

ABSTRACT

Li, Hua. Major gene detection for fusiform rust resistance using segregation analysis and linkage analysis in loblolly pine. (Under the direction of Dr. Bailian Li and Dr. Henry Amerson)

Fusiform rust, a disease caused by *Cronartium quercuum* f.sp. *fusiforme*, is the most economically important disease for loblolly pine (*Pinus taeda* L.) in the southern U.S. Identification and breeding of loblolly pines that are genetically resistant to fusiform rust are important for successful establishment of commercial plantations. This research developed a new analytical approach to detect major genes for rust resistance using complex segregation analysis in a diallel progeny population. Molecular markers were examined for association with the potential rust resistance genes.

Loblolly pines from a six-parent half-diallel mating were planted in a randomized complete block field design at four test sites. Rust infection (gall presence / absence) was recorded annually through age 8. Time trends and genetic differences for rust infection were analyzed based on a polygenic model using a Bayesian logistic approach. For genetic control of rust infection among families, the parental general combining ability (GCA) due to additive effect was much more important than specific combining ability (SCA) due to non-additive effect in full-sib combinations. Large genetic differences among parents and full-sib families were found for rust infection. Among six parents, parent A showed consistent low infection rates over time and across four sites, which

indicated high genetic resistance to fusiform rust due to strong polygenic effects and / or major gene effects.

A Bayesian analysis of a threshold model was developed and used to make inference about a mixed inheritance model (MIM) that included both polygenic effects and major gene effects. The MIM was compared with a pure polygenic model. Marginalizations were achieved by means of Gibbs sampler. A parent block sampling has been implemented to improve mixing. Results showed that the MIM 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 (around 40-50%) indicated the existence of major genes for rust resistance. The major genes would be most likely associated with parent A because it was estimated to have the highest probability carrying two resistance alleles and predicted to have the highest GCA effect for rust resistance among the six parents.

Bulk segregate analyses and marker / trait co-segregation analyses were used to search for major resistance genes. When the progeny of parent A \times F showed an intermediate rust infection level with single spore inoculum, additional RAPD markers were found to be linked with the Fr2 locus in parent A. An improved map (framework linkage map) of the Fr2 linkage group was developed. An effort to define another resistance locus in parent A, using progeny of cross A by F inoculated with an intermediate spore density, with mixed gall inoculum was not successful. This may be due to the complexity of suspected multiple gene interaction effects in parent A and unknown pathogen virulence

composition. Two mixed inocula with extremely high spore density were used to inoculate the diallel progeny population of 12 crosses. Progeny from the cross of parent E by A showed 75% rust infection. The interaction of two pairs of complementary genes was proposed to explain the observed 75% infection based on a gene-for-gene hypothesis. This two-gene model was not confirmed with molecular markers used in this study. Nevertheless, results of this study provided strong statistical and more molecular evidence to support that the parent A is carrying multiple rust resistance genes.

**MAJOR GENE DETECTION FOR FUSIFORM RUST RESISTANCE
USING SEGREGATION ANALYSIS AND LINKAGE ANALYSIS IN
LOBLOLLY PINE**

by

HUA LI

A dissertation submitted to the Graduate Faculty of
North Carolina State University
in partial fulfillment of the
requirements for the Degree of
Doctor of Philosophy

FORESTRY AND BIOINFORMATICS

Raleigh

2003

Approved By

Co-chair of Advisory Committee

Co-chair of Advisory Committee

BIOGRAPHY

Hua Li was born on December 14, 1972 in Handan, Hebei Province, China.

From Sept. 1989 to June 1993, Hua attended HeBei Agriculture Univeristy, Baoding, HeBei Province, China, where she received a Bachelor of Science degree in Agriculture. In Sept. 1993, she started her graduate study in Beijing Forestry University. After three years' study and research in forest tree genetics and improvement, Hua graduated with a Master of Science in Agriculture in 1996. She was employed in Beijing Forest Bureau for about two years. In May 1999, Hua enrolled in North Carolina State University to pursue her doctoral studies in forest genetics. She worked as a graduate research assistant for the NCSU Tree Improvement Program. During her graduate study, she became interested in bioinformatics and was admitted to NCSU Bioinformatics Program for her co-major in 2001.

ACKNOWLEDGEMENTS

I would like to express my deepest gratitude and biggest appreciation to my advisors Dr. Bailian Li, Dr. Henry Amerson and Dr. Sujit Ghosh. Not only has their guidance been immensely helpful in completing this research, but also their constant encouragement has been equally indispensable. My great thanks also go to Dr. Sharon Browning for serving on my committee.

I would like to extend my appreciation to Dr. Steve McKeand, Dr. Bruce Weir and Dr. Barbara Sherry for their help during my graduate study. I am also grateful to people in Tree Improvement Coop and Forestry Biotechnology Group in North Carolina State University for their help and friendship. Special thanks go to Dr. James Grisom, Saul Garcia, Catherine Clark, Jianfeng Li for their valuable helps in computing and laboratory works of my research.

Finally, I want to thank my husband, Jian Li, and my parents. None of this work could have been accomplished without their continuous support and encouragement.

Contents

List of tables.....	vii
List of figures.....	xi

1. Time trends and genetic differences of rust infection in a diallel loblolly pine population across four tests

Abstract.....	1
1.1 Introduction.....	3
1.2 Materials and methods.....	7
1.2.1 Experimental design.....	7
1.2.2 Data collection.....	7
1.2.3 Statistical model.....	8
1.3 Results.....	10
1.3.1 Convergence diagnostics of markov chains.....	10
1.3.2 General trends of rust infection overtime and over site.....	10
1.3.3 Variance components estimates.....	12
1.3.4 GCA and SCA prediction.....	12
1.4 Discussion.....	13
1.5 References.....	18

2. Bayesian inference for presence of major genes affecting rust resistance in loblolly pine

Abstract.....	37
2.1 Introduction.....	39

2.2 Materials and methods	46
2.2.1 Phenotypic data.....	46
2.2.2 Mixed inheritance model for binary response.....	47
2.2.3 The Gibbs Sampler.....	50
2.2.3.1 Latent variables U.....	51
2.2.3.2 Location parameters (β , a, d, g and s).....	51
2.2.3.3 Variance components (σ_g^2 , σ_s^2).....	52
2.2.3.4 Allele frequency.....	52
2.2.3.5 Major genotypes.....	52
2.2.4 Updating scheme	53
2.2.5 Polygenic model and model comparison.....	55
2.3 Results	56
2.3.1 Convergence diagnostics of markov chains.....	56
2.3.2 Parameter estimation in the MIM.....	58
2.3.3 Model comparison with polygenic model.....	60
2.4 Discussion.....	61
2.5 References.....	66
3. Search for molecular markers associated with resistance genes in loblolly pine	
Abstract.....	91
3.1 Introduction.....	93
3.2 Materials and methods	98
3.2.1 Inocula preparation.....	98
3.2.2 Plant materials.....	98

3.2.3 Inoculation and phenotypic assessment.....	99
3.2.4 DNA extraction.....	100
3.2.5 RAPD reactions and marker analysis.....	101
3.2.6 Bulk segregant analysis and cosegregation analysis.....	102
3.2.7 Map construction	103
3.3 Results.....	104
3.3.1 Inoculation results and DNA marker analysis for the diallel progeny population.....	104
3.3.2 Linkage map construction for Fr2	106
3.3.3 Continued efforts to associate DNA markers with field resistance loci in for parent A.....	108
3.3.4 Inoculation results and marker trait association study for the cross- A by F.....	109
3.4 Discussion.....	110
3.5 References.....	115

List of Tables

1-1	Six-parents (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.	25
1-2	Convergence diagnostics of markov chains for 9 parameters (intercept, site effects and variance components). Gandr is Gelman and Rubin reduction factor. Reduction factor is on the corrected score scale.....	26
1-3	Parameter estimates by fitting a logistic model using WINBUGS: intercept, site effects and variance components. Statistics from three single chains and from the concatenated chain are showed. The upper number in each cell is the parameter mean from its posterior distribution. The lower number (in parenthesis) is the standard deviation.....	27
1-4	Summary statistics for posterior distributions of site differences. Using the estimated means of site posterior distributions, site effects were ordered from high to low (site 4, site 2, site 3, site 1), and the statistics of differences between adjacent pairs over time are shown. The upper line in each cell shows the mean and standard deviation of the corresponding site difference, the lower line indicates the estimated 95 % credible interval of the site difference.....	28
1-5	Variance components estimates from the posterior distributions that result from fitting a logistic model using WINBUGS (Spiegelhalter et al. 2000). The upper number in each cell is the mean of posterior distribution, while the lower number (in parenthesis) is the standard deviation.....	29

1-6	GCA predictions for resistance obtained from posterior distributions by fitting a logistic model using WINBUGS. The upper number in each cell is the parameter mean from its posterior distribution, and the lower number (in parenthesis) is the standard deviation.....	30
2-1	Criteria to classify the original dataset to four groups.....	72
2-2	Data that were used in this study. (a) Data in group1, (b) data in group2, (c) data in group3, (d) data in group4. (e) Summary of data in four groups.....	73
2-3	Convergence diagnostics (dependence factor and shrink factor) of the Gibbs sampler for all parameters in the mixed inheritance model. There are three independent chains. Only the results using dataset in group1 are shown here.....	76
2-4	Marginal posterior means (pm) and standard deviations (psd) of polygenic additive variance (V_g), polygenic dominant variance (V_s), major gene additive variance (V_{ma}), major gene dominant variance (V_{md}), and major gene variance (V_m) estimated using the datasets in four groups under the mixed inheritance model. V_t is the total variance which was calculated as: $V_t = 2 V_g + V_s + V_{ma} + V_{md}$	77
2-5	Estimated posterior mean (Pm) and standard deviation (Psd) of the additive effect (a) and dominant effect (d) at the single locus using four groups of data in the mixed inheritance model.....	78
2-6	Estimated parental major gene genotypes. The 0, 1 or 2 in each cell denote rr, Rr(rR) or RR, the genotype of the parent. The number in the parenthesis is the probability of the parent being the correspondent genotype. The parent IDs are A, B, C, D, E, and F.....	79

2-7	Convergence diagnostics (dependence factor and shrink factor) of the Gibbs sampler for all parameters in polygenic model. There were three independent chains. Only the results using dataset in group1 are shown here.....	80
2-8	Estimated posterior mean of deviances for four datasets under the mixed inheritance model and the polygenic model and approximated Bayes factor.....	81
3-1	Overall greenhouse inoculation scheme in this study. L-7-2-85 and NCSU01 are two mixed basidiospore inocula originating from the mixed gall collections of aeciospores obtained from the field-infected trees. Basidiospore inoculum 0-5-32 originated from single aeciospore isolate 0-5-32. 200 progeny were inoculated for each treatment giving 2 replications of 100 each.....	120
3-2	Four and half months phenotypic data expressed as percent galled. (a) The diallel was inoculated with inoculum L-7-2-85 for replication 1; (b) The diallel was inoculated with inoculum L-7-2-85 for replication 2; (c) The diallel was inoculated with inoculum NCSU01 for replication 1; (d) The diallel was inoculated with inoculum NCSU01 for replication 2. Inoculum density was 250,000~300,000 basidiospores/ml for each inoculum.....	121
3-3	Nine month phenotypic data expressed as percent galled for the specific cross E*A. (a) The diallel was inoculated with inoculum L-7-2-85; (b) The diallel was inoculated with inoculum NCSU01. Inoculum density was 250,000~300,000 basidiospores/ml for each inoculum.....	123
3-4	Binomial proportion hypothesis test for $H_0: \text{Proportion} = 0.75$. Two-sided p-values reported from PROC FREQ in SAS are listed in the table. P-values ranged from	

0.32 to 0.83 and suggested that the tested gall percentages are not significantly
different from 75%.....124

3-5 Six month phenotypic data expressed as percent galled for the cross A×F.....125

List of Figures

- 1-1 Time trends of rust percentages over eight years on four testing sites (1-4). The site means were calculated by averaging all family means on each site.....31
- 1-2 Posterior density plots of site effect differences. Posterior densities were estimated using three combined chains with 20,000 iterations in each chain, a 7,000 burn-in period and thinning interval of 2.....32
- 1-3 Time trends of 6 family means (A, B, C, D, E and F) for rust percentage over eight years: site 1 (a), site 2 (b), site 3 (c) and site 4 (d). No data were collected at age 2 for site 1 (a).....33
- 1-4 Time trend of family (A, B, C, D, E and F) means averaged over 4 sites for rust percentage over eight years.....35
- 1-5 Parental GCA predictions for no-galled percentage (deviation from the population mean) of six parents (A-F) from age 3 to age 8.....36
- 2-1 Trace plots of eleven genetic parameters including GCA prediction of six parents (a-f), GCA variance component V_g (g), SCA variance component V_s (h), additive effect (i), dominant effect (j) and the favorable allele frequency (k) for group1 data. There are three independent chains in each plot; three chains are denoted as the straight line, the dot line and the dash line.....82
- 2-2-1 Autocorrelation plots of GCA predictions of six parents (a-s) for group 1 data. There are three independent chains for each parameter and thus three autocorrelation plots of individual chains for each parameter. The parent ID and the chain id are showed on the upper right corner of each plot.....83

2-2-2 Autocorrelation plots of five genetic parameters including GCA variance component (a-c), SCA variance component (d-f). There are three independent chains for each parameter and thus three autocorrelation plots from each independent chain. The parameter name and the chain id are shown on the upper right corner of each plot.....85

2-2-3 Autocorrelation plots of genetic parameters including additive effect (a-c), dominant effect (d-f) and the favorable allele frequency (g-i). There are three independent chains for each parameter and thus three autocorrelation plots from each independent chain. The parameter name and the chain id are shown on the upper right corner of each plot.....86

2-3 Gelman and Robin shrink factor plots of eleven genetic parameters including GCA prediction of six parents (a-f), GCA variance component (g), SCA variance component (h), additive effect (i), dominant effect (j) and the favorable allele frequency (k) for group1 data.....87

2-4 Posterior density plots of eleven genetic parameters including GCA prediction of six parents (a-f), GCA variance component (g), SCA variance component (h), additive effect (i), dominant effect (j) and the favorable allele frequency (k) for group1 data. There are three independent chains generating three posterior density plots for each parameter which are denoted as the straight line, the dot line and the dash line in each plot.....88

2-5 Marginal posterior distributions of major gene genotypes for parent A (left) and parent F (right). The results from four groups of datasets are shown in the plot...89

2-6	Parental GCA predictions for six parents (A-F) by the mixed inheritance model. The GCA was estimated as the deviation from the population mean by four groups of datasets.....	90
3-1	A six-parent half-diallel mating design. 15 crosses were generated in this diallel. Only 12 of them were used in this study because of seed deficiencies in the other three full-sib families.....	126
3-2	An illustration for the approach to identify resistance genes interacting in the proposed two-gene model using dominant molecular markers, (a) bulk segregant analysis, and (b) assessment of candidate markers in a sample progeny set.....	127
3-3	A linkage map of the Fr2 linkage group in parent A. Fr2 and accessory markers are placed around the markers in the framework map.....	128
3-4	A preliminary map for the A19_560 linkage group in parent A.....	129
3-5	A preliminary map for the A11_400 linkage group in parent A.....	130

Chapter I

Time trends and genetic differences of rust infection in a diallel loblolly pine population across four tests

Abstract

Loblolly pines from a six-parent half diallel mating were planted in a randomized complete block field design at four test sites. Fusiform rust infection (gall presence / absence) was recorded annually through age 8. A Bayesian logistic model was developed to estimate genetic parameters and variance components for the half-diallel mating design with binary data. The percentage of rust infection at each site was low in the first two years, and then increased dramatically at age 3 with little change in percentage galled thereafter. Large genetic differences among parents were found for rust infection. The rank of infection for the 6 parents was consistent over time and across sites even though infection rates were significantly different among test sites. No significant genotype by environmental interaction for rust infection was found. For genetic control of rust infection among families, the parental general combining ability (GCA) due to additive genetic effects was much more important than specific combining ability (SCA) due to non-additive genetic effects in full-sib combinations. A time trend showed that additive genetic variance increased from age 3 to 4 and remained stable through age 8. Among the

six parents, parent A showed consistent low infection rates over time and across four sites, which indicated high genetic resistance to fusiform rust due to either strong polygenic effects or major gene effects. Results from this study support selection for rust infection in loblolly pine progeny tests at age 6 or earlier.

Keywords: *Pinus taeda* L, *Cronartium quercuum*, f.sp. *fusiforme*, genetic parameters, variance components, Bayesian method, logistic model, binary data, WINBUGS.

1.1 Introduction

Fusiform rust, a disease of southern pines caused by *Cronartium quercuum* f.sp. *fusiforme*, has been and continues to be the most economically important tree disease in commercial forests in the southern U.S. (Anderson et al., 1986; Kinloch and Stonecypher, 1969; Kinloch and Walkinshaw, 1991; Kuhlman, 1992; Powers et al., 1981; Cabbage et al., 2000). Loblolly pine (*Pinus taeda* L.) and slash pine (*Pinus elliotii* Engelm var *elliotii*) are the two pine species most seriously impacted by fusiform rust disease. Branch and stem galls are symptomatic of the disease. Stem galls reduce wood quality and weaken trees. Infections that occur on the main stem within the first 5 years of a tree's life frequently cause death, especially in slash pine. Because of the relatively short rotation of pine in the south, increased rust resistance translates directly to increased economic value. The estimated annual losses range from \$24M to \$135M in the southeastern United States (Cabbage et al., 2000). Identification of loblolly and slash pine that are genetically resistant to fusiform rust has become very important.

The stable, consistent performance of a deployed candidate resistant family over time and across a large range of regions is highly desirable. One source of highly variable disease performance may be large genotype by environment interactions. Some studies reported that family genotype by environment interaction for fusiform rust infection was of little importance in loblolly pine. Lenhart (1994) observed rust incidences between 1987 and 1990 for different age classes and found no apparent change in overall average incidence for loblolly pine. Blair (1970) found no significant micro site differences regarding the

level of infection, but did find infection differences among planting dates. Kinloch and Stonecypher (1969) found variation in rust incidence among sites within a region, but relative ranking of families remained stable across sites. Hodge et al. (1993) showed that infection ratios of resistant vs. susceptible seedlots were essentially constant in loblolly pine and slash pine across different levels of rust hazard in six locations of the coastal plain.

In contrast with the studies suggesting little genotype by environment interaction, significant variation was found in other studies (Li and McKeand, 1989; Kuhlman et al., 1995; McKeand et al., 1999; Schmidt et al., 1999). Kuhlman et al. (1995) suggested that interactions might be consequential only for specific resistant families. Indeed a recent update of the work by McKeand et al. (1999) now shows that the most highly resistant families in their studies typically were the least predictable across sites (McKeand et al., 2003) and these researchers speculated that the unpredictable nature of the families could be due to family genotype interactions with suspected specific pathogen virulence (differences) among the sites. Variation in pathogen virulence could make a large contribution to genotype by environment interactions and pathogen virulence variation has been noted in many studies (Snow et al., 1975; Powers et al., 1977; Kuhlman and Matthews, 1993; Kuhlman et al., 1997; Nelson et al., 1993; Jordan, 1997; Stelzer et al., 1997, 1999; Wilcox et al., 1996; Amerson, unpublished data).

Estimates for general combining ability (GCA, interpreted as additive genetic variation) and specific combining ability (SCA, interpreted as non-additive genetic variation) for rust

resistance in loblolly pine were found in several studies. Studies by Blair (1970) revealed that non-additive variation was less frequent than additive variation in field experiments, and of lower magnitude even when significant. Prevalent additive variation was also observed in studies by Kinloch and Stonecypher (1969) and Sluder (1993). In contrast, other studies involving greenhouse inoculation experiments (Carson, 1983; Snow et al., 1990) reported that specific combining ability was significant and of higher magnitude than general combining ability.

In field trials, commonly used experimental designs for rust screening are a factorial design or a split-plot design testing full-sib or half-sib families. Variance components analyses in these studies typically are based on a linear mixed model. Because the phenotypic measurements for rust infection are often recorded as binary data (0/1), percent gall is usually calculated and used as the response variable. Some studies used a transformation approach such as arcsins of percentage square root transformations (Sluder, 1993) to meet the assumption that the dependent variable needs to be approximately normally distributed. However, when the number of trees per plot is small, the percent gall may not be normally distributed either on the original scale or after some forms of transformation. In these cases, it may not be appropriate to use a linear model for rust data analysis.

Bayesian inferences based a logistic model provide an alternative way to estimate genetic parameters and variance components in these situations. The Bayesian approach does not rely on a large sample size for its validity. Variance components may be obtained directly

from the analysis. Furthermore, Bayesian analysis combines prior knowledge (prior distributions of the unknown parameters) with a statistical model (likelihood) that defines a probability distribution for the observed data in terms of unknown parameters, generating a posterior distribution for the parameters that represent the knowledge after observing the data. The mean, standard deviation and credible interval of the marginal posterior distribution of the parameter are often used as statistical summaries.

Computation aspects of the Bayesian methods have become simpler by the use of a Gibbs sampling algorithm that simulates samples from conditional distributions of each parameter. Estimates of genetic parameters and variance components using the Bayesian approach with Gibbs sampler have been reported in several studies (Ball et al., 2001; Magnabosco et al., 2000; Burton, 1999; Sorrensen et al., 1995; Janss et al., 1995, 1996; Zeng, 2000)

Most studies of fusiform rust incidence have focused on site/region effects and family performance for specific age classes and have not followed change in infection over time. A few studies have followed rust infection annually (Schmidt et al., 1999; Griggs and Schmidt, 1976). The current study examined 8 years of data on 4 sites in a half-diallel progeny population to investigate rust incidence over time and across locations. Genetic parameters and variance components were evaluated annually after adequate infection. Time trends with regard to the genetic parameters and variance components were estimated using a Bayesian logistic approach.

1.2 Materials and Methods

1.2.1 Experimental design

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 (Table 1-1). All six parents were selected from the upper piedmont of Alabama in plantation forests, and all selections were rust-free at the time of selection. Each progeny test was established using a randomized complete block design with six blocks and six trees per family in each block. The experiments were replicated over four different locations. All four test sites were located in Wilcox County, Alabama and the plantations were established from winter of 1987 to spring of 1989.

1.2.2 Data collection

Trees were evaluated annually from age 1 through 8 for presence or absence of galls on stems and/or branches following the standard procedure to assess rust incidence in loblolly pine (McKeand et al., 1999). Phenotype was recorded as 0 (gall) or 1 (no gall). If a tree was recorded as dead because of the rust infection, this tree was thereafter counted as an infected tree (gall). The number of infected trees and total number of surviving trees were counted within each plot at each age. The infection rate was calculated as the ratio of the number of infected trees to the total number of surviving trees in that plot at the time of assessment.

1.2.3 Statistical model

If y_i represents the number of non-galled trees in the plot i , n_i is the total number of trees in the same plot, then y_i has a binomial distribution with (p_i, n_i) , where p_i is the probability that the tree is gall-free. A logistic model for the probability that an individual is not infected by rust is:

$$\text{logit}(p_i) = \log\left(\frac{p_i}{1-p_i}\right) = \alpha + X_i\beta + Z_i\mu$$

Where α is the intercept parameter; β is a vector of site effects, which is considered as fixed effect; X is the design matrix of site effects; μ is a vector of polygenic effects and interaction effects, including GCA effects, SCA effects, GCA by site effects and SCA by site effects; Z is the incidence matrix of the random polygenic effects and interaction effects.

The normal prior distributions are assigned for the intercept and test effects. Polygenic and interaction effects are assumed to be random effects. Six GCA effects are assumed to be identical, independent, normally distributed (IIND) with mean 0 and variance σ^2_{GCA} ; SCA effects are also assumed to be IIND $(0, \sigma^2_{SCA})$; GCA by site effects are assumed IIND $(0, \sigma^2_{\text{Site} \times \text{GCA}})$; SCA by site effects are assumed IIND $(0, \sigma^2_{\text{Site} \times \text{SCA}})$. Variance components $(\sigma^2_{GCA}, \sigma^2_{SCA}, \sigma^2_{\text{Site} \times \text{GCA}}, \sigma^2_{\text{Site} \times \text{SCA}})$ are given inverted gamma distributions as the prior distributions. The odds ratio of two parents is defined as a ratio of odds (probability of non-infection over probability of infection) for two parents when controlling other variables in the model.

Analyses were performed in WINBUGS (Spiegelhalter et al., 2000). Starting values were randomly chosen and three independent samplers (chains) were implemented with different starting values. Bayesian Output Analysis (BOA version 1.0.0, Smith 2001) was used to analyze the Gibbs chain output. The posterior distributions for test effects, GCA, and variance components were estimated from age3 to age 8.

Convergences were diagnosed by Gelman and Rubin shrink factors for three chains (Gelman and Rubin, 1992; Brooks and Gelman, 1998). If the 0.975 quantile of corrected scale reduction factor is less than 1.2, the sample may be considered to have arisen from the posterior distribution. The burn-in period is the number of sampled values that must be discarded before the samples produced by Gibbs Sampler may be considered to be arising from the true posterior distribution. The burn-in period is chosen to be sufficiently large to eliminate effects of starting values. In consideration of correlations between consecutive samples from a Gibbs chain, thinning interval is used to reduce correlation between samples. Trial chains were run to determine suitable values for the length of chains, burning-in period and thinning interval in order to obtain random Gibbs samples. In summary, the analyses were conducted using three independent chains, with lengths of 20,000 iterations, a 7,000 iteration burn-in period and the thinning interval of 2. Under these conditions, 19,500 samples were used for posterior distribution estimates.

1.3 Results

1.3.1 Convergence diagnostics of markov chains

Only data from age 3 to age 8 with sufficient rust infection levels were analyzed using the Bayesian logistic model. The convergence diagnostics were similar for data in all ages. The results from age 8 data were listed for converge diagnostics of Markov chains (Table 1-2). For all parameters, the 0.975 quantile of corrected scale reduction factors was around 1.0 and indicated good convergences of Markov chains for intercept, site effects and variance components (Table 1-2). Randomly chosen starting values showed no differences on posterior estimates of parameters (Table 1-3). Good convergence and mixing among chains indicated the samples were randomly generated from the posterior distribution and the estimates of parameters, based on the Bayesian logistic model with Gibbs sampling, were valid.

