

Methodology article

## A hierarchical statistical model for estimating population properties of quantitative genes

Samuel S Wu<sup>1</sup>, Chang-Xing Ma<sup>1,2</sup>, Rongling Wu\*<sup>1</sup> and George Casella<sup>1</sup>

Address: <sup>1</sup>Department of Statistics, University of Florida, Gainesville, FL 32611, USA and <sup>2</sup>Department of Statistics, Nankai University, Tianjin 300071, P. R. China

E-mail: Samuel S Wu - samwu@stat.ufl.edu; Chang-Xing Ma - cma@biostat.ufl.edu; Rongling Wu\* - rwu@stat.ufl.edu; George Casella - casella@stat.ufl.edu

\*Corresponding author

Published: 12 June 2002

Received: 15 March 2002

*BMC Genetics* 2002, **3**:10

Accepted: 12 June 2002

This article is available from: <http://www.biomedcentral.com/1471-2156/3/10>

© 2002 Wu et al; licensee BioMed Central Ltd. Verbatim copying and redistribution of this article are permitted in any medium for any purpose, provided this notice is preserved along with the article's original URL.

### Abstract

**Background:** Earlier methods for detecting major genes responsible for a quantitative trait rely critically upon a well-structured pedigree in which the segregation pattern of genes exactly follow Mendelian inheritance laws. However, for many outcrossing species, such pedigrees are not available and genes also display population properties.

**Results:** In this paper, a hierarchical statistical model is proposed to monitor the existence of a major gene based on its segregation and transmission across two successive generations. The model is implemented with an EM algorithm to provide maximum likelihood estimates for genetic parameters of the major locus. This new method is successfully applied to identify an additive gene having a large effect on stem height growth of aspen trees. The estimates of population genetic parameters for this major gene can be generalized to the original breeding population from which the parents were sampled. A simulation study is presented to evaluate finite sample properties of the model.

**Conclusions:** A hierarchical model was derived for detecting major genes affecting a quantitative trait based on progeny tests of outcrossing species. The new model takes into account the population genetic properties of genes and is expected to enhance the accuracy, precision and power of gene detection.

### Background

The identification of individual genes governing phenotypic variation is crucial to understand the mechanistic basis of quantitative inheritance and ultimately provide information about the design of optimal breeding strategies in both plants and animals. Quantitative genetics based on new statistical and computational technologies has advanced to the point at which individual genes can be detected and mapped on chromosomes [1,2]. The un-

derpinning for the detection and mapping of genes is founded on their particular segregation pattern. For a pedigree derived from inbred lines, the segregation pattern of a gene can be exactly predicted. In this situation, quantitative genetic theory can lay a sufficient foundation for gene detection and, thus, the properties of a detected gene can be adequately described by its effect and genomic location. However, for outcrossing populations, such as forest trees, in which inbred lines are not available, genes also

display population genetic properties [3]. Hence, the foundation for gene detection in these populations should be established on a combination of quantitative genetics and population genetics. The estimation of the population genetic properties of genes, such as allele frequencies and the degree of disequilibrium, can not only enhance the precision and power of gene detection, but also is fundamentally important to our understanding of population differentiation and evolution [4,5].

We present a strategy for investigating the existence of genes in an arbitrarily complex progeny test based on phenotype data and estimating the effects of these genes on a quantitative trait and their population properties under the maximum likelihood framework. Only a limited number of statistical studies have been performed to detect genes based on the phenotypic data [6–13], and none of these studies have considered the population and evolutionary genetic properties of genes. Our study population can be derived from a natural population instead of inbred lines as used for general agronomic crops. A finite number of unrelated parents are randomly selected from a natural population and are mated to produce progeny tests under a mating design such as factorial or diallel. In a recent paper, we used such progeny tests to derive a model for detecting major genes affecting a quantitative trait [14]. This model has power to estimate the effect of drift errors on genetic variation during hybridization, thus providing information about dynamic changes of the genetic architecture of a population under artificial hybridization. However, the estimates of population genetic parameters from the progeny test using the above model cannot be generalized to the original population from which the mating design is derived. In this paper, we developed a hierarchical model for analyzing the pattern of gene segregation at both population and family levels, which thus can provide estimates of genetic parameters in two successive generations of the population. Estimates of population genetic parameters using this hierarchical model can reasonably infer the genetic structure of the original population. A forest tree example is used to demonstrate the application of the new statistical model for gene detection.

