Bringing genetics and biochemistry to crop modelling, and vice versa

Yin, X., van der Linden, G., & Struik, P. C.

Yin, X., van der Linden, G., & Struik, P. C. (2018). Bringing genetics and biochemistry to crop modelling, and vice versa. European Journal of Agronomy. DOI: 10.1016/j.eja.2018.02.005

Bringing genetics and biochemistry to crop modelling, and *vice versa*

Xinyou Yin[1]*, C. Gerard van der Linden[2] & Paul C. Struik[1]

[1] Centre for Crop Systems Analysis, Department of Plant Sciences, Wageningen University & Research, P.O. Box 430, 6700 AK Wageningen, The Netherlands

[2] Plant Breeding, Department of Plant Sciences, Wageningen University & Research, P.O. Box 386, 6700 AJ Wageningen, The Netherlands

* The corresponding author; e-mail: Xinyou.Yin@wur.nl

Highlights:

- The rationale to link crop modelling with genetics and biochemistry (the MGB framework) is presented;
- Examples showing the synergy among the three disciplines are highlighted;
- Experiences of practising this MGB framework so far are summarised;
- The MGB framework best serves as a first step towards "Crop Systems Biology".

Abstract

Genetics, biochemistry, and crop modelling are independently evolving disciplines; however, they complement each other in addressing some of the important challenges that crop science faces. One of these challenges is to improve our understanding of crop genotype-to-phenotype relationships in order to assist the development of high-yielding and resource-use efficient genotypes that can adapt to particular (future) target environments. Crop models are successful in predicting the impact of environmental changes on crop productivity. However, when critically tested against real experimental data, crop models have been shown to be less successful in predicting the impact of genotypic variation and genotype-by-environment interactions exhibited in genetic populations. In order to better model gene-trait-crop performance relationships in support of breeding and genetic engineering programmes, crop models need to be improved in terms of both model parameters and model structure. We argue that integration of quantitative genetics and photosynthesis biochemistry with modelling is a first step towards a new generation of improved crop models. With genetic information and biochemical understanding incorporated, crop modelling also generates new insights and concepts that can in turn be used to improve genetic analysis and biochemical modelling of complex traits. This modelling-genetics-biochemistry framework (the MGB triangle framework) stresses the synergy among the three disciplines, and may best serve as a step to achieve the ultimate goal of the more broadly framed “Crop Systems Biology” approach to improve efficiency of both classical breeding and genetic engineering programmes.

Keywords: complex phenotype, crop improvement, G×E, interdisciplinary approach, systems modelling.

Introduction

Since the first plant models, mainly on canopy photosynthesis, were created (de Wit 1959, 1965; Duncan et al. 1967), crop modelling has experienced a phase of rapid development during the period from 1970-1990. Thereafter, progress has been steady, reflected in growing number of models as well as in developing model interfaces to promote model applications. Nowadays, crop models have widely been used in optimising crop management and in predicting the impact of environmental changes on crop productivity. Given the plethora of models, the is a growing interest to compare models and to examine the potential of using an ensemble of multiple models, for example, for better assessment of responses of major crops to climate change variables (e.g. Li et al. 2015).

Another domain of model application is model-based plant breeding (Loomis et al. 1979). These models predict a complex crop trait by quantifying nonlinear responses of its component physiological processes to environmental variables, thereby dissecting the complex trait into its component traits. The component traits correspond to model-input parameters, and reflect effects of genetic origin. The other category of model inputs are soil variables, crop management options and daily weather data. The models, therefore, enable the formula in genetics: phenotype = *f*(genotype, environment) to be manifested, where *f* represents an overall model structure (including both algorithms and the connections between them) allowing for the interactions of component processes related to ontogeny-dependent phenological, morphological, and physiological characteristics.

The most illustrated application of using models in breeding is to propose crop ideotypes (e.g., Penning de Vries 1991; Aggarwal et al. 1997; Dingkuhn et al. 2015). However, most ideotyping studies do not consider their underlying genetic basis of model-input parameters (Stam 1998; Picheny et al. 2017) and, therefore, it remains a question whether the suggested ideotypes could be realised through actual breeding. Also, when critically tested against real experimental data, crop models have been shown to be less successful in predicting the impact of genotypic variation and genotype-by-environment interactions (G×E) on yield in breeding or genetic populations such as recombinant inbred line (RIL) populations (Yin et al. 2000). In order to better model genotype-to-phenotype relationships in support of breeding programmes, crop models need to be improved in terms of both model parameters and model structure (Parent & Tardieu 2014).

To improve model parameters, on the general premise that model parameters are under genetic control, the parameters can be subjected to genetic analysis such as QTL (quantitative trait locus) mapping (Yin et al. 2000) or directly correlated with the allelic information of candidate genes (White & Hoogenboom 1996; White et al. 2008; Boote et al. 2016). Ideally, the individual parameters are under simple and separate genetic control, and one set of distinct parameters can describe the phenotype of a genotype (Tardieu 2003). Achieving this may take several iterations between model parameterization and genetic analysis. Such iterations may yield new parameters or new sets of parameters.

Improved model parameters are often accompanied by improvements in model algorithms, and so, the iterative process between parameter estimation and genetic analysis also involves changes in model structure. Moreover, in recent years, genetic engineering or synthetic biology approaches to modify the genetic composition of crop genotypes have increasingly been put on the research agenda as a complementary approach to conventional breeding in order to improve crops at a faster pace (Long et al. 2015). To accurately assess the impact of genetic modification on the molecular and biochemical processes that underlie the phenotype, it is required to incorporate the understandings of relevant molecular biology and biochemistry into a crop model framework. As complex crop phenotypes are the consequence of multiple biological component processes that interact, the process of incorporating molecular and biochemical knowledge may facilitate an improvement of model structures to better deal with the subtle differences among genotypes within breeding populations.

Integrating crop modelling and genetics in order to quantify gene-trait-crop performance relationships in support of plant breeding has previously been discussed (e.g., Yin et al. 2004; Hammer et al. 2005). Here, we present a framework involving crop modelling, genetics as well as biochemistry, called the MGB-triangle framework (Fig. 1), within which information flows and mutual benefits among the disciplines are emphasised. The rationale for, and present examples of, integration of crop modelling, genetics, and biochemical photosynthesis modelling will be outlined.

Role of genetics in crop modelling

Agronomic traits related to crop yield and resource use efficiency are complex in nature. Plant breeders, commonly relying on their own wit and experience plus some principles of quantitative genetics, identify subtle differences among genotypes exhibited in a genetic population in order to perform selection in moving the population mean towards the target phenotypes. Crop modellers perform simulation exercises to come up with suggestions (e.g., ideotypes) that they recommend breeders to use (Aggarwal et al. 1997). However, except for very few cases (e.g., Sinclair et al. 2016), direct experimental confirmation and objective comparisons of modelled suggestions with those already used in breeding programmes are rare. There are great concerns about this model-based approach from a geneticist's point of view (Stam 1998). Apart from the practical problem that the majority of input parameters in most crop models are difficult to accurately phenotype (Parent & Tardieu 2014), a more fundamental issue is that the genetic basis of the model-input parameters is largely unknown. In designing an ideotype by modelling, model parameter values are largely defined based on the range of observed (phenotypic) variation rather than on their genetic variation. According to quantitative genetics, even the genetic components of a trait are multi-fold (additive effects, dominance, non-additive effects or epistasis), and the expression of these components depends on environmental conditions (G×E) and probably also on developmental stages. If the genetics of the traits have not yet been revealed and one is still relying on phenotypic information for assessing the genetic component, only the additive effects can be fixed by inbred breeding. A modelling approach often ignores the possible existence of constraints arising simply from the fact that little genetic variation exists in the genetic material available for and amenable to selection. Moreover, it is assumed in ideotype modelling that multiple modelled traits can be combined at will in a single genotype. This ignores the potential correlations between the traits, arisen either from a tight linkage between loci or from a single locus that affects multiple traits (pleiotropy). To effectively assist the development of efficient breeding strategies, crop modelling should incorporate the genetic basis of model parameters determining crop productivity (Stam 1998).

White and Hoogenboom (1996) presented a model for bean (*Phaseolus vulgaris* L.), in which the genetic control of model parameters was considered. They applied linear regression to estimate values of more than 20 model-input traits from allelic information on seven candidate genes in the cultivars studied. A similar approach has been used by Messina et al. (2006) for soybean (*Glycine max* (L.) Merr.), by White et al. (2008), Brown et al. (2013) and Zheng et al. (2013) for wheat (*Triticum aestivum* L.), and by Boote et al. (2016) for common bean. Such an approach, however, assumes that all the traits were controlled by pleiotropic effects of the few candidate genes, ignoring the effects of possible additional trait-specific genes. Advances in quantitative genetics, by mapping trait-specific QTL using bi-parental mapping populations, can help to gain insight in the genetic basis of crop traits. Since the first modelling study incorporating QTL genetics using a RIL population (Yin et al. 2000), this QTL-based modelling is now becoming widely practised for a large number of traits across various species (see review of Yin et al. (2016) and references therein).

