

# Major Gene Detection for Fusiform Rust Resistance using Bayesian Complex Segregation Analysis in Loblolly Pine

Hua Li\*, Sujit Ghosh<sup>+</sup>, Henry Amerson<sup>+</sup> and Bailian Li<sup>+</sup>

Institute of Statistics Mimeo Series# 2565

## Abstract

Presence of major genes affecting rust-resistance of loblolly pine was investigated in a progeny population that was generated with a half-diallel mating of six parents. A Bayesian complex segregation analysis was used to make inference about a mixed inheritance model that included polygenic effects and a single major gene effect. Marginalizations were achieved by means of Gibbs sampler. A parent block sampling by which genotypes of a parent and its offspring were sampled jointly was implemented to improve mixing. The mixed inheritance model was compared with a pure polygenic model based on Bayes factor. Results showed that the mixed inheritance model was a better model to explain the inheritance of rust-resistance than the pure polygenic model in the diallel population. A large major gene variance component estimate ( $> 50\%$  of total variance), indicated existence of major genes for rust resistance in this loblolly pine population studied. Based on estimations of parental genotypes, it appears that there may be two or more major genes affecting disease phenotypes in this diallel population.

**Key words:** *Pinus taeda L.*, *Cronartium quercuum*, f.sp. *fusiforme*, Mixed inheritance model, Binary trait, Gibbs sampling, Threshold model.

---

\*Hua Li is an assistant professor at Virginia Bioinformatics Institute, Virginia Tech. 1880 Pratt Drive, RB XV Blacksburg, VA 24061-0477 (hli@vbi.vt.edu)

<sup>+</sup>Sujit Ghosh, Henry Amerson and Bailian Li: Associate professors in North Carolina State University, Raleigh, NC 27695

# 1 INTRODUCTION

Fusiform rust, a disease of southern pines caused by *Cronartium quercuum* f.sp. *fusiforme*, continues to be the most economically important tree disease in commercial forests of the southern U.S. Deployment of genetically resistant trees is viewed as environmentally friendly and the only feasible means of control (Kinloch and Walkinshaw 1991). Recent evidence shows that resistance in loblolly pine, at least in part, is the result of major resistance genes in the host interacting with pathogenicity (avirulence/virulence) genes in the pathogen (Wilcox et al. 1996; Kuhlman et al. 1997; Amerson et al. 1997). Non-infected trees in high hazard rust areas may be candidate carriers for major resistance genes and these trees could be used for rust resistance breeding. In a tree breeding program, the low cost of obtaining phenotypes of a progeny population makes phenotypic observations a worthwhile first step towards detection of major genes using statistical analyses.

To our knowledge, complex segregation analysis is considered to be the most powerful statistical test for major gene detection. It was proposed by Elston and Stewart(1971) and Morton and Maclean(1974), and was further developed by geneticists. Bayesian complex segregation analysis was first introduced for the animal model by Hoeschele (1988). Janss et al. (1995, 1997) applied Gibbs sampling within a Bayesian framework to search for a major gene affecting meat quality traits in a crossed F2 population based on a mixed inheritance model in animal breeding. Another application was based on an investigation for major genes affecting carcass traits in Japanese black cattle populations (Miyake et al. 1999). In forest tree breeding, a major gene affecting height of loblolly pine was detected in some half-diallel progeny populations using Bayesian complex segregation analysis (Zeng 2000).

Most complex segregation analyses have been used to study traits with continuous phenotypic measurement. However, many traits in animal and plant breeding that are postulated to be continuously inherited are categorically scored, such as survival scores, or resistance to insects and diseases. A widely used model for genetic analysis of categorical data is based on the threshold liability concept, first introduced by Wright (1934). In the threshold model, one assumes that there exists a latent or underlying variable (liability) that has a continuous distribution. The threshold concept was applied to complex segregation analysis for a binary trait by Thaller et al. (1996). They used numerical integration methods

to determine the mode of inheritance for two traits in swine. Albert and Chib (1993) developed Bayesian inference and used Gibbs sampling algorithm to obtain parameter estimates based on the posterior distribution. They used latent variables within a data augmentation framework which lead to a computationally simple strategy. Sorensen et al. (1995) applied the above methodology to estimate genetic parameters of the animal model.