**Model**

**Population genetic structure**

Suppose there is a segregating major gene responsible for a quantitative trait in a natural or experimental diploid population. This gene has  $s$  different codominant alleles,  $A_1, \dots, A_s$ , whose frequencies in the population are denoted by  $p_1, \dots, p_s$  with  $\sum_{\kappa=1}^s p_{\kappa} = 1$ . These  $s$  alleles randomly unite to form  $s(s + 1)/2$  distinguishable genotypes including  $s$  homozygotes and  $s(s - 1)/2$  heterozygotes. The population frequencies of the genotypes are denoted by

$p_{\kappa\iota}$  ( $\kappa, \iota = 1, \dots, s$ ). If the population is at Hardy-Weinberg disequilibrium (HWD), the genotypic frequency is the product of the two corresponding allelic frequencies, plus the coefficient of HWD ( $d_{\kappa\iota}$ ), i.e.,  $p_{\kappa\iota} = p_{\kappa}p_{\iota} + d_{\kappa\iota}$ , where  $d_{\kappa\iota}$  has the restrictions,  $\max\{-p_{\kappa}(1 - p_{\iota}), -p_{\iota}(1 - p_{\kappa})\} \leq d_{\kappa\iota} \leq p_{\kappa}p_{\iota}$  and  $\sum_{\nu \neq \kappa} d_{\kappa\nu} - \sum_{\nu \neq \kappa} d_{\nu\nu} \geq -p_{\kappa}^2$  [15].

**Mating design**

A finite number of unrelated individuals are randomly selected from the population as female and male parents to generate a  $I \times J$  factorial mating design. Based on sampling theory, these selected individuals have genotypes at the major locus whose counts follow the same multinomial distribution as those in the original population, with probability vector  $\mathbf{p} = (p_{\kappa\iota}, \kappa, \iota = 1, \dots, n)$ . The segregation of progeny genotypes in the mating design depends on how the major-locus genotypes segregate (1) among the selected female or male parents and (2) within each of the full-sib families (Table 1). The segregation among the  $I$  female or  $J$  male parents having the same pattern as in the original population is viewed as the population-level segregation. The segregation within a full-sib family of size  $n_{ij}$  derived from the  $i$ th female and  $j$ th male parent follows the Mendelian segregation ratio, viewed as the family-level segregation. To understand the genetic structure of the original population, different genetic information derived from this two-level segregation will not be mixed, but rather combined within a maximum likelihood framework.

**Estimation method**

Let  $y_{ijk}$  denote the phenotypic measurement for the  $k$ th offspring from  $i$ th female parent and  $j$ th male parent and  $g_{ijk}$  denote its genotype at the major gene,  $i = 1, 2, \dots, I, j = 1, 2, \dots, J, k = 1, 2, \dots, n_{ij}$ . The statistical model describing the phenotype-genotype relationship can be expressed as

$$y_{ijk} = \mu_{\kappa\iota} I_{\{g_{ijk}=A_{\kappa}A_{\iota}\}} + e_{ijk}$$

where  $\mu_{\kappa\iota}$  is the genotypic value of genotype  $A_{\kappa}A_{\iota}$ ; and  $e_{ijk}$  is the error term with the normal distribution  $N(0, \sigma^2)$ .

Let  $g_{f_i}$  be the genotype for the  $i$ th female parent and  $g_{m_j}$  for the  $j$ th male parent. We denote by  $\mathbf{y}$  the collection of the observed data and by  $\mathbf{z} = (\mathbf{g}, \mathbf{g}_f, \mathbf{g}_m)$  the collection of the major-locus genotypes for offspring, female parents and male parents. We assume the following hierarchical model:

