breeding program for plant and animal as well. For example, molecular marker has been applied in selection for resistance to pathogen in tomato (Barone, 2003), in improvement of quantitative trait for forage crops (Dolstra et al, 2003), improve the efficiency of introgression of cotton fiber quality trait (Lacape, 2003), breeding program for major cereals such as wheat, barley, rice, and maize (Koebner, 2003 and Korzun, 2003), breeding program for pome fruit (Tartarini, 2003), and powdery mildew resistance in grapes (Dalbo et al 2001). In the case of animal breeding, molecular marker has been applied to dairy sheep (Pragnacco and Carta, 2003), detect brodiness trait of Kampung chicken (Sartika et al, 2005) even for fish breeding program (Sonesson, 2003).

Figure 1 shows the principle of inheritance of a marker and a linked QTL to give illustration how to use DNA marker in identifying trait of interest. We can identify the marker genotype (Mm) but not the QTL genotype (Qq). The last is really what we want to know because of its effect on economically important traits. Let the Q allele have a positive effect, therefore being the preferred allele. In the example, the M marker allele is linked to the Q in the sire. Progeny that receive the M allele from the sire, have a high chance of having also received the Q allele, and are therefore the preferred candidates in selection.

As shown in Figure 1 there are 4 types of progeny. All progeny will inherit m alleles and q-alleles from the mother. The sire will provide them with either an M allele or an m allele and either Q or q. In the figure, 90% of the progeny that receive an M allele has also received a Q allele, because M and Q alleles are linked on the same chromosome in the sire. However, in 10% of the cases after the sire reproduced, there has been a recombination between the two loci, and animals that inherited an M allele from their father have received a q allele rather than a Q allele. Therefore, marker

alleles do not always provide certainty about the genotype at the linked QTL.

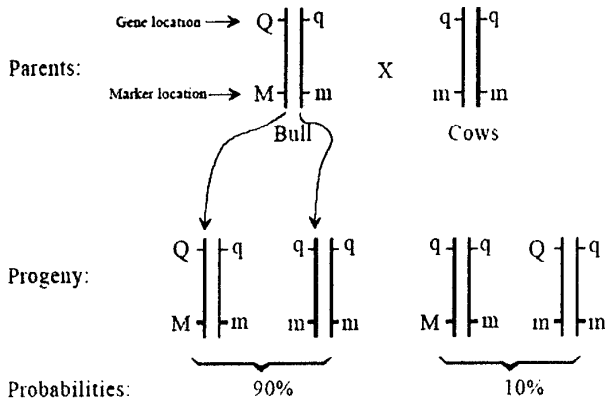


**Figure 1.** The inheritance pattern of DNA marker and QTL

The next problem arises here is that how to evaluate whether certain DNA marker is close to QTL of interest. If certain marker is close to QTL, the next question should be answered is what size is the effect of QTL on the trait of interest. The method to answer these problems is discussed in the next section.

## 3. QTL MAPPING

The questions mentioned above basically refer to what is the location of the QTL relative to DNA marker and what is the size of its effect on trait. All these questions could be answered through a process called QTL mapping.

The history of QTL mapping can be traced back to 1920's. Sax (1923) use the morphological markers to demonstrate an association between seed weight and seed coat color in beans. The method pioneered by Sax is known as single marker. The general idea of single marker is as follow. Suppose that in a certain region of chromosome, QTL of interest is located near the region where the marker is located with distance r as shown in

Figure 2 (here r is recombination unit between marker and QTL).

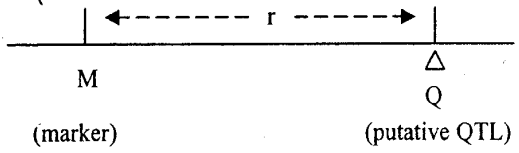


Figure 2. Single marker genetic model

If the QTL and the markers are segregating in a genetically defined population, then the linkage relationships among them may be discoverable by looking at the association between the trait variation and the marker segregation pattern. Here, we use backcross design as for example.

