

Optimizing plant growth model parameters for genetic selection based on QTL mapping

Veronique Letort, Paul Mahe, Paul-Henry Cournède, P. De Reffye, Brigitte Courtois

► **To cite this version:**

Veronique Letort, Paul Mahe, Paul-Henry Cournède, P. De Reffye, Brigitte Courtois. Optimizing plant growth model parameters for genetic selection based on QTL mapping. Fourcaud Thierry, Zhang XiaoPeng. Plant Growth Modeling, Simulation, Visualization and applications, Nov 2006, Beijing, China. IEEE Computer Society Press, pp.16-21, 2006. <hal-00829810>

HAL Id: hal-00829810

<https://hal-ecp.archives-ouvertes.fr/hal-00829810>

Submitted on 3 Jun 2013

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Optimizing plant growth model parameters for genetic selection based on QTL mapping.

Véronique LETORT¹, Paul MAHE¹, Paul-Henry COURNEDE¹, Philippe DE REFFYE^{2,3},
Brigitte COURTOIS⁴

¹ *Ecole Centrale Paris, Laboratory of Applied Mathematics, 92290 Chatenay-Malabry, France.*

² *CIRAD-Amis, TA 40/01 Ave Agropolis, 34398 Montpellier cedex 5 France*

³ *INRIA-Rocquencourt, BP 105, 78153 Le Chesnay cedex, France*

⁴ *CIRAD-Bios, UMR DAP, Montpellier, F34398, France*

veronique.letort@centraliens.net

Abstract

An increasing interest is given to the potential benefits of introducing ecophysiological knowledge in breeding programs. Indeed, crop models provide powerful tools to predict phenotypic traits from new genotypes under untested environmental conditions. But, until now, few attempts have been undertaken to bridge the gap from genes to phenotype with a chain of functional processes. In this paper, we propose a framework for simulating plant growth from its genotype. Thus the genetic correlations between the parameters can be taken into consideration when optimization processes are used to define ideotypes based on model parameters. The example of virtual maize growing under constant environmental conditions is presented using the functional-structural model GreenLab.

1. Introduction

The recent development of marker assisted selection has provided powerful tools to improve breeding efficiency and to investigate genetic contributions to the phenotype. Markers are specific short strands of DNA whose location can be detected: they ideally represent “flags” regularly spaced on the whole genome map. It allows establishing statistical links between target quantitative traits and particular locations on the chromosomes, bordered by two adjacent markers and called quantitative trait loci (QTL). However, a major difficulty are the inextricable genotype×environment interactions when the target traits are complex ones, such as plant height, yield, kernel or grains weight, seedling vigor... [1]. Nowadays, a consensus is emerging concerning the potentials of integrating physiological analysis in

breeding programs ([2], [3], [4], [5], [6]). In particular, Heuvelink *et al.* [7] report that using crop models allows to evaluate new available genotypes, analyze their performances and select the most influential parameters in order to get yield improvement under various environmental conditions. But few models integrate genetic information, although authors generally claim that their parameters are gene-related. Reymond *et al.* [8] built a QTL-based model linking leaf elongation rate of maize to water vapour pressure and soil water potential with satisfactory success. Buck-Sorlin [9][10] integrated QTL for tillering and number of grains per spike of winter barley into a morphological growth model by statistical association but the effect of environment was not taken into account. The interest of this approach for breeding strategies is quantified by Hammer *et al.* [11] who demonstrated that the predictive power and efficiency of marker-assisted selection was enhanced by the contributions of ecophysiological modelling. However, since most of the target traits (e.g. yield) result from interactions between morphological and physiological processes at whole-plant level, it is important to analyze their association in the dynamic context of plant growth. In our paper, we propose to integrate a genetic model into a functional-structural model (FSM), Greenlab.

