Role of crop physiology in predicting gene-to-phenotype relationships

Xinyou Yin, Paul C. Struik and Martin J. Kropff

Crop and Weed Ecology Group, Department of Plant Sciences, Wageningen University, PO Box 430, 6700 AK Wageningen, The Netherlands

Robust crop physiological modelling could become an essential tool in explaining crop behaviour using insights from functional genomics. Current crop models can predict crop performance over a range of environmental conditions. Recently, quantitative trait loci (QTL) information has been incorporated into crop models, which has shown the potential for narrowing genotype–phenotype gaps and for applying QTL-based models to the analysis of genotype-by-environment interactions. For further progress, the model structure must be upgraded to allow more physiological feedback features to be incorporated. Model input parameters should be designed to be grounded potentially in gene-level understanding. Integration of crop modelling into genetic and genomic research should enhance the future position of crop physiology in 'plant breeding by design'.

Many crop traits are quantitative, complex and controlled by multiple, interacting genes. The molecular marker technologies, available since the 1980s, enable the variation in these traits to be dissected into the effects of QUANTITATIVE TRAIT LOCI (see Glossary) [1]. New breeding approaches such as marker-assisted selection [2,3] and 'breeding by design' [4] have emerged with the progress of QTL mapping.

Intrinsic complexities (polygenic control, EPISTASIS and GENOTYPE-BY-ENVIRONMENT INTERACTIONS) complicate the manipulation of quantitative crop traits. Existing QTL-analysis methods do not have the precision required to handle these complexities. Without increasing population sizes, little increase in reliability can be achieved [5]. Support from other disciplines is required. Functional genomics, systems biology and CROP PHYSIOLOGY can jointly improve genetic analysis and breeding efficiencies.

Functional genomics aims to discover the function of all genes. Advances in DNA microarray profiling techniques [6] demonstrate the power of high-throughput studies of multiple gene expression in unravelling developmental processes. Also, proteomics and metabolomics profile many gene products in parallel and reveal gene regulation events. Recent studies have considered gene expression values from microarray profiling for a segregating population as conventional quantitative traits in QTL mapping [7,8]. This combined analysis can unravel the genetic

network and elucidate epistasis, which is often found in phenotypes with interactive and interrelated metabolic and ontogenetic pathways [2]. Understanding the regulation of plant metabolism can provide new avenues for breeders to increase crop yields [9]. Genomics-based approaches and marker-assisted selection allow a

Glossary

Alleles: variant forms of a gene at a particular location on a chromosome.

Cell modelling: mathematical representation of (sub-)cellular systems, by integrating multifaceted information on various levels of the plant systeome from high-throughput technologies, to achieve a mechanistic prediction of the dynamics and behaviours of plant cells and tissues on the basis of understanding of how the different biological layers interact to form higher functional units such as coordinated pathways, regulatory network, or complex structure.

CentiMorgan (cM): genetic map unit (1 cM corresponds to 1% recombination per meiosis).

Crop model: assembled mathematical algorithms to represent crop growth processes as a whole system. Through simulation (usually on a daily basis), a crop model translates input conditions (environmental variables, management practices and genotype-specific coefficients) to an ultimate output crop trait (usually crop yield) via quantifying environmental responses of its underlying physiological processes. Several adjectives (such as 'process-based', 'dynamic', 'mechanistic', 'deterministic', 'ecophysiological') are often used to demarcate crop models from empirical relationships between crop yield and environmental variables established by statistical regression analysis of experimental data.

Crop physiology: the science that studies the life processes of plants in a crop stand as influenced by environmental (abiotic or biotic), management and intrinsic factors and by inter-plant relationships.

Epistasis: Expression of a gene depending on the status of the other gene(s). In genetics, it is specifically defined as the interaction between alleles of different genes (non-allelic or intergenic interaction), as opposed to dominance, which is the interaction between alleles of the same gene (interallelic or intragenic interaction). In an ordinary linear model for the effect of multiple genes on a quantitative trait, epistasis is expressed as interactive terms, following the linear terms for the additive effects of the individual genes.

Genetic coefficients: see 'model input traits'.

Genotype-by-environment interaction (G×E): The phenomenon that the relative performance of genotypes varies with environmental conditions. It is attributed to the dependence of expression of underlying genes or QTL on environments (QTL×E).

Leaf area index: the index used by crop physiologists to indicate the size of the crop canopy. It refers to the ratio of the total area of all green leaves to the area of the ground on which the crop grows.

Model input traits: also named 'genetic coefficients'; a set of input parameters of a crop model, representing genotype-specific characteristics within a crop.

Modular: the property that sub-modules are 'insulated' so that a change or failure in one sub-module does not affect or spread to other parts of the model.

Pleiotropy: multiple effects of a single gene that affects more than one trait.

Quantitative trait locus or loci (QTL): Genomic region(s) responsible for the variation of a trait whose phenotypic values show a continuous distribution on a quantitative scale.

Recombinant inbred lines (RIL): a population of homozygous individuals that is obtained by repeated self crossing from an F1 hybrid and that contains ~50% of each of the parental genomes in different combinations.

Specific leaf area (SLA): an index to indicate leaf thickness, measured as the ratio of leaf area to leaf dry biomass weight.

better design of trait-specific genes and enhance plant breeding [10].

With the rapid progress of genomics, systems biology has emerged to combine complex datasets into useful mathematical models mimicking living cells and organisms *in silico* [11]. Plant systems biology is essential in understanding plant behaviour and for increasing crop production to meet the growing demands for food and energy [12].