In classical backcross design, the population is generated by a heterozygous F1 backcrossed to a homozygous parent (for example, a cross of MmQq x MMQQ)(Figure 3) The rationale behind the single marker analysis can be explained using co-segregation pattern listed in Table 1. Marker M and QTL Q are assumed to be linked with r recombination units apart. The expected frequencies for the four marker-QTL genotypes (MMQQ, MMQq, MmQQ, mmqq) are listed in Table 1. The conditional frequencies of the QTL genotypes (QQ and Qq) on the marker genotypes (MM and Mm) can be obtained by dividing the joint marker-QTL genotypic frequencies by the marginal marker genotypic frequencies. Hence, the expected phenotypic values for the observable marker genotypes can be obtained as

$$\mu_{MM} = P(QQ \mid MM)\,\mu_1 + P(Qq \mid MM)\,\mu_2$$

$$= (1\text{-}r)\,\mu_1 + r\,\mu_2$$

$$\mu_{Mm} = P(QQ \mid Mm)\,\mu_1 + P(Qq \mid Mm)\,\mu_2$$

$$= r\,\mu_1 + (1\text{-}r)\,\mu_2 \qquad (1)$$

where $\mu_1$ and $\mu_2$ are the expected genotypic value for the two QTL genotypes QQ and Qq, respectively. Then, QTL is detected if $\mu_{MM}$ and $\mu_{Mm}$ are not equal. Using this idea, then hypothesis to be tested in single marker analysis is:

$H_0$: $\mu_{MM} = \mu_{Mm}$

To test the hypothesis, we can use several methods such as: T-test, linear regression, and likelihood ratio test.

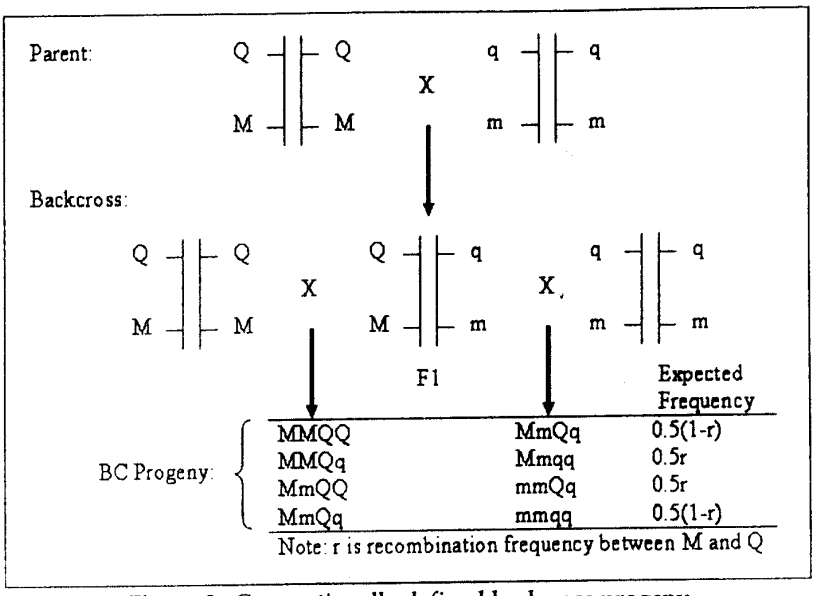


Figure 3. Conventionally defined backcross progeny

Table 1. Co-segregation pattern for backcross design

| Marker Genotype | Observed Count | Marginal frequency | QTL Genotype | | Expected Trait Value |
|---|---|---|---|---|---|
| | | | QQ | Qq | |
| | | | Joint frequency | | |
| MM | $n_1$ | 0.5 | 0.5(1-r) | 0.5r | |
| Mm | $n_2$ | 0.5 | 0.5r | 0.5(1-r) | |
| | | | Conditional frequency | | |
| MM | $n_1$ | 0.5 | 1-r | R | (1-r) $\mu_1$ + r $\mu_2$ |
| Mm | $n_2$ | 0.5 | r | 1-r | r $\mu_1$ + (1-r) $\mu_2$ |

