*Euphytica* **122:** 9–18, 2001.



# Mixed inheritance model for resistance to agromyzid beanfly (*Melanagromyza sojae* Zehntner) in soybean

Jiankang Wang[1,3] & Junyi Gai[2]
[1]*Laboratory Center, Henan Academy of Agricultural Sciences, Zhengzhou, Henan 450002, PR China; E-mail: hnnkywjk@public2.zz.ha.cn;* [2]*Soybean Research Institute, Nanjing Agricultural University, Nanjing 210095, PR China; E-mail: nausri@public1.ptt.js.cn;* [3]*Present address: CIMMYT Wheat Program, Apdo. Postal 6-641, 06600 Mexico, D.F., Mexico; E-mail: j.k.wang@cgiar.org*

Received 5 January 1998; accepted 6 April 2000

*Key words:* agromyzid beanfly (*Melanagromyza sojae* Zehntner), graphical analysis, mixed major gene and polygene inheritance model, segregation analysis, soybean

**Summary**

A quantitative trait could be controlled by a few major genes and many polygenes. Distinguishing the effects of major genes from polygenes and/or environments is important for understanding the expression of a major gene in relation to its genetic background, and for predicting the segregation of a cross in breeding. Our objective was to re-analyze the resistance of soybean to agromyzid beanfly by a mixed inheritance model. Number of insects in stem (NIS) was used as an indicator of resistance. The previous result from the segregation ratio of resistance and susceptibility was that resistance was controlled by one dominant gene. The major results from the mixed inheritance model were (1) the inheritance of resistance was controlled by one major gene along with minor genes; (2) Additive and dominance effects of minor genes were generally less than those of the major gene and varied among crosses, indicating different minor gene systems; (3) Heritability was higher for the major gene than for the minor genes; (4) The $F_2$ plants and $F_{2:3}$ lines were classified into appropriate genotypes according to their posterior probabilities and the critical value to distinguish resistant and susceptible plants was given for NIS based on the classification. These results indicated that mixed major gene and polygene genetic analysis was superior to the frequently used classical Mendelian method.

## Introduction

For quantitative traits showing major gene effects combined with continuous variation, it is assumed that apart from the major gene, polygenes are included in the inheritance system. Many genetic experiments in plants and animals and QTL (quantitative trait locus) mapping have also indicated differences in the magnitude of the effects of individual QTLs. Major genes are those genes in the QTL system with relatively large effects. Therefore, the inheritance system of a quantitative trait might consist of both a few major genes and a number of polygenes. This genetic model is called a mixed major gene and polygene inheritance model (or mixed inheritance model or mixed genetic model). The classic Mendelian method can be used to analyze the inheritance of major genes contained in the system by classifying different segregation types around the valley point or the critical value separating two modes. This approach, however, depends on some artificial factors in the analysis. Furthermore, genetic backgrounds may differ, so there may not be a common critical value for all crosses.

QTL mapping has been widely used to uncover the genetic structure of variation in agronomic and economical traits relevant to breeding. This strategy requires molecular data. However, even without the aid of molecular data, a number of methods have been used to analyze the mixed inheritance model in human and animal populations (Elston & Stewart, 1973; Morton & MacLean, 1974; Elston, 1984; Famula, 1986; Hoeschele, 1988; Knott et al., 1991; Guo &

Thompson, 1992; Shoukri & McLachlan, 1994; Janss et al., 1995). But due to different mating systems and breeding objectives between plants and animals, these methods are not generally applicable in plant quantitative genetic analysis. Wang (1996), Wang & Gai (1998) and Gai & Wang (1998) developed a segregation analysis method to apply the mixed inheritance model in the study of variation for quantitative traits in plants with the aim to estimate genetic parameters describing the variation of a quantitative trait. The principle of the method can be described as follows. Firstly, it is supposed that trait variation in each segregating population is due to the variation in the distribution of major genes modified by polygenes and environments. Secondly, the major gene heritability and polygene heritability are defined, and standard curves of the mixture distributions under various genetic conditions are drawn. Thus, comparing practical frequency distributions with the standard curves can approximate the inheritance model of a quantitative trait. This process is called graphical analysis (Wang, 1996). Thirdly, likelihood functions under various possible genetic models are established, maximum likelihood estimates of parameters contained in each model are calculated through the EM algorithm (Dempster et al., 1977; McLachlan, 1988), and the best fitting genetic model and its parameter estimates are chosen by Akaike's information criterion (Akaike, 1977), likelihood ratio test and tests of goodness of fit. Finally, each individual in segregating generations is classified into a suitable component distribution using Bayesian posterior probabilities. Considering the difficulty of crossing in soybean, joint analyses based on the five generations $P_1$, $F_1$, $P_2$, $F_2$, $F_{2:3}$ (Wang & Gai, 1998) and the six generations P1, $F_1$, $P_2$, $B_1$, $B_2$, $F_2$ (Gai & Wang, 1998) were developed separately.

Agromyzid beanfly (*Melanagromyza sojae* Zehntner) is one of the most important pests in soybean production in regions south of the Great Wall of China. It infects almost all soybean plants and causes great yield losses every year. Breeding new cultivars with resistance to beanfly is the most effective way to control its damage (Chiang & Norris, 1983). Wei et al. (1989) made crosses between resistant lines and susceptible lines and investigated the number of insects in the stem and the total number of insects. They found a common critical value among crosses to classify individuals in $F_2$ and $F_{2:3}$ families into resistant or susceptible. From the segregation ratio of resistance and susceptibility, they concluded that resistance was controlled by one dominant gene. They also showed the absence of any cytoplasmic effect. Quantitative variation within categories indicated that resistance was modified by polygenes. In this paper, the graphical analysis and joint segregation analysis of multiple generations P1, $F_1$, $P_2$, $F_2$ and $F_{2:3}$ will be used to reanalyze the genetic data in Wei et al. (1989), with the objectives to identify whether inheritance of resistance to beanfly is controlled by one major gene or mixed one major gene and minor genes, and furthermore, to estimate the genetic parameters related to both major gene and minor genes.