However, to make genomics work for real-world challenges by manipulating complex traits at the crop level, the science of crop physiology is also required, provided that its traditional quantitative approach is strengthened. Recent studies [13–15] have demonstrated that crop physiology using CROP MODELS can reinforce the QTL analysis of complex traits, thereby improving breeding efficiency and enhancing genetic design. When QTL information is incorporated into crop models, the models should narrow genotype–phenotype gaps and become better tools for analysing the genotype-by-environment interaction (G×E) and therefore should play a pivotal role in the knowledge chain of plant biology in the future. To promote this novel role of crop physiology, robust model structures, new algorithms and input parameters are required. Current crop models can predict crop performance over a range of environmental conditions [16] but are poor in predicting genotype-by-environment interactions [14]. Here, we discuss the opportunities for developing current crop models into tools that can use insights from functional genomics to explain crop behaviour, and incorporate and translate the information from systems biology to crop processes.

Modelling instruments in crop physiology and their added value in genetics

Crop physiology was originally a qualitative, experimental science [17]. It evolved because of the need for instruments that could combine increasing quantities of experimental data. Crop modelling (Figure 1) emerged in the mid-1960s with the pioneering work of Cornelis de Wit [18], who introduced systems theory [19,20] and simulation methodologies [21] into crop physiology.

In modelling research, quantitative algorithms are first developed for individual processes, often based on specific experimental findings. When individual algorithms are assembled into a model, the model can assist in understanding crop behaviour by increasing knowledge about the underlying processes and can guide further experimental research to increase understanding of crop growth at the process level [22]. This heuristic role of crop models and the interplay between experimentation and modelling should guide crop physiology research in the future.

Although current crop models do not simulate the effect of genes, their potential in assisting crop improvement has been realized for years [22]. Indeed, crop models can predict a complex trait by dissecting it into its component attributes and quantifying its underlying component processes. The component attributes correspond to MODEL-INPUT TRAITS or parameters – one set of parameters representing one genotype [23]. Therefore, crop models have the potential to assess a complex trait at a higher

organizational level, via integrating the information known about processes at the lower level. Research to explore the use of crop modelling in plant breeding has ranged from designing new plant types [16,24] to suggesting options to improve selection efficiency [25]. However, studies in which crop physiology and modelling are integrated into active genetic and breeding programmes are rare. Thomas Sinclair *et al.* [26] have identified only three cases where a physiological approach has been applied to develop cultivars: improved water use efficiency of wheat, heat stress tolerance of cowpea and soybean nitrogen fixation tolerance to water stress.

If crop physiology and genetics are combined judiciously, crop physiology and modelling research can reinforce the genetic analysis of complex traits, thereby improving breeding efficiency. Several recent case studies have shown that new knowledge of crop physiology is already showing some promising results. These case studies include: (i) the dissection of a complex trait into component traits based on insight into the ecophysiology (Box 1); (ii) physiological phenotyping of traits having a developmental behaviour (Box 2); and (iii) resolution of G×E (Box 3). Below, we highlight the approach used in the resolution of G×E because of the importance of G×E in crop physiology and genetics [17].

Because crop models represent causality between component processes and yield, they can predict crop performance beyond the environments for which the model parameters were estimated. This singular property allows the models potentially to resolve G×E into underlying processes on a daily basis and predict crop performance in any environment. This model-based approach comprises the following steps.
(i) Create a crop model that predicts complex traits based on relations between elementary processes and environmental variables.
(ii) Evaluate the capability of the model to predict the complex trait across a wide range of G×E combinations.
(iii) Identify QTL for model-input traits using a genetic QTL approach.
(iv) Develop a QTL-based model whereby the original values of model input traits are replaced by QTL-based inputs.
(v) Validate the QTL-based model across environments.

QTL-based modelling was first used to predict barley (*Hordeum vulgare*) grain yield [13,14]. It showed that current crop models need to be improved to predict yield differences among relatively similar RECOMBINANT INBRED LINES. More recently, a similar analysis was conducted for simpler traits: leaf elongation rates in maize (*Zea mays*) [15] and flowering time in barley (Xinyou Yin *et al.*, unpublished), demonstrating the potential of this approach (Box 3).

Fulfilling challenges of crop physiology

The success of the studies on maize leaf elongation and barley flowering time indicates the potential of using the QTL-based modelling approach to predict more complex traits such as yield. Current crop models have to be improved in algorithm structure and input parameters to achieve the level of resolution needed to predict G×E for

Box 1. Dissection of a complex trait into component traits

Potentially, numerous genes contribute to the genetic variation of a complex trait; however, the number of quantitative trait loci (QTL) detected for a trait is rarely more than eight [5]. Instead of looking for the QTL for a complex trait, determining the QTL for its underlying component traits might give more useful information. Using a recombinant inbred line (RIL) population of *Hordeum vulgare*, Xinyou Yin *et al.* [41] dissected barley yield into the three components: number of spikes per unit area, number of grains per spike and grain weight, which when multiplied together give grain yield. Almost all QTL for yield were found at the position of or in close proximity to QTL for its component traits. Several QTL for component traits were not detected when yield itself was subjected to QTL analysis (Table I). Component analysis can provide insight into how differences in the complex trait come about and how the same gene affects different components. It is possible that a QTL can be found for a complex trait but not for its components when the effects of the QTL on the individual components are small but aggregate to be of sufficient significance for the complex trait.