One of the major goals of this study is to detect major rust resistance genes using a Bayesian complex segregation model for binary data in a half-diallel population. A mixed inheritance model and a polygenic model for binary data are presented and compared using Bayes factors in Section 3. In Section 3.2, an efficient Gibbs sampling algorithm is proposed to obtain parameter estimates. In particular, a parent blocking strategy is developed to improve the mixing of the Gibbs iterates. The proposed method is illustrated by major rust-resistance gene detection for a loblolly pine population in Section 4. Some discussions and future directions of the research are presented in Section 5.

## 2 PHENOTYPIC DATA

The progeny population of loblolly pine was generated from a half-diallel mating of six parents with no selfing or reciprocal crosses, and thus there were 15 full-sib families in this diallel population (see Table 1). Rust infection status (disease phenotype) was assessed as no gall/no disease=1 or gall/disease=0. If a tree was recorded as dead due to the rust infection, this tree was counted as an infected tree (gall) instead of a missing value.

[Table 1 about here]

## 3 STATISTICAL MODEL

### 3.1 Mixed Inheritance Model for Binary Response

In the mixed model for diallel mating, it is assumed that the rust resistance is influenced by a single major gene and polygenes. The single locus is assumed to be an additive diallelic locus with Mendelian transmission probabilities. Assuming the progeny population is in Hardy-Weinberg equilibrium,  $R/r$  denote alleles and  $f$

denotes the frequency of the favorable allele R. Three genotypes, RR, Rr(=rR) and rr, are denoted by 2, 1, 0 respectively. The polygenic effects include general combining ability (GCA) caused by additive polygenic effects and specific combining ability (SCA) caused by dominant polygenic effects. The phenotypic measurement  $y_i$  is recorded as 1 (non-diseased) and 0 (diseased).  $y_i$  has Bernoulli distribution with probability of rust-free  $p_i$ . Assume that,

$$p_i = Pr(y_i = 1) = \Phi(H_i^T \theta)$$

where  $\Phi$  is a standard normal cumulative distribution function (probit model),  $H_i$  denotes a vector of explanatory variable and  $\theta$  denotes a vector of parameters.

Equivalently, this model can be represented by introducing a latent variable  $U_i$ . Specifically if  $U_i \sim N(H_i^T \theta, 1)$ , then we define  $y_i = 1$  if  $U_i > 0$  and  $y_i = 0$  if  $U_i \leq 0$ . Notice that  $U_i$ 's are not observed, however they are used to simplify the likelihood. Define  $H = [H_1^T, \dots, H_n^T]^T$  to be the design matrix and we write,

$$H\theta = \mu + X\beta + wLm,$$

where  $\mu$  is the overall mean,  $X$  is the incidence matrix of the polygenic effects for all progenies;  $\beta$  is a vector of random polygenic effects, e.g. *GCA* (additive genetic effects,  $g_1, g_2, \dots$ ) and *SCA* (dominant genetic effects,  $s_1, s_2, \dots$ ). Specifically,  $\beta = (g_1, g_2, \dots, g_6, s_1, s_2, \dots, s_{15})^T$ ;  $w$  be an  $n \times 3$  design matrix of major genes at a single locus and  $L$  be a  $3 \times 2$  indicator matrix of the major gene effects for major

genotypes. In particular,  $\mathbf{L} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \\ -1 & 0 \end{pmatrix}$ . Let  $m$  be a vector of major gene

effects,  $m = (a, d)$ , where  $a$  is an additive major genotypic effect,  $d$  is a dominance major genotypic effect. The product  $wLm$  generates three possible genotypes of progeny ( $a = RR, d = Rr/rR, -a = rr$ ).