1.3.2 General trends of rust infection overtime and over site

The site infection means (Figure 1-1) showed that the rust infection was low in the first two years, increased greatly at age 3, and then tended to stabilize. The average infection rate across sites was 39.2% at age 3 and 44.1% at age 8, ranging from 23.1% to 51.9% at age 3 and from 32% to 59% at age 8. The infection rate at age 8 across sites was 1~8% higher than that at age 3.

From the posterior density plots of site differences between ordered adjacent pairs (Figure 1-2), it was very clear that the means of the site differences deviated from zero,

which indicated significant differences among sites at age 8 (Figure 1-2). Looking at the statistics of posterior distributions of site differences over time (Table 1-4), the means of the differences were larger than zero for most cases. Furthermore, the 95% estimated credible intervals for the difference between site 4 and site 2 effects excluded zero for most times, likewise for the difference between site 3 and site 1. Overall, there were significant site differences overtime. Site 1 had the highest infection, while site 4 had the lowest infection (Figure 1-1). This difference was consistent over time after age 3.

Progeny from six parents showed different amounts of rust infection over time on each site (Figure 1-3). The differences among families started from age 3 and remained unchanged through age 8. There was approximately a 40% difference in gall percentage between the most infected family and the least infected family since age 3 through age 8. Among six families, family A was the least infected in all four tests over time. For example, the percentage of galled trees averaged over 4 sites for family A was 23.9% at age 6, while other families showed higher percentage galled values ranging from 31.5% to 64.2%. Large family differences were evident when infection rates were averaged over sites (Figure 1-4). Even though there were significant site differences in the rust incidence, the rankings of 6 families in terms of rust incidence were consistent across the 4 sites. Family A had the lowest rust infection, while family F had the highest infection at all locations (Figure 1-3). Ranks for families B, C, E and D were very similar overtime and across sites. The lack of rank change among families across testing sites indicated that family by environment interaction was not important for this loblolly pine population on these 4 sites.

1.3.3 Variance components estimates

GCA variance was larger than SCA and other variance components (Table 1-5), indicating larger additive genetic variance than non-additive genetic variance in this study. The variances of site by GCA and site by SCA interaction effects were close to zero, which suggested little family and cross by environment interactions. The additive variances (GCA variance) doubled from age 3 to age 4, and then stabilized through age 8 (Table 1-5). The non-additive variance (SCA variance) changed slightly with age, but was generally unchanged over time. The interaction variance components decreased and moved towards zero as the age increased.

1.3.4 GCA and SCA prediction

By sampling from the posterior distributions of random effects, one can draw inferences for random effects, such as GCA (Table 1-6). Family A had the highest (positive) GCA, and showed the highest non-infected rate in the field, while F had the lowest (negative) GCA, and showed the lowest non-infection rate in the field. Families B and C had similar GCA predictions. These two families showed very little difference for rust infection as both were moderately infected. Discounting the equality of families B and C, the family ranks stayed consistent over time (Figure 1-5). Controlling other variables in the model, the odds ratio of family A to family F equaled 8.9 ($\exp(0.94)/\exp(-1.25)$) which meant there was a 8.94 times better chance of a tree being “resistant” over “susceptible” in family A comparing with a tree in family F (Table 1-6, age 3 data.). In other words, with the overall average infection rate at 39.2% and an odds ratio of 8.94 at age three, the gall

percentage would increase to 80% if family A was substituted by Family F and decrease to 30% if family F was substituted by Family A. With aging, the differences between GCA and the odds ratios increased from age 3 to age 4 and became stable afterward. At age 8, the gall percentage would increase to 82% if family A was substituted by family F and decrease to 25% if family F was substituted by Family A with the overall average infection of 44.1%. The SCA estimates were generally small relative to GCA effects. No cross had relatively stable high SCA over time, even though there were several out of 15 that showed high SCA effects in two or three years.

1.4 Discussion

The logistic model was successfully implemented for rust infection data using a Bayesian method in this study (WINBUGS). For the binary data, the logistic analysis using a Bayesian approach is feasible when transformation approaches are invalid. With availability of the software package (WINBUGS) for Bayesian analysis with Gibbs sampler, the computational difficulty of Bayesian modeling is greatly reduced. The Bayesian method has advantages in terms of ability to input previous knowledge about parameters, and to fit a mixed model without the need for asymptotic and distributional assumptions. Because statistical inferences are based on marginal posterior distributions of the parameters in Bayesian analysis with Gibbs sampling, this method is being increasingly used for estimation of genetic parameters and variance components in animal breeding for both small and large datasets (Noguera et al., 2002; Magnabosco et al., 2000; Luo et al., 2001; Hernandez et al., 1998).

This study provided a good overview of fusiform rust disease development over eight years on multiple sites for a loblolly pine diallel population. The results showed the majority of rust infection occurred by age 3 on these sites. Rust infection increased less than 10% between ages 3 and 8. Overall, the infection rates for the progeny populations were low in the first two years, increased to a slightly sub-maximal point at age three and became essentially stable. This trend was consistent across four sites. (Figure 1-2)

Individual families showed the same profile of increase in rust infection overtime within each site (Figure 1-3). Similar results were reported by Sung and Goddard (1979) in slash pine. In their study, data on rust-resistance were observed at 3 and 5 years after establishment in three locations and two different years for open-pollinated progenies of 92 slash pines. Even though rust incidence ranged from 0% to 100%, the mean infection percentage of the 92 families increased only 10% with the exception of two families. In contrast, Schmidt (1974) reported that the average increase in rust incidence for 19 progeny sites was 14.9% from the third-year measurement to the fifth-year measurement. Only 7 out of 19 had increased less than 10%. The maximum increase of rust infection was 40.6%. Because specific environmental factors (temperature and humidity) are required for teliospore germination and basidiospore dispersion (Siggers, 1947, 1949), periods of low vs. high inocula availability may occur and could account for this dramatic increase of infection at later ages. A typical pattern of disease development would be low infection in first two years and increased infection in years 3-5 as potential susceptible tissue increased (Froelich and Snow, 1986; Schmidt et al., 1999). However, if a family started to show an infection difference at age 3 in a study, this difference typically would remain at least the same or would increase with age even though infection may not have

reached the maximum (Griggs and Schmidt, 1976; Schmidt, 1974; Schmidt et al., 1999). Typically, the rust infection is evaluated at age 5 or age 6 in most breeding and testing programs for both loblolly pine and slash pine (Li et al., 1996). This current study shows that with adequate infection at age 3, family differences could be assessed and in subsequent years they remain relatively reliable in most of cases.

Although significant differences in infection levels among sites were observed in this study, the family rankings for rust percentage were unchanged across the four different hazard sites. Half-sib progeny of parent A were least infected and rust percentages ranged from 10% to 30%, while progeny of parent F were most infected with rust percentages ranging from 50% to 80%. The difference was 40% to 50% irrespective of hazard level. Thus no family by environment interaction was found in this study. This may be due to the fact that these four sites are very close to each other and show no obvious geographic changes. It is possible that the infection differences across sites were due to different spore densities in the fusiform rust microenvironments. Trees suspectedly exposed to high spore density in a high hazard area would expectedly show an elevated infection level relative to a low density in a low hazard site.

In contrast, in other studies (McKeand et al., 1999; Li and McKeand, 1989) it was concluded that although family rank was unchanged, there was a significant genotype by environment interaction variance across a wide range of sites. As reported by Li and McKeand(1989), most genotype by environment interaction from ANOVA was due to heterogeneous regression and no family rank change at the different sites. Later,

McKeand et al. (2003) concluded that resistance to fusiform rust disease at the widely diverse individual sites was relatively unpredictable for the most resistance families. For the observed unpredictable resistance, the postulated explanation was interactions of specific resistance genes in these families and specific pathogen virulence among sites. Kuhlman et al. (1995) also reported a significant interaction effect for family by location in a 7-year old loblolly pine planting. Many families were found to contribute this interaction. They thought the interaction indicated that the relative performance of families was affected by the variation in virulence of the rust pathogen at different locations.

The genetic information about GCA of parents, SCA of crosses, and variance components is very important for selection and breeding. In our study using field data, additive variance was much larger than the non-additive variance over time. The difference increased after age 3 as the infection rate went slightly higher. The predicted GCAs for six parents showed significant differences. The large odds ratio of the best performing family (in terms of resistance) to the worst family indicated that the odds of progeny of family A being “resistant” is much higher than the odds of progeny of family F being resistant. Again, the difference was recognizable from age 3 through age 8.

The current study applied a polygenic model to estimate the variation of rust infection in a six-parent half-diallel cross mating design. Large GCA variance and GCA effects were reported in this study. However, recent evidence shows that resistance in loblolly pine, at least in part, is the result of major genes in the host interacting with virulence genes in the

pathogen (Wilcox et al., 1996; Jordan, 1997; Kuhlman et al., 1997; Amerson et al., 1997). Wilcox et al. (1996) found a single dominant resistance gene Fr1 in tree 10-5. Other resistance genes were also identified in other families by Amerson and coworkers (unpublished data), and Jordan (1997). In the current diallel, progeny of parent A were significantly less infected than progeny from the other five parents in the field. The GCA of this parent was the highest, which indicated strong polygenic effects and/or major gene effect associated with this parent. It is possible that the major gene effect is confounded with GCA effect and could not be distinguished under the polygenic model. A mixed inheritance model including both major gene and polygenes may be used to detect possible major gene. In a previous study, molecular marker mapping has indicated existence of a major rust resistance gene in Family A (Amerson, unpublished data). Both statistical analysis and molecular approaches are needed to further search for and confirm major resistance genes in parent A (See chapter 2 and 3).

1.5 References

1. Amerson, H.V., A.P. Jordan, E.G. Kuhlman, D.M. O'Malley, and R.R. Sederoff. 1997. Genetic basis of fusiform rust disease resistance in loblolly pine. Proc. 24th South For. Tree. Imp. Conf. P. 403.
2. Anderson, R.L., J.P. McCluer, N.D. Cost, and R.J. Uhler. 1986. Estimating fusiform rust losses in five southeast states. South. J. Appl. For. 10: 237-240.
3. Ball, R.D., M. McConchie, and D.J. Cown. 2001. Heritability of internal checking in *pinus radiata*- evidence and preliminary estimates. N. Z. J. For. Sci. 31(1):78-87.
4. Blair, R.L. 1970. Quantitative inheritance of resistance to fusiform rust in loblolly pine. Ph.D thesis. North Carolina State University, Raleigh, NC.
5. Brooks, S.P., and A. Gelman. 1998. General methods for monitoring convergence of iterative simulations. J. Comput. Graph. Stat. 7(4): 434-455.
6. Burton, P.L., K.J. Tiller, and L.C. Gurrin. 1999. Genetic variance components analysis for binary phenotypes using generalized linear mixed models (GLMMS) and gibbs sampling. Genet. Epidemiol. 17: 118-140.
7. Carson, M.J. 1983. Breeding trees for disease resistance in loblolly pine. Ph.D thesis. North Carolina State University, Raleigh, NC.
8. Cabbage F., J. Pye, T. Holmes, and J. Wagner. 2000. An economic analysis of fusiform rust protection research. South. J. Appl. For. 24: 77-85.
9. Froelich, R.C., and G.A. Snow, 1986. Predicting site hazard to fusiform rust. For. Sci. 32(1): 21-35.

10. Gelman, A., and D.B. Rubin. 1992. Inference from iterative simulation using multiple sequences. *Stat. Sci.* 7:457-511.
11. Griggs, M.M., and R.A. Schmidt. 1977. Increase and spread of fusiform rust. In: *Management of fusiform rust in southern pines. Symp. Proc. Univ. Fla., Gainesville.* P. 32-38.
12. Hodge, G.R. , T.L. White, R.A. Schmidt, and J.E. Allen. 1993. Stability of rust infection ratios for resistant and susceptible slash and loblolly pine across rust hazard levels. *South. J. Appl. For.* 17(4): 188-192.
13. Janss, LLG, Van Arendonk JAM, E.W. Brascamp. 1997. Bayesian statistical analyses for presence of single genes affecting meat quality traits in a crossed pig population. *Genetics.* 145: 395-408.
14. Janss, LLG, R. Thompson, and JAM. Arendonk. 1995. Application of Gibbs sampling for inference in a mixed major gene-polygenic inheritance model in animal populations. *Theor. Appl. Genet.* 91:1137-1147.
15. Jordan, A.P. 1997. Fusiform rust disease resistance and genomic mapping in loblolly pine. M.S. thesis. North Carolina State University, Raleigh, NC.
16. Kinloch, B.B., and R.W. Stonecypher. 1969. Genetic variation in susceptibility to fusiform rust in seedlings from a wild population of loblolly pine. *Phytopathology.* 59(9): 1246-1255.
17. Kinloch, B.B., and C.H. Walkinshaw. 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. P. 219-228.

18. Kuhlman, E.G., H. R. Powers, and W.D. Pepper. 1995. Relative fusiform rust resistance of loblolly pine and slash pine sources and families in Georgia and South Carolina. USDA Forest Service Southern Research Station. Research Paper. SE-291.
19. Kuhlman, E.G. 1992. Interaction of virulent single-gall isolates of *Cronatrium quercuum* f.sp. *fusiforme* and resistant families of loblolly pine. For. Sci. 38:641-651.
20. Kuhlman, E.G., and F.R. Matthews. 1993. Variation in virulence among single-aeciospore isolates from single-gall isolate of *Cronatrium quercuum* f.sp. *fusiforme*. Can. J. For. Res. 23: 67-71.
21. Kuhlman, E.G., H.V. Amerson, A.P. Jordan, W.D. Pepper. 1997. Inoculum density and expression of major gene resistance to fusiform rust disease in loblolly pine. Plant Dis. 81 (6): 597-600.
22. Lenhart, J.D. 1994. Characterizing fusiform rust incidence and distribution in east Texas. South. J. Appl. For. 18(1): 29-33.
23. Li, B., S.E. McKeand, and R.J. Weir. 1996. Genetic parameter estimates and selection efficiency for the loblolly pine breeding in the south-eastern US. p. 164-68. In: Tree improvement for sustainable tropical forestry. Proceedings of QFRI-IUFRO Conference. Caloundra, Queensland, Australia.
24. Li, B., and S.E. McKeand. 1989. Stability of loblolly pine families in the southeastern U.S. Silvae Genet. 38: 96-101.

25. Luo, M.F., P.J. Boettcher, L.R. Schaeffer, and JCM. Dekkers. 2001. Bayesian inference for categorical traits with an application to variance component estimation. *J. Dairy Sci.* 2001, 84(3): 694-704.
26. Magnabosco, C.D.U., R. Lobo, and T. Famula. 2000. Bayesian inference for genetic parameter estimation on growth traits for Nelore cattle in Brazil, using Gibbs sampler. *J. Anim. Breed. Genet.* 117: 169-188.
27. McKeand, S.E., B. Li, and H.V. Amerson. 1999. Genetic variation in fusiform rust resistance in loblolly pine across a wide geographic range. *Silvae Genet.* 48(5): 255-260.
28. McKeand, S.E., H.V. Amerson, B. Li, and T.J. Mullin. (2003). Families of loblolly pine that are the most stable for resistance to fusiform rust are the least predictable. *Can. J. For. Res.* 33(7): 1335-1339.
29. Nelson, C.D., R. L. Doudrick, W.L. Nance, J.M. Hamaker, and B. Capo. 1993. Specificity of host pathogen genetic interaction for fusiform rust disease on slash pine. In: *Proc. 22nd South. For. Tree Imp. Conf.* 403-411.
30. Noguera, J.L., L. Varona, D. Babot, and J. Estany. 2002. Multivariate analysis of litter size for multiple parities with production traits in pigs: I. Bayesian variance component estimation. *J. Anim. Sci.* 80(10): 2540-2547.
31. Powers, H.R., F.R. Matthews, and L.O. Dwinell. 1977. Evaluation of pathogenic variability of *Cronartium fusiforme* on loblolly pine. *Phytopathology.* 67:1403-1407.

32. Powers, H.R., R.A. Schmidt, and G.A. Snow. 1981. Current status and management of fusiform rust on southern pines. *Annu. Rev. Phytopathol.* 19:353-371.
33. Schmidt, R.A., R.E. Goddard, and C.A. Hollis. 1974. Incidence and distribution of fusiform rust in slash pine plantations in Florida and Georgia. *Bulletin, Institute of Food and Agriculture Sciences, University of Florida.* No. 763, 21pp.
34. Schmidt, R.A., and D.S. Wilson. 1999. Stability of fusiform rust-resistance loblolly pine families among years and locations in the southeastern USA. *Forest biology research cooperative report #9.* Institute of Food and Agriculture Sciences, University of Florida. 33pp.
35. Siggers, P.V. 1947. Temperature requirements for germination of spores of *Cronartium fusiforme*. *Phytopathol.* 37: 855-864.
36. Siggers, P.V. 1949. Weather and out-breaks of the fusiform rust of southern pines. *J.For.*, 47(10): 802-806.
37. Sluder, E.R. 1993. Results at age 15 years from a half diallel cross among 10 loblolly pines selected for resistance to fusiform rust (*Cronartium quercuum* f. sp. *fusiforme*). *Silvae Genet.* 42:223-230.
38. Smith, B., 2001. BOA version 1.0.0. Available at <http://www.public-health.uiowa.edu/boa/>.
39. Snow, G.A., F.R. Matthews, W.L. Nance, and G.S. Foster. 1990. Effects of pollen source on loblolly pine resistance to *Cronartium quercuum* f.sp. *fusiforme*. *For. Sci.* 36(2): 304-312.

40. Sonw, G.A., R.J. Dinus, and A.G. Kais. 1975. Variation in pathogenicity of diverse sources of *Cronatrium quercuum* f.sp. *fusiforme* on selected slash pine families. *Phytopathology*. 65:170-175.
41. Sorensen, D. A., S. Anderson, D.Gianola, and K. Korsgard. 1995. Bayesian inference in threshold models using Gibbs sampling. *Genet. Sel. Evol.* 27:229-249.
42. Spiegelhalter, D., A. Thomas, and N. Best. 2000. WinBUGS version 1.3 user manual. MRC Biostatistics Unit, Institute of public health, Robinson Way, Cambridge CB2 2SR, UK. <http://www.mrc-bsu.cam.ac.uk/bugs>.
43. Stelzer, H.E, R.L. Doudrick, T.L. Kubisiak, and C.D. Nelson. 1997. Derivation of host and pathogen genotypes in the fusiform rust pathosystem on slash pine using a complementary genetics model and diallel data. *Proc. 24th South For. Tree. Imp. Conf.* 320-330.
44. Stelzer, H.E., R.L. Doudrick, T.L. Kubisiak, and C.D. Nelson. 1999. Prescreening slash pine and *Cronartium* pedigrees for evaluation of complementary gene action in fusiform rust disease. *Plant Dis.* (4): 385-389.
45. Sung, I.S., and R.E. Goddard. 1979. Influence of infection percent on improvement of fusiform rust resistance in slash pine. *Silvae Genet.* 28, 5-6: 173-180.
46. Wilcox, P.L. 1995. Genetic dissection of fusiform rust resistance in loblolly pine. Ph.D thesis, North Carolina State University. Raleigh, NC.
47. Wilcox, P.L., H.V. Amerson, E.G. Kuhlman, B.H. Liu, D.M. O'Malley, and R.R. Sederoff, R.R. 1996. Detection of a major gene for resistance to fusiform rust

disease in loblolly pine by genomic mapping. Proc. Natl. Acad. Sci. USA 93:
3859-3864.

48. 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, NC.

Table 1-1. Six-parents (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 1-2. Convergence diagnostics of Markov chains for 9 parameters (intercept, site effects and variance components). Gandr is Gelman and Rubin reduction factor.

Reduction factor is on the corrected score scale.

Gandr	Intercept	Site1	Site2	Site3	Site4
Estimated	1.00013	1.000524	1.001336	1.001468	1.000175
0.975 quantile	1.000712	1.002008	1.001947	1.001509	1.000399

Gandr	σ^2_{GCA}	σ^2_{SCA}	$\sigma^2_{Site*GCA}$	$\sigma^2_{Site*SCA}$
Estimated	1.000109	1.001629	1.004712	1.003940
0.975 quantile	1.000170	1.004798	1.016177	1.014211

Notes: sites 1-4 denote the site effect of 4 testing locations; σ^2_{GCA} is GCA variance;

σ^2_{SCA} is SCA variance; $\sigma^2_{Site*GCA}$ is GCA by site interaction variance; $\sigma^2_{Site*SCA}$ is SCA

by site interaction variance.

Table 1-3. Parameter estimates by fitting a logistic model using WINBUGS: intercept, site effects and variance components. Statistics from three single chains and from the concatenated chain are showed. The upper number in each cell is the parameter mean from its posterior distribution. The lower number (in parenthesis) is the standard deviation.

	Chain1	Chain2	Chain3	Combined
Intercept	0.293 (0.052)	0.293 (0.052)	0.292 (0.053)	0.292 (0.052)
Site 1	-0.76 (0.11)	-0.76 (0.10)	-0.75 (0.11)	-0.76 (0.11)
Site 2	0.20 (0.11)	0.20 (0.10)	0.20 (0.11)	0.20 (0.11)
Site 3	-0.04 (0.11)	-0.04 (0.10)	-0.05 (0.10)	-0.05 (0.11)
Site 4	0.60 (0.13)	0.60 (0.12)	0.60 (0.12)	0.60 (0.12)
σ^2_{GCA}	0.88 (0.99)	0.91 (1.34)	0.88 (1.03)	0.89 (1.13)
σ^2_{SCA}	0.06 (0.07)	0.06 (0.06)	0.06 (0.07)	0.06 (0.07)
$\sigma^2_{Site*GCA}$	0.006 (0.008)	0.005 (0.006)	0.005 (0.006)	0.006 (0.007)
$\sigma^2_{Site*SCA}$	0.010 (0.014)	0.011 (0.015)	0.012 (0.017)	0.011 (0.016)

Notes: sites 1-4 denote the site effect of 4 testing locations; σ^2_{GCA} is GCA variance; σ^2_{SCA} is SCA variance; $\sigma^2_{Site*GCA}$ is GCA by site interaction variance; $\sigma^2_{Site*SCA}$ is SCA by site interaction variance.

Table 1-4. Summary statistics for posterior distributions of site differences. Using the estimated means of site posterior distributions, site effects were ordered from high to low (site 4, site 2, site 3, site 1), and the statistics of differences between adjacent pairs over time are shown. The upper line in each cell shows the mean and standard deviation of the corresponding site difference, the lower line indicates the estimated 95 % credible interval of the site difference.

	Age3	Age4	Age5	Age6	Age7	Age8
Site 4- site 2	0.92(0.28) [0.37, 1.48]	0.76(0.30) [0.15, 1.33]	0.38(0.20) [-0.03, 0.77]	0.36(0.19) [-0.03, 0.74]	0.25(0.18) [-0.11, 0.61]	0.40(0.19) [0.01, 0.78]
Site 2- site 3	0.17(0.23) [-0.30, 0.63]	-0.06(0.26) [-0.60, 0.46]	0.18(0.18) [-0.18, 0.53]	0.36(0.17) [0.03, 0.69]	0.26(0.16) [-0.06, 0.58]	0.25(0.17) [-0.08, 0.58]
Site 3 - site 1	0.44(0.23) [-0.04, 0.89]	1.19(0.26) [0.66, 1.71]	0.77(0.18) [0.42, 1.12]	0.72(0.17) [0.43, 1.10]	0.87(0.16) [0.55, 1.93]	0.71(0.17) [0.38, 1.04]

Table 1-5. Variance components estimates from the posterior distributions that result from fitting a logistic model using WINBUGS (Spiegelhalter et al., 2000). The upper number in each cell is the mean of posterior distribution, while the lower number (in parenthesis) is the standard deviation.

	Age3	Age4	Age5	Age6	Age7	Age8
σ^2_{GCA}	0.46 (0.61)	0.83 (1.06)	0.91 (1.17)	0.81 (0.90)	0.91 (1.13)	0.89 (1.13)
σ^2_{SCA}	0.10 (0.11)	0.06 (0.08)	0.05 (0.06)	0.02 (0.03)	0.09 (0.09)	0.06 (0.07)
$\sigma^2_{Site*GCA}$	0.03 (0.03)	0.03 (0.04)	0.007 (0.009)	0.006 (0.007)	0.004 (0.005)	0.006 (0.007)
$\sigma^2_{Site*SCA}$	0.02 (0.03)	0.02 (0.03)	0.008 (0.012)	0.008 (0.011)	0.012 (0.018)	0.011 (0.016)

Notes: σ^2_{GCA} is GCA variance; σ^2_{SCA} is SCA variance; $\sigma^2_{Site*GCA}$ is GCA by site interaction variance; $\sigma^2_{Site*SCA}$ is SCA by site interaction variance.

Table 1-6. GCA predictions for resistance obtained from posterior distributions by fitting a logistic model using WINBUGS. The upper number in each cell is the parameter mean from its posterior distribution, and the lower number (in parenthesis) is the standard deviation.