The power of component analysis can be strengthened by a simultaneous study on ontogenetic traits. Zhou *et al.* [42] dissected the pre-heading period in rice (*Oryza sativa*) into two additive components, vegetative and panicle-development phases, and showed that more QTL could be detected based on the two additive components. In cereal crops, periods that determine spike number, kernel number and kernel weight match vegetative development, panicle-development and kernel-filling phases, respectively [29]. Dissection of yield into its components combined with component analysis of the whole crop ontogeny should, therefore, provide a better insight into how the genetics of plant development affects crop performance.

The other merit of component analysis is its ability to manifest an epistatic effect of component-trait QTL on grain yield because yield is the product of its three component traits. Use of three components means a three-way (trigenic) epistasis on yield. Because epistasis is often found in phenotypes that are the result of nonlinear interactions among multiple component processes integrated over ontogenetic stages [2], a more physiologically based embodying of epistasis is to use a robust crop model. Scott Chapman *et al.* [25] showed that complex epistatic effects on yield were generated even though individual genes had been defined as simple additive effects on component model-input traits.

Table I. Position of QTL on different chromosomes[a]

Chromosome	Yield	Spikes	Grains	Grain weight
2	**126.0**	–	**130.7**	–
3	**126.4**	**126.4**	**126.4**	–
5	**62.7**	–	–	**64.4**
6	**57.4**	–	–	**47.4**
1	–	–	54.4	–
1	–	–	–	109.3
2	–	68.4	–	68.4
2	–	–	95.7	–
3	–	–	65.0	72.2
4	–	7.6	–	–
4	–	76.7	–	–
4	–	–	138.1	–
5	–	–	96.3	–
7	–	–	77.8	–
7	–	–	–	95.7

[a]The figures in columns 2–5 of this table indicate positions on a molecular marker map in CENTIMORGANS [13] (counted from the upper terminal marker of the chromosome concerned), of QTL identified for barley yield and its three component traits (reproduced, with permission, from Ref. [41]). QTL shown in bold are common or closely linked, the remaining data represent additional QTL that were not identified when yield itself was subjected to analysis. '–' indicates that there are no QTL identified for those traits on those chromosomes.

such complex traits [27–29]. Model algorithms and structure should capture the conservation and balance of water, crop carbon and nitrogen assimilations (Figure 1); input parameters should reflect the differences between genotypes. This concept is analogous to quantitative genetics whereby the common feature of the parental lines is described by the general mean and the difference is attributed to additive and other types of gene effects. It is also analogous to QTL analysis whereby only QTL that relate to differences between parental lines are identified whereas the effects of other QTL – at which parents carry the identical ALLELES – are attributed to the constant term of the QTL statistics.

Improving model algorithms and structure

To upgrade algorithms for model components, photosynthesis and transpiration have been studied extensively. Robust analytical algorithms [30] for photosynthesis and transpiration can enhance numerical stability and computational efficiency. Although further improvement is necessary, progress has recently been made in developing robust algorithms for respiratory costs [31] and LEAF AREA INDEX [32].

Model algorithms should capture interactive responses to environmental factors and allow all processes to be modelled at a consistent level of detail. Crop growth is determined by the functional balance of contrasting components (source versus sink) or processes (e.g. carbon versus nitrogen metabolism). Moreover, like any other system involving cybernetic mechanisms [20], crop growth is associated with many feedback features (e.g. apparent down-regulation of Rubisco, the primary carboxylating enzyme for photosynthesis, under conditions of high carbohydrate status). To represent realistic crop functions, models have to be structured to embody the interactive physiological causes that drive crop dynamics and generate these feedback, balancing and interaction mechanisms.

Improving model input parameters

Determination of model input parameters for current models often requires substantial destructive sampling of field crop plants, which can involve large random errors and requires large amounts of labour and crop material [14]. Model parameters, known as 'GENETIC COEFFICIENTS', should be defined in a genotype-specific and environment-independent way, and should be readily and precisely measured at the individual plant level. Model design should also allow flexibility in parameters, which are potentially grounded in gene-level understanding. Morphological traits (e.g. plant height under optimal growing conditions) that are directly associated with physiological events, are well suited as input parameters because ample information is already available about their inheritance in many species.

Incorporation of gene information into crop models is increasingly receiving attention because of the rapid progress of functional genomics. Stephen Welch *et al.* [33] presented a quantitative model for transition to flowering in *Arabidopsis* using a genetic neural network approach based on qualitative working models [34]. The approach could be extended to plant height and flowering

Box 2. Physiological phenotyping of traits having a developmental behaviour

Most quantitative trait loci (QTL) studies focused on a crop trait record at a fixed landmark stage, notably at flowering or maturity. However, many traits (e.g. mass weight) have a temporal trajectory and their final value is the consequence of the actions of many QTL expressed during ontogeny. The mapping of dynamic traits is emerging as an important research area because it is relevant for geneticists and breeders to define the genetic architecture of growth traits [43].

Most studies on dynamic traits have been based on the direct use of phenotypic data measured at sequential times [44,45]. Biometric approaches have also been developed to detect dynamic QTL [46]. One tricky aspect for these analyses is the physiological age difference at a phenotyping time among individuals of a mapping population because their flowering or maturity day differed. A logical phenotyping should be conducted at the same physiological age for all individuals. From a practical point of view, this is impossible if phenotyping needs to be made at developmental stages (DS) that are not marked by morphological changes such as flowering. Alternatively, a crop development model could be used to correct the measured data for the same age. None of the studies cited above,

although aiming at gaining insight into the crucial role of development on the analysed trait, considered this 'age' factor.