For the T-test method, the test statistic is

$$t = \frac{\hat{\mu}_{MM} - \hat{\mu}_{Mm}}{\sqrt{s^2\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}} \qquad (2)$$

where $s^2$ is the pooled estimate of the variance within the two classes of marker genotypes. When we use the linear regression method to test the hypothesis, we assumed that the model is

$$Y_i = \beta_0 + \beta_1 X_i + \varepsilon_i \qquad (3)$$

where $Y_i$ is the trait value for the ith individual, and $X_i$ is the dummy variable taking 1 if the individual is MM and -1 for Mm. To test the hypothesis of equality of $\mu_{MM}$ and$\mu_{Mm}$ is equivalent to test the hypothesis of $\beta_1 = 0$. When we use likelihood ratio test method, the likelihood is constructed based on the distribution of trait values usually assumed as normal. Let

$$Y_i \sim N(\mu_i, \sigma^2) \tag{4}$$

where $\mu_i$ takes value $\mu_{MM}$ and $\mu_{Mm}$ if ith individual has MM and Mm of genotypic marker, respectively. Let $Q_1$ and $Q_2$ denote the QQ and Qq of genotypic QTL, respectively. Using Equation 1, we obtain

$$Y_i \sim \sum_{j=1}^{2} P(Q_j \mid M_i) N(\mu_j, \sigma^2) \tag{5}$$

where $M_i$ is the marker genotype of ith individual. Then, the distribution of each individual is mixture of two normal distributions with the mix proportion equal to conditional probability of QTL genotype given marker genotype. Hence the likelihood is

$$L(\mu_1, \mu_2, \sigma^2, r) = \frac{1}{\left(\sqrt{2\pi\sigma^2}\right)^n} \prod_{i=1}^{n} \sum_{j=1}^{2} P(Q_j \mid M_i) \exp\left[ -\frac{(y_i - \mu_j)^2}{2\sigma^2} \right] \tag{6}$$

whereas the logarithm of likelihood is

$$\mathrm{Log}[L(\mu_1, \mu_2, \sigma^2, r)] = \sum_{i=1}^{n} \mathrm{Log}\left\{ \sum_{j=1}^{2} P(Q_j \mid M_i) \exp\left[ -\frac{(y_i - \mu_j)^2}{2\sigma^2} \right] \right\}$$

$$-\frac{n}{2}\mathrm{Log}\left(2\pi\sigma^2\right) \tag{7}$$

Under the null hypothesis $H_0$: $\mu_1 = \mu_2$ or $\mu_1 = \mu_2 = \mu$, the log likelihood is

$$\mathrm{Log}[L(\mu_1 = \mu_2 = \mu)] = -\frac{1}{2\sigma^2} \sum_{i=1}^{n} (Y_i - \mu)^2 - \frac{n}{2} \mathrm{Log}(2\pi o^2)$$

and for the null hypothesis $H_0$: r=0.5, the log likelihood is

$$\text{Log}[L(r=0.5)] = \sum_{i=1}^{n} \text{Log}\left\{\sum_{j=1}^{2} \exp\left[-\frac{(y_i-\mu_j)^2}{2\sigma^2}\right]\right\} - \frac{n}{2}\text{Log}(2\pi o^2)$$

The test statistic for likelihood ratio test is

$$G = 2\{\text{Log}[L(\hat{\mu}_1, \hat{\mu}_2, \hat{\sigma}^2, \hat{r})] - \text{Log}[L(r=0.5)\} \quad (8)$$

where the log likelihoods are evaluated using the maximum likelihood estimates of $\mu_1$, $\mu_2$, $\sigma^2$ and r. it is also common to use lod scores for QTL detection. The differences between the G-statistic and lod score are the bases for the logarithm and the interpretation. G-statistic is computed using natural logarithm and interpreted as a probability of occurrence the data under the null hypothesis. On the other hand, lod score is computed using the base 10 logarithm and interpreted using the concept of an odd ratio. For example, a lod score of 2 means that the alternative hypothesis is $10^2 = 100$ times more likely than the null hypothesis (Liu, 1998).

In evaluating the result of single marker analysis, it has been common to plot of the test statistic against genome position as illustrated in the following Figure 4. The peak in the plot usually treated as genome location of QTL.