	Age3	Age4	Age5	Age6	Age7	Age8
Family A	0.94 (0.21)	1.31 (0.19)	1.33 (0.16)	1.23 (0.13)	1.27 (0.18)	1.26 (0.16)
Family D	0.55 (0.18)	0.65 (0.17)	0.77 (0.14)	0.75 (0.13)	0.85 (0.17)	0.85 (0.15)
Family E	-0.01 (0.18)	0.01 (0.18)	0.01 (0.14)	0.04 (0.12)	-0.07 (0.18)	-0.09 (0.16)
Family C	-0.10 (0.18)	-0.33 (0.18)	-0.39 (0.14)	-0.23 (0.12)	-0.33 (0.17)	-0.29 (0.15)
Family B	-0.13 (0.17)	-0.32 (0.17)	-0.35 (0.14)	-0.47 (0.12)	-0.37 (0.17)	-0.40 (0.15)
Family F	-1.25 (0.22)	-1.31 (0.20)	-1.37 (0.16)	-1.32 (0.14)	-1.36 (0.19)	-1.32 (0.17)
Odds ratio of family A over F	8.94	13.74	14.88	12.81	13.87	13.20

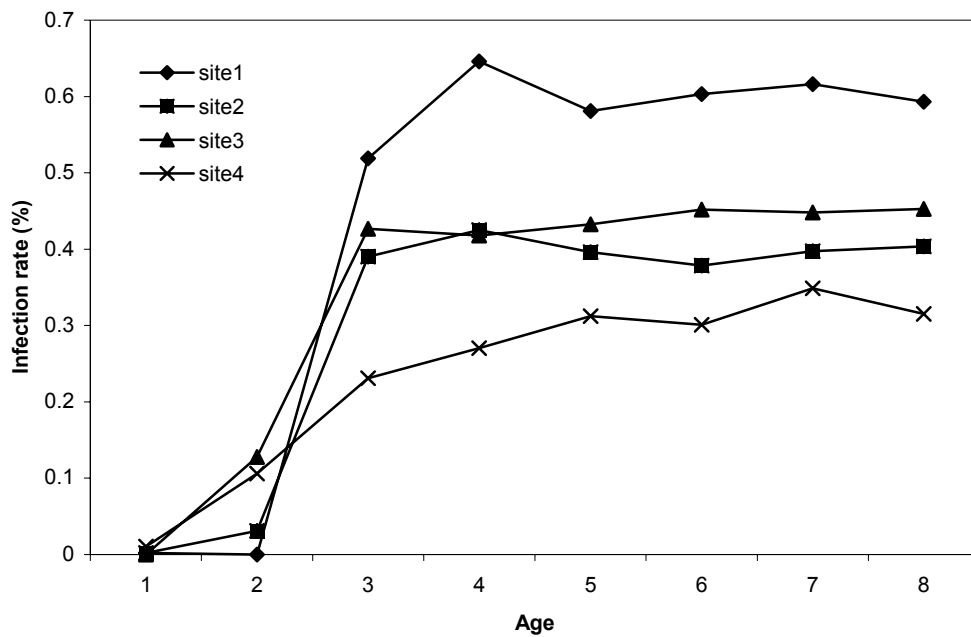


Figure 1-1. Time trends of rust percentages over eight years on four testing sites (1-4).

The site means were calculated by averaging all family means on each site.

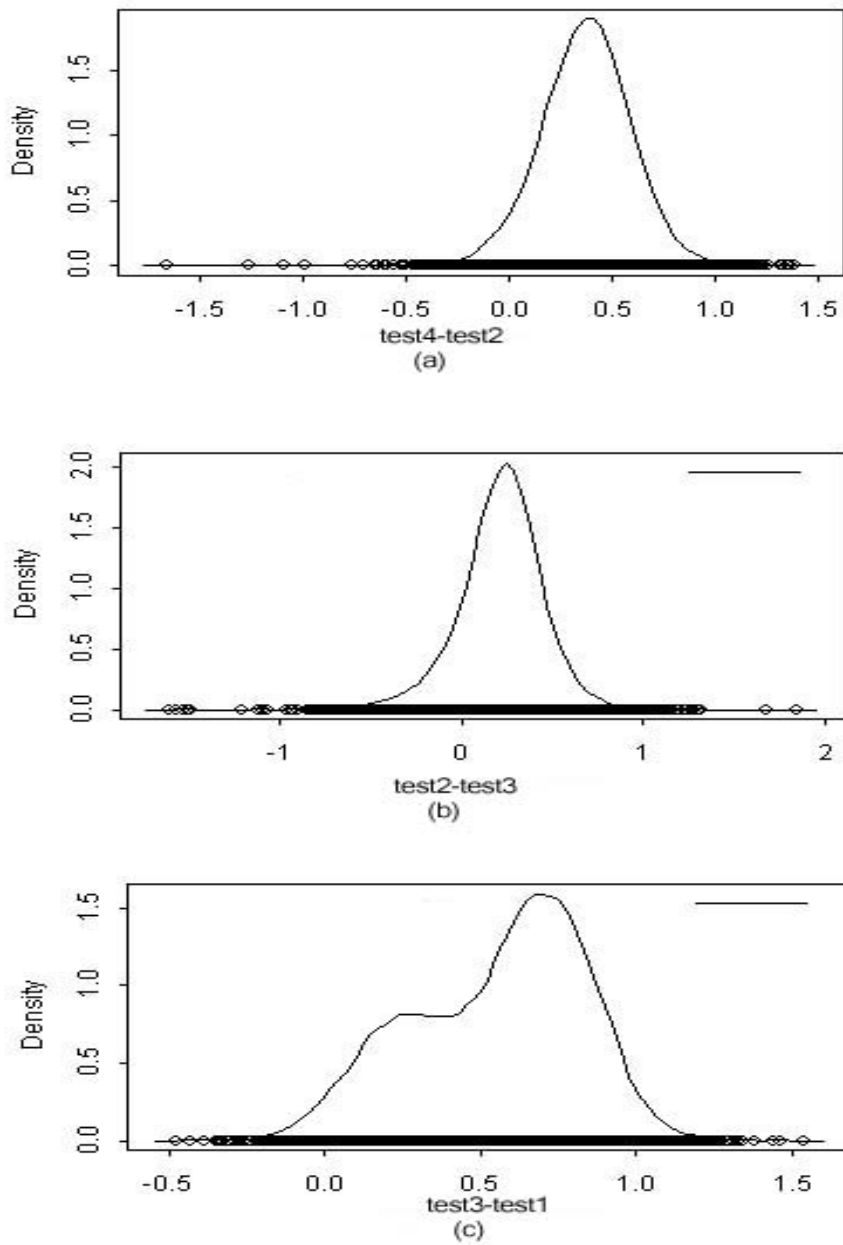
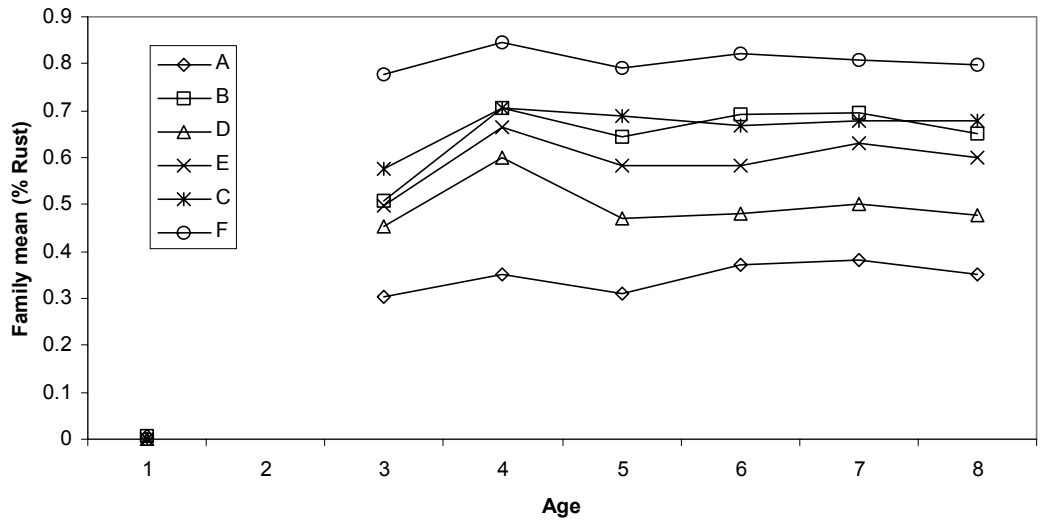
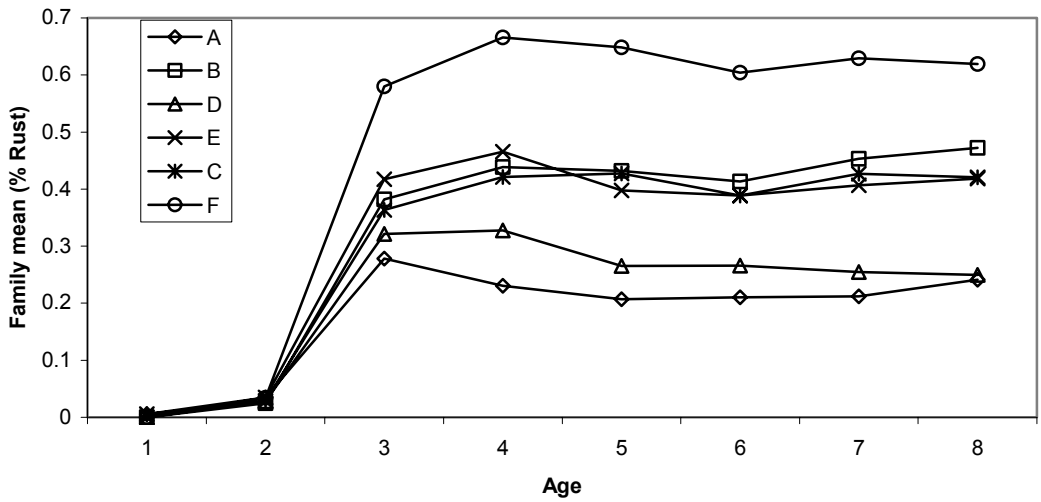


Figure 1-2. Posterior density plots of site effect differences. Posterior densities were estimated using three combined chains with 20,000 iterations in each chain, a 7,000 burn-in period and thinning interval of 2.

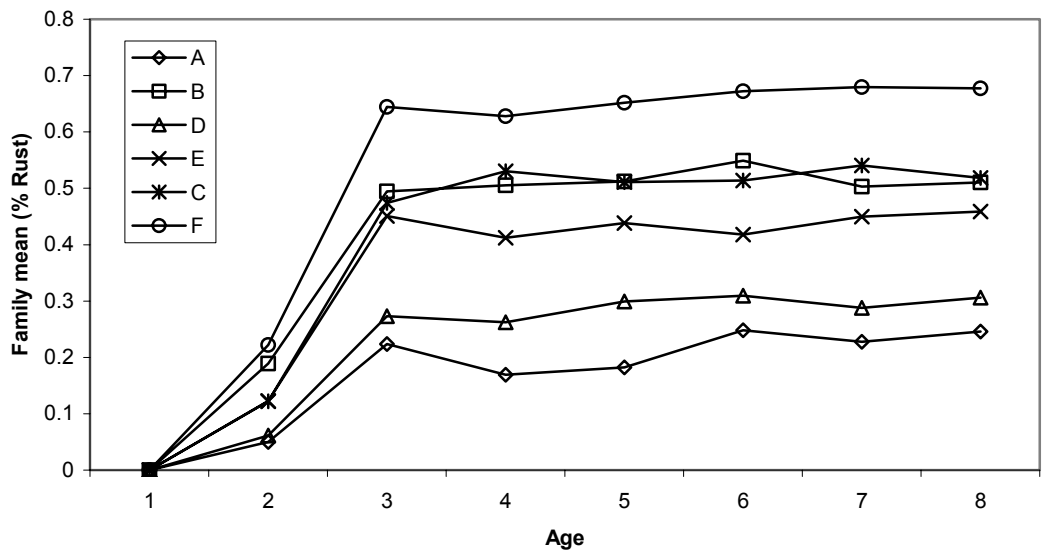


(a) Site 1

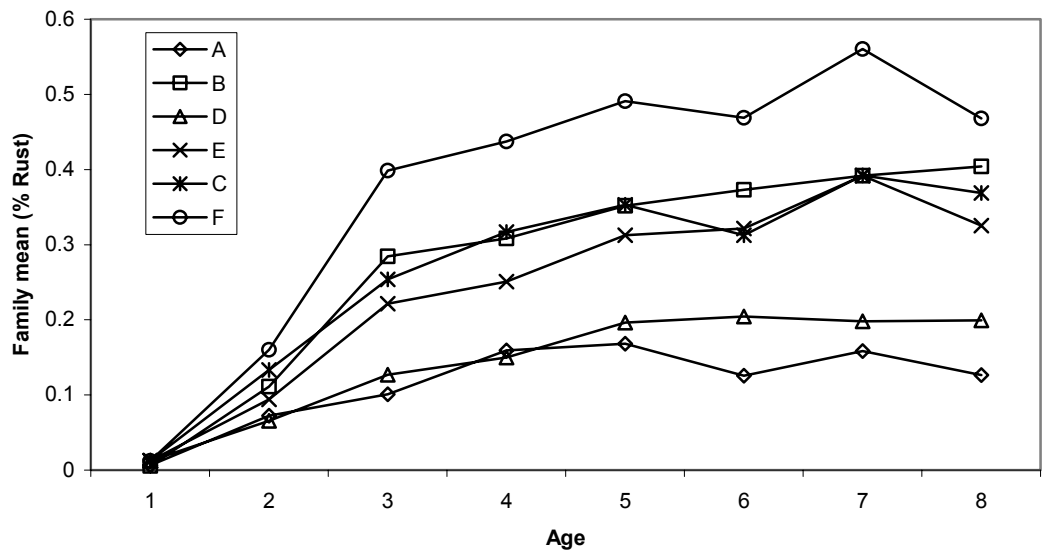


(b) Site 2

Figure 1-3. Time trends of 6 family means (A, B, C, D, E and F) for rust percentage over eight years: site 1 (a), site 2 (b), site 3 (c) and site 4 (d). No data were collected at age 2 for site 1 (a).



(c) Site 3



(d) Site 4

Figure 1-3. (continued)

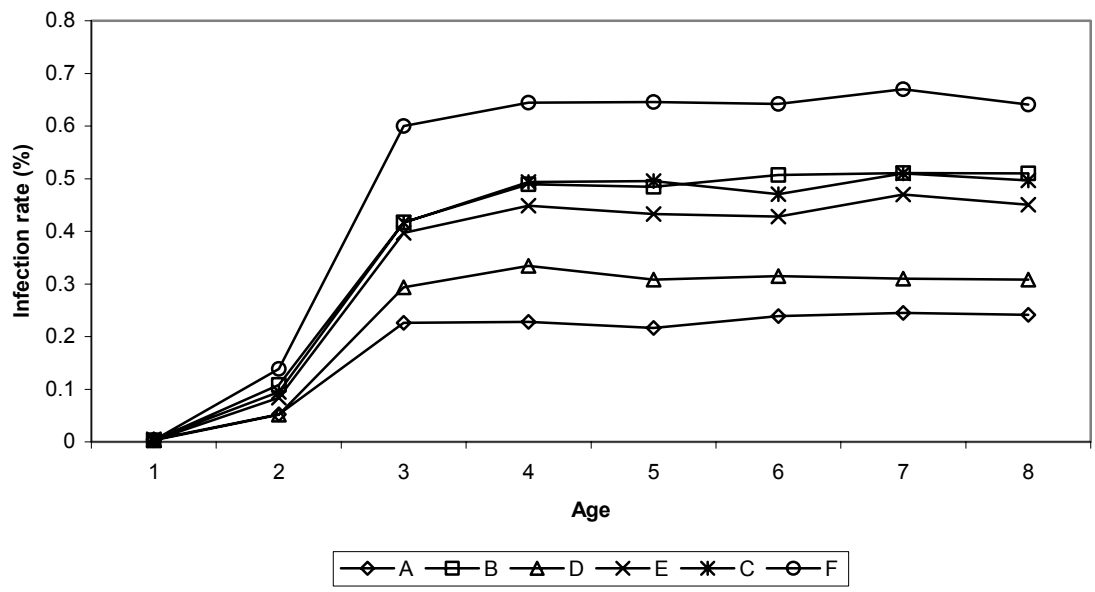


Figure 1-4. Time trend of family (A, B, C, D, E and F) means averaged over 4 sites for rust percentage over eight years.

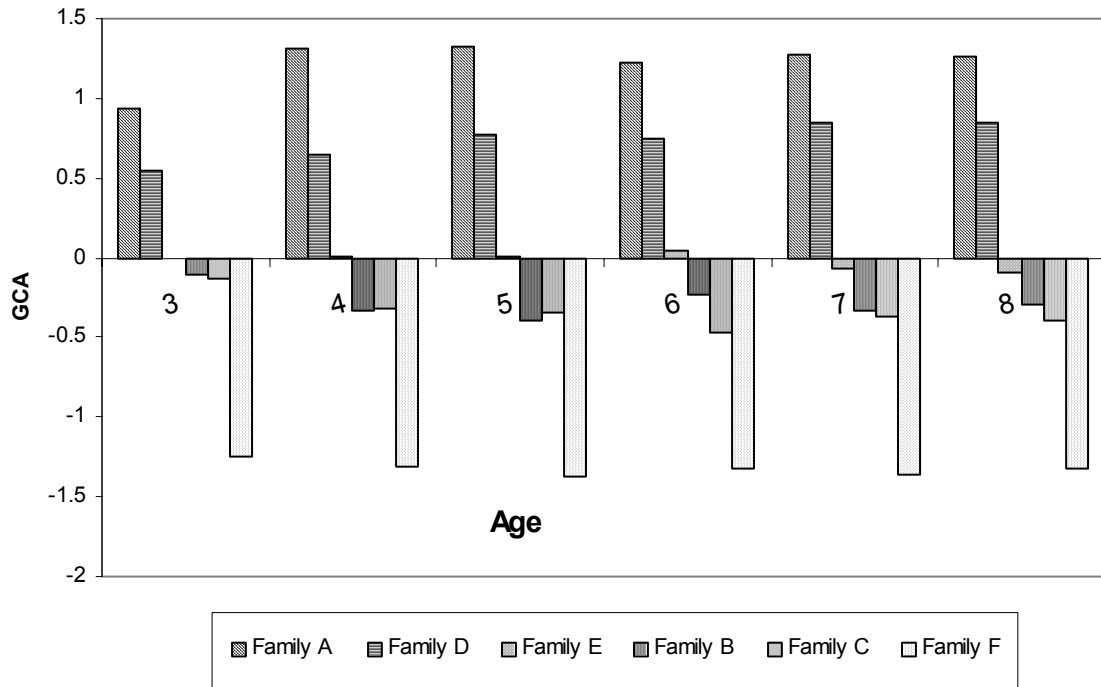


Figure 1-5. Parental GCA predictions for no-galled percentage (deviation from the population mean) of six parents (A-F) from age 3 to age 8.

Chapter II

Bayesian inference for presence of major genes affecting rust resistance in loblolly pine

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 of a threshold model 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 Bayesian Output Analysis was used for convergence diagnostics of multiple chains and characterization of posterior distributions. The mixed inheritance model was compared with a pure polygenic model based on Bayes factor. Results showed that the mixed model with both major gene effects and polygenic effects 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.

Keywords: *Pinus teada* L., *Cronartium quercuum*, f.sp. *fusiforme*, complex segregation analysis, Gibbs sampling, parent block, polygenic effect, diallel mating.

2.1 Introduction

Traditionally, quantitative traits in animal and plant breeding are assumed to be controlled by polygenic models, where many genes with small effects at different loci affect traits simultaneously. However, not all quantitative traits owe their expression to the exclusive action of many genes with small effect. Hanset (1982) and Roberts and Simth (1982) have reviewed examples of single loci (major genes) accounting for an appreciable amount of the genetic variance in quantitative traits. In general, a gene is considered as a major gene if it could be defined individually. Hill and Knott (1990) defined the major gene as one having an effect of at least one phenotypic standard deviation. Separation of the effects by major genes and polygenes is important for understanding major gene expression with polygenic background and for predicting progeny segregation. In a breeding program, identification of major genes could speed up genetic improvement of the traits of interest.

As more efforts have been given to the search and analysis of major genes that control the economically important traits, many genes with large effects have been found. Well-known examples are the Booroola gene in sheep (Piper and Bindon, 1982), the double-muscling gene in cattle (Hanset and Mixhaux, 1985), and the callipyge gene in sheep (Cockett et al., 1994). Examples in plants are the fusiform rust resistance genes in loblolly pine (Wilcox et al., 1996; Amerson et al., 1997), and the dwarfing genes in rice (Jiang et al., 1994). For some major genes, the associated molecular markers have been found, and furthermore, these genes have been mapped. For example, the callipyge gene

in sheep has been mapped to bovine chromosome 18 using bovine minisatellite markers (Cockett et al., 1994).

Several approaches for detection of major genes have been developed. One simple way for searching for major genes is to examine the distribution of phenotypes. When there is a major gene segregating, the phenotypic distribution can depart from normality and exhibit multimodality, skewness and / or kurtosis. Many tests for normality have been developed, but none of these is particularly powerful. Instead of testing normality using random individuals, examining heterogeneity of variance within family by use of known relatives improves the power of tests for major genes. Theoretically, the variance within family (full-sib and half-sib) increases with the segregation of major alleles. Commonly used tests for heterogeneity of variance include the Bratlett test, the log-ANOVA test, the mixture test, and the Fain test. Bartlett's and Fain's test were found to have low power for detecting a major gene with a small number of sibs per family (MacCluer and Kammerer, 1984). The log ANOVA test was reported to be the most robust simple test under a hierarchical data structure (nested design) (Uimari et al., 1996). Except for tests based on normality and variance within family, there are some other approaches for major gene detection, like major-gene indices (MGI), nonparametric line-cross tests and complex segregation analysis.

Up to date, complex segregation analysis is considered to be the most powerful statistical test for detecting major genes affecting quantitative variation. It was proposed by Elston and Stewart(1971) and Morton and Maclean(1974), and developed by human population

geneticists. In complex segregation analysis, the mixed inheritance model, including polygene effects and major gene effects, is considered as a full model. The full model specifies the allele frequency, additive and dominance effects at a single major locus, additive and dominant genetic effects from multiple polygenetic factors, common environmental and random environmental effects. Several other genetics models, such as a pure polygenic model or an oligogenic model would be regarded as the reduced model by suppressing one term in the full model. By testing alternative modes of inheritance, the most likely mode of inheritance is determined and parameter estimates are obtained. There are two kinds of approaches in complex segregation analysis: Maximum likelihood based complex segregation analysis (Elston and Stewart, 1971; Morton and Maclean, 1974) and Bayesian complex segregation analysis (Hoeschele, 1988; Janss et al., 1995; Zeng, 2000). Both of these have been developed and used in animal and plant breeding.

In the maximum likelihood method, the likelihood function of the observed data is maximized with respect to all parameters in each model. Likelihood ratio tests are used for model comparison and hypothesis tests. In the case that alternate hypotheses are not nested, likelihood functions are compared by using Akaike's information criterion (AIC). The model with the smallest AIC is chosen as the best model explaining inheritance.

Likelihood based segregation analysis is widely applied in animal breeding to determine the inheritance mode of litter size at birth in the Brazilian grass mouse (Aulchenko et al., 2002), a disease trait in swine (Thaller and Hoeschele, 1996), and a birth defect in swine (Thaller and Hoeschele, 1996). In plants, Jiang et al. (1994) applied maximum likelihood

based segregation analysis to detect effects of major genes affecting height in rice. They fitted the mixed model including six populations (P1, P2, F1, F2, B1 and B2). The model was extended to include populations of progeny testing such as F₃, B_{1s} and B_{2s} families derived from F2 and backcrosses in a cross between two parental inbred lines. Two recessive dwarf genes were shown to be nonallelic and unlinked using the likelihood ratio method from their results. By combining information from multiple generations, segregation analysis has a higher power to detect major genes (Wang et al., 2001). Wang et al. (2001) applied likelihood ratio tests and supported the fact that the inheritance of resistance to agromyzid beanfly in soybean was controlled by one major gene along with minor genes. In another example, a putative major gene controlling flower color in *Gerbera jamesonii* was defined and accounted for 66% of the total variation in flower hue of a sample representing advanced generations (Tourjee et al., 1995).

Many simulation studies have been carried out for the maximum likelihood approach of segregation analysis (Le Roy et al., 1989; Knott et al., 1992; Borecki et al., 1994; Thaller et al., 1995; Uimari et al., 1996). In general, the larger the effect, the easier it is to detect the gene. If a dominant major gene is present with moderate to low frequency in the population, there is a better chance to detect it (Borecki et al., 1994). By increasing sample size (number of families and size of the families), major genes with smaller effects could be detected. Thaller's simulation study (1995) for rare binary traits showed that major genes could be detected if the half of the homozygote difference of the major gene effects equal to at least one standard deviation of the underlying liability of a rare binary trait.

The maximum likelihood approach for complex segregation analysis involves computation of likelihood by integration and summation. When the design or pedigree becomes more complicated, the computational requirements are extremely demanding, which makes calculation of an exact likelihood very difficult. Ott (1979) suggested estimation of the mixed model using the EM algorithm. However, repeated computations of exact likelihood restrict this method to small pedigrees. Instead of calculating the exact likelihood, many approximation methods have been suggested for estimation (Hasstedt, 1982, 1991; Hoeschele, 1988; Le Roy et al., 1990; Knott et al., 1992 a, b; Janss and Van Der Werg, 1992). Guo and Thompson (1994) proposed the Gibbs sampler and EM algorithm for parameter estimates for mixed models. Along with the likelihood evaluation method by Monte Carlo (Guo and Thompson, 1991), this approach provided an integrated approach for estimation and hypothesis testing of mixed models.

Bayesian complex segregation analysis was first introduced for the animal model by Hoeschele (1988). In her study, estimates of parameters were achieved by a Maximum A-Posteriori Estimation (MAPE) approach using an iterative algorithm. Janss et al. (1995) first applied Gibbs sampling for inferences in a mixed inheritance model in animal breeding. The Gibbs sampler is a special case of the Markov Chain Monte Carlo (MCMC) approach. It is a simulation-based approach to obtain samplers from posterior distributions of parameters. Posterior densities, Bayesian point and interval estimates can be visualized and obtained from the Gibbs sample. Janss et al. (1997) applied this approach for a crossed F2 pig population and found a single gene affecting seven meat quality traits. Another application, using the Bayesian approach with Gibbs sampling,

was on investigation of major genes affecting carcass traits in Japanese black cattle populations (Miyake et al., 1999). They found a large percentage of major gene variance out of the total phenotypic variance (47% and 49%) for the traits back fat thickness and beef marbling and concluded that there might have been a major gene segregating for both traits. In forest tree breeding, a major gene affecting height of loblolly pine was detected in some half-diallel progeny populations (Zeng, 2000).