The importance of the 'age' factor was highlighted by an analysis of time-dependent SPECIFIC LEAF AREA (SLA) in barley [47]. The SLA was measured once at the same DS (i.e. at flowering) for all individuals, and four times at specific days before flowering. When the SLA of each measurement time was directly analysed, a major dwarfing mutation, the *denso* gene mapped at 126 cM of chromosome 3, was found to affect SLA strongly at all pre-flowering times (Figure Ia for 27 days after emergence), except at flowering. If the SLA of the different individuals was re-scaled by a development model (in which DS is defined as 0 at emergence and 1 at flowering) for differences in physiological age, the effect of the *denso* gene was no longer significant (Figure Ib for the DS 0.35). The effect of the *denso* gene on the SLA detected in the first instance was therefore the artefact of its PLEIOTROPIC effect on the pre-flowering duration (see inset in Figure Ib). This analysis underlines the importance of applying crop physiology and modelling in QTL analysis of traits having a dynamic behaviour.
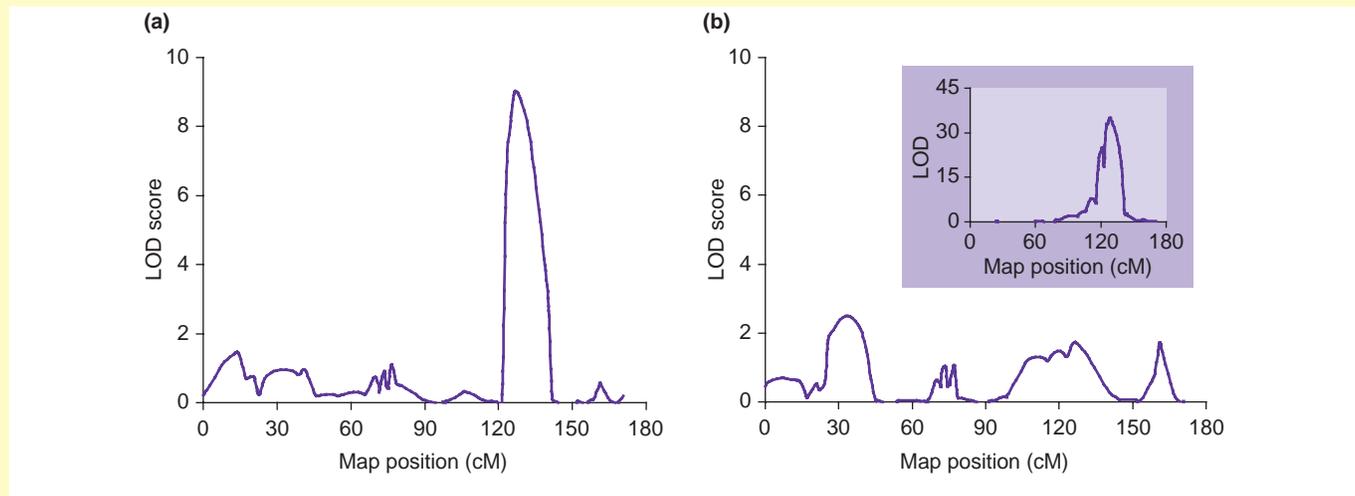
Figure I. Plot of quantitative trait loci (QTL) likelihood (LOD) over chromosome 3 in barley. The trait is specific leaf area (SLA) measured at 27 days after emergence (a), and the SLA corrected at the same developmental stage (DS) of 0.35 (b), which is roughly equivalent to the time of 27 days after emergence. Map position refers to centiMorgans (cM) on a molecular marker map described in [13]. The LOD threshold for the presence of a QTL is 3.0 for this chromosome. The inset in (b) shows the LOD profile on chromosome 3 for the pre-flowering duration, indicating the strong effect of the *denso* gene at 126 cM on this trait. Figure is modified, with permission, from Ref. [47].

Box 3. Resolution of genotype-by-environment interactions

Manipulating quantitative traits is complicated by the genotype-by-environment interaction (G×E) and the associated interaction between quantitative trait loci and environment (QTL×E). In statistical G×E models, the mean phenotype value of genotypes in each environment is used as a measure of environmental quality [48]. Similarly, in statistical QTL × E models [49], the phenotype with any combination of QTL alleles of the mapping population is evaluated within multiple environments. However, such a QTL model cannot predict the phenotype across other independent environments because it lacks an explicit environmental input. Although factorial regression allows QTL×E to be expressed as a function of environmental variables [50], the approach does not account for temporal or spatial variation of environmental variables in the crop, which can be important for predicting G×E on a biological basis. This problem can be circumvented by use of a novel, five-step approach (see main text), in which ecophysiological modelling is combined with QTL analysis. The following two examples show the power of this approach.

A model for maize leaf elongation rate was established as a linear function of both meristem-air vapour pressure deficit and soil water potential [15]. A QTL analysis was performed on the intercept and the slopes of this model. Each parameter was then computed as the sum

of QTL effects, resulting in QTL-based model parameters. Predicted leaf elongation rates using original and QTL-based parameters were comparable. Furthermore, for 11 recombinant inbred lines (RILs) and parental lines not included in QTL analysis, the QTL-based model accounted for 74% of the observed phenotypic variability, indicating that the QTL-based model had a general value for any RIL within the population under any environment.