Genetic mapping approaches now go beyond bi-parental populations with broadening genetic background, for example, by creating the multi-parent advanced generation inter-cross lines using multiple genetically diverse genotypes (Huang et al. 2011). In addition, there is a growing number of studies pursuing GWAS (genome-wide association study) based on a high-density molecular marker set in a broad range of genotypes (e.g., McCouch et al. 2016; Kadam et al. 2017). This will provide unprecedented opportunities, not only to improve breeding efficiency via genomic selection based on genome-wide markers (Spindel et al. 2015) or even via direct genome editing (Bortesi & Fischer 2015), but also to practise QTL- or gene-based crop modelling (Dingkuhn et al. 2017a,b). However, most of the existing modelling studies, either for bi-parental populations or for GWAS panels were implemented for relatively simple crop traits (e.g., Quilot et al. 2005; Bogard et al. 2014; Uptmoor et al. 2017). Using a bi-parental introgression line (IL) population produced through repeated backcrossing with the recurrent parent, Gu et al. (2014b) reported on QTL-based modelling of crop biomass and grain yield in rice (*Oryza sativa* L.) under drought and well-watered conditions, using an upgraded crop model GECROS, which was first described by Yin & van Laar (2005) and designed for modelling G×E interactions. While modelling yield differences among individuals of a genetic population was still a challenge (as first shown by Yin et al. 2000), QTL were identified for each of seven input parameters of GECROS. The majority of these input parameters were very similar to the type of traits breeders usually score, like plant height, grain weight, time to flowering and time to maturity. The model-based dissection approach detected more markers/QTL than the analysis using only yield *per se* (Fig. 2). Every input-parameter correlation was mirrored by the co-localisation of QTL. Model-based sensitivity analysis ranked all markers for their importance in determining yield differences among the IL, and the marker ranking varied between drought and well-watered conditions. Simulation showed that virtual ideotypes based on markers identified by modelling had 19-36% more yield than the virtual ideotypes based on markers for yield *per se*. This suggests that crop modelling can be enhanced by using information for genetic basis of the model-input parameters, thereby yielding a powerful tool for marker design and ideotyping for improved crop yields under contrasting conditions.

Role of crop modelling in genetics

Many crop traits are intrinsically complex as a result from polygenic control, epistasis (interactions between genes), G×E, and age-dependent expression of genes and gene effects. Existing QTL analysis methods do not seem to have the power required to deal with these complexities. Support from other disciplines should be explored for improved QTL analysis.

Following are the areas where crop modelling can enhance genetics in analysing these complex traits.

Dissection of a complex trait into simpler components

For a complex crop trait, potentially numerous genes contribute to its genetic variation. However, the number of QTL for a trait detected with existing QTL-detection approaches is limited, because in mapping populations of moderate size many QTL effects are below the threshold of significance that filters for false positives. Instead of looking for QTL for a complex trait itself, determining QTL for its underlying simpler component traits might help to detect more of the relevant genetic factors. As just stated, Gu et al. (2014b) showed that the model-based dissection approach detected more markers/QTL than the analysis using only yield *per se* (Fig. 2). Similar results have been reported in many other studies (e.g., Prudent et al. 2011; Rebolledo et al. 2015).

Analysis of the genetic basis for trait correlation

Difficulties to manipulate crop traits also arise from the correlation between traits. Often it is difficult with existing genetic approaches to elucidate whether the correlation between the traits is due to the tight linkage of genes or to pleiotropy. However, some pleiotropic traits can be immediately inferred from physiological models. Crop models commonly relate to A_{max} (light saturated leaf photosynthesis) as a function of specific leaf nitrogen (amount of nitrogen in the leaf per unit leaf area), of which specific leaf area (SLA; area per unit of leaf dry matter) is a component; so, any correlation of these traits could be due to pleiotropy and increasing A_{max} could be at the cost of leaf expansion (Boote & Tollenaar 1994). Using crop modelling, Hammer et al. (2016) demonstrated the pleiotropic effect of stay-green and dwarfing in sorghum (*Sorghum bicolor* L.). Taller genotypes required more nitrogen for structural stem tissue, leaving less available for leaves, which was more rapidly diminished by translocation to grain during grain-filling. Hence, the “stay-green” trait was expressed in the shorter genotypes as a result of genetic differences in plant height. In other cases, the trait correlations may be caused by the linkage of genes. For example, given the correlation between the two different traits early vigour and drought tolerance, Luquet et al. (2016) used a crop model to explore the theoretical margins for improving both early vigour and drought tolerance and suggested ideotypes without much trade-off between vigour and tolerance.

Assisting to analyse age-dependent traits

Many quantitative crop traits (e.g., plant height, SLA, tiller number, mass weight) have a dynamic behaviour, and their value at the end of growth is the consequence of the actions and interactions of many genes expressed during the entire ontogeny. Statistical approaches have been developed to detect QTL for such a dynamic process (e.g. Ma et al. 2002), assuming that growth follows a general sigmoidal curve. However, not each dynamic trait follows a sigmoidal trend. Most QTL studies on time-related traits do not rely on a model; instead, phenotypic data measured at sequential times are directly analysed either individually or jointly. One tricky aspect for these analyses is the difference in physiological stage among individuals of a mapping population at the particular time that phenotyping is performed. For instance, when flowering time varies within a population, a particular phenotyping day may correspond to vegetative stage for one individual while others may already be in the reproductive stage. It would be preferred to measure at the same physiological stage for all individuals. This is practically impossible if phenotyping needs to be done at stages that are not marked as clearly by morphological changes as spike initiation or flowering. Use of a crop development model to synchronize the measured data for the same phenological stage is a highly useful alternative for such analyses. This was shown in the study of Yin et al. (1999)

for SLA of barley (*Hordeum vulgare* L.). When the SLA of each measurement time was directly subjected to QTL analysis, a major dwarfing mutation, the *denso* gene segregating in their mapping population (mapped on chromosome 3) was found to affect SLA strongly at all six measurement times. When the SLA of the different RIL was re-scaled for differences in phenological stage at the time of measurement using a development model, the effect of the *denso* gene was no longer significant during the pre-flowering stages. The effect of the *denso* gene on the SLA in the vegetative stage was therefore the result of its direct pleiotropic effect on the pre-flowering duration, which was confirmed by the detection of a flowering time QTL at the same genetic location. Other minor QTL detected also differed between the two approaches. This analysis underlines the importance of applying crop modelling in QTL analysis of this type of time-dependent traits.

Modelling of genotype-by-environment interactions

In the classical G×E models in quantitative genetics, the mean phenotype value of genotypes in each environment is used as a measure of environmental quality. Similar consideration has been made to detect QTL×E using multiple environment data. However, this limits the capability of predicting phenotypes across other independent environments, since these early statistical genetic models lack explicit physical measures of the environment (such as temperature) as input.

To use the information of physical environments, statistical linear models have been proposed that allow G×E and QTL×E to be modelled directly as a function of environmental variables (van Eeuwijk et al. 2001). The success of this approach relies on whether the correct physical environmental factors are included and whether used values of these environmental factors match the relevant growth periods. Obviously, the correct choice of physical environmental factors and their values requires in-depth knowledge of crop physiology for the traits under study. Even when the choice is made correctly, the power of the factorial regression can be limited because the phenotype of complex crop traits, as stated earlier, is achieved through interactive and ontogenetic responses of multiple underlying processes to multiple environmental variables having temporal dynamics and spatial profiles.

Mangin et al. (2017) showed that crop models can help to develop “stress indicators” that explain yield variation across multiple environments, thereby assisting the identification of relevant QTL for yield in response to environmental stresses. However, a more direct use of crop models is to let G×E be predicted in any new environment. The concept for this physiological G×E model approach has been illustrated for a number of simple traits in many species, under diverse environmental scenarios. For example, Reymond et al. (2003) used a simple model for maize (*Zea mays* L.) leaf elongation rate, expressed as mm $(°C.d)^{-1}$ based on meristem temperature, which was established as a linear function of both meristem-air water vapour pressure deficit and soil water potential. A QTL analysis was performed on both the intercept and slopes of this linear model, assuming both additive and epistatic QTL actions. Most QTL were specific for one parameter only. Each parameter was then computed from the sum of its QTL effects, resulting in QTL-based model parameters. The predictions of leaf elongation rates using original parameter values were comparable with those using QTL-based parameter values. For 11 RILs and the two parental lines that were not included for QTL analysis, the QTL-based model accounted for 74% of phenotypic variability of leaf elongation rates. In the context of genomic selection, Onogi et al. (2016) even showed that compared with a pure statistical “genomic prediction” model, the approach that integrates genomic prediction with an ecophysiological model gave more accurate predictions across multiple environments.

Modelling of epistasis

The value of crop models lies also in manifesting epistatic effects of component trait QTL on yield, as epistasis is often found in phenotypes that are the result of nonlinear interactions among multiple component processes integrated over ontogenetic stages. Through model simulation, Chapman et al. (2003) showed that complex epistatic effects for yield were generated even though individual genes had been defined as simple additive effects on component model-input traits. Whether this reflects true epistasis needs to be confirmed in a more specific study, but the simulation does indicate the potential of using crop models to predict epistasis on a physiological basis.