Simulation studies for Bayesian complex segregation analysis using Gibbs sampling were studied by Miyake et al. (1999) in animal populations and Zeng (2000) in plant populations. In the simulated animal outbred population by Miyake, the additive allele effect and total major gene variance were correctly estimated with the existence of a major gene with large effect. The major gene variance is very close to zero if there are no major genes. They showed that the large major gene variance could be an indicator of major gene segregation in a population. Zeng (2000) simulated a half-diallel progeny population and confirmed that all genetic parameters and genotypes of parents were estimated correctly when the additive effect (a) is larger than 1.0. The additive major gene effect and major gene variance component were likely to be underestimated when more than one major gene is affecting the phenotypic variance. Since Bayesian complex segregation analysis with the Gibbs sampler provides a simple way to estimate genetic parameters based on their marginal posterior distributions, it appears that this method generally is more feasible and flexible than the maximum likelihood segregation analysis. 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 used by Wright (1934). In the threshold model, one assumes that there exists a latent or underlying variable (liability) that has a continuous distribution. A response in a given category is observed if the actual value of liability falls between the thresholds defining the appropriate category. The advent of Gibbs sampler avoids the analytical approximations and makes it possible to estimate quantitative genetic parameters by simulation-based approach. Albert and Chib (1993) used the Gibbs sampler in conjunction with data augmentation in binary and polychotomous data for parameter estimates using Bayesian methods, leading to a computationally simple strategy. Sorensen et al. (1995) applied the threshold model and Gibbs sampler to estimate genetic parameters in the animal model using Bayesian methods.

The purpose of this study is to describe a Bayesian complex segregation analysis with Gibbs sampler and threshold model for binary data in a half-diallel population. A mixed inheritance model and a polygenic model for binary data are presented in the Bayesian concept. The threshold model and Gibbs sampler are applied for inference. All conditional posterior distributions needed for running the Gibbs sampler are derived and given. A parent blocking strategy is implemented to improve the move of the Gibbs chains. The method is illustrated in major rust-resistance gene detection in loblolly pine. Bayes factor is used for model comparison to determine the inheritance mode of rust-resistance in this half-diallel population of loblolly pine.

2.2 Materials and methods

2.2.1 Phenotypic data

The detailed description about mating design and data collection was given in chapter one. In brief, the progeny population was generated by a six-parent half-diallel design. The experiments were established at four different sites using a randomized complete block design with 6 replications and 6 tree row plots per replication. Field data were collected from age 1 to 8 years for each individual tree as presence or absence of galls on stems and/or branches. Phenotype was recorded as 0 (gall) or 1 (no-gall). If a tree was recorded as dead because of the rust infection, this tree was counted as an infected tree (gall) instead of a missing value.

Some individuals showed inconsistent phenotype from year to year due to either inaccurate reading or the self-pruning of branch galls in the later years. To deal with this, phenotypes were assigned to 4 groups with different confidence levels for each group. In group 1, only consistent observations were used, i.e., if a tree showed inconsistent phenotypic readings during the eight years, the readings for this tree were discarded and considered as missing. In group 4, if a tree was ever recorded as infected after three years old, then this tree was considered as infected. Group 2 and Group 3 were intermediate between these two groups. Table 2-1 lists detailed assumptions for each of the four groups.

Based on analysis results from the study diallel in chapter one, there were significant sites effects on rust-resistance of study trees, but no genotype by environment interaction was found. Different infection levels of half-sib families in four testing sites were more likely due to spore density differences in the microenvironment. Stated differently, low infection rates on some sites were probably not caused by genetic effects but rather by environmental effects. In order to make the genetic inference more accurate, only the data from the testing site with the highest infection rate was used. Table 2-2 presents the basic summary of the data used in complex segregation analysis according to our grouping criteria.

2.2.2 Mixed inheritance model for binary response

In the mixed model of a 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. In Hardy-Weinberg equilibrium, the favorable allele is denoted as R with frequency f in the population. Three genotypes, RR, Rr(=rR) and rr, are denoted by 2, 1, 0. 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 was recorded as 1 (non-infected) and 0 (infected). y_i has Bernoulli distribution with probability of rust-free p_i . Define $p_i = P(Y_i = 1) = \Phi(H_i^T \theta)$, where Φ is a standard normal cumulative distribution function (probit model).

The variable Y_i represents the expression of an underlying continuous random variable U_i that is the liability of individual i . When U_i is larger than an unknown fixed threshold t ,

then the observed $Y_i = 1$ and $Y_i = 0$ otherwise. The U_i is assumed to be normally distributed with mean $H_i^T \theta$ and unit variance (1) without loss of generality. Then:

$$U_i \sim N(H_i^T \theta, 1), \text{ for } i=1, \dots, n.$$

Where $H\theta = \beta + Z\mu + WLM$; β is the overall mean; Z is the incidence matrix of the random polygenic effects for all progenies; μ is a vector of random polygenic effects, like GCA (additive genetic effects, g_1 - g_6) and SCA (dominance genetic effects, s_1 - s_{15}), then $\mu^T = (g_1, g_2, \dots, g_6, s_1, s_2, \dots, s_{15})$; W is an unknown $n \times 3$ design matrix of major gene at the single locus; L is a 3×2 indicator matrix of the major gene effects for major genotypes, $L = ((1,0), (0,1), (-1, 0))$; m is a vector of major gene effects, $m' = (a, d)$, where a is additive major genotypic effect, d is dominance major genotypic effect. The product of W , L and m generates three possible genotypes of progeny ($a=RR$, $d=Rr/rR$, $a=rr$).

The probability density of Y is

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

According to Bayes theorem, the joint posterior density of the unobservable parameters and U (latent variables) given the data Y is:

$$\begin{aligned} & \pi(\beta, g, s, a, d, w, Wp, f, \sigma_g^2, \sigma_s^2, U | Y) \\ & \propto \pi(Y | U) \pi(U | \beta, g, s, a, d, w, Wp, f, \sigma_g^2, \sigma_s^2) \pi(\beta, g, s, a, d, w, Wp, f, \sigma_g^2, \sigma_s^2) \end{aligned}$$

Where $\pi(\beta, g, s, a, d, w, Wp, f, \sigma_g^2, \sigma_s^2)$ denotes the prior distribution. The parameters β , g , s , a , d , and f are assumed independent. The overall mean (β), additive major gene effect (a), and dominant major gene effect (d) are given normal priors, i.e., $\beta \sim N(0, k_1^2)$.

$a \sim N(0, k_2^2)$, and $d \sim N(0, k_3^2)$. k_1^2 , k_2^2 , and k_3^2 are assigned to fixed values. Polygenic effects are assumed to be random effects. The GCA effects of six parents are assumed to be identical, independent, normally distributed as $N(0, \sigma_g^2)$, σ_g^2 is given an inverted gamma distribution (IG (μ_1, ν_1)) as the prior with hyper-parameters μ_1, ν_1 . Similarly, the SCA effects of 15 crosses are assumed to be identical, independent, normally distributed as $N(0, \sigma_s^2)$, σ_s^2 is given an inverted gamma distribution (IG (μ_2, ν_2)) as the prior with hyper-parameters μ_2, ν_2 . w is a vector of parent genotypes, Wp is a vector of progeny genotypes, and f is the major gene frequency of the favorable allele in the population. The conjugate beta prior is assumed for the allele frequency f , i.e. $p(f) \sim \text{Beta}(\alpha_f, \beta_f)$. Assuming that the progeny population is in Hardy-Weinberg equilibrium, the probability distribution of the parent i is $p(w_i|f)$ given the favorable gene frequency f in the base population. The joint distribution of parental genotype is the product of individual parental genotype distributions assuming the independence among parents. The genotype distribution of progeny can be easily derived following Mendelian transmission probabilities given their parental genotypes. Assuming the major gene effects and polygenic effects are independent, the joint posterior distribution is derived as the following:

$$\begin{aligned}
& \pi(\beta, g, s, a, d, w, Wp, f, \sigma_g^2, \sigma_s^2, U | Y) \\
& \propto \pi(Y | U)\pi(U | \beta, g, s, a, d, w, Wp, f, \sigma_g^2, \sigma_s^2)\pi(\beta) \\
& \pi(a)\pi(d)\pi(g | \sigma_g^2)\pi(\sigma_g^2)\pi(s | \sigma_s^2)\pi(\sigma_s^2)\pi(w | Wp)\pi(Wp | f)\pi(f) \\
& = \prod_{i=1}^N \{I_{(U_i > 0)}I_{(Y_i=1)} + I_{(U_i < 0)}I_{(Y_i=0)}\} \\
& \left(\frac{1}{\sqrt{2\pi}} \right)^n \exp\left\{-\frac{1}{2}(U - \beta - Zu - WLM)'(U - \beta - Zu - WLM)\right\} \\
& \left(\frac{1}{\sqrt{2\pi k_1^2}} \right) \exp\left(-\frac{\beta^2}{2k_1^2}\right) \left(\frac{1}{\sqrt{2\pi k_2^2}} \right) \exp\left(-\frac{a^2}{2k_2^2}\right) \left(\frac{1}{\sqrt{2\pi k_3^2}} \right) \exp\left(-\frac{d^2}{2k_3^2}\right) \\
& \left(\frac{1}{\sqrt{2\pi\sigma_g^2}} \right)^{n_g} \exp\left(-\frac{\sum_{i=1}^{n_g} g_i^2}{2\sigma_g^2}\right) \left(\frac{1}{\sqrt{2\pi\sigma_s^2}} \right)^{n_s} \exp\left(-\frac{\sum_{j=1}^{n_s} s_j^2}{2\sigma_s^2}\right) \\
& (\sigma_g^2)^{-\mu_1+1} \exp\left(-\frac{v_1}{\sigma_g^2}\right) (\sigma_s^2)^{-\mu_2+1} \exp\left(-\frac{v_2}{\sigma_s^2}\right) \\
& \prod_{k=1}^n p(w_k = w_g | w_{p1(k)}, w_{p2(k)}) \prod_{i=1}^{n_g} p(w_{pi} = w_g | f) f^{\alpha_f-1} (1-f)^{\beta_f-1}
\end{aligned}$$

2.2.3 The Gibbs Sampler

The complexity of posterior distribution makes multidimensional integration very difficult. The Gibbs sampler was used to compute an exact distribution of parameters. High dependency of parent and progeny genotypes is a possible factor to slow down the Markov chains. In order to speed up the move from one stage to another stage of the chains, a parent blocking strategy was used. The parent blocking strategy was introduced by Zeng (2000) for implementation of Bayesian-based complex segregation analysis in the diallel mating design of forest trees. In the parent blocking, the genotypes of a parent and its offspring were blocked and updated simultaneously. The full conditional

distributions for each unknown parameter were derived in order to implement the Gibbs sampler.

2.2.3.1 Latent variables U

Conditional on observed values y and all parameters, the conditional distribution of latent variables U has a simple form. The random variables U_1, U_2, \dots, U_N are independent with a truncated normal distribution with

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

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

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

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

2.2.3.2 Location parameters (β, a, d, g and s)

Given U_1, U_2, \dots, U_N , where U_i are independent $N(H_i^T\theta, 1)$, and $H\theta = \beta + Z\mu + WLm$

let θ_j denote the parameter β, a, d, g and $s, j = 1, \dots, p$. The full conditional distribution of θ_j is proportional to terms obtained by just rearranging the terms that contain θ_j .

$$\theta_j | Y, U, \theta_{-j}, w, Wp, f, \sigma_g^2, \sigma_s^2, \sim N (\tilde{\theta}_j, \sigma_{\tilde{\theta}_j}^2)$$

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

H_{kj} is the k^{th} row j^{th} column element in the matrix H , the same as H_{kr} , and σ_j^2 is the corresponding variance component for θ_j , i.e. for $\theta_j = \beta, a, d, g$ and s , the σ_j^2 are $k_1^2, k_2^2, k_3^2, \sigma_g^2, \sigma_s^2$ respectively.

2.2.3.3 Variance components (σ_g^2 , σ_s^2)

$$\pi(\sigma_g^2 | Y, U, \beta, a, d, g, s, \sigma_s^2, w, Wp, f) \propto (\sigma_g^2)^{-\left(\frac{n_g}{2} + \mu_1 - 1\right)} \exp\left(-\frac{1}{\sigma_g^2} \left[\frac{\sum_{i=1}^{n_g} g_i^2}{2} + v_1 \right]\right)$$

$$\text{So, } \pi(\sigma_g^2 | Y, U, \beta, a, d, g, s, \sigma_s^2, w, Wp, f) \sim IG\left(\frac{n_g}{2} + \mu_1, \frac{\sum_{i=1}^{n_g} g_i^2}{2} + v_1\right)$$

Similarly, the full conditional distribution of σ_s^2 is:

$$\pi(\sigma_s^2 | Y, U, \beta, a, d, g, s, \sigma_g^2, w, Wp, f) \sim IG\left(\frac{n_s}{2} + \mu_2, \frac{\sum_{j=1}^{n_s} s_j^2}{2} + v_2\right)$$

2.2.3.4 Allele frequency (f)

$$\pi(f | Y, U, a, d, g, s, w, Wp, \sigma_g^2, \sigma_s^2, \sigma_t^2, \sigma_{b/t}^2) \propto f^{(\alpha_f + n_1 - 1)} (1 - f)^{(\beta_f + n_2 - 1)}$$

This is a Beta distribution with $(\alpha_f + n_1, \beta_f + n_2)$, where n_1 and n_2 are the number of R and r allele in the base population, $n_1 + n_2 = n_p$, n_p is the number of parents.

2.2.3.5 Major genotypes

The joint conditional distribution of a parent and its all offspring is required to use the parent blocking strategy, i.e.

$$\pi(w_{pi}, w_{i(1)}, \dots, w_{i(ni)} | W_{-i(k)}, w_{-pi}, U, Y, \beta, a, d, g, s, \sigma_g^2, \sigma_s^2, f)$$

where n_i denotes the number of offspring of parent i , and the offspring are indexed by $i(1), i(2), \dots, i(n)$, or $i(k)$ generally. W_{-pi} denotes all other parents except for parent w_{pi} . This joint distribution is proportional to the product of the distribution of parent i 's genotype marginalized with all progeny and the joint distribution of its offspring given all parent's genotypes, i.e.:

$$\begin{aligned} \pi(w_{pi}, w_{i(1)}, \dots, w_{i(ni)} | W_{-i(k)}, w_{-pi}, U, Y, \beta, a, d, g, s, \sigma_g^2, \sigma_s^2, f) &\propto \\ \pi(w_{pi} | W, w_{-pi}, U, Y, \beta, a, d, g, s, \sigma_g^2, \sigma_s^2, f) & \\ \pi(w_{i(1)}, \dots, w_{i(ni)} | W_{-i(k)}, w_p, U, Y, \beta, a, d, g, s, \sigma_g^2, \sigma_s^2, f) & \end{aligned}$$

The marginalized parent's full conditional distribution would be:

$$\begin{aligned} \pi(w_{pi} | W_{-i(k)}, w_{-pi}, U, Y, \beta, a, d, g, s, \sigma_g^2, \sigma_s^2, f) &\propto \\ \pi(w_{pi} = w_T | f) \prod_{k \in i(k)} \sum_{b=1}^3 p(w_{pi} = w_b | w_{p1} = w_T, w_{p2}) \pi(U_k - \beta - Z_k u | w_k = w_b) & \end{aligned}$$

The full conditional distribution of the offspring would be:

$$\begin{aligned} \pi(w_k = w_T | W_{-k}, w_p, U, Y, \beta, a, d, g, s, \sigma_g^2, \sigma_s^2, f) &\propto \\ \pi(w_k = w_T | w_{p1(k)}, w_{p2(k)}) \pi(U_k - \beta - Z_k u | w_k = w_T) & \end{aligned}$$

2.2.4 Updating scheme

To start the Gibbs chains, starting values of latent variables and related parameters were needed. The initial values of parental genotypes (W_p) were generated by 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,

σ_g^2 , σ_s^2 and θ also needed to be initiated with reasonable guesses from their support.

Samplers were taken through the following iterations, starting from $t=0$:

1. Sample latent variables $U_i^{(t)}$ given $W^{(t)}$, $w_p^{(t)}$, $\theta^{(t)}$, $\sigma_g^{2(t)}$, $\sigma_s^{2(t)}$.
2. Sample major genotypes from full conditional distributions by parent blocking.
 - (1) Update parent 1 and its offspring, $w_{p1}, \dots, w_{p1(k)}$, in one block with $w_{p2}^{(t)}, \dots, w_{p6}^{(t)}$ known
 - (2) Update parent 2 and its offspring, $w_{p2}, \dots, w_{p2(k)}$, in one block with $w_{p1}^{(t+1)}, w_{p3}^{(t)}, \dots, w_{p6}^{(t)}$ known
 -
 - (6) Update parent 6 and its offspring, $w_{p6}, \dots, w_{p6(k)}$, in one block with $w_{p1}^{(t+1)}, \dots, w_{p5}^{(t+1)}$ known.

Each offspring is updated twice in each cycle.

3. Sample allele frequency $f^{(t)}$ from its full conditional distribution.
4. Sample location parameters $\theta^{(t)}$ respectively from its full conditional distribution.
5. Sample variance components from its full conditional distribution.
6. Repeat 1-5

Three independent chains were run with different sets of initial values. Bayesian Output Analysis (Smith, 2001) was used for convergence diagnostics and posterior distribution characterization. Convergence of a single chain was check by Raftery and Lewis's dependence factors (Raftery et al., 1992); convergence of multiple chains was check by

Gelman and Rubin's (Gelman et al., 1992) shrink factor. Trace plots and autocorrelation plots were examined for consideration of burn-in period and thinning factor.

In the Gibbs sampler, a normal prior was assigned to the additive major gene effect. Consequentially, the full conditional posterior distribution of the additive major gene effect was a normal distribution. The value of "a" could be essentially both positive and negative values in the samples. 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 (σ_m^2) was calculated as the sum of additive major gene variance (σ_a^2) dominant major gene variance (σ_d^2)

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

The total variance was calculated as the sum of the major gene variance and polygenic variance.

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

2.2.5 Polygenic model and model comparison

In order to test if the proposed mixed inheritance model is a good model to explain rust-resistance in the diallel population, the polygenic model was also fitted. The polygenic model is the subset of the full model generated by suppressing the major gene effect part. Similarly, the threshold model and Bayesian inference with Gibbs sampler were also applied for 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 fully conditional distributions were derived according to Bayes Theorem and the Gibbs sampler algorithm. The update scheme for parameters was also similar except that it did not involve updating steps for genotypes, gene frequency and additive and dominant effects.

Bayes factor was used to compare two models. Since the MCMC approach was applied for statistical inference in both cases, Bayes factor could be approximately calculated from the MCMC output using the following formula:

$$B = \frac{p(Y | M_{MIM})}{p(Y | M_{poly})} = \frac{m_1 \left(\sum_{i=1}^{m_1} \frac{1}{L(y | \theta^{(i)}, M_{MIM})} \right)}{m_2 \left(\sum_{i=1}^{m_2} \frac{1}{L(y | \theta^{(i)}, M_{poly})} \right)}$$

where m_1 and m_2 are the lengths of Markov chains under both models, and $L(y|\theta,M)$ is the likelihood function given the corresponded model (M). Deviance function (-2log-likelihood) was calculated to compare the fitness of different group data under both models.

2.3 Results

2.3.1 Convergence diagnostics of markov chains

Both a mixed inheritance model and a polygenic model were fitted for four groups of data. For inference in the mixed inheritance model, trial chains were run to determine suitable values for a burning-in period and a thinning factor in order to obtain independent Gibbs samples. From the trial chains, it was thought that the independent

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 a spacing of 1000 cycles.

Considering the relatively large sample size in group 4, 100,000 more cycles were added with the same burning-in period and thinning. For each dataset, three Gibbs chains with independent starting values were run. 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. This was true for other parameters.

Raftery and Lewis's dependence factor and Gelman and Rubin's shrink factor were used for convergence diagnostics of the single chain and multiple chains respectively. The analysis results of Gibbs chains from BOA for the group1 dataset are shown for illustration. The trace plots show little dependence among iterations for GCA estimates for parents, gene frequency, GCA and SCA variance (Figure 2-1). Several extreme values were observed in the plots of GCA and SCA variance. The Markov chains showed a little dependence for additive effect and dominant effect. This was more evident in autocorrelation plots (Figure 2-2).

The convergence diagnostics used shrink factors and dependence factors for all estimated parameters in the mixed inheritance model (Table 2-3). For a single chain, a dependence factor less than 5 usually indicates good mixing and convergence. For multiple chains, the 0.975 quantile of corrected scale shrink factor less than 1.2 would be considered as good mixing among chains. Most dependence factors of single chains were less than 5 (Table 2-3) and the 0.975 quantiles of corrected scale shrink factors were less than 1.2,

which indicated that the samples were generated from the stationary distribution. The Gelman and Rubin shrink factor plots (Figure 2-3) and posterior density plots (Figure 2-4) also provided alternative ways for convergence check of Markov chains. The posterior densities of all parameters were similar for all three chains. After examining these diagnostic factors, trace plots, autocorrelation plots, and density plots, it appears that chains converged well and samples could be considered as being generated from the stationary distribution and used for statistical inferences.

2.3.2 Parameter estimates in the mixed inheritance model

Estimated marginal distributions for GCA and SCA variance in the mixed inheritance model are shown in figure 2-4 and estimated posterior means and standard deviations are shown in table 2-4. Major gene variance was calculated as the sum of major additive gene variance and major dominant gene variance, which is a function of additive effect (a), dominant effect (d) and gene frequency (f). Consequently, the Gibbs samplers for the major additive gene variance, the major dominant gene variance and the major gene variance were directly generated along with the update of chains, which could lead to estimates for the posterior density of major gene variance (Table 2-4). Overall, the estimates from the different datasets were similar. GCA variance and SCA variance were similar. Major additive variance was much larger than the major dominant variance, and also larger than ploygenic additive and dominant variances (σ^2_{gca} and σ^2_{sca}). The total variance decreased with increasing of sample size in four groups of datasets. Large major gene variance components were observed. The percentage of the major gene variance of the total variance ranged from 40%-50% in four datasets, indicating a large major gene effect.

The estimated posterior means and standard deviations for additive effect (a) and dominant effect (d) were summarized in table 2-5. Again, estimations under four datasets were similar with some minor differences. If the absolute value of d was subtracted from a , the difference tended to be non-zero. This suggested that the dominant effect (d) was likely to be of a different absolute value than the additive effect, which may indicate the additive allele on this single locus. The estimated posterior density of the additive effect tended to be sharper for the larger dataset, which suggested that the variance of additive effects tended to be smaller.

The parental genotypes with the highest probability were listed in Table 2-6. The number 2, 1, and 0 denote the dominant homozygous (RR), heterozygous (Rr) and recessive homozygous (rr) genotypes. Across the 4 groups, six parents were estimated as the same genotypes with the highest probability in four groups of data, even though the actual probability varied slightly among groups. Parent A was estimated as dominant homozygous (RR) with the probability larger than 0.5, while parent F was estimated as recessive homozygous (rr) with the probability larger than 0.5. Other parents were estimated as heterozygous with probabilities ranging from 0.7 to 0.9. Figure 2-5 showed the posterior probability distributions of parental genotypes for parent A and F.

In the mixed inheritance model, the parental GCA effect serves as polygenic additive effect, counting for an important variance component. The marginal posterior distributions of six parents' GCA effects were well shaped (Figure 2-4). Figure 2-6

showed the posterior means of six parent's GCA prediction for four groups of datasets under the mixed inheritance model. There were some differences regarding the values of GCA prediction in different datasets. The differences among families were smaller in the dataset with the largest sample size; however, the family rankings stayed the same. Parent A was predicted to have the largest GCA effect, where parent F was predicted to have the smallest GCA effect.

2.3.3 Model comparison with polygenic model

The polygenic model with polygenic additive effect (GCA) and polygenic dominant effect (SCA) was more easily implemented in Gibbs sampler with less number of parameters. 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. Similarly, three Gibbs chains were run with independent initial values for all four groups of datasets. Single chains and multiple chains mixed well for all genetic parameters according to the dependence factors and corrected shrink factors (Table 2-7).

Along with the Gibbs chains, the likelihood and deviance ($-2\log\text{likelihood}$) were calculated under both models. Comparing the deviance between models allows one to infer the better model that best fits the dataset, while comparing deviances among datasets under the same model allows one to infer the best dataset that fits the best model. Deviance under the mixed inheritance model was consistently smaller than that under the polygenic model for all groups dataset, which indicated larger likelihood (Table 2-8). Enormous Bayes Factors gave strong support for the mixed inheritance model. In

summary, the mixed inheritance model including polygenic effects and a single major gene effect fitted the data better than the polygenic model. This was true regardless of data quality. By dividing the sample size in each dataset, fitness of data was compared under the same model. The dataset in group 1 fit both models better than the dataset in group2 and 3, but no better than dataset in group 4. Comparisons of deviance indicated that criteria for group 1 and 4 were better in this specific data analysis.

2.4 Discussion

The Bayesian complex segregation analysis with Gibbs sampler and the threshold model was developed for binary data in this study. This method was illustrated in the mixed inheritance model and polygenic model in a half-diallel mating design for rust infection. For statistical models and data used in this study, marginal posterior distributions of major gene effects, polygenic effects (General Combining Ability) and variance components were well approximated by normal distributions. Statistical inferences were based on marginal posterior distributions with no 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. Applying the Gibbs sampler and the threshold model into Bayesian complex segregation analysis simplified complicated computation. 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 that was caused by a huge number of possible genotypic configurations 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 analysis for categorical data.