To be of practical interest for breeders, growth models should also provide promising search directions. This leads to the problem of determining the ideal set of parameters and the associated genotypic values to get an objective trait (e.g. yield maximisation) under given constraints. The notion of ideotype has been defined by Donald [12] as the set of desirable traits that a plant should present to enhance yield or any other objective trait under specified climatic conditions. It has been characterized for several species (e.g. bean [13], chrysanthemum [14],

pine [15]). This approach can be widened to ideotype definition based on FSM parameters, such as sink ratios [16]. The perfect ideotype can generally not be obtained in practice due to negative genetic correlations between the parameters (*i.e.* one QTL can have effect on several parameters, with opposite influences on the yield). Nevertheless the best compromise can still be characterized: it only requires introducing genetic information as constraints in the optimization procedure. In this paper, we analyze the parameter influence on the determination of cob mass under constant conditions with GreenLab and we characterize an ideotype based on the GreenLab parameters for a virtual diploid cereal whose parameters were calibrated on maize by Ma *et al.* [17].

2. Linking genetic model and GreenLab

Detailed presentations of the GreenLab model can be found in [18], [19], [17] and [20]. It consists of two interacting parts: a dual-scale automaton to perform organogenesis and a source-sink model for biomass production and allocation. We briefly recall here the meanings of the parameters that are used in the following parts. The biomass production Q_n at cycle n is the ratio of the total blade surface available for light interception over a parameter of resistance to radiation-biomass conversion r . This biomass is allocated to each organ according to its demand. The demand of an organ of a given chronological age j is defined by its sink strength P_o , considered as a constant, multiplied by the value of a function of sink variation which has a beta law density shape $f_o(j)$ (o takes the values of b :blade, i :internode, r :root, f :fruit (or cob)). The demand of the plant D_n is thus calculated as:

$$D_n = \sum_{o \in \{b,i,r,f\}} \sum_{j=1}^n f_o(j) \cdot P_o \quad (1)$$

where $f_o(j)=0$ if $j > T_e$, T_e being the expansion duration of the organ.

A simple genetic model was introduced to link the model parameters to a virtual genotype of the plant. For sake of clarity, it is assumed that the virtual genome consists of one pair of homologous chromosomes. They are represented as vectors whose components are numbers that can take several values, called alleles (as in [9]). From the virtual chromosomes C_1 and C_2 , an application f defines the rules of allele expression (dominance or additivity) and then the ‘genetic’ vector of parameters Y is calculated as a product of matrices:

$$Y = D \times A \times f(C_1, C_2) \quad (2)$$

The components of the vector Y are the endogenous parameters that are assumed to be genetically

determined. A is a matrix defining the influence of genes on each parameter, including pleiotropic rules (one gene has an influence on several parameters) or combinations of several gene effects on one parameter. D is a diagonal matrix whose coefficients are scaling factors to have range compatibility. To illustrate the study, the reference values were taken from the calibration of maize [17]. The genetic parameters were arbitrarily selected and are presented in Table 1.

For practical applications, the coefficients of matrix A must be determined, which is analogous to detecting QTL for the model parameters. The QTL Cartographer ([21]) was used on a virtual mapping population that was generated from recombinant inbred lines (detailed procedure can be found in [1]). It relies on statistical methods to test the presence of a QTL at each marker against the opposite hypothesis. For each individual of the population, the virtual genome provided a direct access to its GreenLab parameters and thus to any phenotypic trait by running the growth simulation software. The results give the QTL associated to each parameter, *i.e.* the non null coefficients of matrix A in our model. The detection accuracy and the number of QTL were higher than when classical phenotypic trait was used [22]. Indeed, the virtual phenotypic measurements are the result of a step by step plant growth process where all the genetic parameters are involved through complex equations.

3. Parameter optimization considering genetic correlations

To be of practical use for breeders, it is crucial to identify and optimize the key parameters influencing the final cob mass. For the virtual maize, the cob mass $W_f(n)$ at cycle n is defined in equation (3):

$$W_f(n) = \sum_{i=1}^n P_f f_f(i) \cdot \frac{Q_{i-1}}{D_i} \quad (3)$$

where the parameter names are given in paragraph 2 and the ratio of supply over demand Q_{i-1}/D_i is calculated from equation (1). Since all the parameters of the model have an effect on the determination of the cob weight, finding a global optimum for the whole set of parameter values is a complex problem (see [23] for a first approach of parameter optimization on maize). The parameters concern as well organogenesis (integers) as physiology (real values) but the use of heuristic methods allows solving the problem.