Recently, Technow et al. (2015) reported on the power of using crop modelling in enhancing whole-genome prediction in plant breeding. Purely statistical methods have been shown to deliver good predictions in the most common settings, such as prediction of across-environment performance for traits with additive gene effects. However, prediction of traits with non-additive (epistatic) effects continues to be challenging. Crop models can potentially explain certain types of non-additive gene effects, in addition to the impact of G×E, on the expressed phenotype. Combined with approximate Bayesian computation, which allows the incorporation of crop models directly into the estimation of whole genome marker effects, Technow et al. (2015) provided a proof of concept study and demonstrated that this novel approach can be considerably more accurate than the statistical method in predicting performance in environments for traits determined by non-additive gene effects. This provides additional evidence that incorporating crop models into whole-genome prediction is a very promising approach to improving prediction accuracy for some of the most challenging scenarios in plant breeding and genetics. However, other approaches can potentially take this a step further: direct resolution of epistatic effects of individual genes was shown to be possible via modelling of the phenotype, such as *Arabidopsis* flowering time, based on inclusion of molecular regulatory networks (e.g., Welch et al. 2003; Chew et al. 2014).

Role of biochemistry in crop modelling

To effectively address some recalcitrant genetic questions such as the effects of epistasis and G×E for complex traits at the crop level, crop models have to be biologically robust. Our experience over the last >15 years in QTL-based crop modelling suggests that crop models built upon traditional agronomic and crop physiological concepts can hardly resolve such subtle differences among genotypes (Yin et al. 2000). There is a common awareness that current models need to implement more mechanistic algorithms (Boote et al. 2013; Hammer et al. 2010), given that crop modelling as a discipline increasingly faces questions that can be solved only in combination with fundamental plant (molecular) biology.

However, opinions differ with regard to the extent of model details that should be brought to crop models. Some tend to stress the need to bring together pieces of biological details to form crop or ecosystem models (e.g. Zhu et al. 2016), following the bottom-up approach for systems biology models that was earlier advocated (Minorsky 2003). Others emphasised a top-down approach, in which a meta-mechanism is designed to exhibit reliable predictive skill at the crop level while also introducing sufficient rigour for complex phenotypic responses to become emergent properties of model dynamics (Hammer et al. 2004; 2016). Chew et al. (2017) proposed an intermediate approach on the basis that both crop science and systems biology have some elements of the other and that they have an overlapping goal in assisting crop improvement. In our view, crop model development itself is an evolving process. On the one hand, we need to make sure that the model is capable of predicting the crop-level traits reliably; on the other hand, models could be improved one step at a time, with the advancement of our understanding of individual processes and how these interact. To that end, use of well-established knowledge in biochemistry is a first step towards crop model improvement. This may not necessarily involve the modelling of biochemical processes *per* *se*; it is more important to incorporate the essential results of biochemical analysis or the end equations of biochemical modelling into crop models.

The idea to use biochemistry in this way for crop modelling is not entirely new. The outcome of the exhaustive work of Penning de Vries et al. (1974) on the costs of biochemical pathways for biosynthesis has long been used as growth respiration coefficients in crop models in which canopy photosynthesis and crop respiration are explicitly modelled. So, it is a bit surprising that the other basic process, photosynthesis, is still modelled in the form of empirical light response curves in most of these crop models. Other crop models bypass both photosynthesis and respiration, and use the coarse-grained concept of radiation use efficiency (RUE), for simulation of biomass production from intercepted light. However, short-term RUEs are hard to measure accurately, and often an overall RUE is determined as the linear slope of plotting biomass versus cumulative light interception in time. This approach ignores the expected subtle variations of RUE with development stage and radiation intensity.

A biochemical model of C_3 photosynthesis was well-described and derived for its details (Farquhar, von Caemmerer & Berry 1980, the FvCB model hereafter). Because its end equations are few and simple, the model has been widely used, from analysing leaf biochemistry to predicting the impact of global change on ecosystem functioning. Crop modellers sometimes argue that the FvCB model has a large number of input parameters to estimate, and estimating these parameters for various species is challenging and time consuming. However, many of these parameters can in most cases be assumed to be conserved among crop species. Yin & Struik (2017a) recently summarised representative parameter values of the FvCB model for C_3 species and its equivalent model for C_4 species coupled with intra-leaf CO_2 diffusion models.

The advantages of the FvCB-type models, compared with empirical light response curve models, are multi-fold. First, the FvCB model can correctly predict the interaction of multiple environmental variables (like ambient CO_2 concentration and temperature) with (sub-) photosynthetic parameters (Yin & Struik 2010), which is the basis for correctly modelling the impact of the interaction of these variables on more general parameters like RUE. Secondly, various model versions exist, and therefore this type of model is flexible, depending on the modelling objectives of users. For example, the simplest version is not to consider the details for CO_2 diffusion into the carboxylation sites of Rubisco (e.g. van Oijen et al. 2004) and this was successfully used in the context of developing a simple framework to analyse constraints to potential solar energy conversion efficiency in annual crops (Yin & Struik 2015). Thirdly, once the FvCB-type biochemical models are incorporated, application of crop models can be broadened to assess the impact of altered biochemical targets via genetic engineering, which is probably the only way to supercharge crop yields to a level of 50% or higher that can hardly be achieved by conventional breeding (Mitchell & Sheehy 2006; Long et al. 2015).

Using the new version of the GECROS crop model in which the FvCB model and its C_4 equivalent coupled with intra-leaf CO_2 diffusion models are included, we recently conducted a comprehensive analysis of the potential of genetic engineering towards improved leaf photosynthesis for increasing C_3-crop productivity (Yin & Struik 2017a). Of the nine engineering routes of enhancing photosynthesis, comprised of exploiting mesophyll conductance, Rubisco specificity, C_4 mechanisms, and cyanobacterial CO_2-concentrating mechanisms (CCM), only the complete mechanism that combines improved CCM, improved photosynthetic capacity and improved quantum efficiency, brings an advantage of $\geq$50% under any environmental conditions (Table 1). The need for the combination of routes and making a combination that works in concert means that supercharging crop productivity via engineering may be a long shot. Additional analysis showed that manipulating photosynthesis may result in unwanted secondary effects on other traits at the crop level, e.g., inducing faster senescence if nutrient uptake is not increased (Yin & Struik 2017a). These results have important implications for crop models in playing a translational role from photosynthesis biology to crop science. On the one hand, there is a need for simple yet robust biochemical model algorithms that are amenable for linking up with crop models. On the other hand, there should be an overarching framework that enables a reliable operation of modules to generate emergent properties of whole-crop dynamics. The previous predictions by photosynthesis physiologists did not use a crop model as such an overarching framework, and may have been overestimating the beneficial effect of the routes to engineer for high photosynthesis on crop productivity (see Discussion of Yin & Struik 2017a).

Role of crop modelling in biochemistry

The simulation study of Yin & Struik (2017a) has already indirectly shown the value of crop modelling for correctly assessing the potential impact of altered photosynthetic biochemical parameters or pathways on crop yield. There are more examples for the impact of crop modelling approaches on addressing questions in biochemistry. One of the crop modelling approaches over the decades is to develop generic models that could be applied to various species or environments (e.g. Spitters et al. 1989; Wang et al. 2002; Yin & van Laar 2005). Inspired by the generic modelling philosophy, Yin et al. (2004) extended the FvCB model with the generalised algorithms describing electron transport limited photosynthetic rates.

The original FvCB model assumes that 100% of electrons transported along the whole chain (the linear electron transport (LET)) are used in support of CO_2 fixation and photorespiration, thereby ignoring possible alternative electron transport routes. The model has two forms of equations for electron transport limited rate of CO_2 assimilation depending on whether it is NADPH or ATP supply that causes electron transport to limit CO_2 assimilation. These two forms were used, randomly by large, in many applications without knowing the underlying assumption of the model forms. Yin et al. (2004) developed analytical algorithms to account for cyclic electron transport around Photosystem I (CET) and pseudocyclic electron transport (PET), the two alternative routes of electron transport that may act in concert with LET to permit flexibility in the ratio of NADPH and ATP synthesis to meet the variable demands of carbon assimilation and photorespiration. These two widely used forms of the original FvCB model represent the most and least efficient electron transport stoichiometry, respectively, among the forms covered by the extended model. In addition, this generalized model accounts for the difference in electron transport efficiency between two photosystems and integrates other basic elements of steady-state photosynthesis.

