Computer Simulation of Marker-Directed Population Improvement

M. Lstibůrek[1], C. D. Nelson[2], C. S. Echt[3], T. J. Mullin[4], Y. A. El-Kassaby[5], and R. Pong-Wong[6]

A wide range of commercially important traits in forest tree breeding may be referred to as complex, where a situation-specific approach will make genetic improvement as efficient as possible. The most common approach in current programs is to treat all traits as purely polygenic, assuming the classical "infinitesimal model" (Fisher 1918). The objective of our research is to quantitatively evaluate breeding strategies using marker-directed population improvement (MDPI) (Nelson and Echt 2004). In these strategies, the complex nature of commercial traits is reflected in predicting both polygenic and quantitative trait loci (QTL) genetic effects, and combining these into a single selection criterion. We review here the development of the computer simulation model that will enable this research. With the help of the model, it is possible to assess impact of several parameters, such as the density and information content of markers flanking the QTL, and the relative effect of the QTL on the trait's phenotype. The effects of these variables can be analyzed within the context of a regular recurrent selection strategy, where the main objective is the genetic response in a production population. Proportional reduction in gene diversity, cost, and time components can also be evaluated.

SIMULATION MODEL

The basic model features were derived from the stochastic simulation model POPSIM™ [initial reference by Mullin and Park (1995), and the review of the most recent version by Lstibůrek et. al (2005)]. We assume a single "breeding population" being managed over a number of successive, non-overlapping generations. The population consists of N_T individuals; the population size is held constant. Mating within each generation is declared, and any possible mating scheme can be implemented. Following mating, N_C individuals (progenies) are generated for each family. The full set of families is referred to as the "recruitment population", which then serves as a pool of candidates for forward selection. Environmental conditions are assumed homogeneous across generations.

Individual phenotypic values can be modeled as a function of independent genetic and environmental components. The genetic component is subdivided into two causal factors, polygenic and QTL, in what is commonly referred to as a "mixed-inheritance model" (Gomez-Raya and Klemetsdal, 1999). The polygenes are assumed to be distributed throughout the

[1] Assistant Professor, Faculty of Forestry and Environment, Czech University of Life Sciences, Prague, Czech Republic
[2] Research Geneticist / Project Leader,Southern Institute of Forest Genetics, U. S. Forest Service, Southern Research Station, Saucier, MS, USA
[3] Research Geneticist, Southern Institute of Forest Genetics, U. S. Forest Service, Southern Research Station, Saucier, MS, USA
[4] President, BioSylve Forest Science NZ Limited, Hamilton, New Zealand
[5] Professor and Senior Industry-NSERC Research Chair, Department of Forest Sciences, Faculty of Forestry, University of British Columbia, Canada
[6] Research Scientist, Roslin Institute, Scotland, UK

genome (following the infinitesimal model assumptions), while the QTL is assumed to reside in a particular genome position. A number of QTL loci can be declared. Individuals in the population are genotyped for any number of polymorphic loci flanking the QLT(s) (neutral markers with respect to the trait). Different types of markers can be simulated, such as SNPs or SSRs, where the variables of interest are the number of alleles per locus, their respective frequency, and the distribution of marker loci across the genome.

A QTL locus may reside anywhere within a given chromosome; the actual position is declared in terms of recombination distance from marker loci residing on the same linkage group. The recombination frequency is used when generating genotypes of individual progenies, using the random-walk algorithm, as implemented by Crosby (1973). Individual polygenic values are simulated assuming the infinitesimal model (the method is presented in detail by Mullin and Park, 1995). The effect of the QTL (α) is calculated (following Lstibůrek et al. 2005) as:

$$\alpha = \sqrt{\lambda h^2 V_p / 2p(1-p)} \qquad (1)$$

where λ is the proportion of the additive genetic variance explained by the QTL, h^2 is the narrow-sense heritability, V_P is the phenotypic variance, and p is the initial frequency of the A allele at the QTL (alleles in the founder population are sampled from a uniform distribution with mean p).

The dissection of the phenotypic observation into causal components is performed within a mixed-model framework introduced by Fernando and Grossman (1989). The following naming conventions as well as basic model assumptions are similar to those used by Villanueva et al. (2002). The mixed-model equations in the matrix formulation are as follows:

$$\begin{pmatrix} X'X & X'Z & X'W \\ Z'X & Z'Z + \gamma_1 A^{-1} & Z'W \\ W'X & W'Z & W'W + \gamma_2 G^{-1} \end{pmatrix} \begin{pmatrix} \hat{b} \\ \hat{u} \\ \hat{v} \end{pmatrix} = \begin{pmatrix} X'Y \\ Z'Y \\ W'Y \end{pmatrix} \qquad (2)$$

where X, Z, and W are known incidence matrices; A is the additive relationship matrix; G is a matrix of marker-QTL effects (genotypic relationship matrix); γ_1 and γ_2 are variance ratios in the founder population; $\hat{b}$ is a vector of estimated fixed effects; and $\hat{u}$ and $\hat{v}$ are vectors of predicted random effects of polygenes and QTL, respectively. We use the rapid method of Pong-Wong et al. (2001) for calculating the gametic IBD matrix, which is later converted to the genotypic matrix, G (see Nagamine, 2005) and supplied as input to the ASReml2 software (Gilmour et al. 2006) along with other required information. Data in advanced generations accumulate and full-pedigree information is used in the genetic evaluation (extending back to founders, as well as including all selection candidates in all generations) by Best Linear Unbiased Prediction (BLUP) (Mrode, 2005).