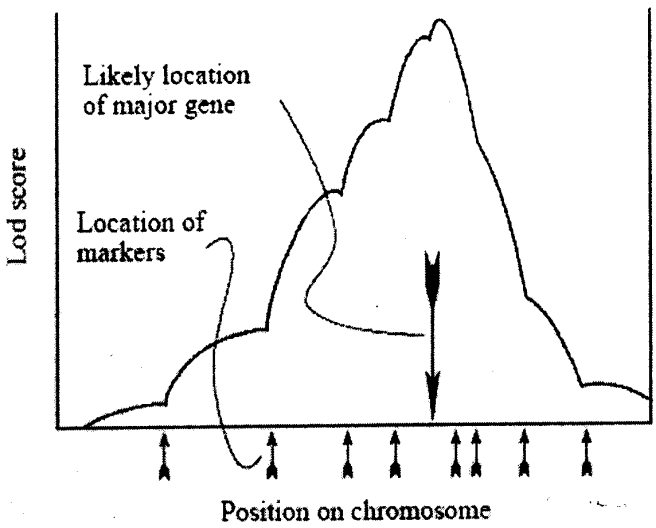


Figure 4. LOD score plotted against genome locations

As mentioned above that the hypothesis to be tested in single marker analysis is $H_0$: $\mu_{MM} - \mu_{Mm} = 0$. The expectation of difference between the two marker classes is

$$E[\mu_{MM}-\mu_{Mm}] = [(1-r)\mu_1 + r\mu_2] - [r\mu_1 + (1-r)\mu_2]$$

$$= (1-2r)(\mu_1-\mu_2)$$

$$= 2g(1-2r) = (a+d)(1-2r) \quad (9)$$

For the null hypothesis $H_0$: $\mu_{MM} - \mu_{Mm} = 0$ there are two possible interpretations: (a+d) = 0 or r = 0.5. The biological meaning for the first one is that there is no genetic effect, for the other, that the QTL and the marker are independent (no linkage). Therefore, although all of the methods used in the single marker analysis are relatively easy to conduct the QTL effect and the QTL location are confounded. In addition, single marker analysis can not estimate the number of QTL.

To overcome this problem, Lander and Botstein (1989) proposed a method called interval mapping. The idea of interval mapping is in investigating the existence of QTL we search upon certain interval in genome flanked by two adjacent markers rather than near one marker as in single marker analysis (see **Error! Reference source not found.** for an illustration of this idea). Markers A and B are linked with recombination fraction r, and Q is located between the two markers with $r_1$ recombination fraction from A and $r_2$ from B.

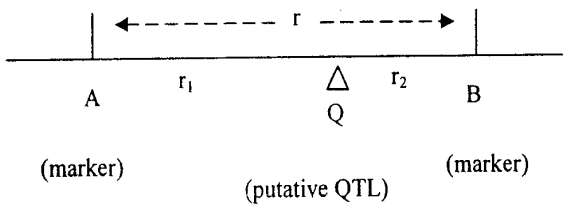


**Figure 5.** Linkage relationship of a QTL and two flanking markers

There are two common methods in analyzing QTL using interval method: likelihood approach and regression approach. Moreover, Liu (1998) mentioned that problem still exist in interval mapping, such as the number of QTL can not be resolved, the location of QTL are sometimes not well resolved and the exact positions of the QTL can not be determined, and the statistical power is still relatively low. These problems occur mainly due to linked QTL, there is interaction among QTL, and limited information

contained in the model. One of the reasons for these shortcomings is that the test used in the interval mapping is not an interval test. An interval test is that the effect of the QTL within a defined interval should be independent of the effects of QTL outside the region. Hence, Jansen and Stam (1994) and Zeng (1994) proposed composite interval mapping as extension of interval mapping by incorporating another marker as cofactor.

All the method previously mentioned, assuming that the trait of interest is in continuous scale. On the other hand, many important traits are obtained in categorical scale, such as resistance from certain disease. If the resistance from the disease is obtained as suscept or resistance, then the trait is in binary scale, whether if the resistance scored on ordered scale varying from unaffected to dead then the trait is in ordinal scale. Another trait could also be obtained in nominal scale such as shapes and colors of flowers, fruits, and seeds in plants, as well as coat colors. From a theoretical point of view, QTL mapping method assuming continuous trait could not be applied to categorical trait.

In dealing with binary trait, Xu and Atchley (1996) proposed likelihood based method by assuming there is continuous distribution called liability underlying binary trait by means of threshold model. Similar approach proposed by Hackett and Weller (1995) in dealing with ordinal trait. On the other hand, Hayashi and Awata (2006) proposed likelihood based approach in analyzing trait in nominal scale. The binary trait is analyzed using threshold model obtained as the following.

In dealing with binary trait, it is assumed that there is continuous distribution, say U, underlying binary trait, say Y, referred to as liability (Xu and Atchley, 1996). In relation between liability and binary trait (such as resistance to certain disease), it is assumed that there is threshold ($\gamma$) in the scale of liability, below which the individual has unaffected phenotype, and above which it is affected (see Figure 6).