## Materials and methods

### *Materials*

Three crosses, JNCWD × HJQDHY (I), WXCJGJ × PXTED (II) and 1138–2 × PXTED (III), between resistant lines JNCWD, WXCJGJ and 1138–2 and susceptible lines HJQDHY and PXTED were made at Jiangpu Experimental Station of Nanjing Agricultural University in the summer of 1985. All the parental materials are pure lines selected from land races in Jiangsu, China. The $F_1$s were planted in Hainan Island, China in the winter of 1985, the $F_2$s were grown in the spring of 1986 also in Hainan Island, and then all the generations P1, $F_1$, $P_2$, $F_2$ and $F_{2:3}$ were planted at Jiangpu Experimental Station in the summer of 1996. An experimental design similar to a split plot was used, with crosses in main plots and parents and hybrid generations in sub-plots. The resistance indices were the numbers of insects in the stem (NIS) and the total number of insects (NIP, which is NIS plus the number of insects in the petiole), which were investigated in the fall at the flowering stage for every generation in the three crosses.

### *Graphical analysis*

For a mixed inheritance model, the phenotypic value ($p$) can be expressed as the summation of population mean ($m$), major gene effect ($g$), polygene effect ($c$) and environmental effect ($e$), i.e., $p = m + g + c + e$ (Morton & MacLean, 1974), where $g$ is different for different major gene genotypes, and $c$ and $e$ are normally distributed variables. So, the phenotypic variation ($\sigma_p^2$) can be expressed as major gene variation ($\sigma_{mg}^2$), polygenic variation ($\sigma_{pg}^2$) and environmental variation ($\sigma_e^2$). Therefore, we can define major gene heritability ($h_{mg}^2$) and polygene heritability ($h_{pg}^2$)

as $h^2_{mg} = \sigma^2_{mg}/\sigma^2_p$ and $h^2_{pg} = \sigma^2_{pg}/\sigma^2_p$, respectively. For simplicity, we suppose $\sigma^2_p$=1. So, for the mixed one major gene and polygene inheritance model, the major gene and polygenic heritabilities in the $F_2$ generation derived from a cross between two inbreds can be represented as:

$$h^2_{mg} = \tfrac{1}{2}a^2 + \tfrac{1}{4}d^2 \text{ and } h^2_{pg} = V^*_A + V^*_D,$$

where $a$ and $d$ are the additive and dominance effects of the major gene, and $V^*_A$ and $V^*_D$ are the sums of the additive and dominance variances of individual minor genes (Kearsey & Pooni, 1996). Hereafter, we will use $A$, $D$ and $E$ to represent $V^*_A$, $V^*_D$ and $\sigma^2_e$, respectively. Supposing r to be the dominance degree of the major gene, and the average dominance ratio of polygenes $\sqrt{2D/A}$ to be 0.5 (the distribution shapes of different generations do not change too much for different values of $\sqrt{2D/A}$, and this value is under the cond ition that each minor gene has the dominance degree 0.5), we have

$$a = \frac{2\sqrt{h^2_{mg}}}{\sqrt{2+r^2}}, d = ra, A = \tfrac{8}{9}h^2_{pg}, D = \tfrac{1}{8}A, \text{ and}$$
$$E = 1 - h^2_{mg} - h^2_{pg}.$$

Thus, we can have the distribution of $F_2$, $B_1$, $B_2$ and $F_{2:3}$ as follows.

$$F_2 : \tfrac{1}{4}N(-a, A + D + E) + \tfrac{1}{2}N(d, A + D + E) + \tfrac{1}{4}N(a, A + D + E)$$

$$B_1 : \tfrac{1}{2}N(-a, \tfrac{1}{2}A + D + E) + \tfrac{1}{2}N(d, \tfrac{1}{2}A + D + E)$$

$$B_2 : \tfrac{1}{2}N(d, \tfrac{1}{2}A + D + E) + \tfrac{1}{2}N(a, \tfrac{1}{2}A + D + E)$$

$$F_{2:3} : \tfrac{1}{4}N(-a, A + \tfrac{1}{4}D + \tfrac{1}{n}E) + \tfrac{1}{2}N(\tfrac{1}{2}d, \tfrac{1}{2}a^2 + \tfrac{1}{4}d^2 + A + \tfrac{1}{4}D + \tfrac{1}{n}E) + \tfrac{1}{4}N(a, A + \tfrac{1}{4}D + \tfrac{1}{n}E)$$

where $N$ is the symbol of the normal distribution, the bracketed items are the mean and variance of the normal distribution, the summation of two or three normal distributions indicates a normal mixture with the number in front of symbol $N$ being the propo rtion of the component distribution in the mixture, and $n$ is the sample size in each $F_{2:3}$ family. For backcross generations, we also suppose the cross product of the additive and the dominance effects of polygenes to be zero. So, we can draw the distribu tion curves of the above generations under different genetic parameters. From the above distributions, we can also calculate the major gene and polygenic heritabilities in all generations except $F_2$. In our research, we set $h^2_{mg}$ = {0.9, 0.8, 0.7, 0.6, 0.5, 0.4, 0.3, 0.2}, $h^2_{pg}$ = {0, 0.1, 0.2, 0.3}, and $r$ = {0, 0.5, 1.0, 1.5}. For $F_{2:3}$, we have three different sample sizes in a family, i.e., 5, 10 and infinity. Figure 1, as an example, shows curves for the mixed one complete dominant major gene and polygene inheritan ce model. These curves are called standard curves.

From those curves, we can have the following impressions. (1) Phenotypically normal distribution does not implicate a polygene system; (2) If skewness and multi-modality are the evidence for the existence of a major gene, different generations have different powers to detect the major gene; (3) Comparing practical frequency distributions with the standard curves can approximate the inheritance model of a quantitative trait.

