CHAPTER 24

A DIALOGUE ON INTERDISCIPLINARY COLLABORATION TO BRIDGE THE GAP BETWEEN PLANT GENOMICS AND CROP SCIENCES

P.C. STRUIK[#], K.G. CASSMAN[##] AND M. KOORNNEEF[###]

[#]*Crop and Weed Ecology Group, Plant Sciences, Wageningen University,*
P.O. Box 430, 6700 AK Wageningen, The Netherlands.
E-mail: paul.struik@wur.nl
[##]*Department of Agronomy and Horticulture, University of Nebraska,*
P.O. Box 830724, 68583-0724 Lincoln, NE, USA.
E-mail: kcassman1@unl.edu
[###]*Max Planck Institute for Plant Breeding Research,*
Carl von Linnéweg 10, 50829 Köln, Germany.
E-mail: koornnee@mpiz-koeln.mpg.de

Abstract. In the future, more food needs to be produced with increasingly scarce natural resources. Genomics can play a key role in accelerating yield gains because it helps to improve our understanding of genetic traits and assists in breeding for better crop performance. The scientific muscle of genomics attracted tremendous research investments, but the efficiency with which these investments are paying off is still low. How can we accelerate the application of molecular genetics to our understanding of crop physiology and subsequently to crop improvement? The missing link is a more detailed understanding of the effects of gene function on crop performance at field level under agronomically relevant conditions captured in robust, physiology-based mechanistic models. With such models the most sensitive processes and mechanisms at whole-crop level that contribute to improved crop performance can be identified. To achieve the detailed understanding necessary to build and feed these models, more research on whole-plant physiology and crop ecology is required, with a focus on the complexity of scaling up knowledge from the molecular level to the farmers' fields and production systems. Such studies assess how the plant is able to integrate the information at different levels of organization into the functioning of the whole plant and predicting the phenotype of transgenic plants engineered for improvement of a complex trait.

More investment is needed in linking whole-plant physiology, crop ecology and crop simulation with molecular biology and genomics. Moreover, long-term progress can be enhanced by the formation of multidisciplinary teams that operate through networks of excellence in developing quantitative tools that integrate complex information and different levels of organization and by the exchange of young scientists between research groups working at different hierarchical levels. On the short term improvement of the characterization of experimental environments (preferably through commonly shared protocols) and of the characterization of parents for creating mapping populations is needed. In addition,

319

joined multi-location trials and advanced physiological and statistical approaches for determining what aspects of the environment are most influential on the genotype × environment interactions are required.

INTRODUCTION

The need to accelerate yield gains

Our human population continues to increase and will reach 8.5 to 9 billion within the next 40 years. Future food security for this growing population will depend on *acceleration* of yield gains per unit of land and per unit of input for the major food crops at rates *well above* the historical trend of the past 50 years. The challenge is to produce more food on limited land resources and with less water because the availability of these natural resources for agriculture is decreasing rapidly as a result of economic development, which diverts these resources for non-agricultural uses. These trends are strongest in developing and emerging countries, where nearly all of future population growth will occur. It has been claimed that (functional) genomics can play a key role in the necessary acceleration of yield gains.

The potential of genomics

Genomics provides a powerful tool for identifying genes of agronomic importance. Genomics implies the study of all genes and their gene products in an organism with respect to their function and their control by environmental and developmental factors. It is suggested that the knowledge arising from genomics not only helps to improve our understanding of complex crop traits (such as yield and yield stability), but will also assist us in breeding for better crop performance and in designing better cultural practices.

For genomics, tools have been developed that allow the detection of the genes (genome sequencing), the study of the expression of these genes (micro-array, gene-chip analysis) and of their ultimate gene products (transcriptomics, proteomics, metabolomics, expression of traits). Moreover, methods have been developed to study the function of genes. The latter methods make use of genetic variation from within the available germplasm of a species, and also from induced mutants and transgenic plants that over- or under-express a specific gene.

Genomics is not paying off yet

The scientific muscle of this relatively new approach has attracted a tremendous research investment in both the private and public sectors. As a result, entire genomes of several crop species have been or will soon be sequenced, and there has been an explosion of new knowledge about genome structure and function. At issue is the efficiency with which this huge investment is paying off in terms of leveraging this genetic knowledge to meet the challenge of global food security.