**Table 1: Factorial mating design using female and male parents sampled from a natural population.**

			Male		
		$A_1A_1$	$A_1A_2$	—	$A_3A_5$
		$(p_1^2)$	$(2p_1p_2)$	—	$(p_3^2)$
		$A_1A_1$	$A_1A_1 : A_1A_2$	—	$A_1A_5$
		$(p_1^2)$	$(1/2 : 1/2)$	—	$(1)$
		$A_1A_2$	$A_1A_1 : A_1A_2 : A_2A_2$	—	$A_1A_5 : A_2A_5$
		$(2p_1p_2)$	$(1/4 : 1/2 : 1/4)$	—	$(1/2 : 1/2)$
Female					
		$A_kA_l$	$A_kA_1 : A_lA_1$	—	$A_kA_5 : A_lA_5$
		$(2p_kp_l)$	$(1/2 : 1/2)$	—	$(1/2 : 1/2)$
	$A_5A_5$	$A_1A_5$	$A_1A_5 : A_2A_5$	—	$A_5A_5$
	$(p_5^2)$	$(1)$	$(1/2 : 1/2)$	—	$(1)$

The segregation patterns of genotypes are indicated at the parental and progeny generation levels. Genotype frequencies in each generation are given in parentheses.

$$\begin{aligned}
 y \mid \mathbf{g} &\sim \prod_{ijk} f(y_{ijk} \mid \mathbf{g}_{ijk}), \\
 \mathbf{g} \mid (\mathbf{g}_f, \mathbf{g}_m) &\sim \prod_{ijk} h(\mathbf{g}_{ijk} \mid (\mathbf{g}_{f_i}, \mathbf{g}_{m_j})), \\
 (\mathbf{g}_f, \mathbf{g}_m) &\sim \prod_i \psi(\mathbf{g}_{f_i}) \prod_j \psi(\mathbf{g}_{m_j}), \quad (1)
 \end{aligned}$$

where  $f$  denotes the normal distribution of the phenotype  $N(\mu_{kl}, \sigma^2)$ ,  $h$  is distributed according to the Mendelian segregation pattern and  $\psi$  is multinomially distributed with probability vector  $p$ . Note that genotypes for siblings are dependent and, given genotypes, the traits are independent of each other. Under the above model  $(y, z)$  has a joint distribution as follows:

$$f(y, z \mid \theta) = \prod_{ijk} \left[ f(y_{ijk} \mid \mathbf{g}_{ijk}) h(\mathbf{g}_{ijk} \mid (\mathbf{g}_{f_i}, \mathbf{g}_{m_j})) \right] \prod_i \psi(\mathbf{g}_{f_i}) \prod_j \psi(\mathbf{g}_{m_j}) \quad (2)$$

Statistically, equation (2) is a mixture model in which each component is represented by a genotype within a full-sib family [16–18]. The maximum likelihood estimates for  $\theta = (\mu, \sigma^2, p)$  can be obtained using the EM algorithm [19,20]. At the  $t^{th}$  step of the EM sequence, the expected log-likelihood

$$Q(\theta \mid \hat{\theta}^{(t)}, \mathbf{y}) = E_{z \mid \hat{\theta}^{(t)}, \mathbf{y}} \log f(\mathbf{y}, z \mid \theta) \text{ is proportional to}$$

$$-\frac{1}{2} \sum_{ijk} \left[ \sum_G \frac{(y_{ijk} - \mu(G))^2}{\sigma^2} \hat{p}_{ijkG} + \log(\sigma^2) \right] + \sum_G \log [p(G)] \left[ \sum_i \hat{p}_{iG} + \sum_j \hat{p}_{jG} \right] \quad (3)$$

where  $\sum_G$  sums over all possible genotypes,  $\hat{p}_{ijkG} = Pr(\mathbf{g}_{ijk} = G \mid \hat{\theta}^{(t)}, \mathbf{y})$  are the posterior distributions of offspring genotypes and  $\hat{p}_{iG}, \hat{p}_{jG}$  are the posterior distributions of parent genotypes. Conditional on  $(\hat{\theta}^{(t)}, \mathbf{y})$ , these posterior probabilities can be evaluated as follows:

$$\begin{aligned}
 p(\mathbf{g}_m, \mathbf{g}_f) &\propto \prod_{ijk} \left[ \sum_G f(y_{ijk} \mid G, \hat{\theta}^{(t)}) h(G \mid (\mathbf{g}_{f_i}, \mathbf{g}_{m_j})) \right] \prod_i \psi(\mathbf{g}_{f_i} \mid \hat{\theta}^{(t)}) \prod_j \psi(\mathbf{g}_{m_j} \mid \hat{\theta}^{(t)}) \\
 \hat{p}_{ijkG} &\propto \sum_{\mathbf{g}_f, \mathbf{g}_m} \frac{f(y_{ijk} \mid G, \hat{\theta}^{(t)}) h(G \mid (\mathbf{g}_{f_i}, \mathbf{g}_{m_j}))}{\sum_G f(y_{ijk} \mid G, \hat{\theta}^{(t)}) h(G \mid (\mathbf{g}_{f_i}, \mathbf{g}_{m_j}))} P_r(\mathbf{g}_{f_i}, \mathbf{g}_{m_j} \mid \hat{\theta}^{(t)}, \mathbf{y}) \quad (4)
 \end{aligned}$$

The EM sequence is defined by

$$\begin{aligned} \hat{\mu}(G)^{(t+1)} &= \sum_{ijk} (\hat{p}_{ijkG} y_{ijk}) / \sum_{ijk} \hat{p}_{ijkG}, \\ (\hat{\sigma}^2)^{(t+1)} &= \sum_{ijkG} (\hat{p}_{ijkG} [y_{ijk} - \hat{\mu}(G)]^2) / \sum_{ijkG} \hat{p}_{ijkG}, \\ \hat{p}(G)^{(t+1)} &= \sum_i \hat{p}_{iG} + \sum_j \hat{p}_{jG}. \end{aligned} \quad (5)$$

Suppose the major gene has two different codominant alleles, the posterior probabilities in equation (4) can be exactly calculated for a usual mating design. However, for large designs with many parents or for the case where the major gene has many different alleles, we have to rely on a Monte Carlo version of the EM algorithm. For example, we can sample from the conditional distribution of  $z =$

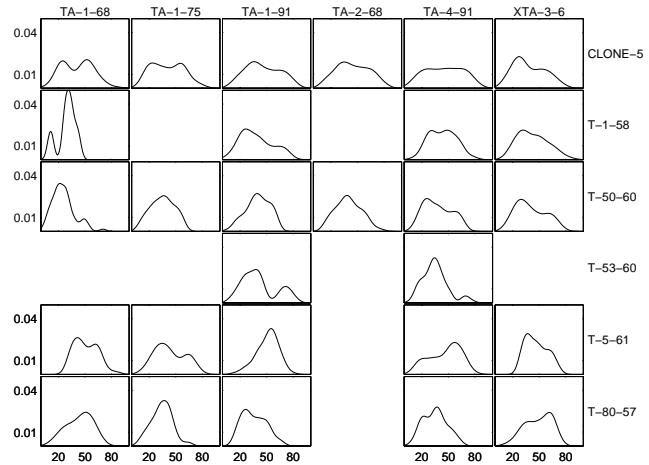
$(g, g_f, g_m)$  given  $(\hat{\theta}^{(t)}, y)$  using the following Gibbs sampler [20]:

$$\begin{aligned} g_m | (g, g_f, \hat{\theta}^{(t)}, y) &\propto \prod_{ijk} \left[ f(y_{ijk} | g_{ijk}, \hat{\theta}^{(t)}) h(g_{ijk} | (g_f, g_m)) \right], \\ g_m | (g, g_f, \hat{\theta}^{(t)}, y) &\propto \prod_i \left[ \psi(g_f | \hat{\theta}^{(t)}) \prod_{jk} h(g_{ijk} | (g_f, g_m)) \right], \\ g_f | (g, g_m, \hat{\theta}^{(t)}, y) &\propto \prod_j \left[ \psi(g_m | \hat{\theta}^{(t)}) \prod_{ik} h(g_{ijk} | (g_f, g_m)) \right]. \end{aligned}$$

**Example**

We use data from aspen trees to illustrate the application of our statistical model for detecting a major gene. The example is derived from a  $6 \times 6$  factorial mating design of *Populus tremuloides* and *P. tremula*. The parents for the crosses were randomly selected from a mixed breeding population of two different species established at the University of Minnesota. Originally, the *P. tremuloides* trees used as the female parents from the Upper Peninsula of Michigan and northern Wisconsin, and the *P. tremula* male parents came from northeastern Poland and Germany [21]. The seedlings from the progeny population were planted in two different locations, one on a cutover forest site near Grand Rapids, Minnesota, and the other on farmland near Rhinelander, Wisconsin. Both trials were laid out in a randomized complete block design with 10 replicates and six seedlings per family in each replicate at  $2.5 \times 2.5$  spacing. At the end of the second year in the field, all trees were measured for stem height.