*Joint segregation analysis*

The main principle and steps in joint segregation analysis of multiple generations P1, $P_2$, $F_1$, $F_2$ and $F_{2:3}$ will be described as follows.

*(1) Genetic model establishment.* Under some assumptions (Gai & Wang, 1998), three classes of genetic model were established, i.e., one major gene inheritance model (1MG), polygenic inheritance model (PG) and mixed one major gene and polygene inheritance model (MX). Considering the additive and dominance relationships of both the major gene and polygenes, we also established different model types within each model class. Table 1 lists these genetic models and the genetic parameters in each model. There are two kinds of parameters, which require clarification in the joint segregation analysis. One is of component parameters, which are used to describe the characteristics of a mixture distribution. Another is of genetic parameters. In the joint segregation analysis, we will first estimate the component parameters and then estimate genetic parameters from component parameters.

*(2) Component parameter estimation.* Maximum likelihood estimates of component parameters in each genetic model were carried out by the EM algorithm. Let's take the MX-class model as an example. Supposing A and a to be the two alleles of the major locus and AA, aa and Aa to be the genotypes of parents and $F_1$, respectively, the $F_2$ genotypes will be a 1:2:1 mixture

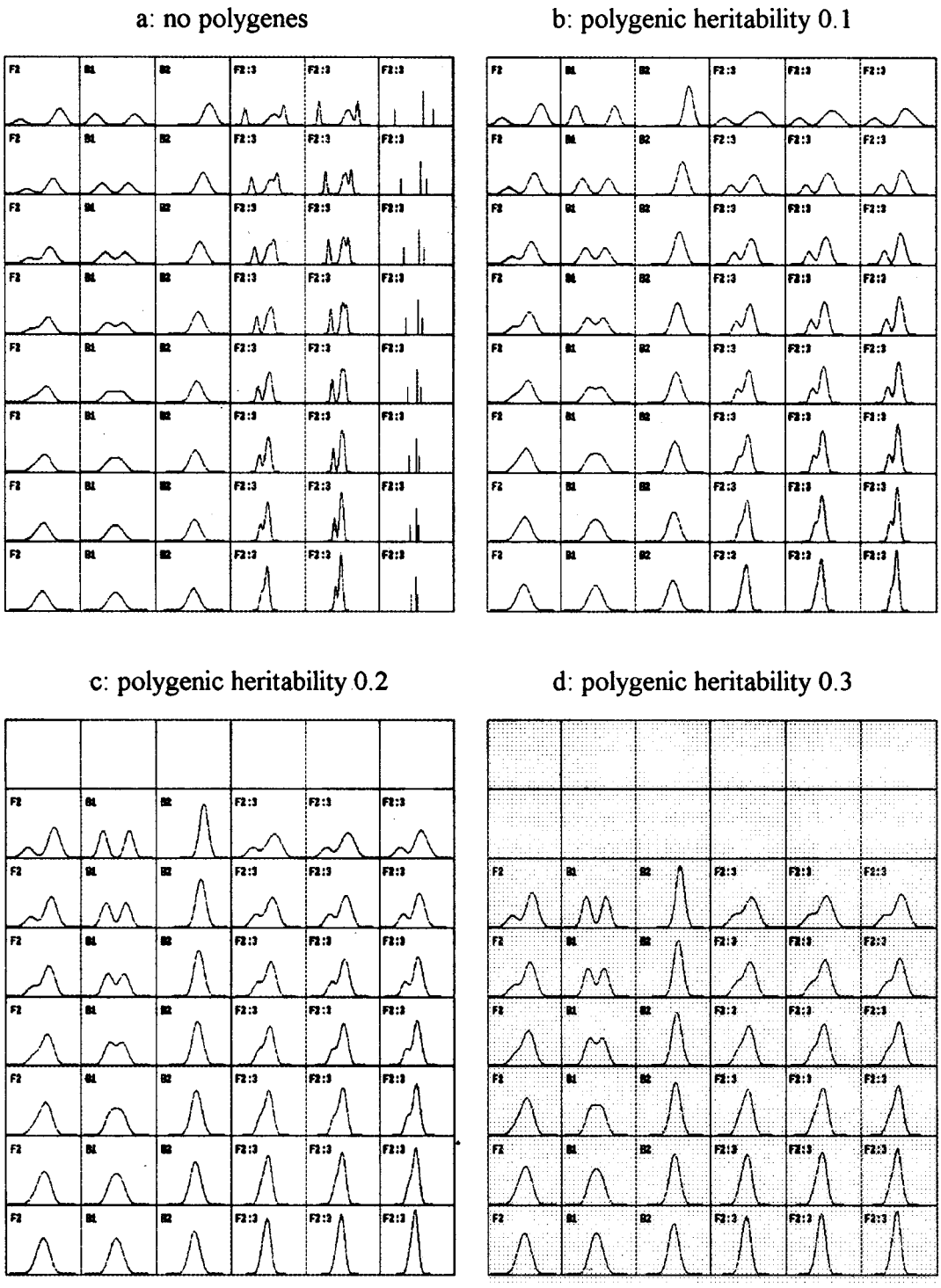


*Figure 1.* Distribution curves of $F_2$, $B_1$, $B_2$ and $F_{2:3}$ for mixed one dominant major gene and polygene inheritance model. The major gene heritabilities in $F_2$ are from 0.9 to 0.2 in rows in each sub-figure.