Xinyou Yin *et al.* (unpublished) used a simple model for predicting barley flowering time based on hourly air temperature and daily photoperiod. Values of four model input traits were derived from a photoperiod-controlled greenhouse experiment. The additive genetic models were used to calculate QTL-based inputs. The QTL-based model accounted for 72% of the variation among 94 RIL and 94% of the variation among the two parents across eight field environments. The QTL-based model predictions correlated well with those using original parameter values (correlation coefficient $r=0.918$; $n=768$). Predicted days using the QTL-based model also agreed with those predicted from per field environment QTL analysis for days to flowering *per se* ($r=0.891$; $n=768$), indicating that the model was capable of extrapolating (QTL-) information obtained in a controlled environment to independent field environments.
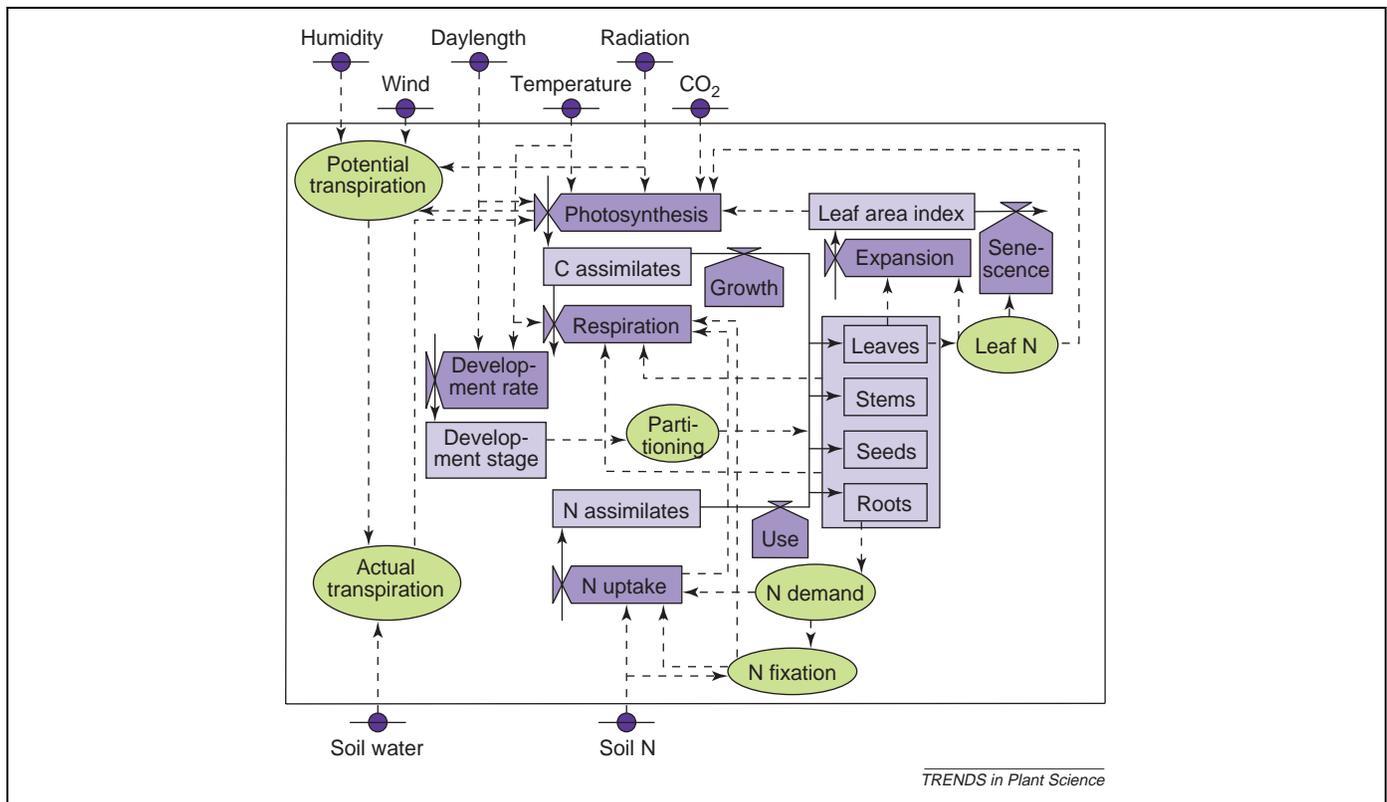
Figure 1. A scheme using standard Forrester's [21] symbols (i.e. boxes are state variables, valves are rate variables, ellipses are intermediate variables, crossed small cycles are environmental inputs, solid lines are flows of material and broken lines are flows of information) to represent a conceptual crop model that captures the conservation and balance of energy, water, carbon (C) and nitrogen (N) assimilations. To understand crop production, one must examine the structure and dynamics of crop growth as a system rather than the characteristics of its isolated parts. Drawing a diagram of their interconnections represents an important first step towards mathematical modelling to understand crop system structures and predict crop dynamics. A crop model captures several physiological processes. Each process is quantified in relation to other processes and to environmental factors. For model computation, the rate of change of processes is assumed constant during a short time-step (usually one day). Crop growth rate at a time step is computed, depending on the actual crop status and current environmental conditions. The biomass formed in a time-step equals the multiple of growth rate and the length of the time-step. This is added to the quantity of biomass already present. The growth rate is then re-calculated for the next time-step. Calculating growth rates and updating the quantity of biomass are repeated in sequence until the entire growing season is predicted to end by the developmental sub-model. The quantity of biomass is a state variable and the growth rate is its rate variable. Both tangible quantities (e.g. biomass weight, nitrogen content and leaf area) and abstract quantities (e.g. development stage) can be considered as state variables. Biomass weight of plant organs is distinguished as a separate state variable; the partitioning of the newly produced assimilates at each time-step is thereof described. The key issue in model development is to describe rate variables, for which one equation might not be enough and many intermediate variables are introduced. Environmental variables as inputs to the crop model include climatic factors (e.g. radiation and temperature) and edaphic variables (water and nutrient availabilities, which are influenced by management practices). Genetic coefficients are another type of model inputs but not shown in the diagram.