The model reveals that even within the electron transport-limited range the relationship between quantum yields of CO_2 assimilation and Photosystem II photochemical efficiency is linear only if the latter varies in proportion with Photosystem I photochemical efficiency. More importantly, the generalised model can be used to assess any occurrence of alternative electron transport and to answer ‘what-if’ questions with respect to uncertain or unmeasured parameters, based on combined gas exchange and biophysical measurements (e.g., chlorophyll fluorescence). As long as current biophysical measurements are accurate, the analysis using the generalised model (Yin et al. 2006) supports the possible *in vivo* occurrence of CET and basal PET even under limiting irradiance (about 10% of the total electron flux for each). Sensitivity analysis, which is a common exercise in crop modelling, showed that CET, rather than PET, is a major ‘brake’ for LET to accommodate the balance between quantum efficiencies of electron transport and of CO_2 assimilation (Fig. 3). These results, while still subject to testing because of the measurement uncertainties with current equipment, are important because unlike LET, CET is cyclic in nature and therefore, cannot be directly detected experimentally. The generalised model therefore is a tool to indirectly infer the stoichiometries, bioenergetics and regulation of photosynthesis under different environmental conditions, based on easily implemented measurements.

The generalised model was found later to be applicable to C_4 photosynthesis, with only two additional parameters required (Yin & Struik 2012). An analysis using the model showed that compared with C_3 photosynthesis, the most striking stoichiometry in the malic enzyme C_4 species (e.g., C_4 crops like sorghum and maize) is its higher fraction for CET, required for the operation of CCM in C_4 photosynthesis. Due to the higher CET, C_4 leaves might have a lower energy loss via dissipation as heat compared with C_3 leaves (Yin & Struik 2015). This contributes to the more linear light response in C_4 than in C_3 photosynthesis, which has a strong implication on canopy photosynthesis. Because of this difference in the curvature of the light response, the canopy-to-leaf photosynthesis ratio, the benefit from the optimum leaf-nitrogen profile in the canopy and the productivity gain from improvement in canopy architecture were shown to be higher in C_3 than in C_4 species (Yin & Struik 2015).

These new insights at both leaf and canopy level all come from the generalised modelling of electron transport limited photosynthesis, and are impossible to obtain with the original FvCB model. A similar gain in insight was shown for mesophyll conductance in leaves (Yin & Struik 2017b). We expect that in future there will be more examples of new biochemical understanding obtained by using crop modelling concepts.

Genetic mapping of biochemical parameters and making the MGB framework work

As stated earlier, many biochemical parameters are most likely conserved among plant species and even more among genotypes within a crop species. However, there is evidence that some biochemical parameters of the FvCB model, V_{cmax} (Rubisco carboxylation capacity) and J_{max} (light saturated capacity of LET), vary among cultivars in wheat (Driever et al. 2014) and soybean (Koester et al. 2016). Similarly, there has been a report of variation in mesophyll conductance (g_{m}) among genotypes in wheat (Jahan et al. 2014; Barbour et al. 2016) and in rice (Adachi et al. 2013). Questions arising then are whether it is possible to genetically map these biochemical parameters, and, if so, whether biochemical models could be explored to assist genetic design for improved photosynthesis and related traits.

A practical problem is that these biochemical parameters cannot be estimated from any high-throughput phenotyping, but only from intensive measurements on CO_2- and light-response curves of leaf photosynthesis and electron transport efficiency; so these parameters cannot be phenotyped for a large number of genotypes. To that end, it is preferable to use a population of ILs that differ in a relatively few number of loci, despite some drawbacks of ILs. Gu et al. (2012b) reported a study using rice IL genotypes, in which combined gas exchange and chlorophyll fluorescence data were collected for entire CO_2- and light-response curves of leaf photosynthesis (A), with which biochemical and physiological parameters of a combined conductance-FvCB biochemical photosynthesis model were estimated. Because measuring entire response curves is time consuming, 13 lines (including the two parents) were carefully selected as representatives of the population, based on the QTL for A_{max} earlier reported by Gu et al. (2012a). The curves were assessed at two stages (flowering and grain filling) for plants grown under moderate drought and well-watered conditions (Gu et al. 2012b). Using these curves, photosynthesis was then quantitatively dissected into six different component traits: stomatal conductance (g_{s}), mesophyll conductance (g_{m}), biochemical capacity parameters (V_{cmax}, J_{max}), electron-converting efficiency under limiting light (κ_{2LL}), and shape factor (θ) for the hyperbolic minimum of light limited and saturated electron transport. Note that “component traits” are a relative term as some of these traits can be further dissected if more measurements are available. Although the effects of development stage and water supply on photosynthesis were predominant, significant genetic variation in the six parameters was found. Genomic regions linked to the variation of these biochemical parameters were identified. Genetic variation in A_{max} and TE (transpiration efficiency) was mainly caused by variation in g_{s} and g_{m}, which suggests more efforts should be focused on g_{s}

and g_{m} in rice breeding programmes for improving photosynthesis and TE. Gu et al. (2012b) thus showed that relationships between these photosynthetic parameters and leaf nitrogen or dry matter per unit area, which were previously found across environmental treatments, were also valid for variation across genotypes.

Gu et al. (2012b) next used the biochemical model to evaluate the potential of utilizing the genetic variation in the six parameters, and to explore the genetic design of ideotypes for improved leaf-level photosynthesis (A) and transpiration efficiency (TE), by combining alleles positively influencing different components of photosynthesis. Model calculations showed that these ideotypes can potentially improve photosynthesis and TE significantly, compared with the best genotype of the 13 lines investigated. It was also shown that if the correlation between g_{m} and g_{s} was not due to pleiotropy but due to a genetic linkage that could be broken, both photosynthesis and TE could be improved simultaneously, despite the common negative correlation between A and TE.

Next, Gu et al. (2014a) examined the extent to which natural genetic variation in biochemical parameters can contribute to increasing rice productivity. Using the crop model GECROS, they analysed the impact of genetic variation in A on crop biomass production, based on the QTL for various biochemical photosynthetic components within their rice IL population. Such an analysis best showcased the MGB framework (Fig. 1) as a whole. The genetic variation in A of 25%, created by 20-50% variation in biochemical parameters, can be scaled up almost equally to crop level, resulting in a potential increase in biomass of 28-29% across different locations and years (Table 2). Using the simulation results, multiple regression analysis showed that the genetic variation in $\kappa_{2\mathrm{LL}}$ contributed most to the variation in total biomass, followed by g_{s} and g_{m}. This was probably because the genetic variation in A of this rice population resulted not only from Rubisco-limited photosynthesis but also from electron transport-limited photosynthesis; as a result, photosynthetic rates could be improved more from increasing both light-saturated and light-limited leaves in the canopy. This study demonstrates the potential of improving rice productivity by mining the natural variation in existing germplasm, especially the variation in parameters determining light-limited photosynthesis. The genetic variation of these parameters still needs more experimentation to investigate as photosynthesis under light limitation is generally conservative. Nevertheless, this result echoes the analysis of Yin & Struik (2017a) of the simulated impact of genetic engineering routes on the importance of improving quantum efficiency in addition to the usually emphasised CCM and photosynthetic capacity.

Concluding remarks

Productivity of major crops has to increase at a greater pace in the coming decades than before, in order to ensure food and energy security for a growing and increasingly demanding population while guarding against the negative impact of increasing threats under global climate change (Fischer et al. 2014). Crop model based simulation studies can provide useful information for breeders to better design their breeding strategies and to improve their selection efficiency. For that, crop models should be robust, being able of handling the subtle gene-trait-crop performance relationships. In this paper we propose a MGB triangle framework to improve both model structure and parameters. The experience so far in practising this framework reveals that:

- Crop models may be structured to allow that model parameters can be easily estimated from the type of traits breeders usually score, and future model design should recognise the potential boost that high-throughput phenotyping techniques can give for estimating model parameters.

- Such parameters are amenable to QTL analysis, allowing the genetic effects to be incorporated into crop models to overcome the limitations of model-based ideotyping exercises that ignore the genetic basis of these parameters.
- Bringing biochemistry to crop models not only creates opportunities to upgrade model structure but also allows identifying genetic engineering targets and optimising engineering routes for improved crops.
- While bringing in biochemical modules, they should function within an overarching model framework or meta-mechanism in the crop model that enables a reliable operation of these modules to generate emergent properties of whole-crop dynamics; *in silico* biochemical modules, when acted alone, can over-estimate the beneficial effect of engineering routes on increasing crop productivity.
- Genetic engineering of a single biochemical route likely alters one photosynthetic parameter, whereas natural variation of photosynthesis can arise from multiple parameters, e.g., either from A_{max} or from light use efficiency. Combined exploitation of both A_{max} and light use efficiency parameters with CCM is more effective than exploring a single route for improving crop yields.
- Principles or philosophy of crop modelling can enhance genetic analysis and elucidation of biochemical mechanisms.

The above experiences suggest that the interdisciplinary MGB framework generates new insights into the control of complex crop traits that can otherwise be hard to obtain if M, G, and B are practised independently. Previously, “Crop Systems Biology” was proposed to integrate applied crop science and a broader range of fundamental plant biology (genetics, biochemistry, genomics, and molecular biology) for achieving the cross-fertilisation among relevant disciplines (Yin & Struik 2007; 2008; 2010; 2016), and this has been identified as one of the three cornerstone scope domains of the European Society of Agronomy (http://www.european-agronomy.org/scope.html). The MGB triangle approach outlined here may serve as a first step for achieving the ultimate goal of “Crop Systems Biology” to better assist crop improvement, given the already shown mutual benefits between crop modelling and genetics and biochemistry.