*Table 1.* Genetic models and their parameters in the joint segregation analysis of the five generations P1, $F_1$, $P_2$, $F_2$ and $F_{2:3}$

| Model class | Model type and its implication | Genetic parameters |
|---|---|---|
| One major gene (1MG) | 1MG-AD: additive and dominant major gene | $m, a, d, \sigma^2_{mg}, \sigma^2_e$ |
| | 1MG-D: dominant major gene | $m, a, d(=a), \sigma^2_{mg}, \sigma^2_e$ |
| | 1MG-A: additive major gene | $m, a, d(=0), \sigma^2_{mg}, \sigma^2_e$ |
| | 1MG-ND: negative dominant major gene | $m, a, d(=-a), \sigma^2_{mg}, \sigma^2_e$ |
| Polygenes (PG) | PG-ADI: additive, dominance and epistasis polygenes | $m, [a], [d], [aa], [dd], \sigma^2_{pg}, \sigma^2_e$ |
| | PG-AD: additive and dominance polygenes | $m, [a], [d], \sigma^2_{pg}, \sigma^2_e$ |
| Mixed one major gene and polygenes (MX) | MX-AD-ADI: mixed one major gene and polygenes | *See text for details* |
| | MX-AD-AD: mixed one major gene and additive-dominant polygenes | $m, a, d, [a], [d], \sigma^2_{mg}, \sigma^2_{pg}, \sigma^2_e$ |
| | MX-D-AD: mixed one dominance major gene and additive-dominant polygenes | $m, a, d(=a), [a], [d], \sigma^2_{mg}, \sigma^2_{pg}, \sigma^2_e$ |
| | MX-A-AD: mixed one additive major gene and additive-dominant polygenes | $m, a, d(=0), [a], [d], \sigma^2_{mg}, \sigma^2_{pg}, \sigma^2_e$ |
| | MX-ND-AD: mixed one negative dominant major gene and additive-dominant polygenes | $m, a, d(=-a), [a], [d], \sigma^2_{mg}, \sigma^2_{pg}, \sigma^2_e$ |

$m$: population mean; $a$: additive effect of the major gene; $d$: dominance effect of the major gene; $[a]$: additive effect of polygenes; $[d]$: dominance effect of polygenes; $\sigma^2_{mg}$: major gene variance; $\sigma^2_{pg}$: polygenic variance; $\sigma^2_e$: environmental variance.

of AA, Aa and aa, and so will be the genotypes for $F_{2:3}$ (the genotype of a $F_{2:3}$ family is represented by that of its $F_2$ parent). Supposing observations from P1, $F_1$ and $P_2$ are distributed as normal distributions $N(\mu_1, \sigma^2)$, $N(\mu_2, \sigma^2)$ and $N(\mu_3, \sigma^2)$, respectively, $F_2$ observations and $F_{2:3}$ family means will be distributed as 1:2:1 mixtures of three normal distributions, which can be denoted as,

$$F_2 : (1/4)N(\mu_{41}, \sigma^2_4) + (1/2)N(\mu_{42}, \sigma^2_4) + (1/4)N(\mu_{43}, \sigma^2_4) \text{ and}$$

$$F_{2:3} : (1/4)N(\mu_{51}, \sigma^2_{51}) + (1/2)N(\mu_{52}, \sigma^2_{52}) + (1/4)N(\mu_{53}, \sigma^2_{53}),$$

where $\mu_{41}$, $\mu_{42}$ and $\mu_{43}$ represent means of the $F_2$ individuals having major genotypes AA, Aa and aa, respectively, and $\mu_{51}$, $\mu_{52}$ and $\mu_{53}$ represent means of the $F_{2:3}$ families derived from AA, Aa and aa $F_2$ individuals, respectively. $\sigma^2$ is the combined variance of the three non-segregating generations P1, $F_1$ and $P_2$, $\sigma^2_4$ is the common variance of components in $F_2$, $\sigma_{51}$ and $\sigma_{53}$ are the common variances of components having mean $\mu_{51}$ and $\mu_{53}$, and $\sigma^2_{52}$ is the variance of the component having mean $\mu_{52}$. For reasons of simplicity, $\sigma^2_{51}$, $\sigma^2_{52}$ and $\sigma^2_{53}$ are assumed to be unrelated with the component means in $F_2$ and $F_{2:3}$. Component parameters in this model consist of $\mu_1$, $\mu_2$, $\mu_3$, $\mu_{41}$, $\mu_{42}$, $\mu_{43}$, $\mu_{51}$, $\mu_{52}$, $\mu_{53}$, $\sigma^2$, $\sigma^2_4$, $\sigma^2_{51}$, $\sigma^2_{52}$ and $\sigma^2_{53}$.

*(3) Model selection and test.* From Akaike's Information Criterion (Akaike, 1977)) the model with the least $AIC$ value is the best fitting model. Here, $AIC = -2L_c(\Phi) + 2N$, where $L_c(\Phi)$ is the maximum logarithm likelihood and $N$ is the number of independent parameters in a genetic model. We first use $AIC$ to select the best fitting model class. Then we use the likelihood ratio test to test whether two genetic models in a model class show significant difference. If there is no significant difference, we will choose the one with lower parameter. By the above two steps, we can always select the best fitting model among the 11 models. But there may be a case where the real model is not included in all of these models. We then use tests of goodness of fit to further test the fitnes s of the selected model. If the selected model passes all the goodness-of-fit tests, we will say this model is suitable for the genetic data. Or, other approaches or genetic models should be considered for the genetic data.

