# Crop growth models for the -omics era: the EU-SPICY project

**R.E. Voorrips[1], A. Palloix, A. Dieleman, M. Bink, E. Heuvelink, G. van der Heijden, M. Vuylsteke, C. Glasbey, A. Barócsi. J. Magán, F. van Eeuwijk**
*1 Plant Research International, P.O. Box 16, 6700 AA Wageningen, the Netherlands. Contact: roeland.voorrips@wur.nl*

## Abstract
The prediction of phenotypic responses from genetic and environmental information is an area of active research in genetics, physiology and statistics. Rapidly increasing amounts of phenotypic information become available as a consequence of high throughput phenotyping techniques, while more and cheaper genotypic data follow from the development of new genotyping platforms. , A wide array of -omics data can be generated linking genotype and phenotype. Continuous monitoring of environmental conditions has become an accessible option. This wealth of data requires a drastic rethinking of the traditional quantitative genetic approach to modeling phenotypic variation in terms of genetic and environmental differences. Where in the past a single phenotypic trait was partitioned in a genetic and environmental component by analysis of variance techniques, nowadays we desire to model multiple, interrelated and often time dependent, phenotypic traits as a function of genes (QTLs) and environmental inputs, while we would like to include transcription information as well. The EU project 'Smart tools for Prediction and Improvement of Crop Yield' (KBBE-2008-211347), or SPICY, aims at the development of genotype-to-phenotype models that fully integrate genetic, genomic, physiological and environmental information to achieve accurate phenotypic predictions across a wide variety of genetic and environmental configurations. Pepper (*Capsicum annuum*) is chosen as the model crop, because of the availability of genetically characterized populations and of generic models for continuous crop growth and greenhouse production. In the presentation the objectives and structure of SPICY as well as its philosophy will be discussed.

## Introduction

Plant breeding has considerably contributed to the increased quality and yield of crops over the last decades. This was initially achieved by a systematic comparison of crosses in an experimental set-up. In the last decade the use of molecular markers has been added as a tool in breeding and this has increased insight in the genetics behind the genotypic differences. By selecting genotypes on the basis of molecular markers, we aim to select the ones having the favorable phenotype. This method of breeding is commonly known as marker assisted breeding and has proven to be especially successful when used for simple traits involving a very limited number of genes, e.g. disease resistance.
For complex traits like development and yield, current molecular breeding still has some severe limitations. By complex traits we mean traits that are the outcome of many underlying genetic factors that mask or accentuate each other and that interact with environmental factors. Prediction of the phenotype for complex traits is difficult due to the many interactions that need to be taken into account and the large variation observed. These traits are however most crucial to face the challenges of the future. In order to select and breed the best genotypes for a large range of diverse conditions, ideally the breeder should test all his crosses under all these conditions. Especially with complex physiological traits like energy content, food quality or yield, which exhibit large

variation, this would require many expensive and large trials. The considerable costs involved hamper this approach.

*How can molecular breeding help to assist breeders for these complex traits?*
The 'traditional' approach to link genetic markers to a trait which is the result of multiple interacting genes, is by quantitative trait loci (QTL) analysis. This analysis is generally conducted for phenotypes observed in a single environment, but this is often not sufficient for complex traits that exhibit considerable genotype x environment interaction. Recently, advances have been made by considering the combination of the QTL under different environments, a so called QTL x E analysis, and new methods are still being developed in this area (Alimi et al, this issue). The occurrence of QTLxE interactions can be discovered by performing experiments at several locations under different conditions. However, in itself this doesn't lead to predictive models. In order to achieve that, it is necessary to know what the important environmental factors are, and how changes in these factors affect the traits studied. This can be approached purely statistically (Van Eeuwijk et al., 2010), e.g. by the inclusion of environmental data as cofactors. However, a different and biologically more meaningful approach is the use of crop growth models.

Crop growth models have proven to be an excellent tool to predict crop yield of a specific variety under different environmental conditions. A crop growth model disentangles the complex trait yield under different conditions in a number of model parameters specific for the crop, based on known physiological principles like photosynthesis, and for the environment, like light and temperature (Figure 1). In this project we want to integrate the two approaches of QTL and crop growth modelling.