time in crops because the 'Green Revolution' genes [35] and those for photoperiodism and vernalization [36] are now well understood. Another candidate trait could be photosynthesis, for which there is detailed quantitative information about the biochemical pathways [37].

Towards a new generation of crop models

With the growing system-level understanding of links between gene function and crop performance, the new generation of crop models should have robust model architecture, numerical consistency and stability, enhanced heuristics and gene–action-based input. Such models would greatly enhance our ability to translate short-time scale gene-level information to the crop performance in continuously changing field environments. This does not necessarily imply an increase in the detail or the complexity of the model structure. However, to speed up the development of model-inputs based on gene-level understanding, models should be 'MODULAR', allowing incorporation of gene network information to the model one trait at a time.

Crop physiology as a part of the knowledge chain from gene to phenotype

High-throughput technologies have revolutionized experimental research, greatly expanded the fundamental domain of plant biology and enhanced opportunities generated by multidisciplinary integrated studies. Combining measurements of physiological components with gene expression profiles should illustrate the function of genes, biochemical pathways and cellular processes that are affected in a coordinated manner. Such studies should lay the groundwork for elucidating regulatory networks and linkages among gene products, biochemistry and whole-plant physiology. Systems biology [11,12] emerged as a result of renewed recognition of systems theory [19,20], aiming at synthesizing complex datasets into useful mathematical models. Plant systems biology, as currently defined, has largely overlooked the rich history of crop modelling [38]. To fulfil real-world challenges for increased crop production, the information available from functional genomics and systems biology needs to be taken to the crop level and crop physiology will have a fundamental role in achieving this.
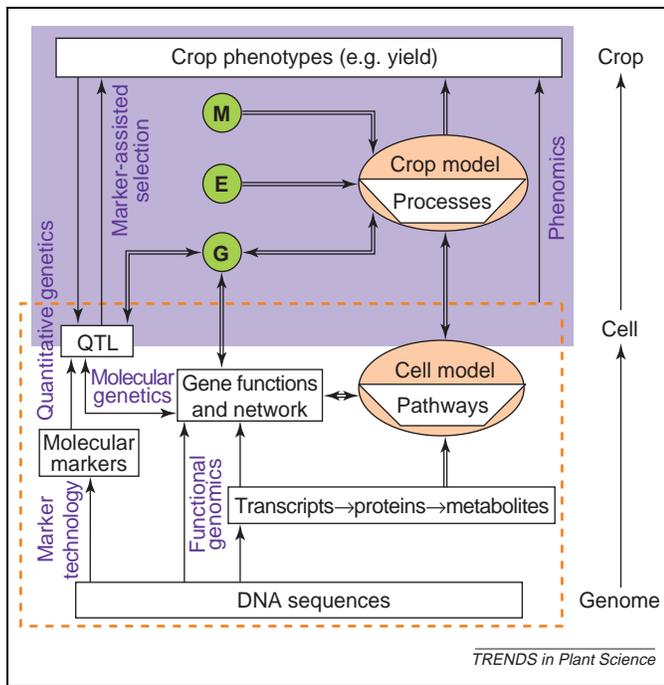
Figure 2. The position of crop physiology in the knowledge chain of plant biology when applied to the study of complex quantitative crop traits. Abbreviations: E, environmental variables; G, genetic coefficients; M, management actions. Arrows with a single arrow shaft indicate the scientific discipline that links the organizational hierarchies. Arrows with a double arrow shaft indicate the inputs or outputs of a model, for scaling of insights between levels of understanding. The area indicated by a broken orange line indicates systems biology – which integrates modern high-throughput experimental technology-based plant sciences. The area highlighted in light purple indicates crop physiology. In both systems biology and crop physiology, modelling plays a central role in integrating information across levels of organization. Recent studies, in which DNA marker allelic variations are coupled into crop models via quantitative trait loci (QTL), highlight the potential of predicting genotype–phenotype relationships. Rapid development of functional genomics and systems biology should generate new insights into the gene functions and regulatory networks that underlie genetic variations. Coupling these new insights into crop models to predict G×E×M for complex quantitative traits should provide new opportunities to bridge gaps between genes and phenotypes (phenomics). However, if systems biology modelling tools are to have a significant impact on the design of complex crop traits, insights from crop modelling should be used to enhance the integration of genomic technologies in crop improvement.