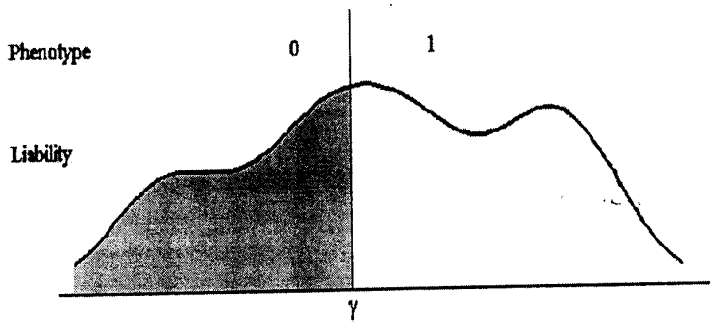


Figure 6. Liability and threshold model for binary trait

The relation can be summarized by

$$y_i = \begin{cases} 1; & \text{if } u_i \geq \gamma \\ 0; & \text{if } u_i < \gamma \end{cases} \qquad (10)$$

### *Maximum likelihood (ML) approach*

Using liability model, the one-QTL ML mapping model for a backcross population can be written as

$$u_i = \mu + bx_i^* + \varepsilon_i, \quad i = 1, 2, \ldots, n \qquad (11)$$

where $u_i$ is the liability value for individual i, $\mu$ is the mean, b is the effect of QTL Q, xi* taking the value of 1 (0) for homozygote QQ (heterozygote Qq), denotes the genotypes of Q, $\varepsilon_i$ is environmental deviation and is assumed to follow $N(0, \sigma^2)$. Since the liability is unobserved, the mean $\mu$ and variance of $\varepsilon$ can be set at any arbitrary value (for simplicity, it is determined that $\mu = 0$ and $\sigma^2 = 1$).

Based on the conditional probability of $u_i$ given $x_i^*$, the conditional probability of $y_i$ given $x_i^*$ is obtained by

$$P(y_i|x_i^*) = \int_{\gamma}^{\infty} f(u_i|x_i^*)d(u_i|x_i^*)$$

$$= 1 - \int_{-\infty}^{\gamma} f(u_i|x_i^*)d(u_i|x_i^*) = 1 - \Phi\left(\gamma - bx_i^*\right) = \Phi\left(bx_i^* - \gamma\right) \qquad (12)$$

where $\Phi(\xi)$ stands for the standardized cumulative normal distribution function and $\xi$ is the argument. Analysis involving $\Phi(\xi)$ is referred to as probit analysis. However, the probit model is difficult to manipulate because numerical integration is required although the parameters are easy to interpret. So, a logistic model is employed to approximate $\Phi(\xi)$ for estimation purpose and is expressed by

$$\psi(\xi) = \frac{\exp(\xi)}{1 + \exp(\xi)} \qquad (13)$$

The relationship between a probit model and a logistic model is $\Phi(\xi) \approx \psi(d\xi)$, where $d = \pi/\sqrt{3}$. Therefore,

$$P(y_i = 1|x_i^*) \approx \frac{\exp\{d(bx_i^* - \gamma)\}}{1 + \exp\{d(bx_i^* - \gamma)\}} \qquad (14)$$

Since the QTL genotype $x_i^*$ could be homozygote (1) or heterozygote (0) for an individual, the likelihood is then a mixture distribution with mixing proportions equivalent to the conditional probabilities of QTL genotypes given two flanking markers, $q_{i1}$ and $q_{i2}$ for the QTL genotypes QQ and Qq respectively (see Table 1). For n individuals in the sample, the likelihood function is

$$L = \prod_{i=1}^{n}[\sum_{j=1}^{2} q_{ij} p_{ij}^{\ y_i} (1 - p_{ij})^{1-y_i}]$$

where $p_{i1}$ and $p_{i2}$ denotes the conditional probability of $y_i = 1$ given the QTL genotypes $x_i^* = 1$ and $x_i^* = 0$, respectively. The log likelihood function is

$$l = \sum_{i=1}^{n} \log(\sum_{j=1}^{2} q_{ij} p_{ij}^{\ y_i} (1 - p_{ij})^{1-y_i}). \qquad (15)$$

On the other hand, ordinal trait is analyzed using threshold model by introducing several thresholds. The example of ordinal trait could be accessed such as in Afendi et al (2006).