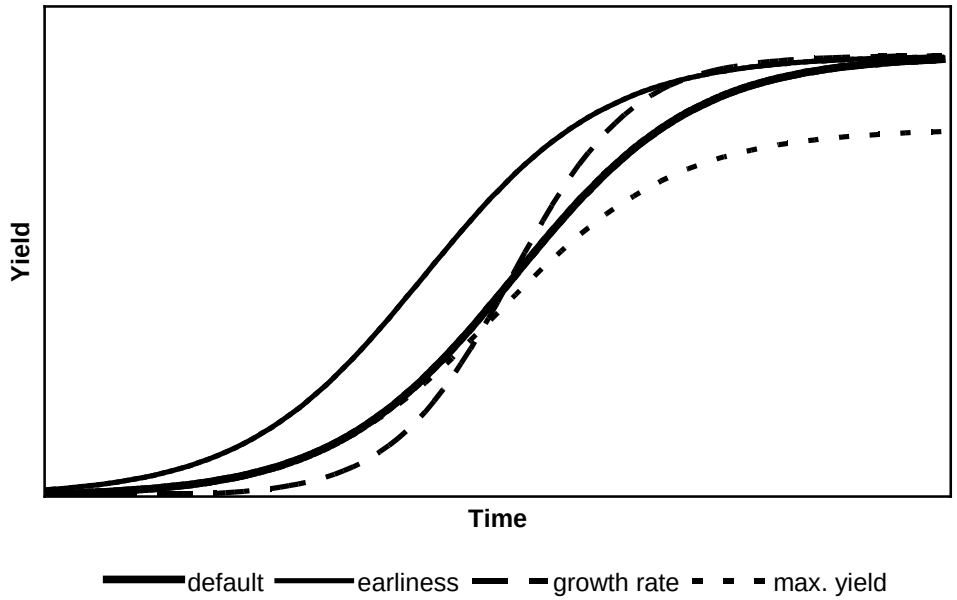


Figure 1. A simple growth model with three parameters describes the development of yield over time. The responses are shown of a "default" genotype and of three other genotypes, each differing from the default in only one parameter: earliness, growth rate or maximum yield. It is expected that QTLs for such parameters are more stable across environments than QTLs for yield itself.

Basically we propose to use explanatory models to disentangle the sink and source components of growth. The hypothesis is that model parameters are more directly linked to genetic information than direct plant measurements (e.g. length, fruit size, leaf area) as the latter are the final result of complex interactions between sink and source. Hence QTL

regions for these model parameters are expected to be more specific and stable over environments than QTL for those directly measured traits (Van Eeuwijk et al., 2010). The potential of this "gene-to-phenotype" modeling approach was illustrated in a simulation study by Chenu et al. (2009). The results of this approach will be compared with those of a QTL study for the measured traits (Barchi et al., 2009) in the same population.

If QTLs can be found for the crop growth model parameters, this will help us to predict the performance of a genotype under a range of environmental conditions, reducing the need for large scale phenotyping. Recent research has shown the potential of this approach (Letort et al, 2006). This approach requires extension of existing crop growth models to better handle the genotype specific parameters and new QTL-analysis tools to link genetic markers / QTL with these model parameters. An illustration of the concept is shown in Figure 2.

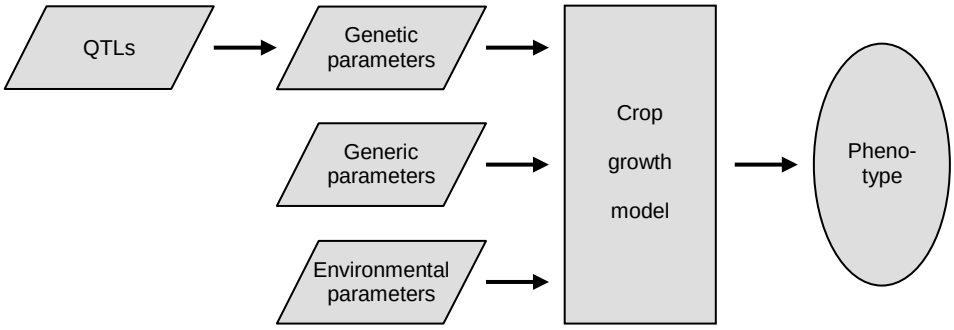


Figure 2. The concept of QTL identification for model parameters instead of for phenotypic traits.