References

Adachi, S., Nakae, T., Uchida, M., Soda, K., Takai, T., Oi, T., Yamamoto, T., Ookawa, T., Miyake, H., Yano, M. and Hirasawa, T., 2013. The mesophyll anatomy enhancing CO_2 diffusion is a key trait for improving rice photosynthesis. Journal of Experimental Botany 64: 1061-1072.

Aggarwal, P.K., Kropff, M.J., Cassman, K.G. and ten Berge, H.F.M., 1997. Simulating genotypic strategies for increasing rice yield potential in irrigated, tropical environments. Field Crops Research 51: 5-17.

Barbour, M., Bachmann, S., Bansal, U., Bariana, H. and Sharp, P., 2016. Genetic control of mesophyll conductance in common wheat. New Phytologist 209: 461-465.

Bogard, M., Ravel, C., Paux, E., Bordes, J., Balfourier, F., Chapman, S.C., Le Gouis, J., Allard, V., 2014. Predictions of heading date in bread wheat (*Triticum aestivum* L.) using QTL-based parameters of an ecophysiological model. Journal Experimental Botany 65: 5849-5865.

Boote, K.J., and M. Tollenaar. 1994. Modeling genetic yield potential. p. 533–565. In: K.J. Boote et al. (editors) Physiology and determination of crop yield. ASA, CSSA, and SSSA, Madison, WI, USA.

Boote, K.J., Jones, J.W., White, J.W., Asseng, S. and Lizaso, J.I., 2013. Putting mechanisms into crop production models. Plant, Cell and Environment 36: 1658-1672.

Boote, K.J., Vallejos, C.E., Jones, J.W. and Correll, M.J., 2016. Crop modeling approaches for predicting phenotype of grain legumes with linkage to genetic information. p. 163-198. In: X. Yin and P.C. Struik (editors) Crop Systems Biology: Narrowing the Gaps between Crop Modeling and Genetics. Springer, Switzerland.

Bortesi, L. and Fischer, R., 2015. The CRISPR/Cas9 system for plant genome editing and beyond. Biotechnology Advances 33: 41-52.

Brown, H.E., Jamieson, P.D., Brooking, I.R., Moot, D.J. and Huth N.I., 2013. Integration of molecular and physiological models to explain time of anthesis in wheat. Annals of Botany 112: 1683-1703.

Chapman, S.C., Cooper, M., Podlich, D. and Hammer, G.L., 2003. Evaluating plant breeding strategies by simulating gene action and dryland environment effects. Agronomy Journal 95: 99-113.

Chew, Y.H., Seaton, D.D. and Millar, A.J., 2017. Multi-scale modelling to synergise plant systems biology and crop science. Field Crops Research 202: 77-83.

Chew, Y.H., Wenden, B., Flis, A., Mengin, V., Taylor, J., Davey, C.L., Tindal, C., Thomas, H., Ougham, H.J., de Reffye, P., Stitt, M., Williams, M., Muetzelfeldt, R., Halliday, K.J. and Millar, A.J., 2014. Multiscale digital *Arabidopsis* predicts individual organ and whole-organism growth. Proceedings of the National Academy of Sciences of the United States of America (PLUS) 111: E4127-E4136.

de Wit, C.T. 1959. Potential photosynthesis of crop surfaces. Netherlands Journal of Agricultural Sciences 7: 141-149.

de Wit, C.T. 1965. Photosynthesis of leaf canopy. Agricultural Research Report 663. Wageningen, The Netherlands.

Dingkuhn, M., Laza, M.R.C., Kumar, U., Mendez, K.S., Collard, B., Jagadish, K., Singh, R.K., Padolina, T., Malabayabas, M., Torres, E., Rebolledo, M.C., Manneh, B. and Sow, A., 2015. Improving yield potential of tropical rice: Achieved levels and perspectives through improved ideotypes. Field Crops Research 182: 43-59.

Dingkuhn, M., Pasco, R., Pasuquin, J., Damo, J., Soulie, J.-C., Raboin, L.-M., Dusserre, J., Sow, A., Manneh, B., Shrestha, S., Balde, A. and Kretzschmar, T., 2017a. Crop-model assisted phenomics and genome-wide association study for climate adaptation of indica rice. 1. Phenology. Journal of Experimental Botany 68: 4369-4388.

Dingkuhn, M., Pasco, R., Pasuquin, J.M., Damo, J., Soulie, J.-C., Raboin, L.-M., Dusserre, J., Sow, A., Manneh, B., Shrestha, S., and Kretzschmar, T., 2017b. Crop-model assisted phenomics and genome-wide association study for climate adaptation of indica rice. 2. Thermal stress and spikelet sterility. Journal of Experimental Botany 68: 4389-4406.

Driever, S.M., Lawson, T., Andralojc, P.J., Raines, C.A. and Parry, M.A.J., 2014. Natural variation in photosynthetic capacity, growth, and yield in 64 field-grown wheat genotypes. Journal of Experimental Botany 65: 4959-4973.

Duncan, W.G., Loomis, R.S., Williams, W.A. and Hanau, R. 1967. A model for simulating photosynthesis in plant communities. Hilgardia 4: 181-205.

Farquhar, G.D., von Caemmerer, S. and Berry, J.A., 1980. A biochemical model of photosynthetic CO_2 assimilation in leaves of C_3 species. Planta 149: 78-90.

Fischer, R.A., Byerlee, D. and Edmeades, G.O., 2014. Crop yields and global food security: will yield increase continue to feed the world? ACIAR Monograph No. 158. Australian Centre for International Agricultural Research: Canberra. xxii + 634 pp.

Gu, J., Yin, X., Struik, P.C., Stomph, T.J. and Wang, H., 2012a. Using chromosome introgression lines to map quantitative trait loci for photosynthesis parameters in rice (*Oryza sativa* L.) leaves under drought and well-watered field conditions. Journal of Experimental Botany 63: 455-469.

Gu, J., Yin, X., Stomph, T.J., Wang, H. and Struik, P.C., 2012b. Physiological basis of genetic variation in leaf photosynthesis among rice (*Oryza sativa* L.) introgression lines under drought and well-watered conditions. Journal of Experimental Botany 63: 5137-5153.

Gu, J., Yin, X., Stomph, T.J., and Struik, P.C., 2014a. Can exploiting natural genetic variation in leaf photosynthesis contribute to increasing rice productivity? A simulation analysis. Plant, Cell and Environment 37: 22-34.

Gu, J., Yin, X., Zhang, C., Wang, H. and Struik, P.C., 2014b. Linking ecophysiological modelling with quantitative genetics to support marker-assisted crop design for improved rice (*Oryza sativa* L.) yields under drought stress. Annals of Botany 114: 499-511.

Hammer, G., Cooper, M. Tardieu, F., Welch, S., Walsh, B., van Eeuwijk, F., Chapman, S. and Podlich, D., 2006. Models for navigating biological complexity in breeding improved crop plants. Trends in Plant Science 11: 587-593.

Hammer, G.L., Sinclair, T.R., Chapman, S.C. and van Oosterom, E., 2004. On systems thinking, systems biology, and the in silico plants. Plant Physiology 134: 909-911.

Hammer, G., Messina, C., van Oosterom, E., Chapman, S., Singh, V., Borrell, A., Jordan, D. and Cooper, M., 2016. Molecular breeding for complex adaptive traits: how integrating crop ecophysiology and modelling can enhance efficiency. p. 147-162. In: X. Yin and P.C. Struik (editors) Crop Systems Biology: Narrowing the Gaps between Crop Modeling and Genetics. Springer, Switzerland.

Hammer, G.L., van Oosterom, E., McLean, G., Chapman, S., Broad, I., Harland, P. and Muchow, R.C., 2010. Adapting APSIM to model the physiology and genetics of complex adaptive traits in field crops. Journal of Experimental Botany 61: 2185-2202.

Huang, X., Paulo, M.J., Boer, M, Effgen, S., Keizer, P., Koornneef, M. and van Eeuwijk, F.A., 2011. Analysis of natural allelic variation in Arabidopsis using a multiparent recombinant inbred line population. Proceedings of the National Academy of Sciences of USA 108: 4488-4493.

Jahan, E., Amthor, J.S., Farquhar, G.D., Trethowan, R. and Barbour, M.M., 2014. Variation in mesophyll conductance among Australian wheat genotypes. Functional Plant Biology 41: 568-580.

Kadam, N., Tamilselvan, A., Lawas, L.M.F., Quinones, C., Bahuguna, R.N., Thomson, M.J., Dingkuhn, M., Muthurajan, R., Struik, P.C., Yin, X. and Jagadish, K.S.V., 2017. Genetic control of plasticity in root morphology and anatomy of rice in response to water-deficit. Plant Physiology174: 2302-2315.

Koester, R.P., Nohl, B.M., Diers, B.W. and Ainsworth, E.A., 2016. Has photosynthetic capacity increased with 80 years of soybean breeding? An examination of historical soybean cultivars. Plant, Cell and Environment 39: 1058-1067.