Even with the parent blocking strategy implemented in the Gibbs sampler, slow mixing that was likely caused by the threshold parameters was still noticed. It might relate to the data augmentation approach in this study. Similar problems were found in other studies (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, other computational strategy and parameterization of model may be needed.

The mixed inheritance model and the polygenic model were well fitted for all four groups of data with some differences in phenotypic scores. The parameter estimates were generally similar overall among groups. By calculating the log-likelihood of the model, it was concluded that the fitness of datasets in group1 and 4 were better, although the data in group 1 may be more reliable but has fewer observations than that in group 4. It is not clear if the minor differences with respect to parameter estimates are due to the data quality or the sample size. Considering the uncertainty of the observed value in the binary data, one way to improve the analysis is to model the observed variable (y) with defined sensitivity and specificity parameters by introducing a new variable that denotes the true state of the outcome (McInturff et al., 2003). This scheme could be easily

implemented in the simple problem. With complex segregation analysis in our study, it may slow the move of the Markov chains by adding more parameters.

From the results of complex segregation analysis, the mixed inheritance model including a major gene effect and polygenic effects fit the data better than the polygenic model with the strong support from the Baye's factor. In the mixed inheritance model, a large percentage of major gene variance of the total variance (40-50%) was estimated, which indicated the existence of major genes. The dominant effect was likely to be different from the additive effect, suggesting the possible additive allele in the potential locus. The differences among parental GCA predictions varied a little among the four datasets, while the rank was same. 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 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 (around 0.6) 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 dominate allele in parent A meant that this parent would be most likely to carry major genes for rust-resistance.

While these analyses strongly support the existence of major gene(s) for rust resistance in this diallel loblolly pine population, it could 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 effect when actually there is 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. The practical examples to support the validity and accuracy of identification of genes using complex segregation analysis are from studies of breast cancer. Complex segregation analysis supported an infrequent autosomal dominant breast cancer locus (Williams and Anderson, 1984; Newman et al., 1988). Later, two dominant breast cancer-susceptibility loci, BRCA1 (Hall et al., 1990) and BRAC2 (Wooster et al., 1994) were mapped and cloned.

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. The possible problems with this new model would be the convergence of the Gibbs chains because the more complex genotype configurations will be involved. Other algorithms or optimal data augmentation approaches may be needed for the extended model

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. As the mixed inheritance model only including one major gene was specified in this study, 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 is crossed with parent F, we might be able to see the segregation in progeny population based on the results of this study. The resistant individuals in the progeny population would likely get the inheritance from the parent A based on analysis results from complex segregation analysis. Mapping major genes in parent A for rust-resistance is suggested for further study.

2.5 References

1. Albert, J.H., S. Chib.1993. Bayesian analysis of binary and polychotomous response data. *J. Am. Stat. Assoc* 88: 669-679.
2. Aulchenko, Y.S., L.O. Araripe, P.S. D'Andrea, A.A. Shishkin, R. Cerqueira, P.M. Borodin, T.I. Axenovich. 2002. Inheritance of litter size at birth in the Brazilian grass mouse (*Akodon cursor*, Sigmodontinae, Rodentia). *Genet. Res.* 80: 55-62.
3. Borecki, L.B., M.A Province, D.C. Rao. 1994. Power of segregation Analysis for Detecting of Major Gene Effects on Quantitative Traits. *Genet. Epidemiol.* 11: 409-418.
4. Cockett, N.E., S.P. Jackson, T.L. Shay, D. Nielsen, S.S. Moore, M.R. Steele, M.R., W. Barendse, R.D. Green, and M.Georges. 1994. Chromosomal localization of the callipyge gene in sheep (*Ovis aries*) using bovine DNA markers. *Proc. Natl. Acad. Sci. USA.* 91(8): 3019-3023.
5. Elston, R.C., and J. Stewart. 1971. A general model for the genetic analysis of pedigree data. *Hum. Hered.* 21:523-542.
6. Gelman, A, and D.B. Rubin. 1992. Inference from iterative simulation using multiple sequences. *Stat. Sci.* 7:457-511.
7. Guo, S.W., and E.A. Thompson. 1994. Monte Carlo estimation of mixed models for large complex pedigrees. *Biometrics.* 50:417-432.
8. Guo, S.W., and E.A. Thompson. 1991. Monte Carlo estimations of variance component models for large complex pedigrees. *IMA J. Math. Appl. Med. Biol.* 8:171-189.

9. Hall, J et al. 1990. Linkage of early onset familial breast cancer to chromosome 17p21. *Science*. 250:1684-1689.
10. Hanset, R. 1982. Major genes in animal production, examples and perspectives: cattle and pigs. 2nd world congress genetics applied livestock products, Vol VI. Publications Agrarias, Madrid (Spain), P439.
11. Hanset, R., and C. Mixhaux. 1985(a). On the genetics determinism of muscular hypertrophy in the Belgian white and Blue cattle breed I: Experimental data. *Genet. Sel. Evol.* 17 (3): 359-368.
12. Hanset, R., and C. Mixhaux. 1985(b). On the genetics determinism of muscular hypertrophy in the Belgian white and Blue cattle breed II: Population data. *Genet. Sel. Evol.* 17 (3): 369-385.
13. Hasstedt, S.J. 1991. A variance components/major locus likelihood approximation on quantitative data. *Genet. Epidemiol.* 8(2): 113-125.
14. Hasstedt, S.J. 1982. A mixed-model likelihood approximation on large pedigrees. *Comput. Biomed. Res.* 15(3): 295-307.
15. Hill, W.G., and S. Knott. 1990. Identification of gene with large effects. *Advances in statistical methods for genetic improvement of livestock*. Springer – verlag. Berlin, 517-538.
16. Hoeschele, I. 1988. Genetic evaluation with data presenting evidence of mixed major gene and polygenic inheritance. *Theor. Appl. Genet.* 76:81-92.
17. Hoeschele, I. 1988. Statistical techniques for detection of major genes in animal breeding data. *Theor. Appl. Genet.* 76:311-319.

18. Janss L.L.G., JAM Van Arendonk, EW. Brascamp. 1997. Bayesian statistical analyses for presence of single genes affecting meat quality traits in a crossed pig population: *Genetics*. 145:395-408.
19. Janss, L.L.G., and Van Der Werf, J.H.J. 1992. Identification of a major gene in F1 and F2 data when alleles are assumed fixed in the parental lines. *Genet. Sel. Evol.* 24:511-526.
20. Janss, L.L.G., R. Thompson, and JAM. Arendonk. 1995. Application of Gibbs sampling for inference in a mixed major gene-polygenic inheritance model in animal populations. *Theor. Appl. Genet.* 91:1137-1147.
21. Jarvik, G.P. 1998. Complex segregation analysis: use and limitations. *Am. J. Hum. Genet.* 63: 942-946.
22. 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.
23. Knott, S.A., C.S. Haley, and R. Thompson. 1992a. Methods of segregation analysis for animal breeding data: a comparison of power. *Hered.* 68:229-311.
24. Knott, S.A., C.S. Haley, and R. Thompson. 1992b. Methods of segregation analysis for animal breeding data: parameter estimates. *Heredity*. 68:313-320.
25. Le Roy, et al. 1990. Evidence for a new major gene influencing meat quality in pigs. *Genet. Res. Camb.* 55: 33-40.
26. Le Roy, P., J.M. Elson, and S. Knott. 1989. Comparison of four statistical methods for detection of a major gene in a progeny test design. *Genet. Sel. Evol.* 21:341-357.

27. Liu, JS, W.H. Wang, A, Kong. 1994. Covariance structure of the Gibbs sampler with application to the comparisons of estimators and augmentation schemes. *Biometrika*, 81: 27-40.
28. MacCluer, J.W., and C.M. Kammerer. 1984. Power of sibship variance tests to detect major genes. In A. Chakravarti (ed.), *Human population genetics: the Pittsburgh symposium*, pp.125-141. Van Nostrand Reinhold, NY.
29. McInturff, P.S., W.O. Johnson, I.A. Gardner, and D. Cowling. 2003. Modeling Risk When Binary Outcomes Are Subject to Error. *Stat. Med.* (Accepted).
30. Miyake, T., C. Gaillard, K. Moriya, and Y. Sasaki. 1999. Accuracy of detection of major genes segregating in outbred population by Gibbs sampling using phenotypic values of quantitative traits. *J. Anim. Breed. Genet.* 116: 281-288.
31. Miyake, T., T. Dogo, K. Moriya, Y. Sasaki. 1999. Bayesian analysis for existence of segregation of major genes affecting carcass traits in Japanese Black cattle population. *J. Anim. Breed. Genet.* 116:207-215.
32. Morton, N.E. and C.J. MacLean. 1974. Analysis of family resemblance III. Complex segregation of quantitative traits. *Am. J. Hum. Genet.* 26: 489-503.
33. Newman, B, and M. Austin. 1988. Inheritance of human breast cancer: evidence for autosomal dominant transmission in high-risk families. *Proc. Natl. Acad. Sci. USA* 85: 3044-3048.
34. Ott, J. 1979. Maximum likelihood estimation by counting methods under polygenic and mixed models in human pedigrees. *Am. J. Hum. Genet.* 31:161-175.
35. Piper, L.R. and B.M. Bindon. 1982. Genetics segregation for fecundity in Booroola merino sheep. *Proceedings of the 1st world congress on sheep and beef*

- cattle breeding. Volume 1: 395-400.
36. Raftery, A.L., and S. Lewis. 1992. How many iterations in the Gibbs sampler? In Bayesian statistics 4, pp 763-774. Oxford University Press.
 37. Roberts, R.C., C. Smith C. 1982. Genes with large effects: theoretical aspects in livestock breeding. 2nd World Congress Genetics Applied Livestock Products, Vol VI. Publications Agrarias, Madrid (Spain), P420.
 38. Sorensen, DA, S. Anderson, D. Gianola, I. Korsgaard. 1995. Bayesian inference in threshold models using Gibbs sampling. Genet. Sel. Evol. 27:229-249.
 39. Thaller, G., L. Dempfle, and I. Hoeschele. 1996. Investigation of the inheritance of birth defects in swine by complex segregation analysis. J. Anim. Breed. Genet. 113: 77-92.
 40. Thaller, G., L. Dempfle, L., and I. Hoeschele. 1996. Maximum likelihood of Rare Binary Traits Under Different Modes of Inheritance. Genetics. 143:1819-1829.
 41. Tourjee, K.R., J. Harding, and T.G. Byrne. 1995. Complex segregation analysis of *Gerbera* flower color. Hered. 74:303-310.
 42. Wang, J., D.W. Podlich, M. Cooper, and 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. 103: 804-816.
 43. Wang, J., J. Gai. 2001. Mixed inheritance model for resistance to agromyzid beanfly in soybean. Euphytica. 122: 9-18.
 44. Wilcox, P.L., H.V. Amerson, E.G. Kuhlman, B.H. Liu, D.M. O'Malley, and R.R. Sederoff, R.R. 1996. Detection of a major gene for resistance to fusiform rust

- disease in loblolly pine by genomic mapping. Proc. Natl. Acad. Sci. USA 93: 3859-3864.
45. Williams W., D.E. Anderson. 1984. Genetics epidemiology of breast cancer: segregation analysis of 200 Danish pedigrees. Genet. Epidemiol. 1:7-20.
46. Wooster R., S. Neuhausen, J. Manigion, Y. Quirk, D. Ford, N. Collins, K. Nguyen, S. Seal, T. Tran, D. Averill, P. Fields, G. Marshall, S. Narod, G.M. Lenoir, H. Lynch, J. Feunteun, P. Devilee, C.J. Cornelisse, F.H. Menko, P.A. Daly, W. Ormiston, R. McManus, C. Pye, C.M. Lewis, L.A. Cannon-Albright, J. Peto, BAJ Ponder, M.H. Skolnick, D.F. Easton, D.E. Goldgar, M.R. Stratton. 1994. Localisation of a breast cancer susceptibility gene (BRCA2) to chromosome 13q by genetic linkage analysis. Science. 265:2088-2090.
47. Wright, S. 1934. Analysis of variability in number of digits in an inbred strains of guinea pigs. Genetics. 19: 506-536.
48. 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 2-1. Criteria to classify the original dataset to four groups.

	Group1	Group2	Group3	Group4
Infected	1. Infected through 8 years 2. Continuously recorded as infected after it was recorded as infected from the first time	1. Infected in year 5,6 and 7. 2. Infected in year 7 and 8. 3. Since the 4 th to 8 th year, 4 out of 5 were recorded as infected. 4. Recorded as infected in the first six years	1. Infected in two successive years 2. Since the 4 th to 8 th year, 4 out of 5 were recorded as infected.	1. Infected in at least one of the following years since year 4.
Non-infected	Non-infected through 8 years	1. Non-infected since year 4. 2. Only one measurement was recorded as infected except for the 8 th year,	1. Non-infected since year 4. 2. Only one measurement was recorded as infected except for the 8 th year 3. recorded as non-infected at least in 4 different years.	Non-infected since year 4
Missing	Recorded as non-infected in first 7 years and infected in the 8 th year	Recorded as non-infected in first 7 years and infected in the 8 th year	Recorded as non-infected in first 7 years and infected in the 8 th year	

- For a particular tree, if it didn't fall into any categories described above, the tree was discarded and treated as missing.

Table 2-2. Data that were used in this study. (a) Data in group1, (b) data in group2, (c) data in group3, (d) data in group4. (e) Summary of data in four groups

(a) Data in group1. In each cell, the number before the slash is the number of non-infected trees in that specific cross, the number after the slash is the total number of trees in that specific cross.

Female/male	A	B	C	D	E	F
B	14/26					
C	10/18	6/28				
D	17/22	8/19	4/18			
E	21/25	7/30	5/27	7/23		
F	4/21	0/36	1/32	3/20	0/25	
	66/112	35/139	26/123	39/102	40/130	8/134

(b) Data in group2. In each cell, the number before the slash is the number of non-infected trees in that specific cross, the number after the slash is the total number of trees in that specific cross.

Female/male	A	B	C	D	E	F
B	18/34					
C	14/25	7/32				
D	24/30	12/27	6/23			
E	26/30	8/34	5/30	10/28		
F	8/30	0/36	1/36	4/27	3/34	
	90/149	45/163	33/146	56/135	52/156	16/163

Table 2-2. (Continued)

(c) Data in group 3. In each cell, the number before the slash is the number of non-infected trees in that specific cross, the number after the slash is the total number of trees in that specific cross.

Female/male	A	B	C	D	E	F
B	18/36					
C	16/34	7/33				
D	24/34	12/33	7/30			
E	27/33	8/36	5/34	11/34		
F	8/32	0/36	1/36	5/33	3/35	
	93/169	45/174	54/172	59/164	36/167	17/172

(d) Data in group 4. In each cell, the number before the slash is the number of non-infected trees in that specific cross, the number after the slash is the total number of trees in that specific cross.

Female/male	A	B	C	D	E	F
B	14/36					
C	11/35	5/34				
D	20/36	8/36	3/33			
E	22/35	7/36	5/35	8/36		
F	6/36	0/36	1/36	3/35	0/35	
	73/178	34/178	25/173	42/176	42/177	10/178

Table 2-2. (Continued)

(e) Summary of data in four groups. The table shows the total number of observations in each group and the percentage of the data of the original complete dataset.

Group	1	2	3	4
#Obs.	370	456	509	530
%	0.69	0.84	0.94	0.98

Table 2-3. Convergence diagnostics (dependence factor and shrink factor) of the Gibbs sampler for all parameters in the mixed inheritance model. There are three independent chains. Only the results using dataset in group1 are shown here.

A: Convergence Diagnostics of the Gibbs sampler for additive effect (a), dominant effect (d), gene frequency (f), GCA variance (σ_g^2), SCA variance (σ_s^2).

		a	d	f	σ_g^2	σ_s^2
Dependence factor	Estimated	1.03	1.08	1.03	1.00	1.00
	0.975	1.07	1.23	1.06	1.00	1.00
Shrink 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

B: Convergence Diagnostics of the Gibbs sampler for parent's GCA. The first row of the table lists the ID's of the parents used in this study.

		A	B	C	D	E	F
Dependence factor	Estimated	1.00	1.04	1.00	1.00	1.02	1.04
	0.975	1.00	1.12	1.01	1.01	1.04	1.12
Shrink factor	Chain1	1.98	2.31	2.24	4.62	3.22	7.59
	Chain2	2.11	2.22	1.95	2.19	2.13	4.36
	Chain3	2.25	2.41	2.22	6.61	11.17	5.93

Table 2-4. Marginal posterior means (pm) and standard deviations (psd) of polygenic additive variance (V_g), polygenic dominant variance (V_s), major gene additive variance (V_{ma}), major gene dominant variance (V_{md}), and major gene variance (V_m) estimated using the datasets in four groups under the mixed inheritance model. V_t is the total variance which was calculated as: $V_t = 2 V_g + V_s + V_{ma} + V_{md}$.

Group	Chain	V_g		V_s		V_{ma}		V_{md}		V_m		V_t	V_m / V_t
		pm	psd	pm	psd	pm	psd	pm	psd	pm	psd		
1	1	1.6	1.4	1.5	1.0	5.0	4.1	0.7	0.9	5.7	4.2	9.5	53%
	2	1.6	1.4	1.5	1.0	5.3	4.3	0.7	0.8	6.0	4.3	10	53%
	3	1.5	1.4	1.5	1.0	4.5	3.7	0.6	0.8	5.1	3.8	9	50%
2	1	1.5	1.3	1.7	1.2	3.8	3.7	0.6	0.8	4.4	3.8	8.5	45%
	2	1.5	1.2	1.7	3.8	3.9	3.8	0.6	0.8	4.5	3.8	8.6	45%
	3	1.5	1.6	1.7	1.4	3.5	3.5	0.6	0.9	4.0	3.6	8.2	43%
3	1	1.5	1.4	1.6	1.2	4.6	3.9	0.7	0.9	5.3	3.9	9.2	50%
	2	1.5	1.3	1.7	1.3	3.7	3.3	0.6	0.9	4.3	3.3	8.4	44%
	3	1.6	1.4	1.6	1.8	4.9	3.9	0.7	0.9	5.6	3.9	9.7	51%
4	1	1.6	1.8	1.6	1.3	3.4	2.8	0.8	1.0	4.2	2.9	8.2	41%
	2	1.6	1.6	1.6	1.1	3.5	2.9	0.8	1.1	4.3	3.0	8.3	42%
	3	1.5	1.3	1.6	1.2	3.4	3.0	0.8	1.0	4.2	3.0	8	43%

Table 2-5. Estimated posterior mean (Pm) and standard deviation (Psd) of the additive effect (a) and dominant effect (d) at the single locus using four groups of data in the mixed inheritance model

		Group1			Group2			Group3			Group4		
Chain		1	2	3	1	2	3	1	2	3	1	2	3
a	Pm	3.0	3.1	2.8	2.5	2.5	2.4	2.8	2.5	2.97	2.3	2.4	2.4
	Psd	1.3	1.3	1.3	1.4	1.4	1.3	1.3	1.3	1.3	1.1	1.1	1.2
d	Pm	-1.2	-1.1	-1.0	-0.9	-1.2	-0.7	-1.2	-0.8	-0.9	-1.4	-1.4	-1.5
	psd	1.5	1.5	1.5	1.6	1.5	1.7	1.6	1.6	1.6	1.5	1.5	1.6

Table 2-6. Estimated parental major gene genotypes. The 0, 1 or 2 in each cell denote rr, Rr(rR) or RR, the genotype of the parent. The number in the parenthesis is the probability of the parent being the correspondent genotype. The parent IDs are A, B, C, D, E, and F.

Group	Chain	A	B	C	D	E	F
1	1	2(0.65)	1(0.80)	1(0.88)	1(0.89)	1(0.81)	0(0.85)
	2	2(0.61)	1(0.81)	1(0.91)	1(0.89)	1(0.88)	0(0.88)
	3	2(0.64)	1(0.77)	1(0.88)	1(0.90)	1(0.64)	0(0.82)
2	1	2(0.58)	1(0.71)	1(0.82)	1(0.83)	1(0.76)	0(0.59)
	2	2(0.63)	1(0.75)	1(0.82)	1(0.71)	1(0.71)	0(0.54)
	3	2(0.52)	1(0.75)	1(0.75)	1(0.80)	1(0.72)	0(0.66)
3	1	2(0.63)	1(0.78)	1(0.91)	1(0.84)	1(0.83)	0(0.61)
	2	2(0.52)	1(0.77)	1(0.72)	1(0.88)	1(0.79)	0(0.57)
	3	2(0.54)	1(0.75)	1(0.88)	1(0.89)	1(0.85)	0(0.70)
4	1	2(0.61)	1(0.85)	1(0.92)	1(0.85)	1(0.87)	0(0.56)
	2	2(0.63)	1(0.86)	1(0.79)	1(0.91)	1(0.77)	0(0.59)
	3	2(0.45)	1(0.76)	1(0.73)	1(0.87)	1(0.81)	0(0.56)

Table 2-7. Convergence diagnostics (dependence factor and shrink factor) of the Gibbs sampler for all parameters in polygenic model. There were three independent chains.

Only the results using dataset in group1 are shown here.

		V_g	V_s	A	B	C	D	E	F
Dependence factor	Estimated	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
	0975	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
Shrink factor	Chain1	1.02	0.98	2.20	2.12	2.28	2.18	1.74	1.09
	Chain2	1.00	1.01	1.92	1.89	1.09	2.09	2.00	1.78
	Chian3	0.99	1.04	2.12	2.05	2.17	3.14	2.30	1.08

- V_g : GCA variance; V_s : SCA variance.
- A-F is GCA effects for parent A-F.

Table 2-8. Estimated posterior mean of deviances for four datasets under the mixed inheritance model and the polygenic model and approximated Bayes factor.

	Group1	Group2	Group3	Group4
D(MIM)	158	245	250	230
D(poly)	327	434	496	452
Bayes Factor	3.01E+3	6.69E+1	5.10E+2	8.74E+3
D(MIM) /n	0.43	0.54	0.49	0.44
D(Poly)/n	0.88	0.95	0.97	0.85

* D(MIM) denotes the posterior mean of deviance under the mixed inheritance model;
D(poly)denotes the posterior mean of deviance under the polygenic model; n denotes the
number of observations in each dataset.

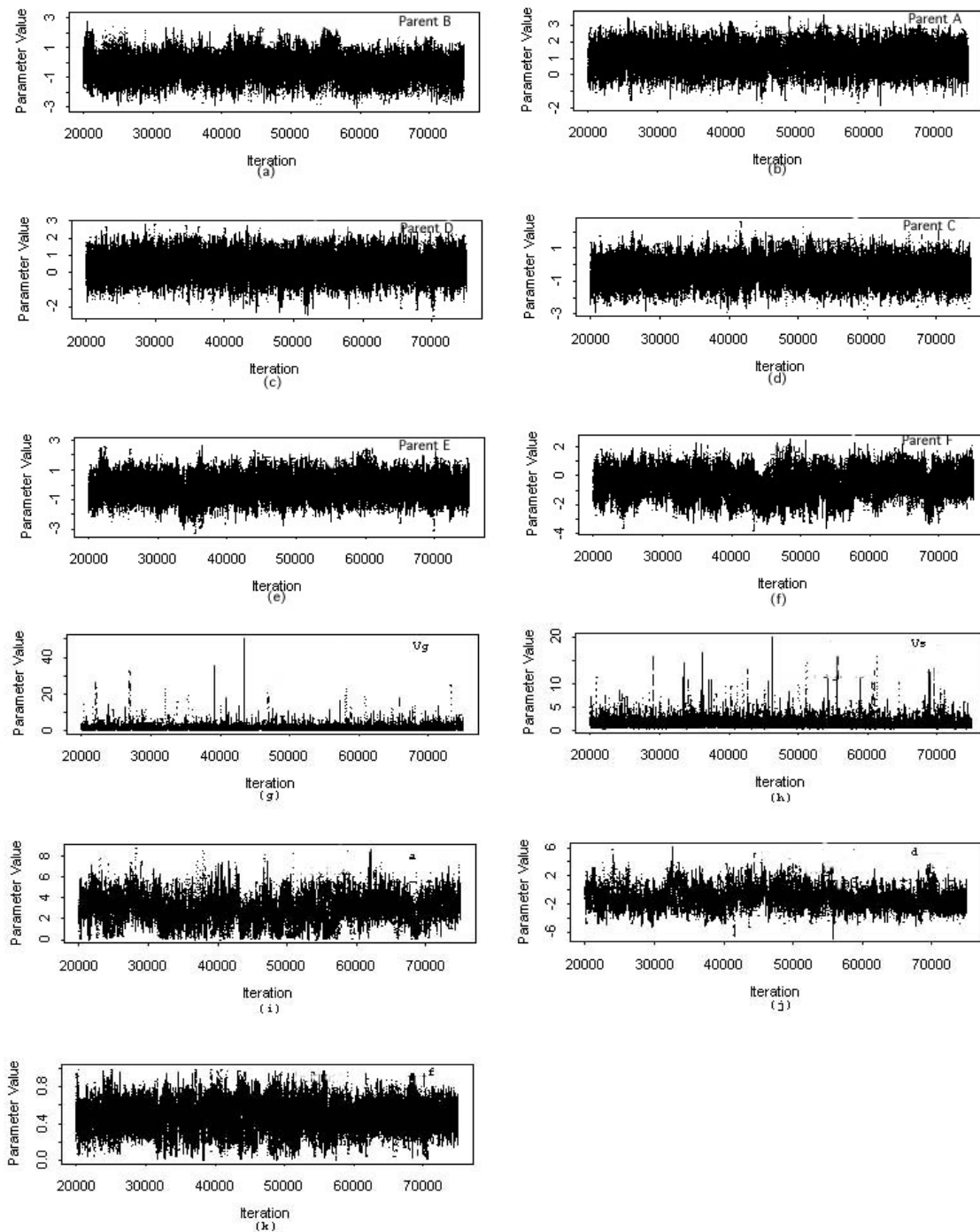


Figure 2-1. Trace plots of eleven genetic parameters including GCA prediction of six parents (a-f), GCA variance component V_g (g), SCA variance component V_s (h), additive effect (i), dominant effect (j) and the favorable allele frequency (k) for group1 data. There are three independent chains in each plot, three chains are denoted as the straight line, the dot line and the dash line.