Figure 1 presents density estimates of the second year height from each family pooled over the two locations. Trees from eight families did not have growth data due to mortality. The phenotypic distributions suggest the existence of a major gene in the progeny test. For example, in



**Figure 1**  
Density estimates of second year stem height (in cm) for all families. The measurement is the stem height at the end of second year. The randomly selected parents are labelled at the top (male) and on the right of plots (female). Trees from eight families did not have growth data due to mortality.

the family derived from Clone-5  $\times$  TA-1-68, the phenotypic distribution is a mixture of at least two components, showing the existence of a segregating gene. Figure 1 also indicates different segregation patterns among the families, with some more smooth and others more waved, thus suggesting that some parents are heterozygous with their alleles segregating in the progeny.

The effect and genotypic frequencies of the major gene are estimated using the hierarchical model assuming two alleles. Since no significant difference is detected between two locations, we only report the result based on pooled data. With probability of almost one, parents TA-1-91, TA-2-68, TA-4-91, XTA-3-6 have genotype  $A_1A_1$ , parents TA-1-68, TA-1-75, CLONE-5, TA-5-61 have genotype  $A_1A_2$  and the rest parents have genotype  $A_2A_2$ . The estimates of the genotypic values of  $A_1A_1$ ,  $A_1A_2$  and  $A_2A_2$  are  $\hat{\mu}_{11} = 56.7 \pm$

$0.80$ ,  $\hat{\mu}_{12} = 40.6 \pm 0.32$  and  $\hat{\mu}_{22} = 30.0 \pm 1.02$ , respectively.

The estimate of residual variance is  $\hat{\sigma}^2 = 218.676 \pm 6.09$ . Here, the estimates of standard errors for the parameter estimators are obtained based on the Fisher information matrix [22]. We also fitted the model under the null hypothesis that there is no major gene with a single normal distribution.

The likelihood ratio test has  $\chi^2_{df=4} = 150$  ( $p < 0.0001$ ), suggesting the existence of a major gene for height growth. The ratio of the dominant to additive effect for this major gene is less than 0.2, which implies that it affects the stem height growth in an additive manner.

The population genetic parameters of this major gene are estimated based on the genotypes of the parents detected. The genotype frequencies estimated are identical for the three genotypes, i.e.  $\hat{p}_{11} = \hat{p}_{12} = \hat{p}_{22} = 1/3$ . In consequence, the allele frequencies are also same ( $\hat{p}_1 = \hat{p}_2 = 1/2$ ). Thus, the mixed breeding population of *P. tremulodes* and *P. tremula* is not at Hardy-Weinberg equilibrium, because the genotype frequencies are not products of the corresponding allele frequencies. The coefficient of Hardy-Weinberg disequilibrium is estimated as  $\hat{d} = 0.06$ . The estimate of allele frequencies and Hardy-Weinberg disequilibrium using our hierarchical model can be well generalized to the original breeding population from which the parents were sampled. However, one should be cautious about the accuracy of the estimator due to the small sample size.

### Simulation

The behavior of the EM estimators is evaluated by two simulation experiments. In the first simulation, a 6 by 6 mating design same as our real example is assumed. The major gene is assumed to have two codominant alleles  $A_1$  and  $A_2$  whose frequencies in the population are  $p_1 = 0.6$  and  $p_2 = 0.4$ . The genotypic values are set to be 65, 40 and 25 with standard deviation 15. Based on the hierarchical model defined in equation (1), we first simulated a set of the parental genotypes ( $A_1A_1, A_1A_1, A_1A_2, A_1A_1, A_1A_2, A_1A_1$  for females and  $A_1A_2, A_2A_2, A_2A_2, A_1A_1, A_1A_2, A_1A_2$  for males). Then we randomly generated 50 offspring genotypes and observations for each family.