Li, T., Hasegawa, T., Yin, X., Zhu, Y., Boote, K., Adam, M., Bregaglio, S., Buis, S., Confalonieri, R., Fumoto, T, Gaydon, D., Marcaida III, M., Nakagawa, H., Oriol, P., Ruane, A.C., Ruget, F., Singh, B., Singh, U., Tang, L., Tao, F., Wilkens, P., Yoshida, H., Zhang, Z. and Bouman, B., 2015. Uncertainties in predicting rice yield by current crop models under a wide range of climatic conditions. Global Change Biology 21: 1328-1341.

Long, S.P., Marshall-Colon, A. and Zhu, X.-G., 2015. Meeting the global food demand of the future by engineering crop photosynthesis and yield potential. Cell 161: 56-66.

Loomis, R.S., Rabbinge, R. and Ng, E., 1979. Explanatory models in crop physiology. Annual Review of Plant Physiology 30: 339-367.

Luquet, D., Rebolledo, C., Rouan, L., Soulie, J.-C. and Dingkuhn, M., 2016. Heuristic exploration of theoretical margins for improving adaptation of rice through crop-assisted phenotyping. p.105-127. In: X. Yin and P.C. Struik (editors) Crop Systems Biology: Narrowing the Gaps between Crop Modeling and Genetics. Springer, Switzerland.

Ma, C.X., Casella, G. and Wu, R.L., 2002. Functional mapping of quantitative trait loci underlying the character process: A theoretical framework. Genetics 161: 1751-1762.

Mangin, B., Casadebaig, P., Cadic, B., Blanchet, N., Boniface, M.-C., Carrere, S., Gouzy, J., Legrand, L., Mayjonade, B., Pouilly, N., Andre, T., Coque, M., Piquemal, J., Laporte, M., Vincourt, P., Munos, S. and Langlade, N.B., 2017. Genetic control of plasticity of oil yield for combined abiotic stresses using a joint approach of crop modeling and genome-wide association. Plant, Cell and Environment 40: 2276-2291.

McCouch, S.R., Wright, M.H., Tung, C.-W., et al. 2016. Open access resources for genome-wide association mapping in rice. Nature Communications 7: 10532 (DOI: 10.1038/ncomms10532).

Messina, C.D., Jones, J.W., Boote, K.J. and Vallejos, C.E., 2006. A gene-based model to simulate soybean development and yield responses to environment. Crop Science 46: 456-466.

Minorsky, P.V. 2003. Achieving the in silico plant. Systems biology and the future of plant biological research. Plant Physiology 132: 404-409.

Mitchell, P.L. and Sheehy, J.E., 2006. Supercharging rice photosynthesis to increase yield. New Phytologist 171: 688-693.

Onogi, A., Watanabe, M., Mochizuki, T., Hayashi, T., Nakagawa, H., Hasegawa, T. and Iwata, H., 2016. Toward integration of genomic selection with crop modelling: the development of an integrated approach to predicting rice heading dates. Theoretical and Applied Genetics 129: 805-817.

Parent, B. and Tardieu, F., 2014. Can current crop models be used in the phenotyping era for predicting the genetic variability of yield of plants subjected to drought or high temperature? Journal of Experimental Botany 65: 6179-6189.

Penning de Vries, F.W.T. 1991. Improving yields: designing and testing VHYVs. p. 13-19. In: F.W.T. Penning de Vries et al. (editors), Systems simulations at IRRI. IRRI Research Paper 151, IRRI, Los Baños, Philippines.

Penning de Vries, F.W.T., Brunsting, A.H.M. and van Laar, H.H., 1974. Products, requirements and efficiency of biosynthesis: a quantitative approach. Journal of Theoretical Biology 45: 339-377.

Picheny, V., Casadebaig, P., Trepos, R., Faivre, R., Da Silva, D., Vincourt, P. and Costes, E., 2017. Using numerical plant models and phenotypic correlation space to design achievable ideotypes. Plant, Cell and Environment 40: 1926-1939.

Prudent, M., Lecomte, A., Bouchet, J.-P., Bertin, N., Causse, M. and Génard, M., 2011. Combining ecophysiological modelling and quantitative trait locus analysis to identify key elementary processes underlying tomato fruit sugar concentration. Journal of experimental Botany 62: 907-919.

Quilot, B., Kervella, J., Génard, M., and Lescourret, F., 2005. Analysing the genetic control of peach fruit quality through an ecophysiological model combined with a QTL approach. Journal of Experimental Botany 56: 3083-3092.

Rebolledo, M.C., Dingkuhn, M., Courtois, B., Gibon, Y., Clément-Vidal, A., Cruz, D.F., Duitama, J., Lorieux, M., Luquet, D., 2015. Phenotypic and genetic dissection of component traits for early vigour in rice using plant growth modelling, sugar content analyses and association mapping. Journal of Experimental Botany 66: 5555-5566.

Reymond, M., Muller, B., Leonardi, A., Charcosset, A. and Tardieu, F., 2003. 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 131: 664-675.

Shikanai, T., 2014. Central role of cyclic electron transport around photosystem I in the regulation of photosynthesis. Current Opinion in Biotechnology 26: 25-30.

Sinclair, T.R., Devi, J.M., Carter, T.E., Jr., 2016. Limited-transpiration trait for increased yield for water-limited soybean: From model to phenotype to genotype to cultivars. P. 129-1476. In: X. Yin and P.C. Struik (editors) Crop Systems Biology: Narrowing the Gaps between Crop Modeling and Genetics. Springer, Switzerland.

Spindel, J., Begun, H., Akdemir, D., Virk, P., Collard, B., Redona, E., Atlin, G., Jannink, J.-L. and McCouch, S.R., 2015. Genomic selection and association mapping in rice (*Oryza sativa*): Effect of trait genetic architecture, trait population composition, marker number and statistical model on accuracy of rice genomic selection in elite, tropical rice breeding lines. PLoS Genetics 11(2): e1004982. DOI:10.1371/journal.pgen.1004982

Spitters, C.J.T., van Keulen, H. and van Kraalingen, D.W.G., 1989. A simple and universal crop growth simulator: SUCROS87. P. 147-181. In: R. Rabbinge, S.A. Ward and H.H. van Laar (editors): Simulation and Systems Management in Crop Protection. Simulation Monographs 32: Pudoc, Wageningen, The Netherlands.

Stam, P. 1998. Crop physiology, QTL analysis and plant breeding. p. 429-440. In: H. Lambers et al. (editors) Inherent Variation in Plant Growth: Physiological Mechanisms and Ecological Consequences. Backhuys Publ., Leiden, Netherlands.

Tardieu, F., 2003. Virtual plants: modelling as a tool for the genomics of tolerance to water deficit. Trends in Plant Science 8: 9-14.

Technow, F., Messina, C.D., Radu Totir, L. and Cooper M., 2015. Integrating crop growth models with whole genome prediction through approximate Bayesian computation. PLoS One: doi:10.1371/journal.pone.0130855.

Uptmoor, R., Pillen, K. and Matschegewski, C., 2017. Combining genome-wide prediction and a phenology model to simulate heading date in spring barley. Field Crops Research 202: 84-93.

van Eeuwijk, F.A., Cooper, M., DeLacy, I.H., Ceccarelli, S. and Grando, S., 2001. Some vocabulary and grammar for the analysis of multi-environment trials, as applied to the analysis of FBB and PPB trials. Euphytica 122: 477-490.

van Oijen, M., Dreccer, M.F., Firsching, K.-H. and Schnieders, B.J., 2004. Simple equations for dynamic models of the effects of CO_2 and O_3 on light use efficiency and growth of crops. Ecological Modelling 179: 39-60.

Wang, E., Robertson, M.J., Hammer, G.L., Carberry, P.S., Holzworth, D., Meinke, H., Chapman, S.C., Hargreaves, J.N.G., Huth, N.I. and McLean, G., 2002. Development of a generic crop model temperate in the cropping system model APSIM. European Journal of Agronomy 18: 121-140.

Welch, S.M., Roe, J.L. and Dong, Z. 2003. Genetic neural network model of flowering time control in *Arabidopsis thaliana*. Agronomy Journal 95: 71-81.

White, J.W. and Hoogenboom, G., 1996. Simulating effects of genes for physiological traits in a process-oriented crop model. Agronomy Journal 88: 416-422.

White, J.W., Herndl, M., Hunt, L.A., Payne, T.S. and Hoogenboom, G., 2008. Simulation-based analysis of effects of *Vrn* and *Ppd* loci on flowering in wheat. Crop Science 48: 678-687.

Yin, X., Chasalow, S., Dourleijn, C.J., Stam, P. and Kropff, M.J., 2000. Coupling estimated effects of QTLs for physiological traits to a crop growth model: Predicting yield variation among recombinant inbred lines in barley. Heredity 85: 539-549.

Yin, X., Kropff, M.J. and Stam, P., 1999. The role of ecophysiological models in QTL analysis: the example of specific leaf area in barley. Heredity 82: 415-421.