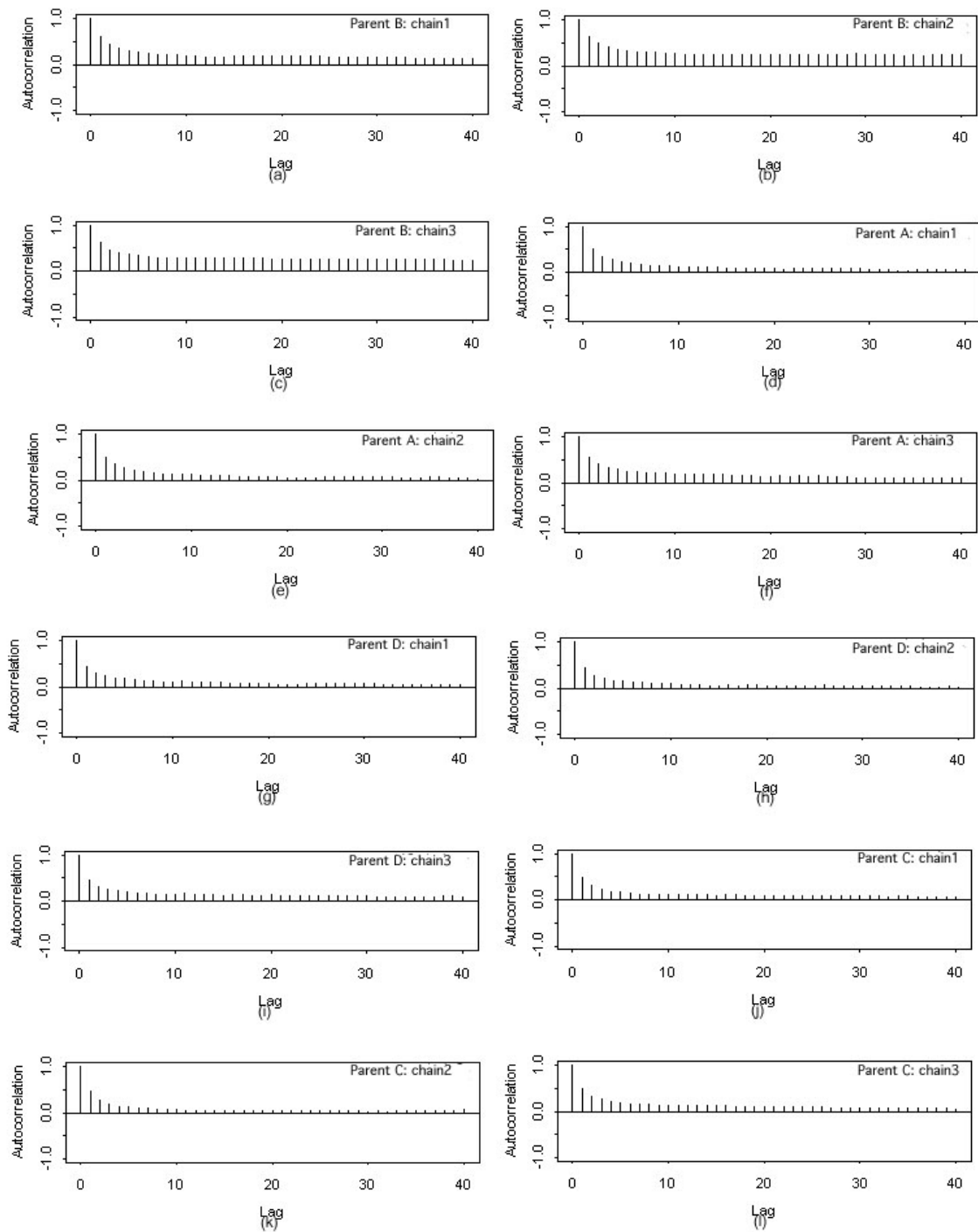


Figure 2-2-1. Autocorrelation plots of GCA predictions of six parents (a-s) for group 1 data. There are three independent chains for each parameter and thus three autocorrelation plots of individual chains for each parameter. The parent ID and the chain ID are shown in the upper right corner of each plot.

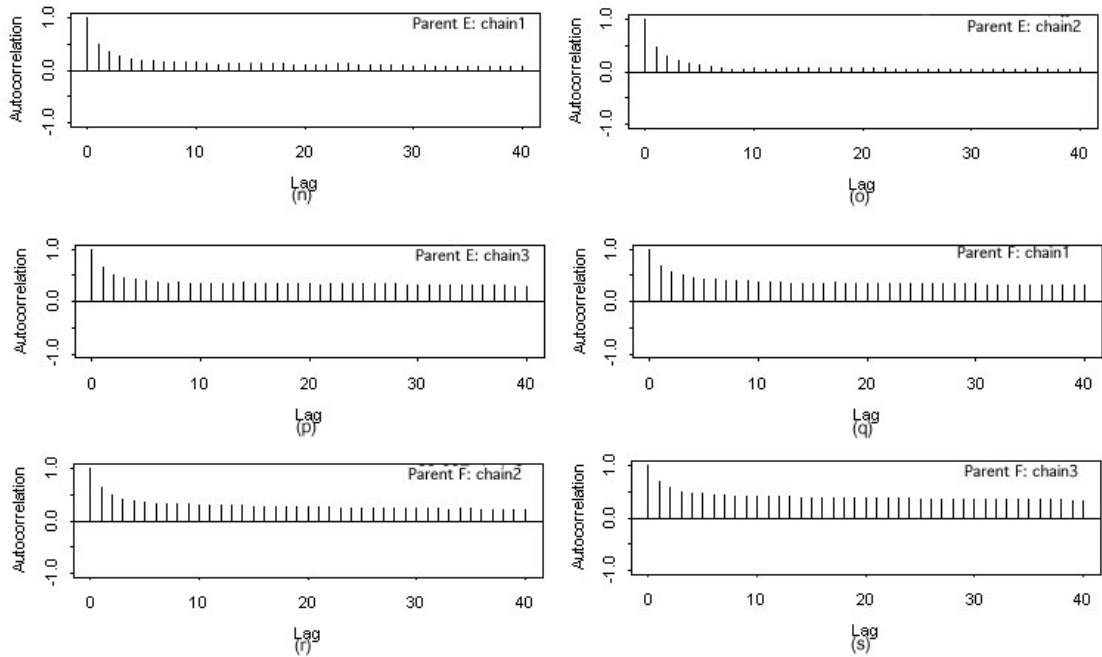


Figure 2-2-1 (continued). Autocorrelation plots of GCA predictions of six parents (a-s) for group 1 data. There are three independent chains for each parameter and thus three autocorrelation plots of individual chains for each parameter. The parent ID and the chain ID are shown in the upper right corner of each plot.

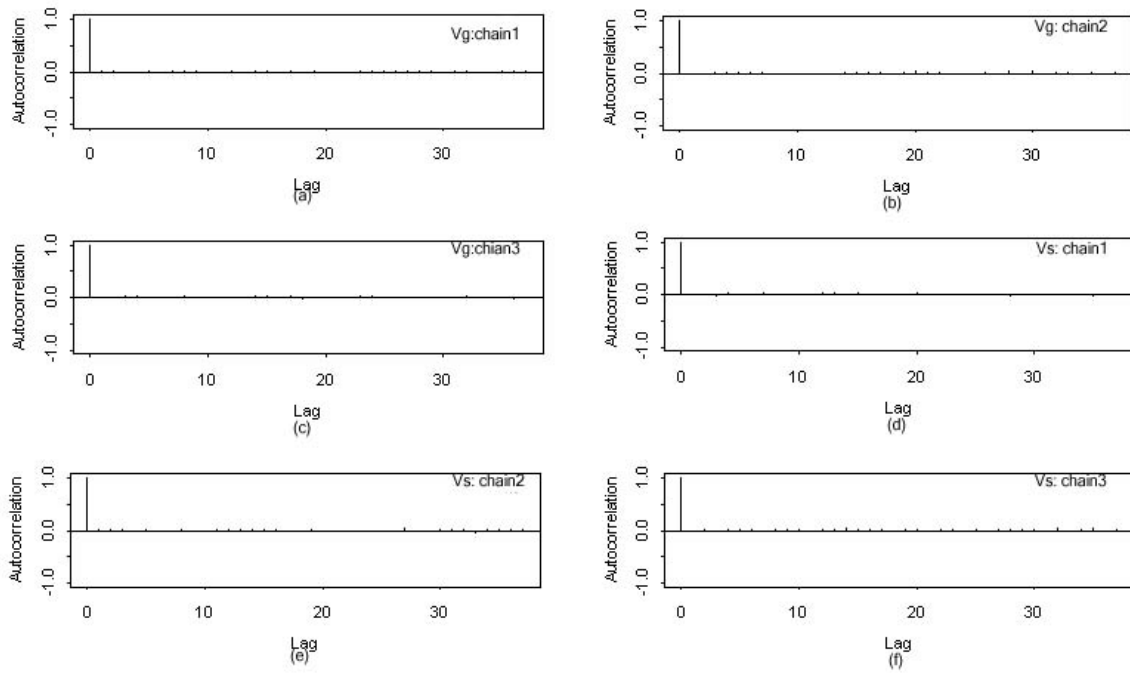


Figure 2-2-2. Autocorrelation plots of five genetic parameters including GCA variance component (a-c), SCA variance component (d-f). There are three independent chains for each parameter and thus three autocorrelation plots from each independent chain. The parameter name and the chain ID are shown in the upper right corner of each plot. Notes: Vg is the GCA variance component, Vs is the SCA variance component.

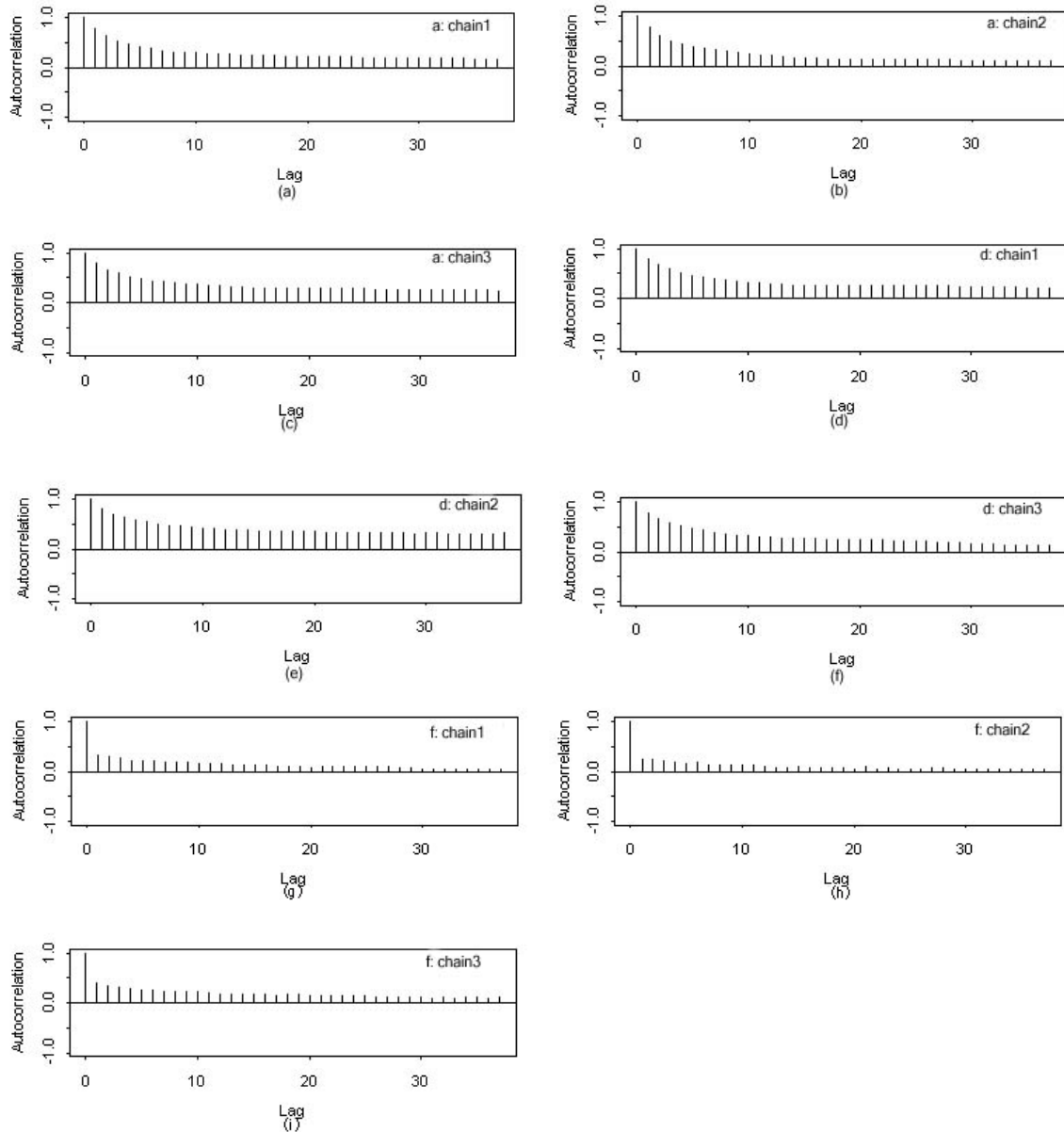


Figure 2-2-3. Autocorrelation plots of genetic parameters including additive effect (a-c), dominant effect (d-f) and the favorable allele frequency (g-i). There are three independent chains for each parameter and thus three autocorrelation plots from each independent chain. The parameter name and the chain ID are shown in the upper right corner of each plot.

Notes: a is the additive effect, d is the dominant effect; f is the favorable allele frequency.

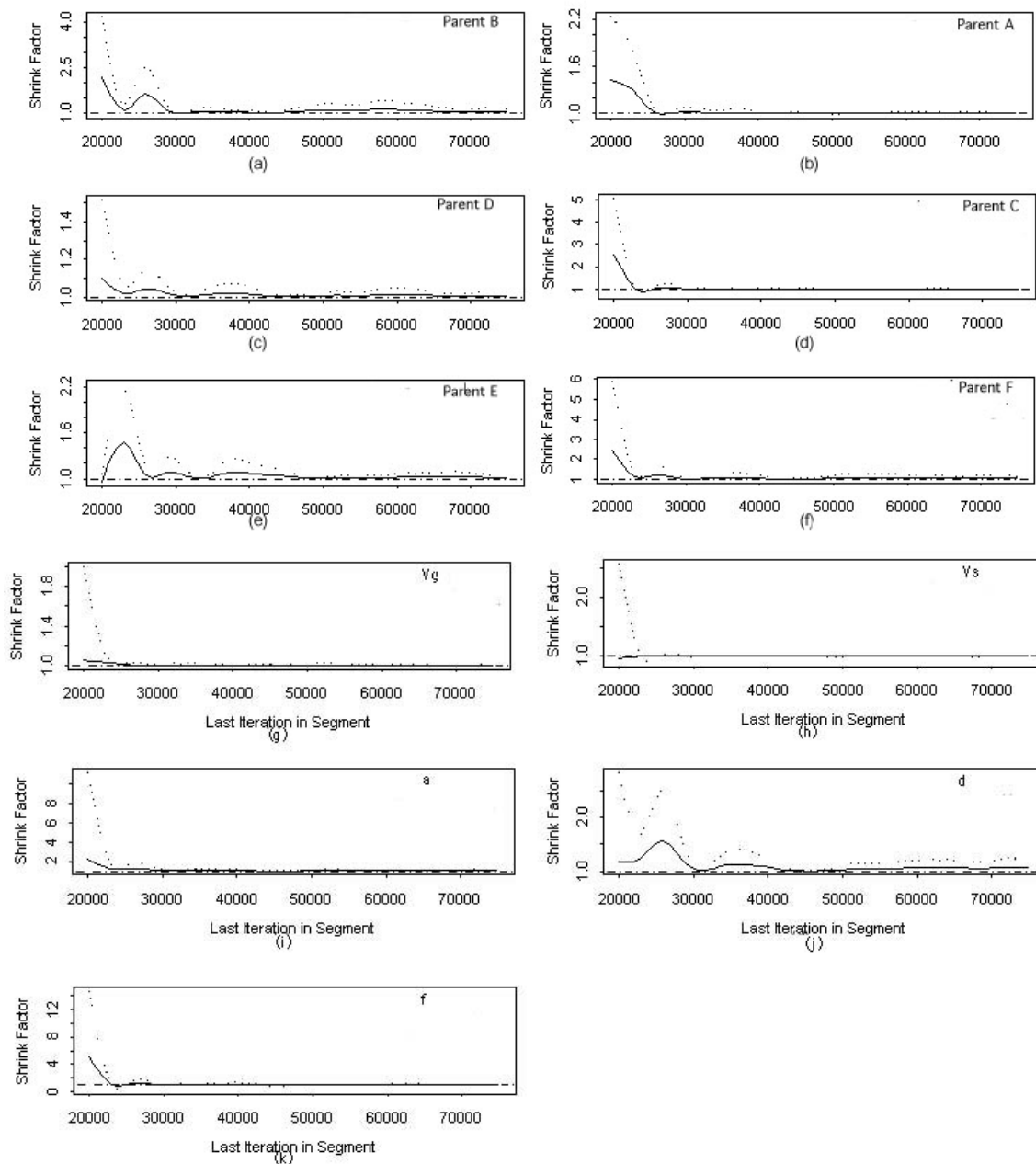


Figure 2-3. Gelman and Robin shrink factor plots of eleven genetic parameters including GCA prediction of six parents (a-f), GCA variance component (g), SCA variance component (h), additive effect (i), dominant effect (j) and the favorable allele frequency (k) for group 1 data.

Notes: notation in the upper right corner of the first six plots represents the parent ID for GCA prediction, V_g is the GCA variance component, V_s is the SCA variance component, a is the additive effect, d is the dominant effect; f is the favorable allele frequency.

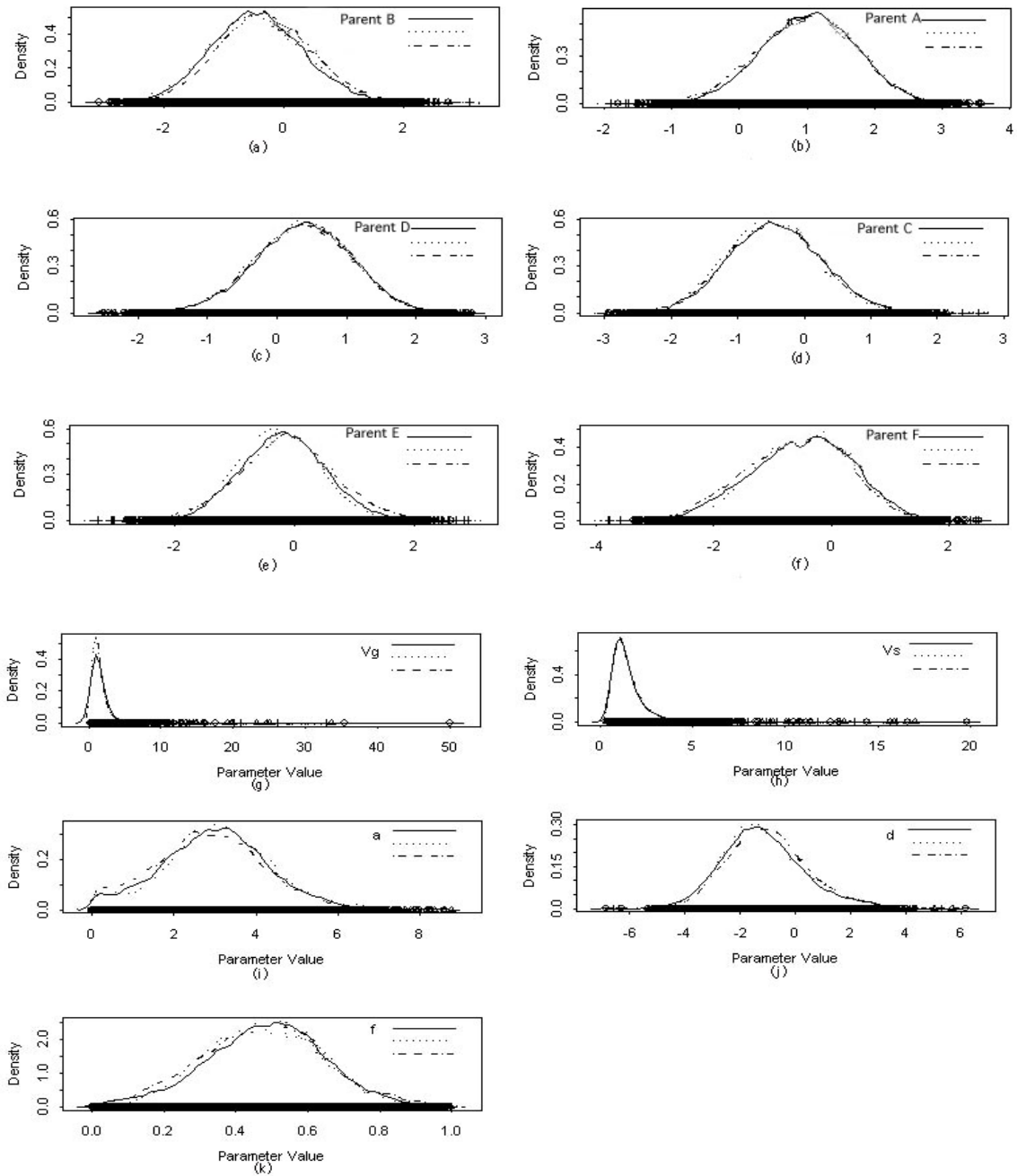


Figure 2-4. Posterior density plots of eleven genetic parameters including GCA prediction of six parents (a-f), GCA variance component (g), SCA variance component (h), additive effect (i), dominant effect (j) and the favorable allele frequency (k) for group1 data. There are three independent chains generating three posterior density plots for each parameter which are denoted as the straight line, the dot line and the dash line in each plot.

Notes: notation in the upper right corner of the first six plots represents the parent ID for GCA prediction, Vg is the GCA variance component, Vs is the SCA variance component, a is the additive effect, d is the dominant effect; f is the favorable allele frequency.

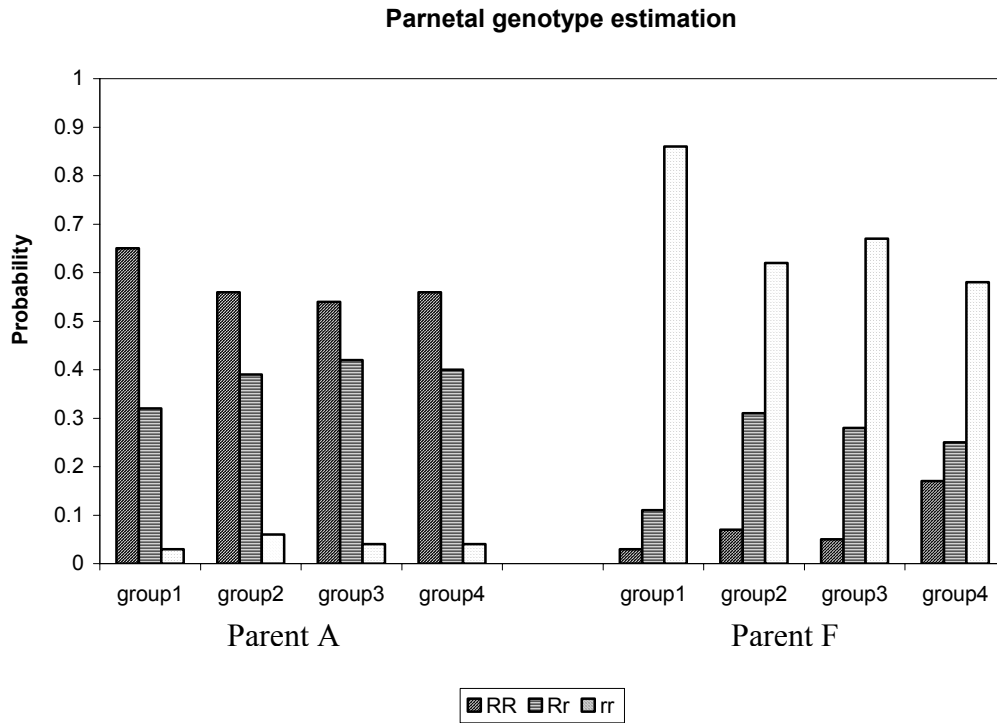


Figure 2-5. Marginal posterior distributions of major gene genotypes for parent A (left) and parent F (right). The results from four groups of datasets are shown in the plot.