Under the MDPI strategy, both polygenic and QTL components enter the BLUP analysis, and the resulting breeding value (*bv*) of an individual *i* is the respective sum of both effects:

$$\widehat{bv}_i = \hat{a}_i + \hat{v}_i \quad (3)$$

When marker data is not utilized in the prediction, the breeding value is predicted in the regular fashion as:

$$\widehat{bv}_i = \hat{a}_i \quad (4)$$

In this latter strategy, the QTL effect is not properly separated from the polygenic genetic effect (therefore not properly accounted for). Thus, one can evaluate the difference due to the added value of marker information contributing to the selection criterion.

The best set of trees is then selected based on predicted breeding values and group coancestry, using group-merit selection (Lindgren and Mullin 1997), where a weighting is applied to group coancestry to control the proportional reduction in gene diversity. A full range of weight factors can be considered, resulting in two extreme situations (strong family selection and balanced within-family selection), and a number of values between these two extremes.

The approach introduced here makes use of the restricted maximum likelihood (REML) implemented by the ASReml2 software; variances are therefore considered unknown, and are estimated based on the phenotypic and marker data supplied with each ASReml2 call (generation). We believe that this approach is more realistic compared to assuming true variances, as performed in many other simulation models published in the scientific literature. In combination with a powerful method to calculate the inverse of *G* matrix, a proportion of the individuals may have missing records. This newly developed simulation model also offers the potential to assess the utility of early-age selection, which is likely where marker-aided selection can most economically enhance the efficiency of forest tree breeding programs.

Acknowledgements: This project was supported by the U. S. Forest Service, Southern Research Station Cooperative Agreement SRS-05-IC-11330126-234, and the Czech University of Life Sciences, Prague. Dr. R. Pong-Wong is finantially supported by the BBSRC. We thank Prof. Julius van der Werf for technical help on genetic evaluation.

LITERATURE CITED

Crosby, J.L. 1973. Computer simulation in genetics. John Wiley & Sons, London.

Fernando, R.L., and Grossman, M. 1989. Marker assisted selection using best linear unbiased prediction, Genet. Sel. Evol. 21(4): 467–477.

Fisher, R.A. 1918. The correlation between relatives on the supposition of Mendelian inheritance. Trans. Roy. Soc. Edinburgh 52: 399-433.

Gilmour, A.R., Gogel, B.J., Cullis, B.R., and Thompson, R. 2006. ASReml User Guide Release 2.0 VSN International Ltd, Hemel Hempstead, HP1 1ES, UK.

Gomez-Raya, L., and Klemetsdal, G. 1999. Two-stage selection strategies utilizing marker-quantitative trait locus information and individual performance. J. Anim. Sci. 77(8): 2008–2018.

Lindgren, D., and Mullin, T.J. 1997. Balancing gain and relatedness in selection. Silvae Genet. 46(2-3): 124–129.

Lstibůrek, M., Mullin, T.J., Mackay, T.F.C., Huber, D., and Li, B. 2005. Positive assortative mating with family size as a function of parental predicted breeding values. Genetics 171(3): 1311-1320.

Mrode, R.A. 2005. Linear models for the prediction of animal breeding values. 2nd Edition. CABI Publishing, Wallingford, Oxfordshire, UK. 344 pp.

Mullin, T.J., and Park, Y.S. 1995. Stochastic simulation of population management strategies for tree breeding: a new decision-support tool for personal computers. Silvae Genet. 44(2-3): 132-141.

Nagamine, Y. 2005. Transformation of QTL genotypic effects to allelic effects. Genet. Sel. Evol. 37(5): 579–584.

Nelson, C.D., and Echt, C.S. 2004. Marker-directed population improvement. *In* Proc. IUFRO Joint Conf. For. Genet. (Div. 2), Nov. 1-4, 2004, Charleston, SC, pp 255.

Pong-Wong, R., George, A.W., Woolliams, J.A., and Haley, C.S. 2001. A simple and rapid method for calculating identity-by-descent matrices using multiple markers. Genet. Sel. Evol. 33(5): 453-471.

Villanueva, B., Pong-Wong, R., and Woolliams, J.A. 2002. Marker assisted selection with optimized contributions of the candidates to selection. Genet. Sel. Evol. 34(6): 679–703.