A genetic algorithm [24] was computed on the basis of the genetic model described above and of simulations of reproduction mechanisms. From two parents, a possible ‘child’ is created that inherits one chromosome from each of its parents. Each of those chromosomes can be the result of a crossing-over

between the homologous chromosomes, with a probability that follows a Poisson law whose parameter is the chromosome length. The genetic algorithm consists of generating the evolution of a population and improving its properties by selecting the best individuals as parents for the following generation. The genetic mixing is insured through crossing-over rate and mutation rate (one allele of the chromosome of an individual is randomly replaced by another allele of the species). Mutations allow keeping genetic diversity to avoid stopping at local optima. The output of the genetic algorithm is the combination of alleles that gives the highest cob weight with the genetic rules defined in the model through matrix A and application f .

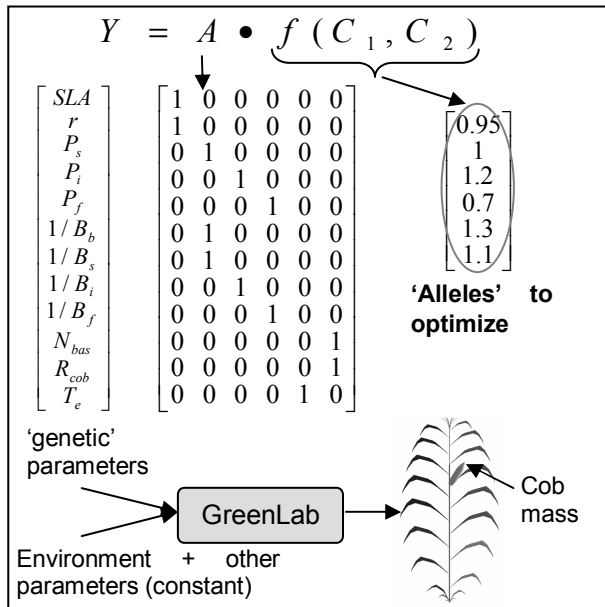


Figure 1. Optimization of allele values for the objective function of cob mass with the genetic correlations defined by matrix A . Parameter names are recalled in Table 1.

Matrix A and application f are defined from the results of the QTL detection on model parameters. Thus it is possible to take into account the genetic correlations between the parameters in the optimization process. In our simulations, as no experimental results were available, the genetic correlations were arbitrarily set as presented in Fig. 1. For example, we can assume that SLA (Specific Leaf Area, $\text{cm}^2 \cdot \text{g}^{-1}$) and resistance to biomass conversion r are determined by the same QTL. The following parameters were also linked through a same genetic control: {sheath sink P_s , sheath sink variation parameter B_s and blade sink variation parameter B_b }; {internode sink P_i and internode sink variation parameter B_i }; {cob sink P_f and cob sink

variation parameter B_f }; {number of basal short internodes N_{bas} and cob rank R_{cob} }. The expansion duration of all organs T_e was set independently.

The case where matrix A is the identity matrix (one gene for one parameter) gives the optimal parameter combinations. It can be compared to the probably more realistic case where matrix A defines correlations between the parameters (Table 1).

Table 1. Results of genetic algorithm without or with genetic correlations (as defined in Fig. 1 and recalled through the superscript indices).

Parameter	Reference value	Optimal value (A=Id)	Optimal value with correlations
SLA ($\text{cm}^2 \cdot \text{g}^{-1}$) ¹	35.7	46.4 (<i>max</i>)	25.0 (<i>min</i>)
Photosynthesis r ¹	354	2478 (<i>min</i>)	248 (<i>min</i>)
Sheath Sink P_s ²	0.7	0.49 (<i>min</i>)	0.91 (<i>max</i>)
Internode Sink P_i ³	2.17	1.52 (<i>min</i>)	2.82 (<i>max</i>)
Cob Sink P_f ⁴	202	222	161
Blade Sink Variation	0.4	0.31 (<i>min</i>)	0.31 (<i>min</i>)
Parameter B_b ²			
Sheath Sink Variation	0.53	0.41 (<i>min</i>)	0.41 (<i>min</i>)
Parameter B_s ²			
Internode Sink Variation	0.79	0.61 (<i>min</i>)	0.61 (<i>min</i>)
Parameter B_i ³			
Cob Sink Variation	0.62	0.43 (<i>min</i>)	0.50
Parameter B_f ⁴			
Number of short basal internodes N_{bas} ⁵	6	8 (<i>max</i>)	6
Cob rank R_{cob} ⁵	15	9 (<i>min</i>)	14
Expansion duration T_e ⁶	12	16 (<i>max</i>)	12
Cob mass (g)	750.2	2325	1428