The likelihood of  $\theta$  is given by,

$$L(\theta) \propto \prod_{i=1}^n (\Phi(H_i^T \theta))^{y_i} (1 - \Phi(H_i^T \theta))^{1-y_i}.$$

Let  $\pi(\theta)$  denotes the prior density of  $\theta$ . According to Bayes theorem, the joint posterior density of the parameters and latent variable  $U = (U_1, \dots, U_n)$  given the data  $Y = (y_1, \dots, y_n)$  is given by

$$\pi(\theta, U|Y) \propto f(Y|U, \theta)f(U|\theta)\pi(\theta)$$

where  $f(Y|U, \theta)$  denotes the conditional density of  $Y$  given  $U$  and  $\theta$  and  $f(U|\theta)$  denotes the conditional density of  $U$  given  $\theta$ .

The overall mean ( $\mu$ ), additive major gene effect ( $a$ ), and dominant major gene effect ( $d$ ) are given independent normal priors, i.e.,  $\mu \sim N(0, k_1^2)$ ,  $a \sim N(0, k_2^2)$ , and  $d \sim N(0, k_3^2)$ .  $k_1^2$ ,  $k_2^2$ ,  $k_3^2$  are chosen based on the simulation studies in Zeng (2000). Polygenic effects are assumed to be random. The GCA effects of six parents are assumed to be identical, independent, normally distributed, i.e.  $g_1, \dots, g_6 \sim N(0, \sigma_g^2)$ , where  $\sigma_g^2$  is assumed to have an inverted gamma distribution,  $IG(\mu_1, \nu_1)$ , with hyper-parameters  $\mu_1, \nu_1$ . Similarly, the SCA effects of 15 crosses are assumed to be identically independently as normally distributed, i.e.,  $s_1, \dots, s_{15} \sim N(0, \sigma_s^2)$ , and  $\sigma_s^2 \sim IG(\mu_2, \nu_2)$ . A conjugate beta prior is assumed for the allele frequency  $f$ , i.e.  $f \sim Beta(\alpha_f, \beta_f)$ . Assuming that the progeny population is in Hardy-Weinberg equilibrium, the distribution of progeny genotype ( $p(w)$ ) is obtained from the parental genotypes ( $Wp$ ) following Mendelian transmission probabilities given the favorable gene frequency ( $f$ ) in the base population. Assuming that the major gene effects and polygenic effects are independent, the joint posterior distribution could be easily derived from the likelihood function and above prior distributions.

### 3.2 The Gibbs Sampler for MIM

A Gibbs sampling algorithm can be used to generate samples from the posterior distribution of target parameters  $\theta$  and variance components. However, high dependence of parent and progeny genotypes causes slow converge of a naive Gibbs sampler and hence parent blocking strategy is needed. The parent blocking strategy, introduced by Zeng (2000) for a complex segregation analysis in the diallel mating design of forest trees, can be adapted here for binary data. In the parent blocking, the genotypes of a parent and its offspring are blocked and updated simultaneously. The full conditional distributions for each unknown parameter are derived in order to implement the Gibbs sampler. Conditional on observed values,  $Y$ 's, and all parameters, the conditional distribution of latent variables  $U$ 's has a standard distributional form. In particular, it follows that,

$$U_i|Y, \mu, g, s, a, d, w, Wp, f, \sigma_g^2, \sigma_s^2 \sim N(\mu + X\beta + wLm, 1)$$

truncated at the left by 0 if  $y_i = 1$

$$U_i|Y, \mu, g, s, a, d, w, Wp, f, \sigma_g^2, \sigma_s^2 \sim N(\mu + X\beta + wLm, 1)$$

truncated at the right by 0 if  $y_i = 0$

(1)

In the parent blocking, the joint conditional distribution of a parent and all its offspring is proportional to the product of the distribution of parent  $i$ 's genotype marginalized over all progeny and the joint distribution of its offspring given all parent's genotypes, i.e.:

$$\begin{aligned} \pi(Wp_i, w_{pi(1)}, \dots, w_{pi(o_i)} | Wp_{-i}, w_{-i(k)}, U, Y, \mu, a, d, g, s, \sigma_g^2, \sigma_s^2, f) \propto \\ \pi(Wp_i | Wp_{-i}, w_{-i(k)}, U, Y, \mu, a, d, g, s, \sigma_g^2, \sigma_s^2, f) \\ \pi(w_{pi(1)}, \dots, w_{pi(o_i)} | Wp_i, w_{-i(k)}, U, Y, \mu, a, d, g, s, \sigma_g^2, \sigma_s^2, f) \end{aligned} \quad (2)$$

Where  $o_i$  denotes the number of offspring of parent  $i$ , and the offspring are indexed by  $i(1), i(2), \dots, i(o_i)$ .  $Wp_{-i}$  denotes all other parents except for parent  $Wp_i$ . The marginalized conditional distribution of a parent and the conditional distribution of its offspring could be easily derived and used for calculation. In the parent blocking strategy, each offspring is updated twice in each cycle.