Yin, X., Harbinson, J. and Struik, P.C., 2006. Mathematical review of literature to assess alternative electron transports and interphotosystem excitation partitioning of steady-state C_3 photosynthesis under limiting light. Plant, Cell and Environment 29: 1771-1782 (corrigendum in PCE 2006. 29: 2252).

Yin, X., Struik, P.C. and Kropff, M.J., 2004. Role of crop physiology in predicting gene-phenotype relations. Trends in Plant Science 9: 426-433

Yin, X. and Struik, P.C., 2007. Crop systems biology: An approach to connect functional genomics with crop modelling. p. 61-71. In: J.H.J. Spiertz, P.C. Struik and H.H. van Laar (editors) Scale and Complexity in Plant Systems Research: Gene-Plant-Crop relations. Springer. Printed in the Netherlands.

Yin, X. and Struik, P.C., 2008. Applying modelling experiences from the past to shape crop systems biology: The need to converge crop physiology and functional genomics. New Phytologist 179: 629-642.

Yin, X. and Struik, P.C., 2010. Modelling the crop: from system dynamics to systems biology. Journal of Experimental Botany 61: 2171-2183.

Yin, X. and Struik, P.C., 2012. Mathematical review of the energy transduction stoichiometries of C_4 leaf photosynthesis under limiting light. Plant, Cell and Environment 35: 1299-1312.

Yin, X. and Struik, P.C., 2015. Constraints to the potential efficiency of converting solar radiation into phytoenergy in annual crops: from leaf biochemistry to canopy physiology and crop ecology. Journal of Experimental Botany 66: 6535-6549.

Yin, X. and Struik, P.C. (editors), 2016. Crop Systems Biology: Narrowing the Gaps between Crop Modelling and Genetics. Springer International Publishing Switzerland, 233pp (ISBN 978-3-319-20561-8).

Yin, X. and Struik, P.C., 2017a. Can increased leaf photosynthesis be converted into higher crop mass production? A simulation study for rice using the crop model GECROS. Journal of Experimental Botany 68: 2345-2360.

Yin, X. and Struik, P.C., 2017b. Simple generalisation of a mesophyll resistance model for various intracellular arrangements of chloroplasts and mitochondria in C_3 leaves. Photosynthesis Research 132: 211-220.

Yin, X., Struik, P.C., Gu, J. and Wang, H., 2016. Modelling QTL-trait-crop relationships: Past experiences and future prospects. p. 193-218. In: X. Yin & P.C. Struik (editors) Crop Systems Biology: Narrowing the Gaps between Crop Modelling and Genetics. Springer International Publishing Switzerland.

Yin, X. and van Laar, H.H., 2005. Crop Systems Dynamics: An Ecophysiological Simulation Model for Genotype-by-Environment Interactions. Wageningen Academic Publishers, Wageningen, The Netherlands, 155pp (ISBN9076998558).

Yin, X., van Oijen, M. and Schapendonk, A.H.C.M., 2004. Extension of a biochemical model for the generalized stoichiometry of electron transport limited C_3 photosynthesis. Plant, Cell and Environment 27: 1211-1222.

Zheng, B., Biddulph, B., Li, D., Kuchel, H. and Chapman S., 2013. Quantification of the effects of *VRN1* and *Ppd-D1* to predict spring wheat (*Triticum aestivum*) heading time across diverse environment. Journal of Experimental Botany 64: 3747-3761.

Zhu, X.-G., Lynch, J.P., LeBauer, D.S., Millar, A.J., Stitt, M. and Long, S.P., 2016. Plants *in silico*: why, why now and what? – an integrative platform for plant systems biology research. Plant, Cell and Environment 39: 1049-1057.

Table 1 The percentage of increase (%) of the 31-year average aboveground biomass by nine photosynthesis-enhancing routes, relative to that for the default C_3-photosynthesis route, in rice crop simulated by the GECROS model for the present climate and the 2050 climate, either under potential or water-stress environments, in three representative sites (tropics: Los Baños in Philippines; subtropics: Nanjing in China; temperate: Shizukuishi in Japan). Based on Yin & Struik (2017a)

Site		Los Baños				Nanjing				Shizukuishi			
Production level		Potential		Water limited		Potential		Water limited		Potential		Water limited	
Climate§		Present	2050	Present	2050	Present	2050	Present	2050	Present	2050	Present	2050
Route†	1	4.3	2.5	4.8	3.1	4.2	2.6	4.5	4.1	4.3	2.7	4.5	4.1
	2	8.8	8.0	7.5	6.8	9.3	8.5	11.7	9.7	9.2	8.1	11.0	9.2
	3	12.9	9.9	13.6	12.5	14.0	10.8	16.8	13.8	13.5	10.2	15.5	14.0
	4	10.4	4.1	12.4	6.4	8.0	3.9	11.8	6.2	14.8	8.3	19.2	10.4
	5	7.6	-0.8	26.6	13.6	5.0	-2.4	24.5	11.6	7.0	-0.7	26.6	14.9
	6	38.0	23.1	51.2	33.8	33.0	21.9	50.5	34.1	39.8	25.4	54.5	36.0
	7	5.4	1.6	9.1	5.2	4.5	0.8	10.6	6.0	5.5	2.1	11.3	7.7
	8	17.9	10.5	39.7	28.7	18.1	10.7	39.9	27.9	19.1	11.3	38.7	28.1
	9	70.1	57.5	78.5	61.2	63.2	51.3	74.8	57.9	60.8	49.0	73.8	57.4

§ Present climate: based on weather data of 1980-2010 (the baseline), with the atmospheric $[CO_2]$ of 400 µmol mol^{-1}; 2050 climate: air temperature 2°C higher than the baseline weather, with the atmospheric $[CO_2]$ of 550 µmol mol^{-1}.

† Route numbers are defined as: 1 – improved mesophyll conductance; 2 – improved Rubisco specificity; 3 – combined routes 1 and 2; 4 – C_4 biochemistry engineered; 5 – C_4 anatomy for CO_2 concentrating mechanism (CCM) engineered; 6 – combined routes 4 and 5 (i.e. the complete C_4 mechanism); 7 – cyanobacterial bicarbonate transporters engineered; 8 – a more elaborate cyanobacterial CCM (based on 'carboxysome') added; 9 – a complete cyanobacterial mechanism engineered that combines the low ATP cost of cyanobacterial CCM and the high photosynthetic capacity per unit leaf nitrogen.

Table 2. The minimum, maximum, and population mean of the traits: biochemical FvCB-model parameters, leaf photosynthesis, canopy photosynthesis, and crop biomass yield simulated by using the GECROS model. Based on Gu et al. (2014a)

	Trait	Min	Max	Mean	Variation (%) [a]
Biochemical level	κ_{2LL} (mol mol^{-1})	0.27	0.37	0.32	29.8
	J_{max} (µmol m^{-2}s^{-1})	136.5	167.1	147.5	20.7
	θ	0.72	0.92	0.79	24.5
	V_{cmax} (µmol m^{-2}s^{-1})	109.6	152.7	124.8	34.6
	δ_m [b]	0.73	0.91	0.88	20.3
	δ_s [b]	0.81	1.36	1.11	49.5
Leaf level (µmol m^{-2}s^{-1})	A_{100} [c]	2.9	4.1	3.6	31.4
	A_{2000} [c]	19.3	25.0	22.4	25.6
Canopy level (g CO_2 m^{-2} d^{-1})	$A_{c,LAI=1}$ [d]	25.5	33.2	28.9	26.4
	$A_{c,LAI=5}$ [d]	78.7	101.7	89.2	25.8
Crop level (g m^{-2})	BY_{BJ} [e]	2092	2775	2436	28.0
	BY_{LB} [e]	2049	2748	2409	29.0

[a] Variation (%) is defined as ((max – min)/mean) × 100;

[b] δ_m and δ_s are intra-leaf CO_2 diffusion model parameters proportional to mesophyll conductance (g_m) and stomatal conductance (g_s), respectively;

[c] A_{100} and A_{2000} are leaf photosynthesis rates when incoming irradiance is 100 and 2000 µmol m^{-2} s^{-1}, respectively;

[d] $A_{c,LAI=1}$ and $A_{c,LAI=5}$ are daily canopy photosynthesis rates under high-light conditions when LAI is 1 and 5 m^2 m^{-2}, respectively;

[e] BY_{BJ} and BY_{LB} are biomass yield simulated for growing season in Beijing, China (temperate climate) and Los Baños, Philippines (tropical climate), respectively.