We propose an integrated research platform to apply the results of genomic research to unravelling complex traits at the crop level (Figure 2). A new generation of crop models combined with systems biology studies should enable us to narrow the gap between genes and complex phenotypes significantly by integrating the knowledge at lower hierarchies to predict field performance of crop genotypes. The 'G' input component of crop models provides the key link between crop physiology and systems biology that combines the results from genetics and functional genomics (Figure 2). This link provides a powerful tool to study G×E×M, where 'M' refers to the agricultural management practices modifying the environment (e.g. fertilization and irrigation), to broaden the original G×E concept. The five steps for QTL-based modelling that are used to analyse G×E are still valid here; but in this new framework, the 'G' component is broadened to also include the information about the genes and gene network elucidated by functional genomics.

The framework illustrated in Figure 2 underlines that systems modelling is central to integration issues in plant biology. CELL MODELLING as embedded in systems biology

can provide insight into biochemical regulation of coordinated pathways underlying physiological processes. Crop models contribute to the higher integration level, directly linking physiological processes to complex crop phenotypes. Crop physiology can contribute by linking cell modelling with genotype-specific crop modelling. The key insights from cell modelling can help to improve the description of processes in crop models, and vice versa. Through these integration steps, an entire knowledge chain in plant biology can be established, thereby enabling the gaps between variation in DNA sequences and variation at the level of phenotype (phenomics) to be narrowed. Unlike conventional quantitative genetics, which is concerned with the direct association between genes and complex phenotypes, the phenomics revealed by the proposed approach considers the underlying processes of the phenotype and their genetic and genomic basis. This chain framework is valuable because complex crop phenotypes with quantitative features and their response to environmental variation cannot be understood and predicted based solely on the genes specific to a single macroscopic phenotype *per se*. Armed with the proposed approach, crop physiology and functional genomics complement each other, enabling the *in silico* assessment of crop responses to genetic fine-tuning under defined environmental scenarios. Therefore, new exploitation of the potential of traditional crop sciences can help to obtain a realistic perspective on the much-touted potential benefit to agricultural production from modern plant functional genomics.

Conclusions

Advances in QTL analysis have led to a suggestion that physiologists take a genetic approach to tackle difficult physiological questions [39]. The examples shown in Boxes 1–3 suggest that crop physiology can assist geneticists in unravelling complex traits. Geneticists and physiologists have interests in common, such as G×E [17]. Traditionally, physiologists cannot afford to carry out detailed experiments on complex traits with many genotypes, whereas geneticists working with many genotypes cannot afford to go into much physiological detail. The rapid development of new research instruments is providing scientists with opportunities to work together more closely. This should result in more insights into G×E, narrowed genotype-phenotype gaps, and effective 'breeding by design'. Future breeding paradigms could be characterized by multidisciplinary 'joint' efforts [26,40]. The joint activity could be used to create a common platform for addressing complex questions, where modelling would play a crucial role in integrating the knowledge chain in plant biology at either the lower point (systems biology) or the higher point (crop physiology). For this to happen, scientists from crop physiology, molecular genetics and genomics should rethink the research agenda to create opportunities to make genomics work to increase crop production.

Acknowledgements

References

1 Paterson, A.H. *et al*. (1988) Resolution of quantitative factors by using a complete linkage map of restriction fragment length polymorphisms. *Nature* 335, 721–726

2 Lee, M. (1995) DNA markers and plant breeding programs. *Adv. Agron.* 55, 265–344

3 Ribaut, J-M. and Hoisington, D. (1998) Marker-assisted selection: new tools and strategies. *Trends Plant Sci.* 3, 236–239

4 Peleman, J.D. and van der Voort, J.R. (2003) Breeding by design. *Trends Plant Sci.* 8, 330–334

5 Kearsey, M.J. and Farquhar, A.G.L. (1998) QTL analysis in plants; where are we now? *Heredity* 80, 137–142

6 DeRisi, J.L. *et al*. (1997) Exploring the metabolic and genetic control of gene expression on a genomic scale. *Science* 278, 680–686

7 Brem, R.B. *et al*. (2002) Genetic dissection of transcriptional regulation in budding yeast. *Science* 296, 752–755

8 Schadt, E.E. *et al*. (2003) Genetics of gene expression surveyed in maize, mouse and man. *Nature* 422, 297–302

9 Morandini, P. and Salamini, F. (2003) Plant biotechnology and breeding: allied for years to come. *Trends Plant Sci.* 8, 70–75

10 Knight, J. (2003) A dying breed. *Nature* 421, 568–570

11 Kitano, H. (2002) Systems biology: a brief overview. *Science* 295, 1662–1664

12 Minorsky, P.V. (2003) Frontiers of plant cell biology: signals and pathways, system-based approaches 22nd symposium in plant biology (University of California-Riverside). *Plant Physiol.* 132, 428–435

13 Yin, X. *et al*. (1999) AFLP mapping of quantitative trait loci for yield-determining physiological characters in spring barley. *Theor. Appl. Genet.* 99, 244–253

14 Yin, X. *et al*. (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

15 Reymond, M. *et al*. (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 Physiol.* 131, 664–675

16 Kropff, M.J. *et al*. (1995) Using systems approaches to design and evaluate ideotypes for specific environments. In *Eco-regional Approaches for Sustainable Land Use and Food Production* (Bouma, J. *et al*., eds), pp. 417–435, Kluwer Academic Publishers

17 Kropff, M.J. and Struik, P.C. (2002) Developments in crop ecology. *NJAS-Wageningen J. Life Sci.* 50, 223–237

18 De Wit, C.T. (1965) *Photosynthesis of Leaf Canopies – Agricultural Research Report 663*, Pudoc, Wageningen, The Netherlands