GCA predictions in the mixed inheritance model

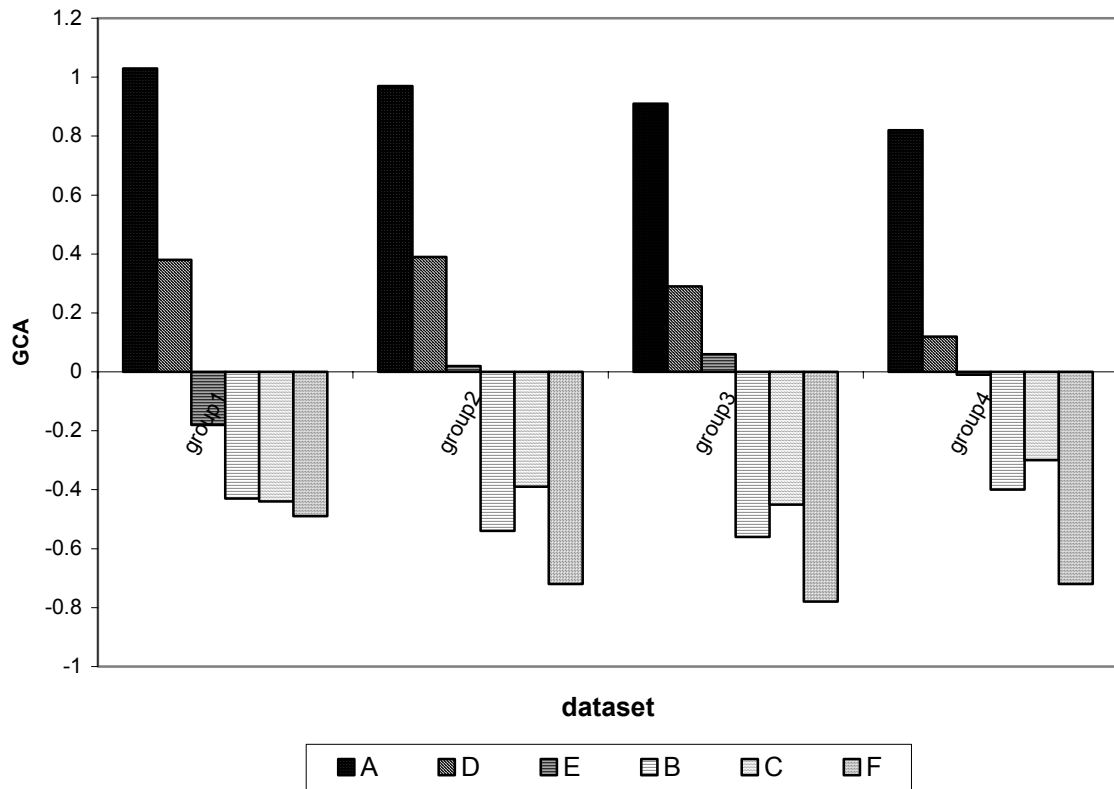


Figure 2-6. Parental GCA predictions for six parents (A-F) by the mixed inheritance model. The GCA was estimated as the deviation from the population mean by four groups of datasets.

Chapter III

Search for molecular markers associated with resistance genes in loblolly pine

Abstract

Bulk segregant analysis (BSA) and marker / trait co-segregation analysis were successfully adopted in mapping a rust resistance locus (Fr2) in parent A. Replicated maternal haploid DNA bulks were constructed based on progeny disease phenotype when the progeny of parent A × F showed an intermediate rust infection level with single spore inoculation. Candidate markers from BSA studies were assessed across parent A megagametophyte DNA from the progeny to reveal RAPD markers linked with the Fr2 locus in parent A. These markers were used to develop an improved map (framework linkage map) of the Fr2 linkage group. Previous work by Wilcox (1995) identified two markers (A19_560 and A11_400) in parent A associated with field resistance (gall vs. no gall phenotype) in the same diallel was evaluated in this current work. Additional markers linked with A19_560 and A11_400 were obtained in this study but they did not clarify understanding for the basis of field resistance. In the current diallel study (a six parent half diallel), two mixed inocula were used to inoculate the progeny population of 12 crosses. From the diallel study, progeny from the cross of parent E by A showed 75% rust infection with extremely high spore density (250,000sp/ml ~ 300,000sp/ml) using two pooled inocula. The interaction of two pairs of complementary genes was proposed

to explain the observed percent infection levels (75%) based on a gene-for-gene hypothesis. Parents A and E were postulated to each have an independent resistance gene, contributing to the observed resistance. Efforts to confirm this two-gene model with molecular markers were not successful. Progeny from the control pollinated cross of parent A by F were inoculated with an intermediate spore density using mixed gall inoculum. Intermediate rust infection was observed. However, the complexity of suspected multiple gene interaction effects and unknown virulence composition confounded marker investigations.

3.1 Introduction

After 40 years of phenotypic selection and breeding in loblolly pine (*Pinus taeda* L.) against the fusiform rust fungus (*Cronartium quercuum* f.sp. *fusiforme*), the genetic basis of inheritance of resistance is still not completely clear. Traditionally, polygenic inheritance is based on an apparently continuous disease distribution, where resistance is the aggregate of many genes with small effects. (Powers et al., 1981; Robinson, 1987; von Weissenberg, 1990; Zobel and Talbert, 1984). Alternatively, the noted continuous distribution of disease could be the result of a few major resistance genes in the host interacting with virulence genes in pathogen (Carson and Carson, 1989; Kinloch and Walkinshaw, 1991) in a gene-for-gene fashion (Flor, 1956).

In gene-for-gene models, success or failure of infection is the interaction of resistance genes in the host and avirulence genes in the pathogen. The "resistant" outcome is the result and only the result of the interaction of avirulence alleles and resistance alleles. When multiple loci are involved in the host-pathogen interaction, a host would be disease free (low infection type) if the host has a resistance allele at any locus for which the pathogen has an avirulence allele at the corresponding locus. When two or more pairs of complementary genes are present, the gene pair that imparts the lowest infection type (e.g. no gall in this study) is epistatic to all other gene pairs. The phenotype is always as low as, or lower than the more incompatible (disease free) of the individual interactions.

The gene-for-gene model has been established or assumed for a number of parasitically affected species (Crute et al., 1997). In 1984, the first pathogen avirulence gene was

cloned, and provided molecular evidence to support the gene-for-gene model (Staskawicz et al., 1984). The first gene-for-gene resistance gene to *Pseudomonas syringae* pv. Tomota has been cloned and sequenced (Martin et al., 1993). Four examples of R gene – Avr gene interactions were reviewed by Hammond and Jones (1996) discussing functions of resistance gene products. To date, over 40 R genes have been reported and cloned (Martin et al., 2003)

In pines, studies investigating the genetic control of interactions for the hosts and *Cqf* fungal inocula provided strong evidence for major gene inheritance of rust resistance, and significant families by inocula interactions have been reported (Kinloch and Walkinshaw, 1991; Nelson et al., 1993; Stelzer et al., 1997; Kuhlman, 1992; Amerson et al., 1997). A dominant resistance gene (Fr1) was first identified in loblolly pine by Wilcox et al. (1996), using specific single spore isolates of *Cqf*. Subsequently, 7 other host resistance genes were identified using multiple host (loblolly) families challenged with several different single spore inocula in a greenhouse (Jordan, 1997; Amerson et al., 1997; Amerson and co-workers, unpublished data). These interactions of defined rust-resistance genes in loblolly pine and single spore inocula sources strongly suggested a gene-for-gene relationship between resistance loci and avirulence loci in the loblolly pine: fusiform rust pathosystem. Similar interaction studies, without the use of molecular markers, also suggest gene for gene interactions for the slash pine: *Cqf* pathosystem (Nelson et al., 1993; Stelzer et al., 1997, 1999).

In pine:*Cqf* interactions, the commonly used approach to define resistance genes (factors) is to examine the disease phenotypes for progeny of host families interacting with various single spore inocula sources. In conifers, megagametophytes are haploid and can be used to assess genes inherited from the maternal parent as depicted by Wilcox et al. (1996). Amerson's work utilized a megagametophyte DNA approach that can discern only heterozygous genes in the maternal host parent. By artificial inoculation of a progeny population, with the tree of interest being the female parent, using different single spore fungal isolates, molecular markers associated with a resistance gene in the maternal parent can be found if the host has a heterozygous resistance gene and the strain is avirulent for that resistance gene.

Although considerable variation for *Cqf* pathogen virulence among rust collections has been recognized (Kuhlman and Matthews, 1993; Kuhlman and Sluder, 1995; Snow et al., 1975; Powers, 1980; Stelzer et al., 1997, 1999), along with geographic and local environmental variation, host individuals with two or more resistance genes should still be comparatively more difficult for the pathogen to overcome and more stable over time. Selecting pines with multiple major resistance genes is potentially important because of the long life cycle and wide plantation areas covered in pine improvement programs. The interaction effects of multi-complementary gene pairs in the pathosystem could be studied more efficiently in complex mating schemes such as diallel designs where one family is crossed with several other families. (Stelzer et al., 1997, 1999). In slash pine, Stelzer et al. (1997) postulated three pairs of complementary genes accounting for observed infection levels among host families in a five-parent diallel, challenged with

four single urediniospore *Caf* inocula. They also derived hypothetical genotypes for parents and two inocula and reported that low infection rates in some of crosses were likely due to epistatic effects. In the wild pathosystem, the plants / trees are exposed to various inocula and the resistant phenotypes could be due to the interaction effects of several complementary gene pairs. These interactions may be complex and difficult to dissect. However, even with pooled spore inoculum, it is in theory possible to find host resistance genes and study interactions between host resistance genes and avirulence genes in pathogen.

In the previous chapter, complex segregation analysis showed that: in a specific six-parent half diallel, there was a major rust resistance gene(s) segregating in the progeny population; most likely, the resistance in the offspring was inherited from parent A and parent F was assumed to have no contribution to the offspring in terms of resistance. Since parent A progeny have been noticed to have low rust infection rates in field tests, there have been some efforts in defining resistance genes in this tree. In the same diallel population used for this study, Wilcox (1995) searched for association between a resistance trait locus (loci) and marker loci inherited from parent A. He found two markers (A19_560, A11_400) that were significantly associated with disease phenotype (no gall vs. gall) but unlinked with each other. He suggested that either the markers flanked a single segregating resistance locus or that two unlinked resistance loci were segregating. In another study using megagametophyte DNA from OP progeny of parent A, Amerson and coworkers (unpublished data), mapped a heterozygous resistance locus termed (Fr2) in parent A. One RAPD marker and 5 AFLP markers were found to be

significantly associated with the Fr2 resistance locus that conferred resistance against single spore isolate (0-5-32). However, analysis of Wilcox's family A diallel samples (part of the same diallel used in the current study) by Amerson for an Fr2 linked marker (AK6_850), revealed that the marker and hence Fr2 were not significantly associated with field phenotype and would not account for the observed resistance in the field test used by Wilcox and the current work (Amerson, unpublished data). Given the existence of Fr2 and its lack of association with phenotype in the diallel and the likelihood of major genes recognized in chapter 2, parent A is believed to have multiple resistance genes (factors). This would make parent A a good candidate parent for a long-term rust resistance breeding program.

The purposes of this study were to search for resistance genes in a half-diallel mating design including parent A, to investigate the epistasis between resistance loci based on the gene for gene model, to improve and expand mapping data for Fr2 in parent A, and to search for more evidence to support or deny the hypothesis that the two markers associated with disease phenotype in parent A (Wilcox, 1995) denote independent resistance loci.

3.2 Materials and methods

3.2.1 Inocula preparation

Two mixed inocula with pooled spores and one single spore isolate were used in this study. One mixed basidiospore inoculum (L-7-2-85) originated from a 10 gall collection of aeciospores made in 1985 from field-infected trees. The other one (NCSU01) originated from an 11 gall collection of aeciospores obtained from field-infected trees in spring 2001. All galls were collected in Wilcox County, Alabama. The single aeciospore inoculum (0-5-32) originated from a single gall collection in Greene county, GA (Kuhlman, 1990) and was obtained courtesy of Dr. E.G. Kuhlman, USDA-FS (retired), Athens, GA.

3.2.2 Plant materials

Seeds from a half-diallel mating of six parents (obtained from the Tree Improvement Program at NC State University) and a controlled pollinated loblolly pine family (A by F, obtained from Dr. H. Amerson, NCSU Department of Forestry) were used in this study. A detailed description of the half-diallel mating design is in chapter I. Briefly, there are 15 crosses with no reciprocal crosses nor selfing in this diallel population. Seeds from only 12 crosses in this diallel were used in this study due to deficiency of seeds (Figure 3-1).

For stratification, seeds were immersed in water for 48 hours at 4°C, followed by removal of standing water and storage in plastic bags at 4°C for a minimum of 6 weeks.

Seeds from the 12 crosses in the diallel were germinated at the USDA-FS Resistance Screening Center (RSC) in Ashville NC, while the seeds from cross A × F were germinated at North Carolina State University to permit megagametophyte collection during germination. At both locations, seeds were placed in germination trays and covered by vermiculite. After 7- 14 days, germinated seedlings were transplanted to tubes when they were 2-3 centimeters high. Seedling potting mix was peat: vermiculite: perlite at either 6:2:1 or 3:4:3. For the cross A × F, before shedding of the seed coat, the megagametophyte was physically removed from the cotyledons of each germinating seedling, placed in a 1.5ml microfuge tube and frozen at –80 C until DNA extraction. Seedlings were inoculated with *Cqf* basidiospores 8 weeks after seeds were sown.

3.2.3 Inoculation and phenotypic assessment:

Northern red oak seedlings were inoculated with aeciospores from sources L-7-2-85 and NCSU01 (120mg/100ml), and urediospores of isolate 0-5-32 (20mg/50ml) at the RSC following screening center protocol (Knighten et al., 1988). Basidiospores were harvested from telia on infected oak leaves following RSC protocol and spore densities were adjusted at the RSC using a Coulter counter (Knighten et al., 1988) for inoculation to pines seedlings. Pine seedlings were inoculated at the RSC with basidiospores of *Cqf*. at different spore densities (Table 3-1) following the concentrated basidiospore spray method of Mattews and Rowan (1972). In summary, for every combination of inoculum and inoculation density, 200 seedlings from a specific full-sib pine family were challenged.

The pine seedlings were evaluated for presence or absence of galls on stems at four and half months for an early reading and at nine months for a final reading, after inoculation. Phenotypes were measured as no gall (NG) and gall (G). The percent gall was calculated as the ratio of the number of galled individuals to the sum of the number of galled individuals and the number of non-galled individuals. PROC FREQ in SAS was used to compare the proportions and analyze the inoculation results.

3.2.4 DNA extraction

Total genomic DNA was extracted from the megagametophytes and needles for different crosses using modified QIAGEN DNeasy® extraction protocols, similar to that reported by Knoth (2002). Frozen megagametophyte or needle tissues were grounded in 2ml screw cap centrifuge tubes, using a 1/4 inch ceramic ball and cylinder (Quibiogene®) with warm lyses AP1/RnaseA buffer (400ul for megagametophytes, 600ul for needles). Ground specimens were incubated in a 65°C water bath for 10 minutes. A 130ul aliquot of buffer AP2 (200ul for needle tissue) was added to each tube for protein precipitation. Protein was precipitated by putting the tubes in an ice-water bath for a minimum of 15 minutes. The samples were then centrifuged at 13,000 x g for 10 minutes to pellet the cellular debris and a 400 ul aliquot of clear lysate was transferred to a 96 well format block with 1.2ml collection tubes. 600ul of AP3 buffer with ethanol were added to each sample. The AP3/E sample mixture was transferred to a 96 well DNeasy® plate placed on a 2ml 96 well square block. The Dneasy® plate / square block assembly was centrifuged at 5635 x g (6000RPM) to bind DNA to the silica matrix of the DNeasy® plate. Contaminants were rinsed from the silica matrix by washing with 800ul of buffer

AW (Etoh already added) and centrifuging the “AirPore sealedTM” plates at 5635 x g (6000RPM) for 15 minutes. 50ul warm Buffer AE was added to each DNeasy® well and DNA was eluted to a clean 96 well 1.2ml block by spinning the DNeasy® plate / block assembly at 5635 x g (6000RPM) for 2 minutes. Another 50ul aliquot of warm buffer AE was added to each DNeasy® well, again followed by centrifugation of the DNeasy® plate/block assembly. In total, a 100ul volume of genomic DNA was obtained for each sample. The DNA concentration was estimated by comparison with known lambda standards in 0.8% agarose gels. Haploid DNA (from megagametophytes) was diluted to 1.0ng/ul and diploid DNA (from needles) was diluted to 1.5ng/ul for use in molecular marker studies.

3.2.5 RAPD reactions and marker analysis

Polymerase chain reactions (PCR) for Random Amplified Polymorphic DNA marker (RAPDs) used ten base-pair single stranded oligonucleotide sequence primers. Primers were typically obtained from Genosis biotechnologies, Inc (The Woodlands, TX), but followed by Operon (Alameda, CA) nomenclature. All PCR reactions were conducted in 96-well Falcon microtitre plates (Becton Dickinson and Co., Oxnard, CA). 15ul volumes were used for PCR reactions. Each reaction contained 15ug non-acetylated bovine serum albumin (New England Biolabs), 1.5ul 10X reaction buffer with MgCl₂ (Promega), 2.97mM dNTPS (Invitrogen), 0.75-1.0 unit Taq polymerase (Roche), 5-10 ng template DNA and 30ng 10 base primer. The mixture in each well of the plate was covered with 50ul mineral oil and the plates were centrifuged for approximately 15 seconds at 2500rpm. Amplifications were carried out in a model PTC 100, 96-well thermal cycler

(MJ research), using the thermal profile of Jordan (1997). After amplification, 2 ul of loading buffer (0.05% bromophenol blue in 40% sucrose with 20ng/ul of ethidium bromide) were added to each PCR reaction. Samples were loaded into 1.5% agarose gels containing 0.2ug/ml of ethidium bromide and electrophoresed for 4.5-5 hours at around 150V in 1X TBE buffer, using model A3 gel rigs (Owl scientific, Cambridge, MA). Polymorphisms were visualized under ultra violet light and were recorded onto thermal paper using an Eagle Eye TM video imaging system (Stratagene).

RAPD markers were scored as 1 for band present and 2 for band absent. Bands that could not be unambiguously scored were scored as 0 and treated as missing data. A marker was named according to primer identity (Operon nomenclature) and the approximate size of the band (estimated by comparison with a known size standard, 1Kb+).

3.2.6 Bulk segregant analysis and cosegregation analysis

Bulk Segregant Analysis (BSA) was used to screen the large numbers of primers to search for RAPD markers potentially associated with resistance genes. BSA is based on the comparison between two DNA bulks, each comprised of DNA from the individuals exhibiting the extreme phenotypes of a particular trait in a segregating population. (Giovannoni et al., 1991). The presence of polymorphism between the amplification patterns of the two bulks is expected only for those bands that are genetically linked to the trait of interest. In this study, replicated bulks were constructed containing pooled DNA from 16 individuals in each bulk based on the phenotypes of seedlings (gall vs no gall) in the segregating population. Two galled and two non-galled bulks represented the

replication. Bulk segregant analysis was also used to find other markers potentially linked to A19_560 or A11_400 in parent A. Replicated bulks (16 samples per bulk) were constructed based on these two RAPD markers. Progeny megagametophyte DNA of cross A by F were genotyped for both markers and then used to construct the bulks based on “+” phase of the marker vs “-“ phase of the marker. The candidate polymorphism markers found in the bulking studies were tested on the appropriate progeny population for cosegregation with the bulking target.

3.2.7 Map construction

Linkage maps including RAPD markers and/or AFLP markers (Amerson, unpublished data) were constructed using MAPMAKER Macintosh 2.0 Software (Lander et al., 1987) and MAPPOP 1.0 (Vision et al., 2000). Datasets containing the raw data and the recoded data were created and analyzed by MAPMAKER for preliminary map construction for the Fr2 locus and for linkage groups containing marker loci A19_560 or A11_400. For mapping the Fr2 locus, a framework map was derived from the preliminary map (obtained from Mapmaker) with the following methods. RIPPLE command was used to compare the likelihood of all adjacent triple permutations for the preliminary marker sequence. For each adjacent triple permutation, a marker sequence with an interval support >3.0 was achieved by dropping markers one by one. Then, the DROP function was used to identify markers that increased the group map size. Markers were dropped one at a time until the exclusion of any remaining markers did not shrink the group map size by more than 2 cM, except for terminal markers. After the framework map for Fr2 was constructed, two datasets were created in order to put the accessory markers and

phenotype onto the framework map using MAPPOP. One dataset contained the markers in the framework map; the other one contained all other markers in the preliminary map and phenotype. Both datasets were loaded to MAPPOP 1.0 and using MAPPOP, the accessory markers and phenotype were positioned relative to the framework map.

3.3 Results

3.3.1 Inoculation results and DNA marker analysis for the diallel progeny population

Four and half month results for inoculation of the diallel mating families are shown in table 3-2. The rust infection rates were high for every cross and were above 90% for most full-sib families for both mixed inocula. No significant differences were observed between two replications and between two inocula. Comparing the half-sib families, half-sibs of A showed approximately 85% galled for both inocula, while infection in all others was 90% or greater with some close to 100%. Notice that the cross E*A had consistent relatively low infection rates. On average, the half-sib infection rates of A and E were 84.5% and 91%, while the full sibs of this cross-showed galled percentages of 71%, 74%, 68% and 76% in different replications (Table 3-2). When two replications were pooled for each inoculum, the galled percentage for L-7-2-85 became 72.5%, while it was 72% for NCSU01. Nine-month readings for cross E×A continued to show the same pattern with slightly increase galled percentages (Table 3-3). Results from PROC FREQ suggested that there were no significant differences of these proportions from 75% (Table 3-4). Based on the observed 75% rust infection rates for the cross E by A and high

infection rates for other full-sib families, it was postulated that this diallel might have two independent resistance genes segregating in the progeny population.

In this diallel based on infection data (Tables 3-2 and 3-3), parents A and E are hypothesized to each have a resistance gene, respectively, at different loci. The pathogen population is a mixture of spores, hypothesized to include spores virulent to both individual R genes but not virulent for the two R-genes combined in a single individual. Two host parental genotypes could be derived as $R_1r_1r_2r_2$ and $r_1r_1R_2r_2$, where R is a dominant resistance allele and r is the recessive allele for lack of rust resistance. Pathogen diploid genotypes are speculated to contain $Avr_1avr_1Avr_2Avr_2$, $Avr_1Avr_1Avr_2avr_2$, or $Avr_1Avr_1Avr_2Avr_2$. Haploid basidiospore genotypes expectedly would be Avr_1avr_2 , Avr_2avr_1 , and Avr_1Avr_2 . When the inoculation density is high, individuals with a single R allele are overcome by the corresponding virulence allele and show no resistance to rust. Individuals without an R allele would similarly be galled. Only individuals with both R alleles would expectedly show no gall (rust-free). Assuming each parent has a heterozygous resistance gene and these two genes are unlinked or very weakly linked, a quarter of progeny will have both R alleles according to Mendelian segregation. So, 25% no gall in the progeny set is the expected result from the interaction of $R_1r_1R_2r_2$ progeny and Avr_1Avr_2 , Avr_1avr_2 and Avr_2avr_1 basidiospores. If avr_1avr_2 basidiospores were present at a significant frequency, all progeny expectedly would have been galled.

This hypothesis was investigated using marker / trait association analysis. By searching for markers polymorphic between no gall and gall groups based on BSA, it is possible to

identify candidate polymorphic markers that may be linked to resistance loci. Because the RAPD markers are typically dominant markers, specific marker configurations have to be observed in order to find R-genes for the proposed two-gene model using phenotype based BSA (Figure 3-2). The “-” phase of the marker is expected to be associated with the no gall bulk and the “+” phase of the marker is expected to be associated with the gall bulk. Then when the progeny population is screened, 2/3 of galled individuals are expected to have the “+” phase of the marker, the other 1/3 of galled individuals and all no gall individuals should have the “-” phase of the marker. Non-galled individuals, those with both postulated R genes should be “-“ for both markers 1 and 2 as depicted in figure 3-2b.

Based on the approach proposed above and in figure 3-2, replicated bulks using diploid DNA (gall vs no gall) were constructed and 420 RAPD primers were screened. Five candidate markers were observed to be polymorphic with the “-“ phase of the marker associated with the no gall phenotype in replicated bulks. However, none of these candidates cosegregated with phenotype (gall vs no gall) as projected in figure 3.2b when tested across a sample progeny population.

3.3.2 Linkage map construction for Fr2

The Fr2 locus was initially detected by Amerson (unpublished data) in parent A by inoculating progeny from its OP family with isolate 0-5-32. In that work, a RAPD marker (AK6_850) was found by Amerson to be significantly associated with phenotypic expression using diploid DNA. Subsequently, 5 AFLP markers and a RAPD marker

AK6_850 were placed around the Fr2 locus in a 64 sample progeny set challenged with isolate 0-5-32. Haploid DNA was used for preliminary mapping of Fr2 (Amerson, unpublished data). In order to better map the Fr2 linkage group, seedlings from a controlled pollinated family A * F were inoculated with single isolate 0-5-32 for the current study. Nine month readings showed 60.5% rust infection. Replicated haploid DNA Bulks (16 samples/bulk) based on phenotypes (gall vs no gall) were assembled and 420 RAPD primers were screened. Candidate polymorphic markers observed in both bulks were tested across 96 individual samples. Six of these candidate markers cosegregated with phenotypes (based on $LOD > 3$).

In order to integrate these new RAPD markers with previously found RAPD and AFLP markers (Amerson, unpublished data), these new RAPD markers were also tested on Amerson's 64 sample progeny set. RAPD marker genotypes of all 160 (96+64) sample progeny of A were used to construct a framework linkage map of the Fr2 resistance linkage group in parent A (Figure 3-3). AFLP markers (Amerson, unpublished data), RAPD markers not in the framework map, and phenotype (Fr2) were placed on the framework map using Mappop software. Mappop placed Fr2 in the interval between RAPD markers AY12_1300 and C13_580 with a probability of 1.0. All framework markers were significantly associated with Fr2 at a LOD level of 6.9 or greater.