QTLs for crop growth model parameters are of use in marker assisted breeding, but they still pose some drawbacks: QTLs identified in one population may not be useful in another, due to differences in parental alleles in markers and/or genes, possible loss of linkage and their interaction with the genetic background. Besides QTLs do not increase insight in the true genetic and metabolic processes involved. It would be more interesting to find the gene(s) underlying the QTL for crop growth model parameters. This would help to identify their mode of action, and also allow multiple alleles to be found in other genetic material. Therefore we will apply and develop tools to localize the responsible genes within the QTL (Nicolaï et al, this issue).

Large scale phenotyping is needed to provide the data for these analyses, and will also remain necessary in breeding. Therefore we will also develop automated and fast high-throughput tools for large scale phenotyping, thereby reducing the amount of manual labour necessary in phenotyping experiments.

Solanaceous species are among the major EU-grown crops (EPSO, 2004). Pepper was selected as a model crop as suitable genetic material (a genotyped set of Recombinant Inbred Lines) was available, as well as a genetic map and a suitable, although not genotype specific, crop growth model. Furthermore the crop is grown indoors, allowing better crop management, hence limiting the environmental variation. The tools developed in this study have the potential to be applied to other crops as well.

## Scientific approach

*Plant material and phenotyping experiments*

For this project a *Capsicum annuum* intraspecific Recombinant Inbred Line (RIL) population of the cross "Yolo Wonder" x "Criollo de Morelos 334" (Barchi et al, 2007) is

used, which was already genotyped. The parents of this population differ markedly in leaf size and shape, stem length, fruit size and shape and other traits (Barchi et al., 2009), allowing to study the segregation of many traits involved in crop growth models.

The main phenotyping s done in four large experiments in 2009, two in Wageningen, the Netherlands and two in Almeria, Spain. In each experiment the RIL population, including controls and replicates, is grown. Phenotyping is done both manually for plant and leaf morphology and fruit number and size, and by using the phenotyping tools described in the next paragraph. Apart from these experiments a pilot experiment was performed in 2008, and a validation experiment will be performed in 2011.

*Large-scale phenotyping tools*

We have developed two phenotyping tools: an imaging tool for capturing and analyzing images of the plants growing in a greenhouse, and a tool for measuring chlorophyll fluorescence as a parameter for photosynthetic potential.

The imaging tool consists of a trolley with 4 color cameras, 4 infrared cameras and 4 range finder cameras, mounted on a vertical frame to capture the entire plant height. The plants are labeled with a bar code that is also included in the image. We are developing software that estimates the leaf area, the amount of stem tissue and the number and size of fruits from the captured images.

The chlorophyll fluorescence tool consists of a mobile setup with several (currently two) sensor heads, each containing a chamber to hold a leaf equipped with multi-wavelength illumination and detection, temperature sensor and humidity sensor, allowing several plants to be monitored simultaneously.

*Genotype specific crop growth and yield models*

Three models are compared within this project. The simplest model (SPICY 1; 7 parameters) simulates growth of vegetative and generative biomass based on light use efficiency. Partitioning to the fruits (harvest index) is assumed to be constant. The second model (SPICY 2; 20 parameters) resembles the simplest model, but includes a boxcar train method to simulate fruit development. The most complex model is INTKAM (> 50 parameters; Marcelis et al., 2006), which contains many submodels for e.g. light interception, photosynthesis, respiration, dry matter partitioning and fruit growth.

It is an important research question in this project, to determine which model will best serve our goals. A simple model with only a few parameters that can all be determined for all genotypes, or a complex model with many parameters. Such a complex model is more flexible and 'physiologically sound'. However, it contains many parameters which cannot be determined for each genotype and hence have to be assumed equal for all genotypes. Furthermore, some of the parameters will hardly influence the model output. Based on probabilistic sensitivity analysis (Oakley and O'Hagan, 2004), the most relevant parameters in such a complex model will be determined and will be measured on all genotypes.