*(4) Genetic parameter estimation.* If all kinds of bigenic interactions are considered for model MX-AD-ADI, we will have the following nine equations.

$$\mu_1 = m + a + [a] + [aa]' + [aa],$$

$$\mu_2 = m + d + [d] + [dd]' + [dd],$$

$$\mu_3 = m - a - [a] + [aa]' + [aa],$$

$$\mu_{41} = m + a + (1/2)[d] + (1/2)[ad]' + (1/4)[dd],$$

*Table 2.* Frequency distribution of the number of insects in stem (NIS)

| Cross | Generation | NIS | | | | | | | | | | Sample size | Mean | Variance |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | 0 | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | | | |
| I | P1 | 3 | 7 | 4 | 6 | | | | | | | 20 | 1.65 | 1.13 |
| JNCWD | $F_1$ | 4 | 4 | 6 | 6 | | | | | | | 20 | 1.70 | 1.21 |
| ×HJQDHY | $P_2$ | | | | | | 4 | 6 | 3 | 3 | 4 | 20 | 6.85 | 2.03 |
| | $F_2$ | 26 | 36 | 45 | 34 | 25 | 12 | 5 | 2 | 1 | 3 | 189 | 2.47 | 3.46 |
| | $F_{2:3}$ | 7 | 61 | 5 | | 9 | 17 | 1 | | | | 100 | 2.36 | 2.99 |
| II | P1 | 4 | 7 | 7 | 2 | | | | | | | 20 | 1.35 | 0.83 |
| WXCJGJ | $F_1$ | 1 | 11 | 5 | 3 | | | | | | | 20 | 1.50 | 0.65 |
| ×PXTED | $P_2$ | | | | | 5 | 3 | 6 | 4 | | 2 | 20 | 5.85 | 2.23 |
| | $F_2$ | 18 | 52 | 28 | 14 | 11 | 10 | 8 | 5 | | | 146 | 2.24 | 3.54 |
| | $F_{2:3}$ | 5 | 52 | 5 | | 5 | 15 | 9 | | | | 91 | 2.71 | 3.90 |
| III | P1 | 2 | 5 | 5 | 8 | | | | | | | 20 | 1.90 | 1.19 |
| 1138-2 | $F_1$ | 5 | 7 | 5 | 3 | | | | | | | 20 | 1.30 | 1.01 |
| ×PXTED | $P_2$ | | | | | 6 | 6 | 5 | 2 | 1 | | 20 | 5.30 | 1.31 |
| | $F_2$ | 24 | 36 | 39 | 37 | 20 | 20 | 12 | 9 | 3 | | 200 | 2.82 | 4.10 |
| | $F_{2:3}$ | 16 | 68 | 15 | | 5 | 18 | 5 | | | | 127 | 2.26 | 3.08 |

$$\mu_{42} = m + d + (1/2)[d] + (1/2)[ad]' + (1/4)[dd],$$

$$\mu_{43} = m - a + (1/2)[d] - (1/2)[ad]' + (1/4)[dd],$$

$$\mu_{51} = m + a + (1/4)[d] + (1/4)[ad]' + (1/16)[dd],$$

$$\mu_{52} = m + d + (1/4)[d] + (1/4)[dd]' + (1/16)[dd], \text{ and}$$

$$\mu_{53} = m - a + (1/4)[d] - (1/4)[ad]' + (1/16)[dd],$$

where $m, a$ and $d$ have the same meaning as before, $[a]$ and $[d]$ are the additive and dominance effects of polygenes, $[aa]'$, $[ad]'$ and $[dd]'$ are the three kinds of interaction between the major gene and the polygene system, and $[aa]$ and $[dd]$ are the two kinds of interaction among polygenes (Kearsey & Pooni, 1996). It is impossible to estimate these 10 parameters from those equations. So, other generations will be required to estimate all the ten genetic effects. Here, we ignore all the interaction effects and only estimate $m, a, d, [a]$ and $[d]$ from estimates of component means by the least squares method.

$\sigma^2$, the weighted average of variances of generations P1, $F_1$ and $P_2$, can be used to estimate the environmental variance in segregating generations, i.e., $\sigma^2$ can be treated as the environmental variance in $F_2$ and $\sigma^2/n$ ($n$ is the sample size of a $F_{2:3}$ family and $n = 5$ in this genetic experiment) as the environmental variance of $F_{2:3}$ family means. Moreover, $\sigma_4^2$ can simply be viewed as the $F_2$ variation excluding the major gene variation. So, $\sigma_4^2 - \sigma^2$ can be considered as the polygenic variance in $F_2$, when $\sigma_4^2 > \sigma^2$; otherwise, the polygenic variance in $F_2$ is set to 0. Similarly, $\sigma_{51}^2 - \sigma^2/n$ can simply be viewed as the estimate of polygenic variance in $F_{2:3}$, when $\sigma_{51}^2 > \sigma^2/n$; otherwise, the polygenic variance in $F_{2:3}$ is set to 0.

*(5) Posterior classification.* While the best fitting genetic model is selected, the posterior probabilities of an $F_2$ individual and $F_{2:3}$ family belonging to different components can be calculated at the same time. This allows classification of each $F_2$ individual and $F_{2:3}$ family by Bayesian rules.

## Results

### *Distribution characteristics of NIS and the graphical analysis of the inheritance*

Both NIS and NIP can be used as an index of resistance to beanfly. Because NIS showed higher correlations among years, lower error variance and higher

*Table 3.* The best fitting genetic model and its *AIC* value and estimates of component parameters

| Cross | Model | *AIC* | $\mu_1$ | $\mu_2$ | $\mu_3$ | $\mu_{41}$ | $\mu_{42}$ | $\mu_{43}$ | $\mu_{51}$ | $\mu_{52}$ | $\mu_{53}$ | $\sigma^2$ | $\sigma_4^2$ | $\sigma_{51}^2$ | $\sigma_{52}^2$ |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| I | MX-AD-ADI | 1220.75 | 1.65 | 1.70 | 6.85 | 1.82 | 1.83 | 4.84 | 1.30 | 1.36 | 5.13 | 1.48 | 1.93 | 0.15 | 0.58 |
| II | MX-AD-ADI | 1029.10 | 1.35 | 1.50 | 5.85 | 1.32 | 1.37 | 5.16 | 1.39 | 1.47 | 5.54 | 1.26 | 0.96 | 0.27 | 0.95 |
| III | MX-AD-ADI | 1364.05 | 1.90 | 1.30 | 5.30 | 2.25 | 1.72 | 5.49 | 1.30 | 1.41 | 5.44 | 1.19 | 1.59 | 0.18 | 0.79 |

heritability than NIP, Gai et al. (1989) proposed the use NIS as the major index of resistance. Both NIS and NIP were investigated and analyzed in the present research, and similar results were attained. Therefore, the NIS index will mainly be used in the paper.