The full conditional distribution for allele frequency ( $f$ ) is a Beta distribution with parameters  $(\alpha_f + n_1, \beta_f + n_2)$ , where  $n_1$  and  $n_2$  are the number of  $R$  and  $r$  allele in the base population, respectively.

$$f | Y, U, \mu, a, d, g, s, w, Wp, \sigma_g^2, \sigma_s^2 \propto f^{\alpha_f + n_1 - 1} (1 - f)^{\beta_f + n_2 - 1} \quad (3)$$

Let  $\eta = (\mu, a, d, g_1, \dots, g_6, s_1, \dots, s_{15})_{p \times 1}^T$ , where  $p = 24$ . The full conditional distribution of  $\eta_j$  ( $j = 1$  to  $p$ ) is normally distributed, i.e.

$$\eta_j | Y, U, \eta_{-j}, w, Wp, f, \sigma_g^2, \sigma_s^2 \sim N(\tilde{\eta}_j, \sigma_{\tilde{\eta}_j}^2) \quad (4)$$

$$\text{where } \tilde{\eta}_j = \frac{\sum_{k=1}^n H_{kj} \left( U_k - \sum_{r=1, r \neq j}^p H_{kr} \eta_r \right)}{\sum_{k=1}^n H_{kj}^2 + \frac{1}{\sigma_j^2}}, \quad \sigma_{\tilde{\eta}_j}^2 = \frac{1}{\sum_{k=1}^n H_{kj}^2 + \frac{1}{\sigma_j^2}}.$$

$H_{kj}$  is the  $k^{th}$  row and  $j^{th}$  column element of the matrix  $H$ , and  $\sigma_j^2$  is the corresponding variance for  $\eta_j$ . The full condition distributions for GCA variance and SCA variance are given by

$$\begin{aligned} \sigma_g^2 | Y, U, \mu, a, d, g, s, \sigma_s^2, w, Wp, f \sim IG \left( \frac{n_g/2}{2} + \mu_1, \frac{\sum_{i=1}^{n_g} g_i^2}{2} + \nu_1 \right) \\ \sigma_s^2 | Y, U, \mu, a, d, g, s, \sigma_g^2, w, Wp, f \sim IG \left( \frac{n_s/2}{2} + \mu_2, \frac{\sum_{j=1}^{n_s} s_j^2}{2} + \nu_2 \right) \end{aligned} \quad (5)$$

In order to initiate the Gibbs sampler, the starting values of parental genotypes ( $Wp$ ) were generated from the initial values of the favorable allele frequency ( $f$ ) assuming Hardy-Weinberg equilibrium in the population. Progeny genotypes were generated based on Mendelian transmission probability given the initial values of related parental genotypes. The other parameters,  $\sigma_g^2, \sigma_s^2$  and  $\eta$  are also generated to be initiated with reasonable guesses from their support. Samples were drawn using the following scheme, starting with  $t=0$ :

1. Sample latent variables  $U_i^{(t)}$  given  $w^{(t)}, \eta^{(t)}, \sigma_g^{2(t)}, \sigma_s^{2(t)}$  using equation (1).
2. Sample genotypes  $w^{(t+1)}$  from by parent blocking using equation (2).
3. Sample allele frequency  $f^{(t+1)}$  using equation(3).
4. Sample location parameters  $\eta^{(t+1)}$  using equation 4.
5. Sample variance components  $\sigma_g^{2(t+1)}$  and  $\sigma_s^{2(t+1)}$  from their full conditional distributions using equation (5).
6. Repeat step 1-5 until the sufficient samples are generated to achieve convergence to the stationary distribution.