The allele values were set in the interval [0.7 ; 1.3], that is to say the parameter variation range was $\pm 30\%$ around the reference value. The results show that sinks of non productive organs (except cob) should take minimal values to avoid waste in biomass partitioning. The number of short internodes should be as large as possible since it lets the plant allocate biomass uppermost to the blades that are the future source of assimilate production. But for other parameters or if matrix A is not the identity matrix, the influence is more complex and must be found through the algorithm. When genetic correlations are introduced, the potential cob mass is reduced of 38% and the dynamics of biomass partitioning are radically different: the optimal coordinates are not anymore positioned at the search domain boundaries for parameters concerning cob sink and plant topology.

4. Sensitivity analysis

A sensitivity analysis helps interpreting those results since it allows understanding the influence of each parameter on the objective function. Fig. 2A shows the effect of the variation of some parameters on

the virtual cob mass, all the other parameters being constant. The total number of phytomers (21) does not vary and the maximal cob rank is 19.

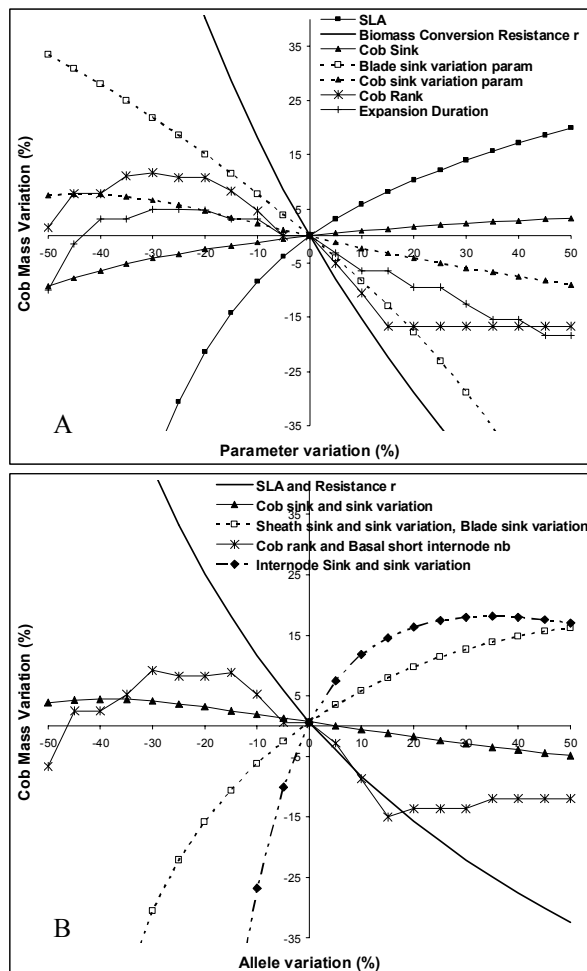


Figure 2. Sensitivity analysis on cob mass when A is identity matrix (A) and with genetic correlations between the parameters (B).

The parameter having the highest impact on the cob mass is the resistance r (intervening in biomass production equation) since its diminution increases the plant's ability to perform photosynthesis. Increasing SLA has a positive effect on the cob mass but it tends to stabilize because of leaf area index saturation (no need to increase the SLA if the light interception is already maximal). The little influence of variation of cob sink is due to its high value compared to other sinks. An optimum can be detected for cob rank. Its position depends on the shape of the cob sink variation function and on the biomass partitioning dynamics. Indeed, if the cob appears too late, it has not enough time to reach maturity; whereas if it appears too early, it penalizes blade growth, thus inducing a global

decrease in plant biomass production. The expansion duration parameter drives the growth duration of all organs except for cob whose expansion duration is five-cycle higher. It shows an optimum when a balance is found between decreasing the expansion duration of the cob (negative effect on cob mass) and decreasing the expansion duration of other organs (positive effect).