From frequency distributions of NIS in Table 2, we can see that the $F_1$ shows a trend toward the resistant parent in all the three crosses. $F_2$ and $F_{2:3}$ are genetically segregating populations, but the variation in $F_2$ is greater than that in $F_{2:3}$ except for cross II. All three $F_2$'s demonstrate single mode distributions, in contrast to $F_{2:3}$ distributions showing bi-modality. The component distribution with a lower mean occupies a larger proportion in the $F_{2:3}$ mixture than that with the higher mean. By comparing these distributions with the standard curves in Figure 1, we found that the fifth and sixth rows of Figure 1.a, and the fourth and fifth rows of Figure 1.b are similar to the $F_2$ and $F_{2:3}$ frequency distributions in Table 2. No curves in Figure 1.c and Figure 1.d are similar to these distributions. This means the polygenic heritability in $F_2$ cannot exceed 20%. We also found other similar standard curves where the major gene displays partial or over dominance. So, we can conclude from graphical analysis that a mixed one major gene and polygene inheritance model fits the inheritance of resistance to agromyzid beanfly. There exist dominant effects, but the degree of dominance cannot be determined from graphical analysis. The major gene heritability ($h_{mg}^2$) in $F_2$ is about 50%, and the polygenic heritability ($h_{pg}^2$) in $F_2$ is less than 20%; the major gene heritability ($h_{mg}^2$) in $F_{2:3}$ is about 80%, and the polygenic heritability ($h_{pg}^2$) in $F_{2:3}$ is less than 10%.

*Mixed inheritance model for the resistance of soybean to agromyzid beanfly*

From the estimates of *AIC*, we can see that model MX-AD-ADI in all three crosses shows the lowest *AIC* value among the 11 models. So, it is the best fitting genetic model to explain the inheritance of resistance to beanfly according to Akaike's information criterion. The results from likelihood ratio tests between model MX-AD-ADI and other MX models indicate that MX-AD-ADI is more suitable than MX-AD-AD, MX-D-AD, MX-A-AD and MX-ND-AD. The results from tests of goodness of fit also indicate its fitness (results not shown). So we can reasonably deduce that the trait of resistance in the three crosses is determined predominantly by one major gene in combination with polygenes. And, there may exist interaction among polygenes and even between the major gene and the polygene system. This result has extended the previous one from classical Mendelian analysis. The estimates of component parameters of model MX-AD-ADI are given in Table 3.

*Table 4.* Genetic parameters of resistance of soybean to beanfly from the joint analysis of P1, $P_2$, $F_1$, $F_2$ and $F_{2:3}$

| Generation | Estimate | Cross I | Cross II | Cross III |
|---|---|---|---|---|
| All | $a$ | –1.71 | –2.00 | –1.85 |
| | $d$ | –1.79 | –2.20 | –2.57 |
| | $d/a$ | 1.05 | 1.10 | 1.39 |
| | $[a]$ | –0.88 | –0.25 | 0.15 |
| | $[d]$ | –0.49 | 0.14 | 0.65 |
| $F_2$ | $\sigma_p^2$ | 3.46 | 3.54 | 4.10 |
| | $h_{mg}^2$ (%) | 44.2 | 72.9 | 61.2 |
| | $h_{pg}^2$ (%) | 13.0 | 0 | 9.8 |
| $F_{2:3}$ | $\sigma_p^2$ | 2.99 | 3.90 | 3.08 |
| | $h_{mg}^2$ (%) | 95.0 | 93.1 | 94.2 |
| | $h_{pg}^2$ (%) | 0 | 0.5 | 0 |

$a$: additive effect of the major gene; $d$: dominance effect of the major gene; $[a]$: additive effect of polygenes; $[d]$: dominance effect of polygenes; $\sigma_p^2$: phenotypic variance; $h_{mg}^2$: major gene heritability; $h_{pg}^2$: polygenic heritability.

*Genetic parameters of the resistance to agromyzid beanfly*

From Table 3, we can obtain estimates of genetic parameters in model MX-AD-ADI (Table 4). From Table 4, we see that additive effects of the major gene are –2.00∼–1.71 heads/plant, and the degrees of domin-

*Table 5.* Posterior probability and estimated genotype of $F_2$ and $F_{2:3}$ of cross II

| NIS | Frequency | Posterior probability | | | Estimated genotype |
|---|---|---|---|---|---|
| | | AA | Aa | aa | |
| $F_2$ | | | | | |
| 0–2 | 98 | 0.35–0.32 | 0.65–0.67 | 0.00 | Aa+AA |
| 3 | 14 | 0.28 | 0.61 | 0.11 | Aa+AA+aa |
| 4 | 11 | 0.04 | 0.09 | 0.86 | aa+Aa |
| 5–7 | 23 | 0.00 | 0.00 | 1.00 | aa |
| $F_{2:3}$ | | | | | |
| 0.80 | 5 | 0.45 | 0.55 | 0.00 | Aa+AA |
| 1.00–1.60 | 43 | 0.53–0.58–0.55 | 0.47–0.42–0.45 | 0.00 | AA+Aa |
| 1.80–2.00 | 14 | 0.47–0.36 | 0.53–0.64 | 0.00 | Aa+AA |
| 4.20–4.60 | 3 | 0.00 | 0.95–0.57 | 0.07–0.43 | Aa+aa |
| 4.80–5.00 | 5 | 0.00 | 0.26–0.06 | 0.74–0.94 | aa+Aa |
| 5.20–6.60 | 20 | 0.00 | 0.02–0.00 | 0.98–1.00 | aa |

ance are 1.05∼1.39. The additive ([$d$]) and dominance ([$h$]) effects of polygenes can be either positive or negative among crosses, which shows differences in their polygenic background. But, in general, the additive and dominance effects of polygenes are less than those of the major gene. The major gene heritabilities in $F_2$ were 44.2∼72.9% and those in $F_{2:3}$ were 93.1∼95.0%, which are greater than those in $F_2$. The polygenic heritabilities in $F_2$ were 0∼13% and those in $F_{2:3}$ were 0∼0.5%, which are less than those in $F_2$.