Three independent chains were generated with dispersed sets of initial values. Bayesian Output Analysis (BOA version 1.0.0, Smith 2001, <http://www.public-health.uiowa.edu>) was used for convergence diagnostics and posterior distribution summarization. Convergence of a single chain was checked by Raftery and Lewis dependence factors (Raftery et al., 1992); Mixing of multiple chains was checked by Gelman and Rubin (Gelman et al., 1992) shrink factors.

The negative value of the additive effect ( $a$ ) was artificially changed to be positive with the consideration of allele  $R$  being the favorable allele. The predicted parental genotypes were also changed along with the additive effect for consistency purpose. Major gene variance ( $\sigma_m^2$ ) was calculated as the sum of additive major gene variance ( $\sigma_a^2$ ) dominant major gene variance ( $\sigma_d^2$ ), i.e.;

$$\sigma_m^2 = \sigma_a^2 + \sigma_d^2 = 2f(1-f)[(1-2f)d+a]^2 + [2f(1-f)d]^2$$

The total variance was calculated as the sum of the major gene variance and polygenic variance, i.e.;

$$\sigma_t^2 = 2\sigma_g^2 + \sigma_s^2 + \sigma_m^2$$

### 3.3 Polygenic Model (PM)

The polygenic model is the subset of the full model generated by suppressing the major gene effect part. A similar Gibbs sampler algorithm (much simpler) was used to obtain parameter estimates. The prior distributions for parameters in the polygenic model were the same as in the mixed inheritance model. The joint posterior distribution and full conditional distributions were derived using Bayes Theorem. The updating scheme for parameters was also similar except that it did not involve updating steps for genotypes, gene frequency and additive and dominant effects.

### 3.4 Bayes Factor(BF) for Model Comparison

Bayes factor was approximated using the MCMC output with the following formula:

$$B = \frac{P(Y|M_{MIM})}{P(Y|M_{PM})} = \frac{m_1 / \sum_{i=1}^{m_1} \frac{1}{L(\theta_i|Y, M_{MIM})}}{m_2 / \sum_{i=1}^{m_2} \frac{1}{L(\theta_i|Y, M_{PM})}}$$

where  $m_1$  and  $m_2$  are lengths of Markov chains under each model, and  $L(\theta|Y, M)$  denote the likelihood function under the corresponding model.

## 4 RESULTS

### 4.1 Convergence Diagnostics of Markov Chains

For both models, trial chains were run to determine suitable starting values for a burning-in period and a thinning factor. From the trial chains, we decided that the random samples for all parameters could be obtained from 750,000 cycles of the chain, with discarding of the first 100,000 samples and using lag of 1000 cycles. Three Gibbs chains with independent starting values produced 19,500 final samples. For example, the initial values of gene frequency for three chains were taken as 0.25, 0.5 and 0.75, which covered the support of this parameter very well. Raftery and Lewis dependence factor and Gelman and Rubin shrink factor has been presented in Table 2. Most dependence factors of single chains were less than 5 and the 0.975 quantiles of corrected scale shrink factors were less than 1.2, which indicated that our samples were adequate for convergence and mixing.

[Table 2 about here]

## 4.2 Parameter Estimates From The Mixed Inheritance Model

Estimated posterior means and standard deviations of variance components including GCA variance ( $\sigma_g^2$ ), SCA variance ( $\sigma_s^2$ ) and major gene variance ( $\sigma_t^2$ ) are shown in Table 3a. The additive polygenic variance (GCA variance) is similar to the dominant polygenic variance (SCA variance), while the major gene additive variance ( $\sigma_{ma}^2$ ) is much larger than the major gene dominant variance ( $\sigma_{md}^2$ ). The ratio of the total major gene variance to the total genetic variance reached 0.55. The ratio of the major gene additive variance to the total additive genetic variance was 0.61. These high percentages of major gene variance components suggest the presence of at least one major resistance gene segregating in this half-diallel loblolly pine population.