*New QTL analysis tools*

A major component in the SPICY project is the development of QTL mapping methodology for the identification of crop growth parameters. As mentioned before, we will model the phenotypic traits over time (longitudinally), and more specifically the changes (increase/decrease and acceleration/deceleration) that these traits show. Furthermore, this analysis should not be done for each growth trait separately, but for all traits simultaneously (Alimi et al, this issue).

The mapping of QTL for longitudinal traits may be done by a two step approach comprising the fitting of a suitable growth curve (e.g., logistic, exponential, Gompertz) and subsequently treating the curve parameter estimates as trait records (e.g., Malosetti et al., 2006). However, here we aim to integrate these two steps into one flexible method that, for example, takes into account the uncertainty in parameter estimates.
A statistical framework that allows explicit specification of prior knowledge (or prior uncertainty) about model parameters is the Bayesian paradigm. In a Bayesian approach the prior knowledge on model parameters is integrated with the information contained by the experimental data. After this integration, conclusions are based on the posterior knowledge that also quantifies the degree of certainty on the model parameters after the analyses. Bayesian approaches for QTL mapping have been successfully applied to analyze complex traits (e.g., Bink et al., 2002; Bink et al., 2008; Yi & Shriner, 2008; Bink & Van Eeuwijk, 2009; Liu & Wu 2009). The Bayesian approach will likely build upon the R packages R/qtl and R/qtlbim (Yandell et al., 2007) as the R language is flexible and publicly available.

*Candidate gene identification*
QTL regions are generally large, containing many hundreds of genes. In order to pinpoint genes in the QTL regions that are likely to be causally related to the QTL effect we will follow two approaches (Nicolaï et al, this issue). The first is to focus on known genes for similar traits that have already been validated in other crops. We will generate SNP markers in the corresponding Capsicum homologues and check whether these are mapped to the QTL regions in the RIL population.

Another approach to identify the genes involved in the growth of pepper is by studying the differential gene expression between contrasting QTL-genotypes (Clark et al. 2006; Clop et al. 2006; Frary et al. 2000). We will assay variation in gene expression of thousands of loci in the pepper genome. By combining QTL mapping with expression profiling, called eQTL mapping, one can identify and locate on a linkage map positional candidate genes for a phenotype of interest whose expression segregates in the progeny. Those genes that are located in a growth model QTL region and whose eQTL also coincides with that QTL (so-called cis-acting eQTLs) will be interesting genes for further study.

## Conclusion

The European SPICY project is a major approach to develop tools for the genetic analysis of, and breeding for complex traits like growth and yield. It is multi-disciplinary, involving contributions from electronics and engineering, crop husbandry, plant physiology and molecular and quantitative genetics. Most major pepper breeding companies are represented on the Industrial Advisory Board. All results of this project will be in the public domain, made available through scientific publication, presentations and through the project website: www.spicyweb.eu. This project is therefore likely to have a significant impact on European pepper breeding.

## Acknowledgement

The research leading to these results has received funding from the European Community's Seventh Framework Programme (FP7/2007-2013) under grant agreement n° 211347.

## References


Alimi, N.; Bink, M.; Dieleman, A.; Voorrips, R.; Palloix, A.; Van Eeuwijk, F. 2010. QTL methodology for correlated physiological traits in pepper. (This issue)

Barchi, L.; Lefebvre, V.; Lanterri, S.; Nagy, I.; Grandbastien, M.A.; Palloix, A. 2007. A high resolution intra-specific linkage map of pepper (*Capsicum annuum* L.) and the selection of reduced RILs subsets for fast mapping. Genome 50: 51-60.

Barchi, L.; Lefebvre, V.; Sage-Palloix, A.M.; Lanteri, S.; Palloix, A. 2009. QTL analysis of plant development and fruit traits in pepper and performance of selective phenotyping. Theoretical and Applied Genetics 118:1157-1171.

Bink, M.C.A.M.; Uimari, P.; Sillanpaa, J.; Janss, L.L.G.; Jansen, R.C. 2002. Multiple QTL mapping in related plant populations via a pedigree-analysis approach. Theoretical and Applied Genetics 104:751-762.