Despite the remarkable recent advance in basic knowledge of plant genes and gene networks, there has been relatively little impact on crop improvement from the application of genomics and recombinant-DNA technology. Insect-resistant (*Bt*) and

Roundup-Ready® herbicide-resistant crops are the exceptions, but these technologies were developed on the basis of studies of single genes a decade before the birth of genomics. Progress in harnessing the power of genomics is still limited, despite all the promises and claims. It is not even clear yet what will be the impact of genomics on the rate of crop improvement by plant breeding.

Results are not yet visible. This certainly can be explained in part by the juvenile stage of the '-omic' technologies. More time is needed. For example, at this moment rice is the only major staple crop for which a complete genome sequence – an important tool in genomics – has been published.

The view that progress is limited may also be obscured by a focus on the use of transgenics, the easy but certainly not the only way to apply '-omics'.

Finally, '-omic' technologies are mainly applied by private research laboratories carrying out their work in secrecy and evaluating physiological processes responsible for genotype × environment interactions regarding complex traits according to their own standards. This research is not dictated by an agenda aimed at solving important scientific issues for the public good, and much of it does not undergo peer review or publication in scientific journals.

Scientific limits to genomics

There are substantial advances in understanding the function of single genes that control agronomic traits (such as pest resistance and grain quality) and several examples of traits under control of *linear* gene cascades or small gene networks (such as flowering response). Even with these impressive advances, the use of this knowledge for the improvement of our major food crops has been relatively slow. In addition, there has been much less progress in elucidating the genetic control of traits for which the genetic variation accessible to breeding is under complex genetic control involving many genes and strong genotype × environment interactions. Fecundity, effective drought resistance and nitrogen use efficiency are examples of such complex traits that are influenced by numerous compensatory feedback mechanisms and for which plant evolution has worked millions of years to perfect. The latter, by the way, could also imply that the genetic variation is limited.

There are also scientific challenges in the application of genomics research. First, in genomics priorities have to be set, as resources are limited, with regard to genetic variation: not all genes involved in traits of interest show relevant genetic variation, not all variation can be identified, for example, by QTL analysis, and costs of generating desirable variation may be too high. Second, genes do not function on their own, and knowing the molecular characteristics (biochemical function, expression regulation, etc.) often does not elucidate the controls on a complex trait such as yield. Third, molecular biology is progressing much faster than the theoretical and experimental framework connecting genes, plants and crops.

Given this situation, what is needed to accelerate the application of molecular genetics to our understanding of crop physiology and subsequently to crop improvement, especially for traits under complex genetic control? The answer to this question is, in our view, a detailed physiological analysis of the genetic variation

and of the controls of the expression of genes in an agronomically relevant environment. The plea for such research at crop level is the main issue of this dialogue, because progress in this field is still limited.

THE MISSING LINK

We believe the critical missing link is a more detailed understanding of the effects of gene function on crop performance at the field level under a relevant range of environmental conditions, and capturing this knowledge in robust simulation models. Such models would facilitate identification of the most sensitive processes and mechanisms at the whole-plant and plant-community levels that contribute to improved crop performance. They would also allow prediction of phenotype from genotype in transgenic plants. To achieve this capability, a greater public investment in plant physiology and crop ecology, and a much closer collaboration between scientists in these disciplines and those in basic and applied plant genetics will be required. Lack of such collaboration has resulted in a number of spurious reports published in influential scientific journals that claim progress on improving complex traits such as crop yield potential based on molecular genetic approaches, but still await confirmation under agronomically relevant conditions. The issues and questions given above indicate that these claims are to date either unjustified or at least not supported by published results. Public research in whole-plant physiology and crop ecology must be strengthened to realize the potential of publicly funded functional genomics. Privately funded research may soon yield significant results, but these need to be embedded in scientific theory and require independent verification, confirmation and testing.

Crop physiology and ecology at whole-plant and plant-community levels are needed for the following reasons:

1. The complexity of scaling up knowledge from the molecular level to the field ecosystem level will require powerful new quantitative tools and approaches, including modularized multi-scale models, proper interfaces between hierarchical levels, specific software allowing up- and downscaling, and mathematical solutions for integration of steps differing in scale but belonging to the same process. Genes that control developmental processes and rates need to be identified with priority, and their effects will be among the first candidates to be included in whole-plant and crop models.

2. Gene function tested on the basis of comparing genetic variants (either transgenic or classical) should not only be measured in artificial growth systems (e.g., small pots in greenhouses or growth chambers) as this may not be relevant in the real world of production agriculture at the field level. It is, therefore, crucial to understand better how to test genotypes in relevant environments that can predict performance in the field.