19 Von Bertalanffy, L. (1933) *Modern Theories of Development: An Introduction to Theoretical Biology*, Oxford University Press

20 Weiner, N. (1948) *Cybernetics or Control and Communication in the Animal and the Machine*, MIT Press, Cambridge, MS, USA

21 Forrester, J.W. (1961) *Industrial Dynamics*, MIT Press, Cambridge, MS, USA

22 Loomis, R.S. *et al*. (1979) Explanatory models in crop physiology. *Annu. Rev. Plant Physiol.* 30, 339–367

23 Tardieu, F. (2003) Virtual plants: modelling as a tool for the genomics of tolerance to water deficit. *Trends Plant Sci.* 8, 9–14

24 Penning de Vries, F.W.T. (1991) Improving yields: designing and testing VHYVs. In *Systems Simulations at IRRI – IRRI Research Paper 151* (Penning de Vries, F.W.T. *et al*., eds), pp. 13–19, International Rice Research Institute (IRRI), Los Baños, Philippines

25 Chapman, S.C. *et al*. (2003) Evaluating plant breeding strategies by simulating gene action and dryland environment effects. *Agron. J.* 95, 99–113

26 Sinclair, T.R. *et al*. (2004) Crop transformation and the challenge to increase yield potential. *Trends Plant Sci.* 9, 70–75

27 Boote, K.J. *et al*. (2001) Physiology and modelling of traits in plants: implications for genetic improvement. *Agric. Syst.* 70, 395–420

28 Hammer, G.L. *et al*. (2002) Future contributions of crop modeling – from heuristics and supporting decision making to understanding genetic regulation and aiding crop improvement. *Eur. J. Agron.* 18, 15–31

29 Slafer, G.A. (2003) Genetic basis of yield as viewed from a crop physiologist's perspective. *Ann. Appl. Biol.* 142, 117–128

30 De Pury, D.G.G. and Farquhar, G.D. (1997) Simple scaling of photosynthesis from leaves to canopies without the errors of big-leaf models. *Plant Cell Environ.* 20, 537–557

31 Cannell, M.G.R. and Thornley, J.H.M. (2000) Modelling the components of plant respiration: some guiding principles. *Ann Bot.* 85, 45–54

32 Yin, X. *et al*. (2003) Some quantitative relationships between leaf area index and canopy nitrogen content and distribution. *Ann Bot.* 91, 893–903

33 Welch, S.M. *et al*. (2003) Genetic neural network model of flowering time control in *Arabidopsis thaliana*. *Agron. J.* 95, 71–81

34 Koornneef, M. *et al*. (1998) Genetic control of flowering time in *Arabidopsis*. *Annu. Rev. Plant Physiol. Plant Mol. Biol.* 49, 345–370

35 Hedden, P. (2003) The genes of the Green Revolution. *Trends Genet.* 19, 5–9

36 Griffiths, S. *et al*. (2003) The evolution of CONSTANS-like gene families in barley, rice and *Arabidopsis*. *Plant Physiol.* 131, 1855–1867

37 Zhu, X-G. *et al*. (2004) Would transformation of C_3 crop plants with foreign Rubisco increase productivity? A computational analysis extrapolating from kinetic properties to canopy photosynthesis. *Plant Cell Environ.* 27, 155–165

38 Hammer, G.L. *et al*. (2004) On systems thinking, systems biology, and the *in silico* plants. *Plant Physiol.* 134, 909–911

39 Prioul, J.L. *et al*. (1997) Dissecting complex physiological functions through the use of molecular quantitative genetics. *J. Exp. Bot.* 48, 1151–1163

40 Koornneef, M. and Stam, P. (2001) Changing paradigms in plant breeding. *Plant Physiol.* 125, 156–159

41 Yin, X. *et al*. (2002) Use of component analysis in QTL mapping of complex crop traits: a case study on yield in barley. *Plant Breed.* 121, 314–319

42 Zhou, Y. *et al*. (2001) Genetic dissection of heading time and its components in rice. *Theor. Appl. Genet.* 102, 1236–1242

43 Maloof, J.N. (2003) QTL for plant growth and morphology. *Curr. Opin. Plant Biol.* 6, 85–90

44 Yan, J. *et al*. (1998) Molecular dissection of developmental behavior of plant height in rice (*Oryza sativa* L.). *Genetics* 150, 1257–1265

45 Wu, W. *et al*. (2002) Mapping of quantitative trait loci based on growth models. *Theor. Appl. Genet.* 105, 1043–1049

46 Ma, C.X. *et al*. (2002) Functional mapping of quantitative trait loci underlying the character process: a theoretical framework. *Genetics* 161, 1751–1762

47 Yin, X. *et al*. (1999) The role of ecophysiological models in QTL analysis: the example of specific leaf area in barley. *Heredity* 82, 415–421

48 Kearsey, M.J. and Pooni, H.S. (1996) *The Genetical Analysis of Quantitative Traits*, Chapman & Hall

49 Piepho, H-P. (2000) A mixture-model approach to mapping quantitative trait loci in barley on the basis of multiple environment data. *Genetics* 156, 2043–2050

50 Van Eeuwijk, F.A. *et al*. (2002) Analysing QTL by environment interaction by factorial regression, with an application to the CIMMYT drought and low nitrogen stress programme in maize. In *Quantitative Genetics, Genomics and Plant Breeding* (Kang, M.S., ed.), pp. 245–256, CAB International