Our EM estimating procedure converged to the correct parental genotypes after only a few iterations (normally 2 or 3). A loglikelihood test is highly significant with log-likelihood ratio equals 125 (df = 4). The estimates of genotypic values from a single simulation replicate are  $64.1 \pm 0.74$  for  $A_1A_1$ ,  $39.5 \pm 0.59$  for  $A_1A_2$  and  $26.8 \pm 1.64$  for  $A_2A_2$ . The estimate of residual variance is  $241.6 \pm 10.30$ .

In our second simulation we studied a 5 by 5 mating design. We used 25 replications to investigate the empirical accuracy of the estimators. We assume the major gene has 2 alleles and randomly generated parental genotypes. The parental genotypes were randomly generated based on allele frequency  $p_1 = 0.5$ . That resulted in female genotypes ( $A_1A_1, A_1A_2, A_2A_2, A_1A_2, A_2A_2$ ) and male genotypes ( $A_1A_2, A_1A_1, A_2A_2, A_1A_1, A_1A_2$ ). For each family we simulated the 50 genotypes for its offsprings and then simulated its trait measurement using a normal and log-normal distribution.

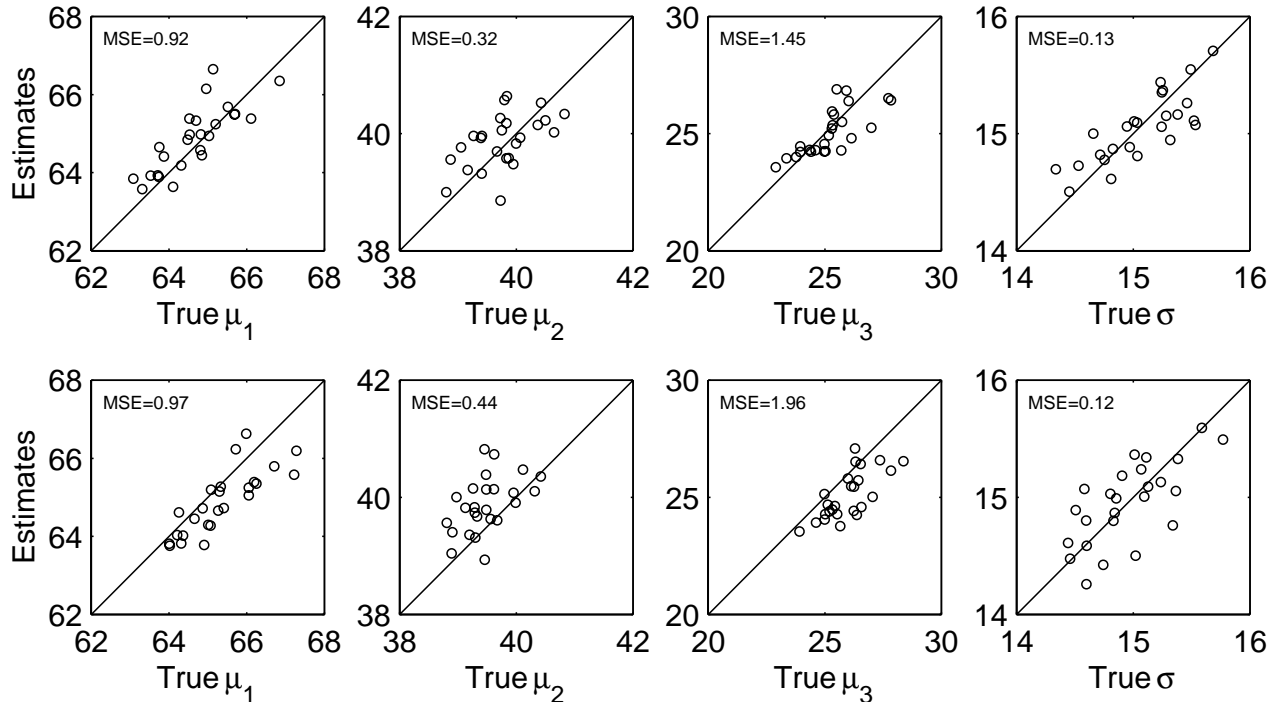
The results from the second simulation study are summarized in Figure 2. The top panel of the plots is for the simulation where trait measurements are simulated from a normal distribution. The bottom panel corresponds to lognormal data. In each plot we graphed the ideal estimates of the parameter (assuming that we know the genotypes of each offspring) versus our estimates based on the EM algorithm. Mean squared errors are also indicated in the plots. It is seen from Figure 2 that all four parameters can be estimated quite accurately in the normal case while they are slightly biased for the lognormal data.