## 4. PROSPECT AND PREPARATION

As explained above, MAS could help the breeding program by effectively selecting genes which significantly affect the trait of interest. Moreover, the selection of the genes is performed through a process in QTL mapping which is deeply full of statistics concept. To give the clear picture, let take a look again at QTL mapping. In QTL mapping, the genotype of QTL is unobserved. But by using the genotype of two adjacent DNA markers flanking the putative QTL, the conditional probability of QTL taking certain genotype could be determined. Here, we need the concept of probability theory. Furthermore, in testing the QTL effect on the trait of interest, it is again we face the deep statistics concept. First, in estimating the effect of QTL we could use likelihood method (Lander and Botstein, 1989), least square method (Haley and Knott, 1992; Martinez and Curnow,

1992), or iteratively reweighted least square method (Xu, 1998; Xu, 1998) which are found in point estimation theory in statistics concept. This part is also applicable for advance optimization method such as Expectation Maximization Method (Xu, 2003; Xu, et al 2005), or Expectation Conditional Maximization (Xu, et al, 2005). Second, evaluation of the effect of QTL on the trait is conducted through modeling the effect into mathematical equation plus random effect as the representative of environment effect which is again full of statistics concept especially statistical modeling. Here, we use linear model if the trait is observed in numerical scale (Lander and Botstein, 1989; Haley and Knott, 1992; Martinez and Curnow, 1992; Jansen and Stam, 1994; Zeng, 1994); and generalized linear model if the trait is observed in categorical scale (Xu and Atchley, 1996; Hackett and Weller, 1995; Hayashi and Awata, 2006). The development in statistics concept could also be applied here such as Bayesian concept (Yi, et al, 2004) as we need prior/posterior information.

From the above explanation, statistics plays important role in MAS especially in QTL mapping part. However, MAS could not be performed by statistics alone. Biology especially genetics also play important role. The process in obtaining the genotype of DNA marker (genotyping) could be performed by using the help of advanced molecular biology. After the genotype is obtained, basic genetics as well as quantitative genetics play important role by giving us the theoretical base in understanding the inheritance pattern of genes in population under study. Moreover, MAS is used in optimalizing the breeding program. Hence, the concept of plant breeding as well as animal breeding also play important role. Finally, the breeding program is designed to answer the problem in many areas such as food and health. To obtain the optimal result, the concept in those areas could not be forgotten. For example, the knowledge of inheritance pattern of certain disease could help much the breeding program by giving us the guidance on how the breeding program is conducted. As an illustration, if the disease is sex linked, then the breeding program should be conducted differently compared to the disease that is not sex linked.

As a result, MAS is a concept which is collaboration of several fields, i.e. statistics, biology especially genetics, breeding, and the area where the breeding program is applied. Hence, MAS is potential as a means in collaborate those fields. Department of Statistics of IPB, due to the nature of the higher education institution, could play its role in the development of statistical concept in MAS. Of course, the collaboration with other institution such as Department of Agronomy, department related to animal science and health science as well as government institution such as BB Biogen is needed. However, breeding program is a global issue. Hence, these collaborations should be conducted also with other institution outside

IPB, such as: ITB (Indonesia), UNISBA (Indonesia), UPM (Malaysia), UKM (Malaysia), and many other universities in Indonesia and Malaysia.

## 5. CONCLUDING REMARK

The future of MAS in agriculture, food, and health problem is obvious. In food and agriculture, the technology will increase agricultural productivity, strengthen diseases resistance, reduce failure, and obtained desired result. In health sector, MAS can detect major genes as early as possible. With help of tools in molecular biology, statistics, as well as IT with computer at high level memory, MAS program is possible. To optimize potential expertise in this area, a collaboration is needed in conducting MAS involving IPB (Indonesia), ITB (Indonesia), UNISBA (Indonesia), UPM (Malaysia), UKM (Malaysia), and many other universities in Indonesia and Malaysia.

## 6. REFERENCES

Afendi, F. M., Saefuddin, A., Jusuf, M., and Martono, T., 2006, Ordinal Logistic Regression for Quantitative Trait Loci Mapping on Trait in Ordinal Scale, Presented in Statistics Seminar, Universitas Padjadjaran, 22 April 2006.