This sensitivity analysis can be extended to the case of complex associations at genotypic level (Fig. 2B). The correlations are given through the non null coefficients of matrix A (see Fig. 1). Globally, the parameter influences decrease due to the opposite interactions, which confirms the results found from the genetic algorithm (Table 1). For example, the strong impact of resistance on cob mass determination is weakened if the corresponding allelic value also drives the SLA. The optimal value is the minimal one (Table 1) since resistance influence is much higher than SLA ones (Fig. 2A). However, the optimum cannot be estimated directly from the sensitivity analysis since it requires a simultaneous optimization of all the parameters. For example, the optimal value for cob sink and sink variation parameter is not the minimal ones as the curve of Fig. 2B would tend to suggest it.

5. Discussion

In this paper, the potential of linking growth models to quantitative genetics are studied using simulation of a GreenLab maize. According to Martin *et al.* [15], "the successful incorporation of ideotypes into breeding programs probably depends on identification of only a few, critical traits that are closely correlated or linked with (...) yield". Concerning model parameters, the sensitivity analysis reveals the most influential ones under a given climatic scenario and optimization procedures allow to get the optimal combination. Genetic correlations are not included at the first approach but once they have been detected, they can be taken into consideration and included as constraints into the optimization process. A genetic algorithm was used to optimize the parameters to get the highest cob weight for maize and to find the best compromise when negative genetic correlations were included. This method has the advantage of allowing optimization on both real and integer values at the same time (for the real values, the accuracy of the solution is set by the length of the discretization step chosen). It requires no *a priori* knowledge on the objective function and no derivation processes. But the convergence time is high (about half an hour) and one major difficulty remains to determine whether a local optimum is also a global optimum of the whole search

domain. However, for practical purpose, even a local optimum would provide an interesting result since the corresponding changes in parameter values would increase the yield. More parameters should be included in a complete study and new constraints should be added to have more realistic optimized values.

6. Conclusion

Model parameters should have higher heritability since they are expected to be less dependent on the environmental conditions and to be more direct gene expression. Optimization processes allows determining the key parameters influencing the yield, even when complex genetic correlations are introduced. Thus QTL detection on model parameters is worth to be tested since it provides new promising selection criteria.