In terms of parental genotype estimates as in Table 3b, parent A was estimated as dominant homozygous (RR) with a probability of 0.63, suggesting that parent A is probably carrying two R alleles. Parent F was estimated as recessive homozygous (rr) with a probability of 0.85, suggesting parent F is more likely to have no resistance allele. From posterior distributions of parental genotypes for parent A and F (Figure 1), parent A is almost certain to have at least one resistance allele given the probability of genotypes RR and Rr being 0.97. Other parents were estimated as heterozygous with probabilities ranging from 0.77 to 0.89 (Table 3b).

[Table 3 about here]

[Figure 1 about here]

In the mixed inheritance model, the parental GCA effect serves as polygenic additive effect, counting for an important variance component. Six parent's GCA predictions under the mixed inheritance model showed that parent A has the largest GCA effect, while parent F has the smallest GCA effect (Figure 2). Apparently, GCA effects are associated with the favorable resistance allele. The predicted high GCA effect and the high probability of parent A carrying at least one resistance allele suggest that the parent A is the most likely candidate among the six parents for further investigations to examine major resistance gene(s).

[Figure 2 about here]

### 4.3 Model Comparison with Polygenic Model

The polygenic model that has less number of parameters with polygenic additive effect (GCA) and polygenic dominant effect (SCA) was more easily fitted using Gibbs sampler. The Gibbs chains of the polygenic model were run 500,000 cycles with much faster speed than that of the mixed inheritance model. The first 100,000 samples were discarded and the thinning factor was taken as 500 cycles. Again, three independent Gibbs chains were run and thus produced 2400 samples. Single chains and multiple chains mixed well for all genetic parameters according to the dependence factors and corrected shrink factors. Enormous Bayes Factors (3000) gave strong evidence for the mixed inheritance model.

## 5 DISCUSSIONS

The Bayesian complex segregation analysis with a block Gibbs sampler was developed for binary data in this study. The method was illustrated for the mixed inheritance model and polygenic model in a half-diallel mating design for rust infection. Statistical inferences were based on the marginal posterior distributions without any asymptotic approximations. The probability distributions of parental genotypes were directly obtained from the output of Gibbs sampler, which could provide valuable reference for the future study. The use of latent variables with Gibbs sampler simplified complicated computations. Further, this study applied a parent blocking strategy (Zeng 2000) into the Gibbs sampler-updating scheme for the half-diallel mating design to improve mixing and eliminate the slow convergence due to the dependency of parents and progeny in the mixed inheritance model. Because of the low cost of phenotypic measurements compared with genotypic data collection, the method described in this study is particularly useful as the first step in disease resistance gene detection with binary phenotypic measurements, and it can be easily adapted to other genetic analyses for categorical data.

Even with the parent blocking strategy implemented in the Gibbs sampler, slow mixing that was likely caused by the latent variables was still noticed. It might be related to the data augmentation approach in this study. Similar problems were found by other authors (Liu et al. 1994; Sorensen et al. 1995). One possible scheme

to accelerate mixing could be implementation of another blocking strategy, for example, sampling jointly from the liability and selected parameters. In that case, another computational strategy and parameterization of models may be needed.

From the results of complex segregation analysis, the mixed inheritance model with a major gene effect and polygenic effects fits the data better than the polygenic model, as asserted by Bayes factor. In the mixed inheritance model, a large percentage of major gene variance of the total variance ( $> 50\%$ ) was estimated, which indicated the existence of major genes. Under the mixed inheritance model, parent A was predicted to have the highest GCA, with parent F having the lowest GCA. The estimated genotypes of parents showed that 5 out of 6 parents have at least one dominant allele with high probabilities ( $> 70\%$ ). Among these five parents, parent A was estimated to carry at least one dominant allele with probability larger than 0.9. Further, the estimated genotype of this parent was dominant homozygous (RR) with largest probability (0.63) among the three genotypes. The estimated genotype of parent F was recessive homozygous (rr) with the largest probability among three genotypes. This result suggests that GCA prediction could be associated with estimated genotypes with the favorable allele. That is, high GCA and high probability of carrying dominant allele in parent A meant that this parent would be most likely to carry major genes for rust-resistance.