3. A plant can only adapt successfully to changing conditions when it is able to integrate the information at different levels of organization into the functioning of the whole plant. Therefore, it must have a finely tuned coordinated control of all individual genes that contribute to the desired phenotype. Recent research on

the regulation of flowering (e.g., in wheat) provides an exciting example of how such a coordinated control system might work for a specific process. For other traits under more complex genetic control, knowledge of the coordinating control function is still lacking.

4. Achieving finely tuned coordination of introgressed genetic variation including transgenes in a gene cascade or network is a difficult task because of a number of factors that affect gene expression, including transgene copy number, RNA silencing, transgene insertion site and the employment of certain regulatory sequences to drive transgene expression. Therefore, predicting phenotype of a transgenic plant on the basis of whether transgenes are present is a major challenge and a costly undertaking. Overcoming this limitation by screening large numbers of transgenic plants becomes less efficient as the number of genes controlling the trait increases. Greater efficiency in achieving the desired level of transgene expression will be critical to improving prediction of the phenotype of transgenic plants engineered for improvement of a complex trait. Even when successful, these predictions can only be based on the expression of genes for which genetic variants differ.

TOP-DOWN OR BOTTUM-UP?

The lack of collaboration between scientists in the fields of genomics and biotechnology on the one hand and scientists in whole-plant physiology and crop ecology on the other hand is probably best illustrated by the debate on how to make use of the wealth of new information obtained by molecular biologists in computational systems analysis. Basically there are two approaches: the top-down approach and the bottom-up approach. Both approaches are facing fundamental problems.

Ecophysiological modelling is a *top-down* approach that predicts crop function based on generic relationships that describe the fundamental processes governing plant growth in relation to environmental conditions. Photosynthesis, respiration, assimilates partitioning to organs, and ontogenic development are key drivers of such models. Individual genotypes can then be represented by a set of response parameters that are valid under a wide range of conditions. The phenotype and its response to environmental conditions are broken down into simpler processes that explicitly take into account actual environmental conditions and behaviour. Such models do not have the detail necessary to simulate expression of single genes or gene networks although such capabilities could be included if the function of single genes or gene networks is known and their coordinated expression can be quantified in relation to environmental conditions.

In contrast, *the bottom-up* approach integrates knowledge at the molecular and cellular level, and a new scientific discipline – systems biology – has been developed for such research and successfully applied in single-cell organisms or relatively simple processes in plants. Examples of the latter are the explanation of phyllotaxis on the basis of gene-regulated accumulation of auxins and the explanation of the progress to flowering based on the knowledge of the expression

level of a set of genes with known function. However, to extend this approach to more complex traits in higher plants and plant communities, we need greater knowledge of how to scale up prediction of gene function at the field level under a range of environmental conditions using information from quantitative estimates of gene expression preferentially obtained under these conditions.

Given these fundamental problems in both approaches we need to re-think the way green plants are organized. The organization of green plants arises as a sequence of developmental processes that allow the plant to behave as an integrated system with multiple feedback controls and cascades to coordinate the growth process. This coordinated integration is achieved by a communication system based on various types of signals and messengers. The plant as a whole also perceives changes in its abiotic and biotic environment, which then evoke responses based on signals. These signals must function across levels of organization, from the genome, cell, tissue and organ levels to the plant and plant community levels. These levels of organization or functional control systems have different principles but yet interact. Response to drought stress provides an example. Drought will induce changes in gene expression, electron transport pathways in photosystems, tissue turgor, specific leaf area, root:shoot ratio and plant-to-plant interaction. But changes in plant-to-plant interaction will affect root:shoot ratio, specific leaf area, photosynthesis of the individual leaf, etc. More insight into the functional interaction between the different levels of organization is needed – something which cannot be easily achieved by a top-down or a bottom-up approach. Understanding gene expression under agronomic conditions is virtually impossible.

THE MIDDLE-OUT APPROACH

We, therefore, need more research that starts from the different levels of biological organization for which we have detailed existing data and understanding, and then use this information to reach up and down to other levels. In human physiology this has been called *the middle-out approach*. Such integral, quantitative studies, on the one hand, integrate knowledge and understanding at the lower level of organization, and, on the other hand, are optimally embedded as an essential component in plant systems at the next-higher hierarchical level. In this way, a knowledge chain can be created that will integrate plant processes in a coherent way, supported by a chain of models or modules that can communicate with each other across levels of organization. A first example may be the modelling of fruit quality, which has been based on modules for daily changes in the available assimilate, hourly changes in water relations and daily partitioning of carbon into different types of sugars. Model parameters have also been linked to genetic variation (QTLs, mutants, transgenics).