3.3.3 Continued efforts to associate DNA markers with field resistance loci in parent A

Given the results from studies by Wilcox (1995) and Amerson (unpublished), the Fr2 locus apparently is not the only resistance locus in parent A. Wilcox (1995) found markers A19_560 and A11_400 in parent A to be significantly associated with phenotype in the field and Amerson's unpublished data showed that the Fr2 locus would not account for observed field resistance in family A in Wilcox's daillel samples. Continued efforts to find markers associated with field resistance loci in parent A were undertaken using the A19_560 and A11_400 markers of Wilcox (1995) as bulking targets. 240 RAPD primers were screened using BSA based on the 2 markers of Wilcox (1995). Six candidate markers were tested across 96 samples. Two of them (BG12_570 and BE16_1300) were significantly associated with marker A19_560 with LOD values of (26.19 and 18.06) (Figure 3-4). The other four (AZ18_530, BE_1200, BH19_550, AY20_1300) were significantly associated with marker A11_400 with LOD values of at least 5. Markers BH19_550 and AY20_1300 showed no recombination with A11_400. An additional marker U1_800 that was known to be associated with A11_400 (Wilcox, 1995) was also again found to be associated with A11_400. A preliminary map of the A11_400 linkage group in parent A is given in figure 3-5. The two linkage groups around A19_560 and A11_400 were independent and could not be mapped together based on the analysis using MAPMAKER.

3.3.4 Inoculation results and marker trait association study for the cross A by F

Seedlings of A by F were inoculated with the two mixed inocula used in the diallel study, but at inoculation densities much lower than that used in the diallel. Results from the lowest inoculation density (20,000sp/ml) showed a lower infection rate than that from the moderate inoculation density (50,000sp/ml) for both inocula (Table 3-5). At the same inoculation density, the infection rate for one inoculum (NCSU01) was consistently higher than the other one (L-7-2-85). The difference in rust infection between the two inocula suggests that virulence levels were likely to be different in two inocula. With the moderate infection level using inoculum L-7-2-85 at 50,000sp/ml, it is possible that a heterozygous resistant locus is segregating in the progeny population. Previous analysis and results suggested that the resistance was very likely inherited from the maternal parent A.

Haploid DNA (Megagametophyte DNA) was utilized to seek markers that may be associated with a presumptive maternally inherited resistance locus. Replicated Bulks based on phenotypes (gall vs no gall) were established and a total of 214 RAPD primers were screened. 10 candidate markers were found to be polymorphic in the replicated bulks (gall vs no gall). However, cosegregation analysis showed no association between the trait locus (phenotype) and the candidate markers when tested across a 96 progeny population.

3.4 Discussion

This study applied bulk segregant analysis and RAPD markers to search for rust resistance genes. With an intermediate infection rate of 60.5% in the progeny population of parent A \times F challenged with single spore isolate 0-5-32, 6 RAPD markers were found associated with the Fr2 locus in parent A using replicated haploid DNA (megagametophytes) bulks based on phenotype (gall vs. no gall). An improved map of Fr2 linkage group was constructed with RAPD markers found in this study and previously discovered markers (Amerson, unpublished data). Since most of the works by Amerson and co-workers mapping rust resistance loci have involved, and will continue to involve, RAPD markers and since full RAPD maps of some loblolly pine clones are available (Wilcox, 1995; Jordan, 1997), the additional RAPD markers included here may be useful for establishing regions of synteny between parent A and other clones. Such comparisons await investigation.

The BSA, co-segregation analysis approach successfully used to find markers for the Fr2 locus succeeded in large part because the inoculum was avirulent at the complementary locus and apparently virulent at all other loci for which parent A has resistance alleles. Using single spore inoculum offers the benefit of a simplified virulence composition but restriction of observations to a single spore inoculum may cause one to overlook other resistance loci which are not evident due to the virulence in the single inoculum. Such is the case with parent A.

The Fr2 locus found in this study and previously by Amerson using inoculum 0-5-32 is not the only resistance locus in parent A, as the marker AK6_850 linked with Fr2 locus was not significantly associated with phenotype (field resistance expression) observed in parent A progeny using materials from this diallel (Amerson, unpublished data). Also, statistical analyses in chapter II would support the presence of more than one major R-gene in parent A apart from Fr2. Efforts to marker identify another resistance locus in parent A using progeny of cross A×F inoculated with 50,000sp/ml of mixed gall isolate L-7-2-85 were not successful despite an intermediate (49%) infection rate at the 50,000 spore level. This 49% rate is suggestive of single gene regulation, but the high infection of cross F × A (95%) and half sibs of family A (85%) by inoculum L-7-2-85 at 250,000-300,000sp/ml indicated that there probably was some level of virulence in L-7-2-85 for any single gene present in parent A. The possibility of finding a single R gene in parent A at a reduced inoculum level depended on lowering the virulence level with the reduced spore numbers to the point that the inoculum would essentially perform as if avirulent at some complimentary locus where parent A had a resistance allele. This apparently did not occur and the 51% no gall response likely involves a number of as yet undefined genes.

A portion of this study investigated mixed spore inocula of the fusiform rust fungus interacting with progeny of 12 full-sib families from a six-parent half-diallel mating design. With the extremely high inoculation density (250,000 – 300,000 sp/ml), progeny of most crosses were above 90% galled, with the exception of cross E × A. Seventy-five percent of the progeny of cross E × A were galled. It was hypothesized that this level of infection was explainable by two independent, epistatic resistance genes in the two

parents and a spore population where individual spores could be virulent for one or neither of the proposed R-genes, but not both. Such a spore population should allow only the individuals with both R-alleles (in theory 25% of progeny) to show the resistance (no gall) phenotype. Bulk segregant analysis and co-segregation analysis were used to seek RAPD markers associated with the proposed resistance genes in the two parents, with a bulking strategy based on phenotype (gall vs. no gall). In the approach used to search for resistance genes in cross E × A, only polymorphisms with the “-“ phase of the marker linked with “resistance” (NG) bulks were considered to be useful due to the dominance of RAPD markers (see Figure 3-2 for details). This specific marker configuration requirement greatly reduced the potential number of useful polymorphisms that could be observed in the bulks. Also, a very small proportion of unexpected dual virulence (both hypothesized virulence alleles in the same spore) in the pooled inocula could potentially confound cosegregation of bulk candidate markers and phenotype in the progeny set. A total of 420 RAPD primers were screened for this two-gene BSA work. Five candidate markers with the appropriate marker configuration were observed based on BSA; however, no markers were found to co-segregate with phenotype. Theoretically, two complementary gene pairs and the proposed spore population could explain the observed rust infection in the half-diallel study. However, with the complexity of interactions between the host and pathogen complementary loci and the constraints of the BSA approach, no molecular marker evidence was found to support the two-gene pair interaction hypothesis.

Wilcox (1995) used BSA and co-segregation analysis with parent A progeny to find two RAPD markers that were significantly associated with field resistance (gall vs. no gall phenotype) in parent A. Since the two markers were not associated with each other, Wilcox was uncertain if the markers denoted two different R loci or if they broadly flanked a single locus. This current work used the markers found by Wilcox (A19_560, A11_400) as bulking targets and attempted to place additional markers around them. Several markers were added to the regions around A11_400 and A19_560; however, these markers did not pull the mapped regions together. The uncertainty faced by Wilcox still exists and this study could not determine if the resistance that he observed resided at one or two loci. Had the markers overlapped to form a single linkage group, one locus would have been a very likely answer.

One can not be sure why efforts with mixed inocula failed to yield marker definable R-genes, but the complexity of mixed inocula and the almost certain presence of two or more resistance genes in the host population is not ideal for marker investigations. Ideally, one would like to have inoculum with a single avirulence locus interacting with a single corresponding host resistance locus. The best way to achieve this is through pathogen management and the use of single spore inoculum. Multiple, different inocula each with different avirulence factors could be used to search a particular host for the presence of more than one R-gene. The proposed multiple resistance genes in parent A could be defined one by one by studying the interactions of multiple single spore inocula and progeny of parent A. Amerson and coworkers have used this approach in loblolly pine families to demonstrate two distinct R-genes in a given mother tree (Amerson et al.,

1997; Amerson unpublished data). Any further molecular studies undertaken with parent A should utilize this approach with multiple single spore isolates.

3.5 References

1. Amerson, H.V., A.P. Jordan, E.G. Kuhlman, D.M. O'Malley, and R.R. Sederoff. 1997. Genetic basis of fusiform rust disease resistance in loblolly pine. Proc. 24th South For. Tree. Imp. Conf. P. 403.
2. Carson, S.D., and M.J. Carson. 1989. Breeding for resistance in forest trees – a quantitative genetics approach. *Annu. Rev. Phytopathol.* 27: 373-395.
3. Crute, I. R., E. B. Holub, J. J. Burdon, Eds. 1997. The gene-for-gene relationship in plant-parasite interactions. CAB International, Wallingford, UK.
4. Flor, H.H. 1956. The complementary genetic systems in flax and flax rust. *Adv. Genet.* 8: 29-54.
5. Giovannoni, J.J., R.A. Wing, M.W. Ganai, and S.D. Tanksley. 1991. Isolation of molecular markers from specific chromosomal intervals using DNA pools from existing mapping populations. *Nucleic. Acids. Res.* 19: 6553–6558.
6. Hammond-Kosack, K.E., and J.D.G. Jones. 1996. Resistance gene dependent plant defense response. *Plant Cell.* 8: 1773-1791.
7. Jordan, A.P. 1997. Fusiform rust disease resistance and genomic mapping in loblolly pine. M.S. thesis. North Carolina State University. Raleigh, NC.
8. Kinloch, B.B. Jr. and C.H. Walkinshaw. 1991. Resistance to fusiform rust in Southern pines: how is it inherited? In: Proceedings of the IUFRO Rusts of Pine Working Party Conference, Banff, Alberta Canada. Northwest Region., North For. Cent Edmonton, Alberta. Inf. Rep. NOR-X-317, pp 219-228.
9. Knighten, J.L., C.H. Young, T.C. McCartney, and R.L. Anderson. 1988. Resistance screening center procedures manual: a step-by-step guide used in the

- operational screening of southern pines for resistance to fusiform rust. USDA forest service, Forest pest management, Asheville, NC. P.1-62.
10. Kuhlman, E.G. 1990. Frequency of single gall isolates of *Cronartium quercuum* f. sp. *fusiforme* with virulence toward three resistant loblolly pine families. *Phytopathology*. 80:614-617.
 11. Kuhlman, E.G. 1992. Interaction of virulent single-gall isolates of *Cronartium quercuum* f. sp. *fusiforme* and resistant families of loblolly pine. *For. Sci.* 38:641-651.
 12. Kuhlman, E.G. and E.R. Sluder. 1995. Interaction of single-aeciospore isolates of *Cronartium quercuum* f. sp. *fusiforme* with control pollinated loblolly pine families. In: proceedings of the 4th IUFRI Rusts of pines Working Party Conference. Tsukuba. Pp. 139-142.
 13. Kuhlman, E.G., and F.R. Matthews. 1993. Variation in virulence among single-aeciospore isolates from single-gall isolates of *Cronartium quercuum* f. sp. *fusiforme*. *Can. J. For. Res.* 23:67-71.
 14. Lander, E.S., P. Green, J. Abrahamson, A. Barlow, M.J. Daly, S.E. Lincoln, and L. Newburg. 1987. MAPMARKER: an interactive computer package for constructing primary genetic linkage maps of experimental and natural populations. *Genomics* 1: 174-181.
 15. Martin, G.B., A.J. Bogdanove, and G.Sessa. 2003. Understanding the functions of plant disease resistance proteins. *Annual Review of Plant Biology*. Jun. Vol. 54: 23-61.

16. Martin, G.B., S.H. Brommonschenkel, J. Chunwongse, A. Frary, and M.W. Ganal. 1993. Map-based cloning of a protein kinase gene conferring disease resistance in tomato. *Science*. 262:1432-36.
17. Matthews, F. W., and S.J. Rowan. 1972. *Plant Dis Rep.* 56: 931-934.
18. Nelson, C.D., R.L. Doudrick, W.L. Nance, J.M. Hamaker, and B. Capo. 1993. Specificity of host:pathogen genetic interaction for fusiform rust disease on slash pine. In: *Proceedings of the 22nd Southern Forest Tree Improvement Conference*. P. 403-410.
19. Newton, A.C., Andrivon, D., 1995. Assumptions and implications of current gene-for-gene hypotheses in existing mapping population. *Plant Pathol.* 44: 4 607-618.
20. Powers, H.R. 1980. Pathogenic variation among single-aeciospore isolates of *Cronartium quercuum* f. sp. *fusiforme*. *For. Sci.* 26: 280-282.
21. Powers, H.R., R.A. Schmidt, and G.A. Snow. 1981. Current status and management of fusiform rust on southern pines. *Annual Review. Phytopathology* 19: 353-371.
22. Robinson, R.A. 1987. *Host management in crop pathosystems* (1st ed.) New York: MacMillan Publishing Co.
23. Snow, G.A., R.J. Dinus, and A.G. Kais. 1975. Variation in pathogenicity of diverse sources of *Cronartium fusiforme* on selected slash pine families. *Phytopathology*. 65: 170-175.

24. Staskawicz BJ, D. Dahlbeck, and N.T. Keen. 1984. Cloned avirulence gene of *Pseudomonas syringae* pv. *glycinea* determines racespecific incompatibility of *Glycine max* (L) Merr. Proc Natl. Acad. Sci. USA 81: 6024-6028.
25. Stelzer, H.E., and R.L. Doudrick. 1997. Derivation for host and pathogen genotypes in the fusiform rust pathosystem on slash pine using a complementary genetics model and diallel data. The 24th southern forest tree improvement conference. 320-330.
26. Stelzer, H.E., and R.L. Doudrick. 1999. Prescreening slash pine and Cronatium pedigrees for evolution of complementary gene action in fusiform rust disease. Plant Dis. April. 385-388.
27. Vision, T.J., D.G. Brown, D.B. Shmoys, R.T. Durrett, and S.D. Tanksley. 2000. Selective mapping: a strategy for optimizing the construction of high-density linkage maps. Genetics. 155: 407-420.
28. Von Weissenberg, K. 1990. Host-parasite relationships in forest ecosystems: a review. Silva Fennica. 24 (91): 129-139.
29. W.E. van de Weg, 1997. A gene-for-gene model to explain interactions between cultivars of strawberry and races of *phytophthora fragariae var fragariae*. Theor. Appl. Genet. 94: 445-451.
30. Wilcox, P. 1995. Genetic Dissection of fusiform rust resistance in loblolly pine. Ph.D thesis. North Carolina State University. Raleigh, NC.
31. Wilcox, P.L., H.V. Amerson, E.G. Kuhlman, B.H. Liu, D.M. O'Malley, and R.R. Sederoff, R.R. 1996. Detection of a major gene for resistance to fusiform rust

disease in loblolly pine by genomic mapping. Proc. Natl. Acad. Sci. USA 93:
3859-3864.

32. Zobel, B., and J. Talbert. 1984. Applied Forest Tree Improvement. New York:
John Wiley and Sons.

Table 3-1. Overall greenhouse inoculation scheme in this study. L-7-2-85 and NCSU01 are two mixed basidiospore inocula originating from the mixed gall collections of aeciospores obtained from the field-infected trees. Basidiospore inoculum 0-5-32 originated from single aeciospore isolate 0-5-32. 200 progeny were inoculated for each treatment giving 2 replications of 100 each.

Crosses	Inocula	Inoculation Density (sp/ml)	seedlings
C*A, C*B, D*C, E*A, E*B, E*C, E*D, F*A, F*B, F*C, F*D, F*E	L-7-2-85,	250,000~300,000	200 for each full-sib families
	NCSU01	250,000~300,000	200 for each full-sib families
A*F	L-7-2-85	20,000	200
		50,000	200
	NCSU01	20,000	200
		50,000	200
0-5-32	100,000	200	

Table 3-2. Four and half months phenotypic data expressed as percent galled. (a) The diallel was inoculated with inoculum L-7-2-85 for replication 1; (b) The diallel was inoculated with inoculum L-7-2-85 for replication 2; (c) The diallel was inoculated with inoculum NCSU01 for replication 1; (d) The diallel was inoculated with inoculum NCSU01 for replication 2. Inoculum density was 250,000~300,000 basidiospores/ml for each inoculum.

(a) Replication 1 for L-7-2-85

Female/male	A	B	C	D	E	F
B						
C	0.86	0.97				
D			0.99			
E	0.71	0.98	0.97	0.96		
F	0.96	1.00	0.98	1.00	1.00	
Mean	0.84	0.98	0.95	0.98	0.92	0.99

(b) Replication 2 for L-7-2-85

Female/male	A	B	C	D	E	F
B						
C	0.84	0.99				
D			0.97			
E	0.74	0.99	0.96	0.95		
F	0.96	1.00	0.98	0.94	0.94	
Mean	0.85	0.99	0.95	0.95	0.92	0.96

Table 3-2 (Continued).

(c) Replication 1 for NCSU01

Female/male	A	B	C	D	E	F
B						
C	0.90	0.98				
D			0.95			
E	0.68	0.93	0.97	0.98		
F	0.99	1.00	0.98	1.00	0.99	
Mean	0.86	0.97	0.96	0.98	0.91	0.99

(d) Replication 2 for NCSU01

Female/male	A	B	C	D	E	F
B						
C	0.75	0.97				
D			0.88			
E	0.76	0.87	0.93	0.90		
F	0.97	1.00	1.00	1.00	1.00	
Mean	0.83	0.95	0.91	0.93	0.89	0.99

Table 3-3. Nine month phenotypic data expressed as percent galled for the specific cross E*A. (a) The diallel was inoculated with inoculum L-7-2-85; (b) The diallel was inoculated with inoculum NCSU01. Inoculum density was 250,000~300,000 basidiospores/ml for each inoculum.

	G	NG	Total	Percent galled *
Rep1	71	24	95	74.7%
Rep2	75	23	98	76.5%

(a)

	G	NG	Total	Percent galled *
Rep1	75	17	92	81.5%
Rep2	71	25	96	74%

(b)

*"G" denotes gall, "NG" denoted no gall.

*Percent galled was calculated as the number of G/(the number of G + the number of NG). Individuals with uncertain phenotype at 9 months were deleted, accounting for the sample sizes less than 100.

Table 3-4. Binomial proportion hypothesis test for H_0 : Proportion = 0.75. Two-sided p-values reported from PROC FREQ in SAS are listed in the table. P-values ranged from 0.32 to 0.83 and suggested that the tested gall percentages are not significantly different from 75%.

	4.5 months reading		9 months reading	
	Percent galled (%)	Two-sided P-value	Percent galled (%)	Two-sided P-value
L-7-2-85	72.5	0.4142	75.65	0.8354
NCSU01	72.0	0.3272	77.66	0.3997

*:Percent galled was calculated using all samples in two replications. If the phenotype was uncertain at 9 months, the sample was excluded.

Table 3-5. Six month phenotypic data expressed as percent galled for the cross A×F.

Inocula/density	20,000sp/ml	50,000sp/ml
L-7-2-85	50G, 148NG 50/198=25.3%	96G, 99NG 96/195=49.2%
NCSU01	85G, 104NG 85/199=42.7%	144G, 53NG 144/197=73.1%

*"G" denotes gall, "NG" denoted no gall.

*Percent galled was calculated as the number of G/(the number of G + the number of NG). Individuals with uncertain phenotype were excluded for the analysis accounting for sample sizes less than 200.

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

Figure 3-1. A six-parent half-diallel mating design. 15 crosses were generated in this diallel. Only 12 of them were used in this study because of seed deficiencies in the other three full-sib families.

(a) Bulk Segregant Analysis

	Parent1	Parent2	No gall progeny	Gall progeny	
Marker1	$\begin{array}{c} M_{1p1} \\ \\ r_1 \end{array}$ $\begin{array}{c} m_{1p1} \\ \\ R_1 \end{array}$	$\begin{array}{c} m_{1p2} \\ \\ r_1 \end{array}$ $\begin{array}{c} m_{1p2} \\ \\ R_1 \end{array}$	$\begin{array}{c} m_{1p1} \\ \\ R_1 \end{array}$ $\begin{array}{c} m_{1p2} \\ \\ r_1 \end{array}$	$\begin{array}{c} M_{1p1} \\ \\ r_1 \end{array}$ $\begin{array}{c} m_{1p2} \\ \\ r_1 \end{array}$	$\begin{array}{c} m_{1p1} \\ \\ R_1 \end{array}$ $\begin{array}{c} m_{1p2} \\ \\ r_1 \end{array}$
Marker Phase	+	-	-	+	-
Marker2	$\begin{array}{c} m_{2p1} \\ \\ r_2 \end{array}$ $\begin{array}{c} m_{2p1} \\ \\ r_2 \end{array}$	$\begin{array}{c} M_{2p2} \\ \\ r_2 \end{array}$ $\begin{array}{c} m_{2p2} \\ \\ R_2 \end{array}$	$\begin{array}{c} m_{2p1} \\ \\ r_2 \end{array}$ $\begin{array}{c} m_{2p2} \\ \\ R_2 \end{array}$	$\begin{array}{c} m_{2p1} \\ \\ r_2 \end{array}$ $\begin{array}{c} M_{2p2} \\ \\ r_2 \end{array}$	$\begin{array}{c} m_{2p1} \\ \\ r_2 \end{array}$ $\begin{array}{c} m_{2p2} \\ \\ R_2 \end{array}$
Marker Phase	-	+	-	+	-

* M represents the “+” phase of a given marker while m is the recessive “-“ phase.

Polymorphic markers in (a) with the “-“ marker allele associated with no gall become candidate markers.

(b) Marker configuration in the progeny set. Individuals that are “-“ for both markers 1+2 should possess both R alleles and be non-galled.

	No gall	Gall		
	1,2,324 (R ₁ r ₁ R ₂ r ₂)	25, 2648 (R ₁ r ₁ r ₂ r ₂)	49, 5072 (r ₁ r ₁ R ₂ r ₂)	73, 7496 (r ₁ r ₁ r ₂ r ₂)
Marker1 phase	-----	-----	+++	+++
Marker2 phase	-----	+++	-----	+++

Figure 3-2. An illustration for the approach to identify resistance genes interacting in the proposed two-gene model using dominant molecular markers, (a) bulk segregant analysis, and (b) assessment of candidate markers in a sample progeny set.

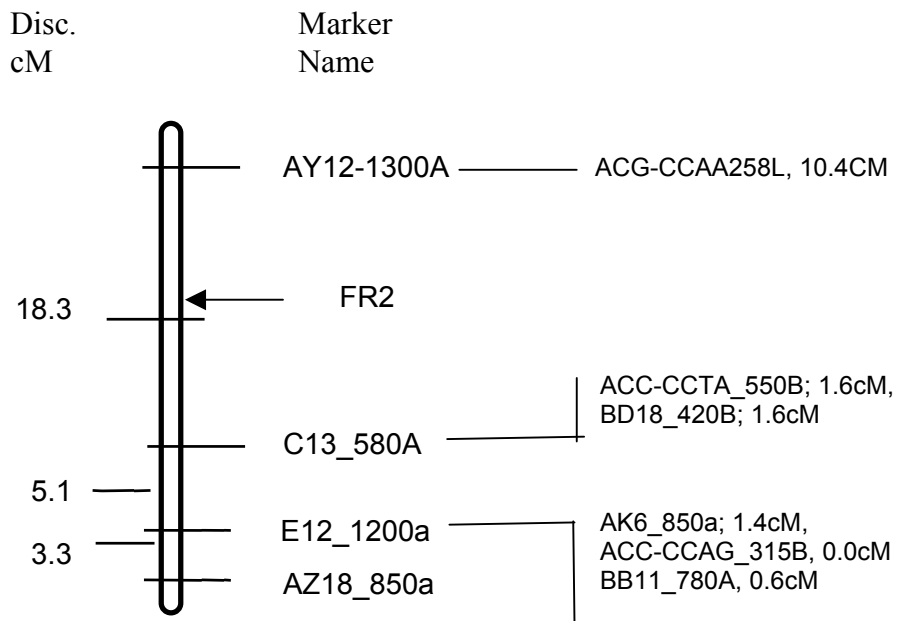


Figure 3-3. A linkage map of the Fr2 linkage group in parent A. Fr2 and accessory markers are placed around the markers in the framework map.

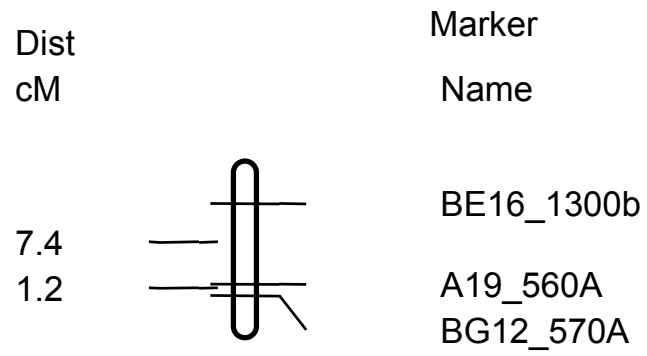


Figure 3-4. A preliminary map for the A19_560 linkage group in parent A

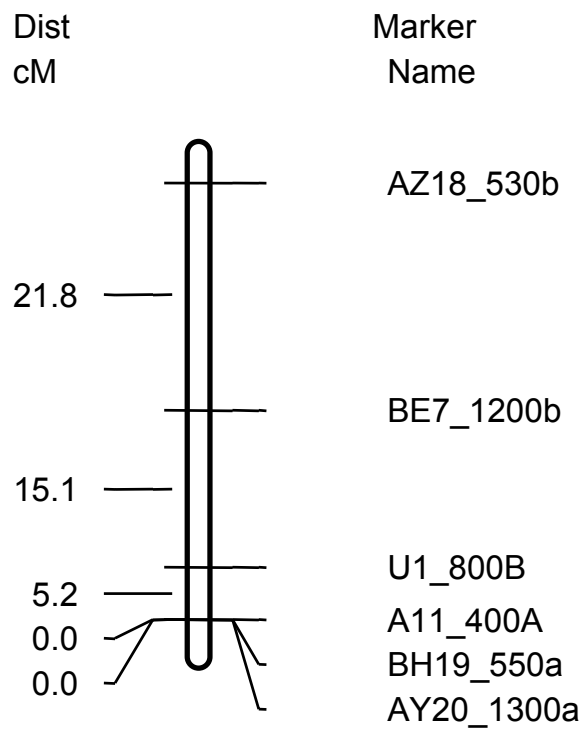


Figure 3-5. A preliminary map for the A11_400 linkage group in parent A.