The mixed model used here assumes only one major gene and while these analyses strongly support the existence of major gene effects for rust resistance in this diallel loblolly pine population, this does not exclude the possibility of two or more major genes in the diallel. Simulation studies showed that major gene effects could be recognized using the mixed inheritance model with a single gene assumption when actually there were more than one major gene affecting the phenotypes (Zeng 2000). The first one or two major genes determined the parental genotype estimations and high GCA estimates, plus the genotype estimates with favorable allele were always associated with a good parent. With only a single major gene included in the mixed inheritance model, it is impossible to distinguish between the effect of a single locus and the effects of two or more independently acting loci with similar transmission patterns. If there are multiple genes, it is postulated that the major gene effects would be detected as if they were a single locus with an allele frequency and effect equaling to the sum of several alleles. A similar situation could be expected for multiple fusiform rust resistance genes, all of which regulate gall presence vs. absence. In this study, the estimated high

frequency of the favorable allele might be the sum of allele frequencies in several loci. With the current model, the above hypothesis was not verified. However, the method presented here could be easily extended to a mixed model with two major genes that essentially involve more complex genotype configurations.

Complex segregation analysis can be applied to any pedigree structure and works with both qualitative and quantitative traits. The results from complex segregation analysis could be a useful starting point for defining major genes or detecting QTL in human genetics, animal and plant breeding (Jarvik 1998). In this study, the estimated large major gene variance component strongly supports the existence of major gene effects. The best performing parent with regard to rust disease was parent A, which was predicted to have the highest GCA effect and was estimated as RR (dominant homozygous) genotype with the highest probability among three genotypes. The estimated dominant homozygous genotype of parent A may be the confounded effects of multiple rust resistance genes. Parent F was the worst performing parent for rust disease with the lowest GCA and the lowest probability of carrying the dominant resistant allele. If parent A were crossed with parent F, we might expect segregation for resistance in the progeny population based on the results of this study. The resistant individuals in the progeny population would be likely to inherit resistance from parent A based on current complex segregation analysis. Mapping major genes for rust resistance in parent A is the subject for a manuscript now in preparation by the current authors.

## REFERENCES

- Albert, J.H., and Chib, S.(1993), "Bayesian analysis of binary and polychotomous response data," *Journal of the American Statistical Association*, 88, 669-679.
- Amerson, H.V., Jordan, A.P., Kuhlman, E.G. O'Malley, D.M., and Sederoff, R.R.(1997), "Genetic basis of fusiform rust disease resistance in loblolly pine," *Proceedings of the 24th South Forrest Tree Improvement Conference*, pp.403.
- Elston, R.C., and Stewart,J.(1971), "A general model for the genetic analysis of pedigree data," *Human Heredity*, 21, 523-542.

- Gelman, A, and Rubin, D.B.(1992), "Inference from iterative simulation using multiple sequences." *Statistical Science*, 7, 457-511.
- Hoeschele, I.(1988), "Genetic evaluation with data presenting evidence of mixed major gene and polygenic inheritance," *Theoretical and Applied Genetics*, 76, 81-92.
- Janss L.L.G., Arendonk, J.V., and Brascamp, E.W.(1997), "Bayesian statistical analyses for presence of single genes affecting meat quality traits in a crossed pig population," *Genetics*, 145, 395-408.
- Janss, L.L.G., Thompson, R., and Arendonk, J.V.(1995), "Application of Gibbs sampling for inference in a mixed major gene-polygenic inheritance model in animal populations," *Theoretical and Applied Genetics*, 91, 1137-1147.
- Jarvik, G.P.(1998), "Complex segregation analysis: use and limitations," *American Journal of Human Genetics*, 63, 942-946.
- Kinloch, B.B., and Walkinshaw, C.H.(1991), "Resistance to fusiform rust in southern pines: How is it inherited? ," *Proceedings IUFRO Rusts of pine working party conference. Banff, Alberta. Inf. Rep. NOR-X-317*, 219-228.
- Kuhlman, E.G., Amerson, H.V., Jordan, A.P., and Pepper, W.D.(1997), "Inoculum density and expression of major gene resistance to fusiform rust disease in loblolly pine," *Plant Disease*, 81(6), 597-600.
- Liu, JS, Wang, W.H., and Kong, A(1994), "Covariance structure of the Gibbs sampler with application to the comparisons of estimators and augmentation schemes," *Biometrika*, 81, 27-40.
- Miyake, T., Dogo, T., Moriya, K., and Sasaki, Y.(1999), "Bayesian analysis for existence of segregation of major genes affecting carcass traits in Japanese Black cattle population," *Journal of Animal Breeding and Genetics*, 116, 207-215.
- Morton, N.E., and MacLean, C.J.(1974), "Analysis of family resemblance III. Complex segregation of quantitative traits," *American Journal of Human Genetics*, 26, 489-503.