In this middle-out approach the proper choice of level of detail is essential. Fine detail might not be required in all cases, robustness (especially across environments) might be more important. This can even be true when this would mean that the models will be rather coarse-grained. Keeping it as simple as possible is a must. How simple relations are, may be best assessed at the middle level. We need crop physiologists well trained in molecular physiology and systems analysis to assess

the proper level of detail. Only they can judge the trade-off between resolution and robustness, between detail and rigour.

New physiology-based mechanistic models will be needed to integrate and quantify functional relationships across levels of organization. These models should allow us to discard obsolete details at each level of organization. They should also be developed in such a way that higher systems-level models can impose the type of lower-level information needed to improve the inputs from low to high, while the lower-level models serve to inspire the higher-level models to seek to stick to the true way plants regulate themselves.

THE CHALLENGES IN SCALING UP

Gene expression studies performed under controlled conditions can create knowledge that is less affected by environment than crop performance data, which is essential for a basic understanding of crop physiology. However, the bottom-up approach in systems biology requires a proper upscaling, linking and interfacing of the following steps: DNA – RNA expression – RNA stability – protein – protein modification – protein stability – protein functioning – metabolites. From the metabolites to traits under variable environmental conditions is then the next, giant step. Given the complexity of this chain of knowledge, it will be difficult to make use of '-omics' (based on large scale gene expression, proteomics, etc.) to improve our understanding of crop physiology; the more so as all processes can be tissue-specific and metabolites move around through the plant and interact with each other. A focus on the single-cell level, as is the case in systems biology, is already a tremendous challenge. In this respect, some of the work published in top international journals, in which claims were made of unravelling simple traits strongly associated with yield potential, should be considered with proper caution and questioned by crop physiologists. As argued before, more progress is needed before crop physiologists can make use of the '-omics' potential. Other technological breakthroughs, such as hybrid breeding in maize, also took a long time before they were widely accepted and utilized in crop improvement. Nevertheless, the proportion of the available resources allocated to crop physiology and ecophysiology is worrying to many crop scientists as the total amount of funding for plant sciences will most likely not increase in the foreseeable future.

Trying to understand the entire organism at all levels of aggregation might also be the wrong approach. Understanding the specific effects of environmental changes based on molecular information is easier to achieve. Even easier is to try to understand the molecular-physiological basis of genetic differences in such specific effects. The latter is currently the most important as this is amenable to crop improvement through breeding. Of course one can also select for high values of end traits (for example yield) without knowing how yield formation works through the brute force of mass selection in relevant target environments.

Scientists active in the field of systems analysis sometimes argue that for scaling up one does not need all the details from the lowest or intermediate levels of aggregation. In general, scaling up across several levels of aggregation simply

results in the loss of impact of mechanisms or relationships at the lower levels, because they are diminished by the most influential mechanisms that operate at higher levels of aggregation. For example, to understand the effects of the *Rht* dwarf genes in wheat on yield, it is sufficient to compare the alleles of these genes in isogenic background, which does not require knowledge about the molecular function of the gene. It is sufficient to carry out well-designed experiments to unravel the crop physiological behaviour of various, well-defined genetic materials (isogenic lines).

Moreover, despite the large increase in detailed knowledge, we do not necessarily need models of increasing complexity. Scaling is about summarizing important knowledge that captures what needs to be taken to the next scale. An important question then is: How much detail is needed to get from gene or molecule to phenotype? In some cases, one can model processes at the crop level based on the information of the effect of the genetic variant and simply build relationships that circumvent the intermediate levels, thus, ignoring the consequences of lower-level traits at intermediate levels (such as circumventing the cell). This approach has been successfully applied for QTL-based modelling of flowering in barley and leaf expansion rate in maize using data from populations of recombinant inbred lines. But for more complex traits, such as grain yield, this modelling approach was not successful. Mutants and transgenic plants, which are even better near-isogenic lines, can also be used, and this is how at least breeding or biotech companies move from '-omics' to crop production.

THE CHALLENGES IN SCALING DOWN

A top-down approach with a keen focus on the 'bottom' to allow further understanding seems most feasible, provided we clearly understand the complexity of the traits involved and have detailed insight about the processes that operate at lower scales. We may wish to start with the identification of genes that are critical (rate limiting) for basic, well-studied processes (such as flower induction, cell elongation), to initiate the links between crop physiology and basic sciences. But that might only work well for developmental processes such as flowering or simple growth processes such as leaf expansion, which are easy to quantify precisely and for which the effects of environmental factors are well known and described. However, even for simple traits, top-down approaches may not always be successful. Although there is no reason for gloomy pessimism on the longer term, at this point this top-down approach seems too ambitious for complex traits such as yield. Some small successes have been reported: research on rice has shown fairly simple inheritance and relatively large effects of QTLs for important yield components, such as seed number and seed size, but the relationships between these traits and seed yield are complex, influenced by feedback mechanisms and dependent on genotype, environment and management.