Some polygenic heritabilities are estimated as 0, but the model test shows polygenic variation. There might be two reasons for such outcomes. One is that it is not appropriate to view $\sigma^2$ as the estimate of the environmental variation in segregating genera tions. The other is due to epistasis effects between the major gene and minor genes.

*Table 6.* The estimated $F_2$ genotypes from posterior probability analysis for $F_2$ and $F_{2:3}$

| Cross | Generation | NIS | Number of plants (lines) | Estimated genotype |
|---|---|---|---|---|
| I | $F_2$ | 0–3 | 141 | Aa+AA |
| | | 4–9 | 48 | aa |
| | $F_{2:3}$ | 0.6–2.6 | 73 | Aa+AA |
| | | 4.4–6.0 | 27 | aa |
| II | $F_2$ | 0–3 | 112 | Aa+AA |
| | | 4–7 | 34 | aa |
| | $F_{2:3}$ | 0.8–4.6 | 65 | Aa+AA |
| | | 4.8–6.6 | 26 | aa |
| III | $F_2$ | 0–3 | 136 | Aa+AA |
| | | 4–8 | 64 | aa |
| | $F_{2:3}$ | 0.0–2.8 | 99 | Aa+AA |
| | | 4.8–6.6 | 28 | aa |

*Classification of major gene genotypes in segregating generations*

The classification process is the same for each cross. As an example, the classification of cross II is given in Table 5. Results of all the three crosses are summarized in Table 6. In Table 5, taking the 0.05 probability as an impossible event, in dividuals in $F_2$ having 0∼3 insects can be classified as genotype Aa, but some of them may be genotype AA or aa with a probability more than 0.05; individuals having insect numbers more than 3 can be classified as aa, but some of them having 4 insects have a probability more than 0.05 of being Aa. For $F_{2:3}$, lines having the mean insect number 0.8∼2.0 can be viewed as coming from AA or Aa, and it is difficult to tell whether such families come from AA or Aa $F_2$ individuals; lines with a mean insect number more than 4.6 may be derived from aa $F_2$ individuals, but some with mean insect number 4.6∼5.0 still have a small chance coming from Aa $F_2$ individuals.

In Table 6, the ratios of Aa+AA to aa classified by posterior probabilities conform to the ratio 3:1 except the $F_2$ in cross III. From the $F_2$ classification data, a resistance threshold value can be attained. For trait NIS, the critical value was 3 heads/plant.

**Discussion and conclusions**

By imposing the graphical analysis and the statistical method of mixed inheritance model on the genetic data of crosses between resistant and susceptible parents, the following conclusions were made. (1) Inheritance of the resistance indicated by NIS can be fitted using the mixed one major gene and polygene inheritance model. (2) Resistance is almost completely dominant with the degree of dominance 1.05∼1.39. (3) The estimated additive effects of the major gene in each cross ranged from –2.00∼–1.71 he ads/plant. Differences might reflect variation in the polygenic background. (4) The additive and dominance effects of polygenes can be either positive or negative among crosses, showing variation in the genetic background. In general, the additive and dominance effects of polygenes were less than those of the major gene. (5) Major gene heritabilities in $F_2$ are 44.2%∼72.9%, and are 93.1%∼95.0% in $F_{2:3}$. In contrast, polygenes contribute only a small fraction to the phenotypic variation, resulting in low polygenic heritabilities of 0∼13.0% in $F_2$, and 0∼0.5% in $F_{2:3}$, respectively. (6) $F_2$ individuals and $F_{2:3}$ families can be classified into different major gene genotypes by Bayesian rules from posterior probabilities. From the classification, a critical line to distinguish resistant and susceptible components can be determined. A common critical value ($3 < x < 4$) was obtained in the three crosses for indicator NIS. Segregation of the effects of major genes from polygenes and/or environments is important for understanding the expression of a major gene in relation to its genetic background, and for predicting the segregation of a cross in breeding (Jiang et al., 1994). For the joint segregation analysis of the mixed major gene and polygene inheritance model based on five generations including P1, $F_1$, $P_2$, $F_2$ and $F_{2:3}$, manual crossing work is needed only once to generate such a set of materials, i.e. hybridization between the two parents. So the method is especially suitable for the quantitative genetic analysis of those crops where it is not easy to obtain backcross seeds. For those where it is easy to obtain hybrid seeds, the joint segregation analysis based on the six generations P1, $P_2$, $F_1$, $F_2$, $B_1$ and $B_2$ is preferred (Gai & Wang, 1998; Wang et al., 2001).

The current investigations indicate that more information can be obtained from the mixed genetic analysis. In contrast to the traditional genetic analysis method to find a critical resistance threshold value, the joint segregation analysis allows the detection of a less subjective criterion resulting in more results. Two methods are included in the current paper. One is graphical analysis. By comparing frequency distributions with the standard curves drawn by Wang (1996), whether the major gene exists, the heritabilities of major gene and polygenes can be approximated. The other is the joint segregation analysis, by which not only the genetic model can be identified but genetic parameters included in the genetic model can also be estimated. When applying the two methods to the same genetic data, graphical analysis can be looked upon as an auxiliary to joint segregation analysis. Genetic data used in the joint segregation analysis consist of plant-level data from $P_1$, $F_1$, $P_2$ an d $F_2$ and family-level data from $F_{2:3}$. Generally, family data are more precise than plant data because less environmental errors are included in family means. In order to improve overall precision, the experimental errors should be minimized. The sample size is also required when using the segregation analysis. From experience, non-segregating generations require at least 30 plants and segregating generations, 100 plant samples; the lower the heritability of a trait, the greater the number of samples required for drawing correct conclusions.