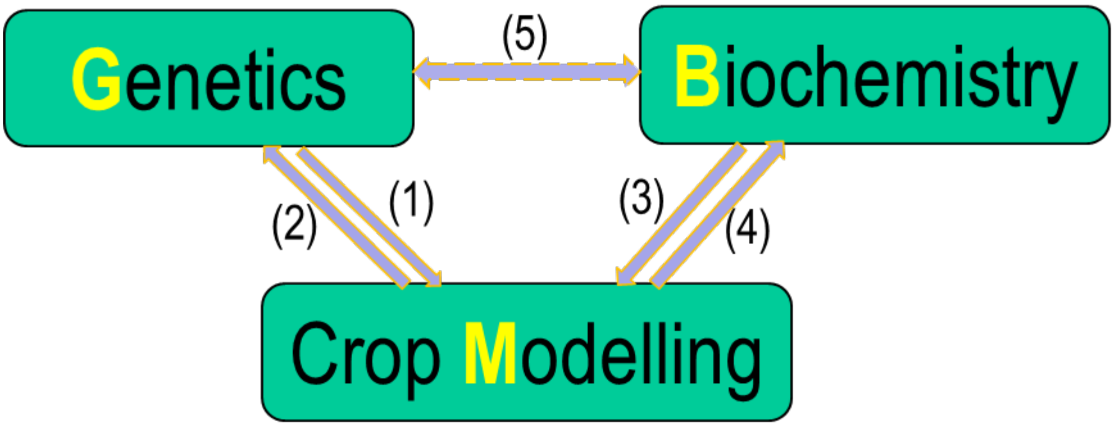

Figure 1 The MGB triangle framework linking crop modelling (M), quantitative genetics (G), and biochemistry (B). Single-headed arrows indicate roles of one discipline for another and the double headed arrow indicates the mutual role of the two disciplines involved. Specific aspects of these roles and their rationales, as indicated in numbers in the framework, include: (1) using approaches in genetics to elucidate genetic basis of model-input parameters, (2) using models to assist genetic analyses of complex traits and trait correlations, (3) introducing biochemical modules to make crop models more biologically rigorous, (4) employing crop modelling philosophy to extend biochemical models for enhanced biological insights, and (5) genetic mapping of biochemical parameters and possible use of biochemical or molecular regulatory network to unravel genetic epistasis. Details are discussed in the text.

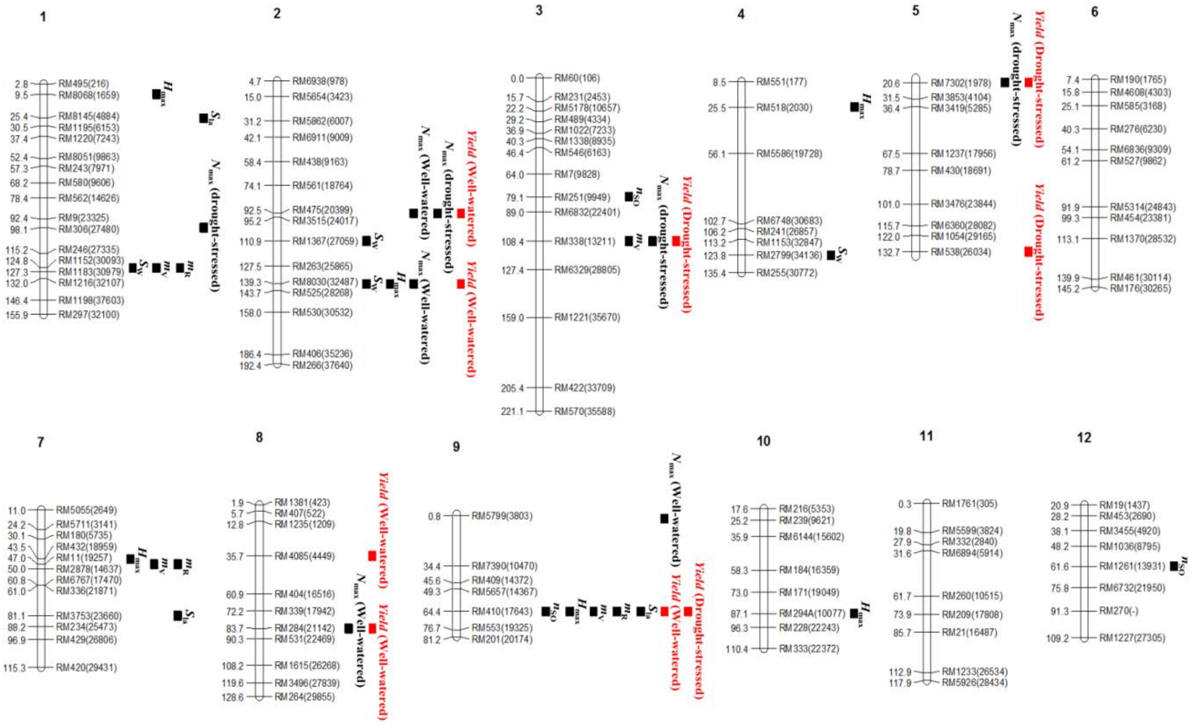

Figure 2 Genome positions of QTL for yield *per se* (red) under either well-watered or drought conditions and of QTL for input parameters of the GECROS crop model (S_W, seed dry weight; n_{SO}, seed (storage organ) N concentration; H_{max}, maximum plant height; m_V, minimum days for vegetative growth phase; m_R, minimum days for reproductive growth phase; S_{la}, specific leaf area for newly appearing leaves; N_{max}, total crop N uptake), mapped on 12 chromosomes using a rice introgression line (IL) population. Based on Supplementary material of Gu et al. (2014b).

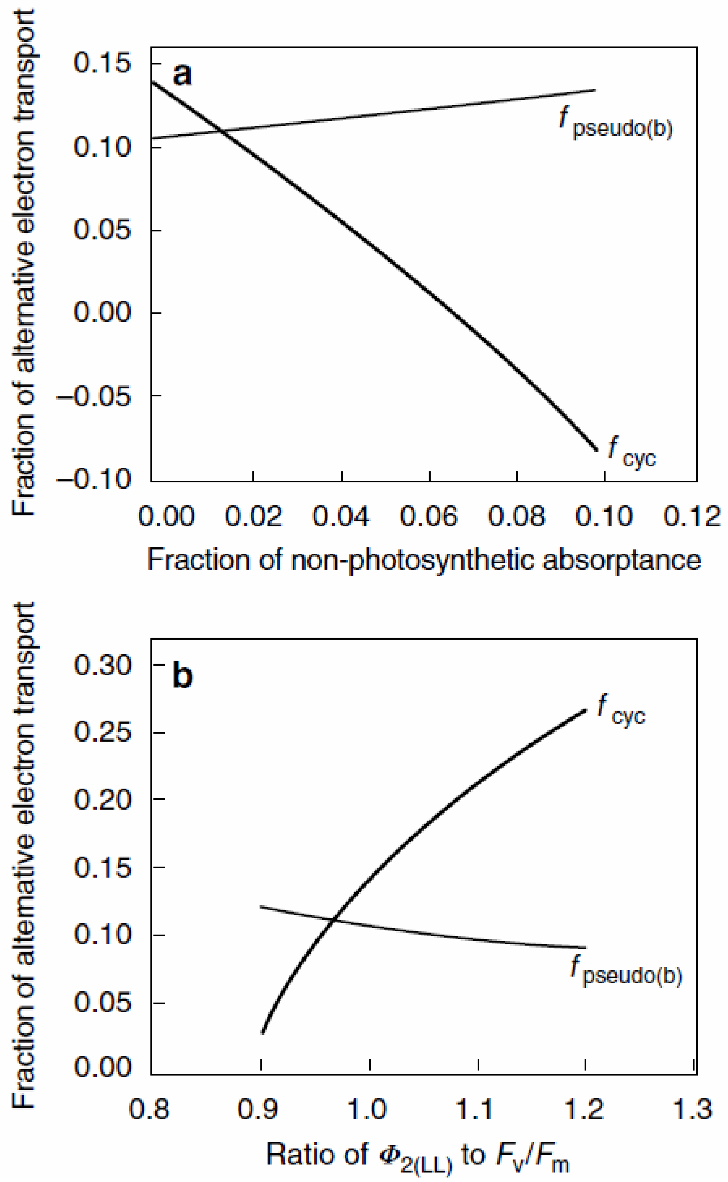

Figure 3 Analysis of sensitivity for the fraction of cyclic electron transport (f_{cyc}) and fraction of basal pseudocyclic electron flow ($f_{\text{pseudo(b)}}$), estimated using the generalised FvCB model, in response to some uncertain factors: (a) the fraction of non-photosynthetic quantum absorption in the leaves, and (b) the ratio of the true photosystem II efficiency under limiting light ($\Phi_{2\text{(LL)}}$) to the apparent maximum photosystem II efficiency (F_v/F_m) measured by chlorophyll fluorescence – here the ratio used to quantify the extent, to which F_v/F_m represents $\Phi_{2\text{(LL)}}$. Reproduced from Yin et al. (2006). The sensitivity of the estimated f_{cyc} was always higher than that of $f_{\text{pseudo(b)}}$ to a change in these uncertainties, suggesting that the estimate of cyclic electron transport (CET) will be more prone to uncertainties in the inputs than the estimate of pseudocyclic electron transport (PET). This also suggests that that if CET and PET run in concert, CET probably acts as a more active brake for LET to avoid or minimise any mismatch between measured quantum efficiency for CO_2 assimilation (Φ_{CO2}) and $\Phi_{2\text{(LL)}}$, even under the limiting light condition. Nowadays, the importance of CET in regulation of photosynthesis, even in C_3 species, is increasingly becoming recognized (e.g., Shikanai 2014).