Barone, A., 2003, Molecular marker assisted selection for resistance to pathogen in tomato, Presented in MAS workshop, Univ. of Turin, 17-18 Oct 2003.

Dalbó, M.A., Ye, G. N., Weeden, N. F., Wilcox, W. F., and Reisch, B. I., 2001, Marker-assisted selection for powdery mildew resistance in grapes, J. Amer. Soc. Hort. Sci. 126(1):83-89

Dolstra, O., Denneboom, C., de Vos, A. L. F., van Loo, E. N., 2003, Marker assisted selection in improvement of quantitative traits of forage crops, Presented in MAS workshop, Univ. of Turin, 17-18 Oct 2003.

Hackett, C. A., and Weller, J. L., 1995, Genetic mapping of quantitative trait loci for traits with ordinal distributions, *Biometrics* **51**: 1252–1263.

Hayashi, T. and Awata, T., 2006, Interval mapping for loci affecting unordered categorical traits, *Heredity* **96**: 185-194.

Jansen, R. C. and Stam, P., 1994, High resolution of quantitative traits into multiple loci via interval mapping, *Genetics* **136**, 1447-1455.

Koebner, R., 2003, MAS in cereals: green for maize, amber for rice, still red for wheat and barley, Presented in MAS workshop, Univ. of Turin, 17-18 Oct 2003.

Korzun, V., 2003, Molecular markers and their applications in cereals breeding, Presented in MAS workshop, Univ. of Turin, 17-18 Oct 2003.

Lander, E. S. & Botstein, D., 1989, Mapping Mendelian factors underlying quantitative traits using RFLP linkage maps, *Genetics* **121**, 185-199.

Lacape, J. M., 2003, Targeted introgression of cotton fiber quality QTLs using molecular markers, Presented in MAS workshop, Univ. of Turin, 17-18 Oct 2003.

Liu, B. H., 1998, Statistical Genomics, CRC Press.

Martinez, O. and Curnow, R. N., 1992, Estimating the Locations and the Sizes of the Effects of Quantitative Trait Loci Using Flanking Markers, *Theor. Appl. Genet.* **85**: 480-488.

Pagnacco, G., Carta, A., 2003, Animal breeding from infinitesimal model to MAS: the case of a backcross design in dairy sheep (Sarda x Lacaune) and its possible impact on selection, Presented in MAS workshop, Univ. of Turin, 17-18 Oct 2003.

Sartika, T., Duryadi, D., Mansjoer, S. S., Saefuddin, A., Martojo, H., 2003, Gen promotor prolaktin sebagai penanda pembantu seleksi untuk mengontrol sifat mengeram pada ayam Kampung, Jurnal Ilmu Ternak dan Veteriner Vol.9(4): 239-245

Sax, K., 1923, The association of size differences with seed-coat pattern and pigmentation in *Phaseolus vulgaris*, *Genetics* **8**: 552–560.

Sonesson, A. K., 2003, Possibilities for marker assisted selection in fish breeding schemes, Presented in MAS workshop, Univ. of Turin, 17-18 Oct 2003.

Tartarini, S., 2003, Marker assisted selection in pome fruit breeding, Presented in MAS workshop, Univ. of Turin, 17-18 Oct 2003.

Xu, C., Zhang, Y. M., Xu, S., 2005, An EM algorithm for mapping quantitative resistance loci, *Heredity*, **94**: 119-128.

Xu, S., and Atchley, W.R., 1996, Mapping quantitative trait loci for complex binary disease using line crosses, *Genetics* **143**: 1417–1424.

Xu, S., 1998, Further Investigation on the Regression Method of Mapping Quantitative Trait Loci, *Heredity*, 80: 364-373.

Xu, S., 1998, Iteratively reweighted least squares mapping of quantitative trait loci, *Behavior Genetics*, **28**: 341-355.

Xu, S., Yi, N., Burke, D., Galecki, A., Miller, R. A., 2003, An EM algorithm for mapping binary disease loci: application to fibrosarcoma in a four way cross mouse family, *Genet. Res.*, **82**:127-138.

Yi, N., Xu, S., George, V., Allison, D. B., 2004, Mapping multiple quantitative trait loci for ordinal traits, Behavior Genetics, **34**: 3-15.

Zeng, Z. B., 1994, Precision mapping of quantitative trait loci, *Genetics* **136**: 1457-1468.