In many genetic experiments, the $F_{2:3}$ generation is also planted to verify results from the $F_2$ generation. Each $F_{2:3}$ line originates from one $F_2$ plant, and the family mean will be used for genetic analysis. For model MX-AD-ADI discussed in this paper, $F_{2:3}$ lines derived from AA or aa $F_2$ individuals will result in offspring showing a normal distribution, but lines derived from an Aa $F_2$ will show a distribution as if a normal mixture. Nevertheless, we use family mean data instead of individual data in each family for genetic analysis. The family mean of an AA or aa line will still be distributed as a normal curve. Through simulation study, if the sample size in a family is larger than 3, the family mean will approximate a normal curve. In practical experiments, each family will be planted in a line, so this condition can be easily satisfied. Thus, the population consisting of $F_{2:3}$ family means can also be viewed as a 1:2:1 normal mixture.

The joint segregation analysis of multiple generations used in the current paper is different from the joint scaling test present in Kearsey & Pooni (1996). The joint segregation analysis views the genetic system of a quantitative trait as mixed major gene(s) and minor genes model and then identifies the existence of the major gene(s) and estimates related genetic parameters for both major gene(s) and minor genes. In contrast, the joint scaling test views the genetic sys-

tem of a quantitative trait as a polygenic system. It cannot tell the genetic behavior of individual genes in the system. However, based only on phenotypic data, the joint segregation analysis can only identify one QTL, whereas the modern QTL mapping methods allow identification many QTLs. The joint segregation analysis cannot locate the major gene on a particular chromosome. As shown in the paper, the joint segregation analysis is capable of identifying a quantitative trait locus with large genetic effects without the aid of marker information. But for further research, QTL mapping should be used to identify more QTLs and locate these QTLs on chromosomes.

## Acknowledgements

This research was supported in part by the National 863 Program of China, Henan Committee of Science & Technology and China Scholarship Council. We thank the reviewers for their relevant comments and suggestions to the original manuscript. We also thank Dr Rex Bernardo of Purdue University for his critical reading of the revised manuscript.

## References

Akaike, H., 1977. On entropy maximum principle. In: P.R. Krishnaiah (Ed.), Applications of Statistics, pp. 27–41. North-Holland Publishing Company, Amsterdam.

Chiang, H.S. & D.M. Norris, 1983. Morphological and physiological parameters of soybean resistance to agromyzid beanflies. Envir Entomol 12: 260–265.

Dempster, A.P., N.M. Laird & D.B. Rubin, 1977. Maximum likelihood from incomplete data via the EM algorithm. J Royal Stat Soc, Series B 39: 1–38.

Elston, R.C. & J. Stewart, 1973. The analysis of quantitative traits for simple genetic models from parental, $F_1$ and backcross data. Genetics 73: 695–711.

Elston, R.C., 1984. The genetic analysis of quantitative trait differences between two homozygous lines. Genetics 108: 733–744.

Famula, T.R., 1986. Identifying single genes of large effect in quantitative traits using best linear unbiased prediction. J Animal Sci 63: 68–76.

Gai, J., J. Xia, Z. Cui et al., 1989. A study on resistance of soybean from southern China to soybean agromyzid fly (*Melanagromyza sojae* Zehntner). Soybean Sci 8(2): 115–121.

Gai, J. & J. Wang, 1998. Identification and estimation of QTL model and effects. Theor Appl Genet 97: 1162–1168.

Guo, S.W. & E.A. Thompson, 1992. A Monte Carlo method for combined segregation and linkage analysis. Amer J Human Genet 51: 1111–1126.

Hoeschele, I., 1988. Genetic evaluation with data presenting evidence of mixed major gene and polygenic inheritance. Theor Appl Genet 76: 81–92.

Janss, L.L.G., R. Thompson & J.A.M. van Arendonk, 1995. Application of Gibbs sampling for inference in a mixed major gene-polygene inheritance model in animal populations. Theor Appl Genet 91: 1137–1147.

Jiang, C., X. Pan & M. Gu, 1994. The use of mixture models to detect effects of major genes on quantitative characters in a plant breeding experiment. Genetics 136: 383–394.

Kearsey, M.J. & H.S. Pooni, 1996. The Genetical Analysis of Quantitative traits. Chapman & Hall, London.

Knott, S.A., C.S. Haley & R. Thompson, 1991. Methods of segregation analysis for animal breeding data: a comparison of power. Heredity 68: 299–311.

McLachlan, G.J., 1988. Mixture Models: Inference and Applications to Clustering. Marcel Dekker, Inc.

Morton, M.E. & C.J. MacLean, 1974. Analysis of family resemblance. III Complex segregation analysis of quantitative traits. Amer J Human Genet 26: 489–503.

Shoukri, M.M. & G.J. McLachlan, 1994. Parametric estimation in a genetic mixture model with application to nuclear family data. Biometrics 50: 128–139.

Wang, J., 1996. Studies on Identification of Major-polygene Mixed Inheritance of Quantitative Traits and Estimation of Genetic Parameters (doctorate dissertation). Department of Agronomy, Nanjing Agricultural University.

Wang, J. & J. Gai, 1998. The segregation analysis of genetic system and effects of QTLs – Joint analysis of P1, $F_1$, $P_2$, $F_2$ and $F_{2:3}$. Acta Agronomica Sinica 24(6): 651–659.

Wang, J., D.W. Podlich, M. Cooper & I.H. DeLacy, 2001. Power of the joint segregation analysis method for testing mixed major gene and polygene inheritance models of quantitative traits. Theor Appl Genet (in press)

Wei, T., J. Gai, J. Xia et al., 1989. Inheritance of resistance to beanfly (*Melanagromyza sojae* Zehntner) in soybean. Acta Genetica Sinica 16(6): 436–441.