### Discussion

We have derived a hierarchical model for detecting major genes affecting a quantitative trait based on progeny tests of outcrossing species. The new model takes into account the population genetic properties of genes and is expected to enhance the accuracy, precision and power of gene detection. The information about the population behavior of genes estimated from this model is fundamentally important to our understanding of the genetic architecture of a natural or experimental population and is also useful for the design of sound breeding strategies for agricultural crops and forest trees.

We used an example from interspecific aspen hybrids to illustrate the application of our gene-detecting model. The model has successfully identified a major gene with additive effects on second year stem growth in the hybrid aspens. The additive effect of genes on height growth in young poplars was also observed in a molecular marker experiment [23]. In this marker experiment, an additive quantitative trait locus was found to explain almost a quarter of the total phenotypic variation in second year height growth. The result from the aspen hybrids was consistent with that from a simulation experiment based on a mating scheme analogous to that used in our real example. The consistency of the results from the simulation study suggests that the model proposed here can be adequately used to detect genes of large effect on the phenotypes.

The model will have immediate applications for many species in which progeny tests have been established for a number of years [24]. Most of these tests are maintained in different locations and measured annually, thus offering desirable opportunities to address two important questions: (1) How does the genes interact with environment? (2) What is the developmental plasticity of gene expression? In our real example, no evidence has been obtained for the change of gene expression at the major locus detected over the two field trials, despite their dramatic differences in climate, soil properties and silvicultural measures [21]. If the stability of the expression of this gene over a range of environments is confirmed by



**Figure 2**

Summary of a simulation study based on a  $5 \times 5$  mating design. In each plot we graphed the ideal estimates of the parameter (assuming that we know the genotypes of each offspring) versus the estimates based on the EM algorithm. The top panel corresponds to normal data and the bottom panel to the lognormal case

more accurate genotype determination methods, e.g. based on genetic markers. It will have a tremendous application in breeding for stable cultivars of forest tree species.

Our analysis and simulation is based on diallelic inheritance. But multiallelic inheritance can be similarly considered, although more parameters should be introduced. Also, our model assumes the segregation of a single gene in progeny tests. The principle behind the model can be extended to consider two or more genes. Modelling multiple genes may be closer to biological reality, because the linkage, epistatic interaction and genetic association among genes are considered. However, the consideration of these relationships needs to estimate an increased number of unknown genetic parameters. Finally, our model is based on a balanced factorial design. For some populations, like mammals, mating designs are frequently unbalanced because of limited resources or reproductive difficulties. It is very important to extend our analysis and model to an unbalanced mating design of any complexity. For all of these extensions mentioned above, maximum likelihood approaches may not be sufficient owing to an increasing dimension of parameter space. A more powerful computational tool, e.g., Markov chain Monte

Carol methods, may be needed [20] to make our model more tractable in real applications.

### Acknowledgements

We thank Dr. Jim Hobert, Dr. Dudley Huber, Dr. Bailian Li and Dr. Tim White for helpful discussions and reviews regarding this study. This manuscript was approved as Journal Series No. R-08032 by the Florida Agricultural Experiment Station.