7. References

- [1] Vienne (de) D., Santoni S., Causse M., Kremer A., Charcosset A. and Gallais A. and E.F. Roberts, *Les marqueurs moléculaires en génétique et biotechnologies végétales*, INRA Editions, Paris, 1998.
- [2] Chapman S, Cooper M, Podlich D, Hammer G. "Evaluating plant breeding strategies by simulating gene action and dryland environment effects". *Agronomy Journal*. 2003. 95: 99–113.
- [3] Hammer G., Cooper M., Tardieu F., Welch S., Walsh B., Van Eeuwijk F., Chapman S., Podlich D. "Models for navigating biological complexity in breeding improved crop plants". *Trends in Plant Science*, 2006, 11(12): 587–593.
- [4] Tardieu F. "Virtual plants: modelling as a tool for the genomics of tolerance to water deficit". *Trends in Plant Science*, 2003, 8(1): 9-14.
- [5] Yin X., Stam P., Kropff M.J., Schapendonk Ad H.C.M. "Crop modeling, QTL mapping, and their complementary role in plant breeding". *Agronomy Journal*, 2003, 95: 90–98.
- [6] Yin X., Struik P., Kropff M. "Role of crop physiology in predicting gene-to-phenotype relationships". *Trends in Plant Science*, 2004, 9(9): 426–432.
- [7] Heuvelink E., Marcelis L.F.M., Bakker M.J., Van der Ploeg A. "Use of plant growth models to evaluate physiological traits in genotypes of horticultural crops." In: Spiertz J.H.J, Struik P.C., Van Laar H.H., eds. Scale and complexity in plant systems research: gene-plant-crop relations 223–233.
- [8] Reymond M., Muller B., Leonardi A., Charcosset A., Tardieu F. « Combining Quantitative Trait Loci analysis and an ecophysiological model to analyze the genetic variability of the responses of maize leaf growth to temperature and water deficit". *Plant Physiology*, 2003, 131: 664–675.
- [9] Buck-Sorlin G.H. and Bachmann K. "Simulating the morphology of barley spike phenotypes using genotype information". *Agronomie*, 2000, 20: 691-702.
- [10] Buck-Sorlin G.H. "The search for QTL in Barley (*Hordeum Vulgare* L) using a new mapping population". *Cellular & molecular biology letters*, 2002, 7: 523-535.
- [11] Hammer G.L., Chapman S., Van Oosterom E., Podlich D.W. "Trait physiology and crop modelling as a framework to link phenotypic complexity to underlying genetic systems". *Australian Journal of Agricultural Research* 2005, 56: 947-960.
- [12] Donald CM. "The breeding of crop ideotypes". *Euphytica* 1968, 17: 385-403.
- [13] Beattie A.D., Larsen J., Michaels T.E., Pauls K.P. "Mapping quantitative trait loci for a common bean (*Phaseolus vulgaris* L.) ideotype". *Genome*, 2003, 46: 411-422.
- [14] Langton F.A. and Cockshull K.E. "An ideotype of chrysanthemum (*C. Morifolium* Ramat.)". *Acta Horticultura* (ISHS), 1976, 63: 165-176.
- [15] Martin T.A., Johnsen K.H., White T.L. "Ideotype development in southern pine: rationale and strategies for overcoming scale related obstacles." *Forest Science*, 2001, 47(1): 21-28.
- [16] Dingkuhn M., Luquet D., Clément-Vidal A., Tambour L., Kim H.K., Song Y.H. "Is plant growth driven by sink regulation?" In: Spiertz J.H.J, Struik P.C., Van Laar H.H., eds. Scale and complexity in plant systems research: gene-plant-crop relations 157–170.
- [17] Ma Y., Li B., Zhan Z., Guo Y., Luquet D., de Reffye P., Dingkuhn M. «Parameter Stability of the Structural-Functional Plant Model GREENLAB as affected by Variation within Populations, among Seasons and among Growth Stages". *Annals of Botany*, 2007, 99: 61–73.
- [18] Cournède P.-H., Kang M.-Z., Mathieu A., Barczy J.-F., Yan H.-P., Hu B.-G., Reffye (de) P. « Structural factorization of plants to compute their functional and architectural growth". *Simulation*, 2006, 82 (7): 427-438.
- [19] Guo, Y., Ma, Y.T., Zhan, Z.G., Li, B.G., Dingkuhn, M., Luquet, D., and de Reffye, P. "Parameter optimization and field validation of the functional–structural model GREENLAB for maize". *Annals of Botany*, 2006, 97, 217-230.
- [20] Yan H.-P., Kang M.Z., Reffye (de) P, Dingkuhn M. "A dynamic, architectural plant model simulating resource-dependent growth". *Annals of Botany*, 2004, 93: 591–602.

[21] Basten C.J., Weir B.S., Zeng Z.B. "QTL Cartographer v1.17". North Carolina State University. 2005. <ftp://statgen.ncsu.edu/pub/qtlcart>

[22] Letort V., Mahe P., Cournede P.H., de Reffye P., Courtois B. "Quantitative Genetics and Functional-Structural Plant Growth Models: Simulation of Quantitative Trait Loci Detection for Model Parameters and Application to Potential Yield Optimization". *Annals of Botany. In press.*

[23] Wu L., de Reffye P., Le Dimet F.X., Hu B.G. "Optimization of source-sink relationships based on a plant functional-structural model: a case study on maize". In: Hu BG, Jaeger M, eds. Proc. Plant Growth Modeling and Applications (PMA'03), Beijing, China. Tsinghua University Press and Springer, 2003, 285-294.

[24] Sastry K, Kendall G and Goldberg D. "Genetic algorithms". In: Burke E K and Kendall G (eds). Search Methodologies: Introductory Tutorials in Optimisation, Decision Support and Search Techniques, 2005, 97-125.