Bink, M.C.A.M.; Boer, M.P.; Ter Braak, C.J.F.; Jansen, J.; Voorrips, R.E.; Van de Weg, W.E. 2008. Bayesian analysis of complex traits in pedigreed plant populations. Euphytica 161:85 - 96.

Bink, M.C.A.M.; Van Eeuwijk, F.A. 2009. A Bayesian QTL linkage analysis of the common dataset from the 12th QTLMAS workshop. BMC Proceedings 3:S4.

Chenu, K.; Chapman, S.C.; Tardieu, F.; McLean, G.; Welcker, C.; Hammer, G.L. 2009. Simulating the yield impacts of organ-level quantitative trait loci associated with drought response in maize: a “gene-to-phenotype” modeling approach. Genetics 183: 1507-1533.

Clark, R.M.; Wagler, T.N.; Quijada, P.; Doebley J. 2006. A distant upstream enhancer at the maize domestication gene tb1 has pleiotropic effects on plant and inflorescent architecture. Nature Genetics 38: 594-597.

Clop, A.; Marcq, F.; Takeda, H.; Pirottin, D.; Tordoir, X.; Bibe, B.; Bouix, J.; Caiment, F.; Elsen, J.M.; Eychenne, F.; Larzul, C.; Laville, E.; Meish, F.; Milenkovic, D.; Tobin, J.; Charlier, C.; Georges, M. 2006. A mutation creating a potential illegitimate microRNA target site in the myostatin gene affects muscularity in sheep. Nature Genetics 38: 813-818.

EPSO (2004) Plants for the Future. 2025 a European vision for plant genomics and biotechnology. EUR 21359 EN. http://www.epsoweb.org/catalog/tp/

Frary, A.; Nesbitt, T.C.; Grandillo, S.; Knaap, E.; Cong, B.; Liu, J.; Meller, J.; Elber, R.; Alpert, K.B.; Tanksley, S.D. 2000. fw2.2: a quantitative trait locus key to the evolution of tomato fruit size. Science 289: 85-88.

Letort, V.; Mahe, P.; Cournède P.-H.; Courtois, B.; De Reffye, P. 2006. Quantitative genetics and plant growth simulation: a theoretical study of linking quantitative trait Loci (QTL) to model parameters, in: PMA06 The Second International Symposium on Plant Growth Modeling, Simulation, Visualization and Applications, Beijing, China, November 13-17, 2006.

Liu, T.; Wu, R. 2009. A Bayesian algorithm for functional mapping of dynamic complex traits. Algorithms 2:667-691.

Malosetti, M.; Visser, R.G.F.; Celis-Gamboa, C.; Van Eeuwijk, F.A. 2006. QTL methodology for response curves on the basis of non-linear mixed models, with an illustration to senescence in potato. Theoretical and Applied Genetics 113: 288-300.

Marcelis, L.F.M.; Elings, A.; Bakker, M.J.; Brajeul, E.; Dieleman, J.A.; De Visser, P.H.B.; Heuvelink, E. 2006. Modelling dry matter production and partitioning in sweet pepper. Acta Horticulturae 718:121-128.

Nicolaï, M.; Sage-Palloix, A.M.; Nemouchi, G.; Savio, B.; Lefebvre, V.; Vuylsteke, M.; Palloix, A. 2010. Providing genomic tools to increase the efficiency of molecular breeding for complex traits in pepper. (This issue).

Oakley, J.E.; O'Hagan, A. 2004. Probabilistic sensitivity analysis of complex models: a Bayesian approach. Journal of the Royal Statistical Society. Series B (Statistical Methodology) 66:751–69.

Van Eeuwijk, F.A.; Bink, M.C.A.M.; Chenu, K.; Chapman, S.C. 2010. Detection and use of QTL for complex traits in multiple environments. Current Opinion in Plant Biology (in press).

Yandell, B.S.; Mehta, T.; Banerjee, S.; Shriner, D.; Venkataraman, R.; Moon, J.Y.; Neely W.W.; Wu, H.; Von Smith, R.; Yi, N. 2007. R/qtlbim: QTL with Bayesian Interval Mapping in experimental crosses. Bioinformatics 23: 641-643.

Yi, N.; Shriner, D. 2008. Advances in Bayesian multiple quantitative trait loci mapping in experimental crosses. Heredity 100:240-252.