- Raftery, A.L., and Lewis, S.(1992), "How many iterations in the Gibbs sampler?", *Bayesian statistics* 4, 763-774. Oxford University Press.
- Sorensen, D.A., Anderson, S., Gianola, D., and Korsgaard, I.(1995), "Bayesian inference in threshold models using Gibbs sampling," *Genetics Selection Evolution*, 27, 229-249.
- Thaller, G., Dempfle, L., and Hoeschele, I.(1996), "Investigation of the inheritance of birth defects in swine by complex segregation analysis," *Journal of Animal Breeding and Genetics*, 113, 77-92.
- Thaller, G., Dempfle, L., and Hoeschele, I.(1996), "Maximum likelihood of Rare Binary Traits Under Different Modes of Inheritance," *Genetics*, 143, 1819-1829.
- Wilcox, P.L., Amerson, H.V., Kuhlman, E.G., Liu, B.H., O'Malley, D.M., and Sederoff, R.R.(1996), "Detection of a major gene for resistance to fusiform rust disease in loblolly pine by genomic mapping," *Proceedings of National Academy of Sciences of the United States of American*, 93, 3859-3864.
- Wright, S.(1934), "Analysis of variability in number of digits in an inbred strains of guinea pigs," *Genetics*, 19, 506-536.
- Zeng, W.(2000), "Statistical methods for detecting major genes of quantitative traits using phenotypic data of a diallel mating," PH.D thesis, North Carolina State University. Raleigh.

Table 1: Six-Parent(A-F) half-diallel mating design used in this study. Each parent was crossed with five others with no selfing or reciprocal crosses, generating 15 full-sib families in each diallel population.

Female/Male	A	B	C	D	E
B	*				
C	*	*			
D	*	*	*		
E	*	*	*	*	
F	*	*	*	*	*

\*: Crosses that were made.

Table 2: Convergence diagnostics (dependence factor and shrink factor) of the Gibbs sampler for additive effect ( $a$ ), dominant effect ( $d$ ), gene frequency ( $f$ ), GCA variance ( $\sigma_g^2$ ), SCA variance( $\sigma_s^2$ ) in the mixed inheritance model. There are three independent chains.

		$a$	$d$	$f$	$\sigma_g^2$	$\sigma_s^2$
Shrink factor	Estimated	1.03	1.08	1.03	1.00	1.00
	0.975	1.07	1.23	1.06	1.00	1.00
Dependence factor	Chain1	7.55	3.36	7.80	0.97	1.01
	Chain2	3.25	2.45	3.85	1.04	0.97
	Chain3	4.23	3.69	3.92	0.97	1.03

Table 3: Estimated posterior statistics from the mixed inheritance model.

(a) Marginal posterior means and standard deviations of variance components.<sup>1</sup>

$\sigma_{ma}^2$	$\sigma_{md}^2$	$\sigma_{ma}^2$	$\sigma_{md}^2$	$\sigma_m^2$	$\sigma_t^2$	$\sigma_m^2/\sigma_t^2$
1.6±1.4	1.5±1.0	5.0±4.1	0.7±0.9	5.7±4.2	10.4	55%

(b) Estimated parental major gene genotypes.

	Parent A	Parent B	Parent C	Parent D	Parent E	Parent F
Genotype	RR	Rr	Rr	Rr	Rr	rr
Probability	0.63	0.79	0.89	0.89	0.77	0.85

Figure 1: Marginal posterior distributions of major gene genotypes for parent A and parent F.

Figure 2: Parental GCA predictions for six parents (A-F) by the mixed inheritance model. The GCA was estimated as the deviation from the population mean.