### References

1. Hoeschele I, Uimari P, Grignola FE, Zhang Q, Gage KM: **Advances in statistical methods to map quantitative trait loci in outbred populations.** *Genetics* 1997, **147**:1445-1457
2. Balding JD, Bishop M, Cannings C: **Handbook of statistical genetics** John Wiley & Sons, New York 2001
3. Wu RL: **Mapping quantitative trait loci by genotyping haploid tissues.** *Genetics* 1999, **152**:1741-1752
4. Wright S: **Evolution and the genetic of populations. II. The theory of gene frequencies.** Univ. of Chicago Press, Chicago 1969
5. Elston RC: **Segregation analysis.** *Adv Hum Genet* 1981, **11**:372-3
6. Hoeschele I: **Genetic evaluation with data presenting evidence of mixed major gene and polygenic inheritance.** *Theoretical and Applied Genetics* 1988, **76**:81-92
7. Hoeschele I: **Statistical techniques for detection of major genes in animal breeding data.** *Theoretical and Applied Genetics* 1988, **76**:311-319
8. Le Roy P, Naveau J, Elsen JM, Sellier P: **Evidence for a new major gene influencing meat quality in pigs.** *Genetical Research* 1990, **55**:33-40

9. Knott SA, Haley CS, Thompson R: **Methods of segregation analysis for animal breeding data: a comparison of power.** *Heredity* 1992, **68**:299-311
10. Le Roy P, Elsen JM: **Simple test statistics for major gene detection: a numerical comparison.** *Theoretical and Applied Genetics* 1992, **83**:635-644
11. Loisel R, Goffinet B, Monod H, De Oca GM: **Detecting a major gene in an F2 population.** *Biometrics* 1994, **50**:512-516
12. Janss LL, Thompson R, Van Arendonk JAM: **Application of Gibbs sampling for inference in a mixed major gene-polygenic inheritance model in animal population.** *Theoretical and Applied Genetics* 1995, **91**:1137-1147
13. Janss LL, Van Arendonk JAM, Brascamp EVW: **Bayesian statistical analysis for presence of single genes affecting meat quality traits in a crossed pig population.** *Genetics* 1997, **145**:395-408
14. Wu RL, Li BL, Wu SS, Casella G: **A maximum likelihood-based method for mining major genes affecting a quantitative character.** *Biometrics* 2001, **57**:764-768
15. Weir BS: **Genetic Data Analysis II.** Sinauer Associates, Sunderland, MA 1996
16. Titterton DM, Smith AFM, Makov UE: **Statistical Analysis of Finite Mixture Distributions.** John Wiley & Sons, NY 1985
17. Feng ZD, McCulloch CE: **On the likelihood ratio test statistic for the number of components in a normal mixture with unequal variances.** *Biometrics* 1994, **50**:1158-1162
18. Fernando RL, Stricker C, Elston RC: **The finite polygenic mixed model: an alternative formulation for the mixed model of inheritance.** *Theoretical and Applied Genetics* 1994, **88**:573-580
19. Dempster AP, Laird NM, Rubin DB: **Maximum likelihood from incomplete data via EM algorithm.** *Journal of Royal Statistical Society Serial B* 1977, **39**:1-38
20. Robert C, Casella G: **Monte Carlo Statistical Methods** Springer, NY 1999
21. Li BL, Wu RL: **Heterosis and genotype x environment interactions of juvenile aspens in two contrasting sites.** *Can. J. Forest Res.* 1997, **27**:1525-1537
22. Oakes D: **Direct calculation of the information matrix via the EM algorithm.** *J. R. Statist. Soc. B* 1999, **61**:479-482
23. Bradshaw HD, Stettler RF: **Molecular-genetics of growth and development in Populus. 4 Mapping QTLs with large effects on growth, form, and phenology traits in a forest tree.** *Genetics* 1995, **139**:963-973
24. McKeand SE, Bridgwater FE: **A strategy for the third breeding cycle of loblolly pine in the southeastern US.** *Silvae Genetica* 1998, **47**:223-234

Publish with **BioMed Central** and every scientist can read your work free of charge

"BioMedcentral will be the most significant development for disseminating the results of biomedical research in our lifetime."

Paul Nurse, Director-General, Imperial Cancer Research Fund

Publish with **BMC** and your research papers will be:

- available free of charge to the entire biomedical community
- peer reviewed and published immediately upon acceptance
- cited in PubMed and archived on PubMed Central
- yours - you keep the copyright



Submit your manuscript here:

<http://www.biomedcentral.com/manuscript/>

[editorial@biomedcentral.com](mailto:editorial@biomedcentral.com)