The immediate challenge is, therefore, to assess the level of detail needed to bridge the gap between physiological approaches (from the crop level) and molecular approaches (to the molecule or gene) depending on the research

objectives. An international effort to elaborate in one case study how this can be achieved using the input of a transdisciplinary team is advocated. Such an effort would be the best way to establish the required working relationships and mutual understanding of problems that is required of both crop physiologists and molecular geneticists, and also to demonstrate the value of this approach. A case study related to genotype × environment interactions is most suitable to achieve these goals.

ACTIONS REQUIRED

Some new approaches are needed, and we urge molecular, plant and crop scientists to collaborate more strongly. We recommend the following general, long-term actions:

1. Establishing private–public partnerships to enhance the role of genomics and its application by applying '-omics' to genetic diversity tested in field conditions.
2. More investment in research on whole-plant physiology, crop ecology and crop simulation to allow efficient integration of knowledge on molecular biology. An interesting complication is that application of genomics can best be tested with transgenics but tests are hardly allowed and/or very expensive and risky for public institutions, at least in Europe.
3. The formation of multidisciplinary teams that operate through networks of excellence in developing quantitative tools that integrate complex information at different levels of organization.
4. The exchange of young scientists between research groups that work at different hierarchical levels to develop a wider set of (T-shaped) skills to deal with complexity and levels of organization in crop science.

In order to be able to identify QTLs and candidate genes that drive complex traits so that they can be included in simulation models, we propose the following short-term actions:

1. Improving the level of detail in characterizing experimental environments, preferably through commonly shared protocols. This will allow modellers to analyse the genotype × environment interaction in a more consistent and precise way.
2. Improving the characterization of parents used for creating mapping populations. For example, careful characterization is needed in terms of the genes involved in developmental requirement (e.g., vernalization (*Vrn*) and photoperiod (*Ppd*) requirement). This will allow the design of populations with no significant genotype × environment interaction for phenology, thus, avoiding this strongly confounding effect in cases where this is desired. The use of near-isogenic materials for the study of the effect of major QTL and genes (mutants or transgenic lines) is highly recommended in specific cases. In other cases, the genotype × environment interactions in phenology might be of particular interest and can then be quantified using QTL-based crop models.
3. International collaboration to carry out a number of multi-location trials with well-designed and characterized populations and with proper characterization of

experimental environments to analyse genotype × environment interactions for other plant characteristics, not related to phenology.

4. Advanced physiological and statistical approaches for determining what aspects of the environment are most influential on the genotype × environment interaction that affect the trait in question, and the stages of crop development at which these interactions are most important.

5. A search for funding to finance international, transdisciplinary teams which will carry out a case study in which scaling across several levels of organization is achieved to identify which level of detail is needed to bridge the gap between molecular approaches and crop physiological approaches and between the genotype and the phenotype.

ACKNOWLEDGEMENTS

FURTHER READING

Gutiérrez, R.A., Shasha, D.E. and Coruzzi, G.M., 2005. Systems biology for the virtual plant. *Plant Physiology,* 138 (2), 550-554.

Hammer, G.L., Sinclair, T.R., Chapman, S.C., et al., 2004. On systems thinking, systems biology and the in silico plant. *Plant Physiology,* 134 (3), 909-911. [http://www.plantphysiol.org/cgi/reprint/134/3/909.pdf]

Hammer, G.L., Chapman, S., Van Oosterom, E., et al., 2005. Trait physiology and crop modelling as a framework to link phenotypic complexity to underlying genetic systems. *Australian Journal of Agricultural Research,* 56 (9), 947-960.

Noble, D., 2002. Modeling the heart: from genes to cells to the whole organ. *Science,* 295 (5560), 1678-1682.

Reymond, M., Muller, B., Leonardi, A., 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 Physiology,* 131 (2), 664-675.

Trewavas, A., 2005. Green plants as intelligent organisms. *Trends in Plant Science,* 10 (9), 413-419.

Wollenweber, B., Porter, J.R. and Lubberstedt, T., 2005. Need for multidisciplinary research towards a second Green Revolution. *Current Opinion in Plant Biology,* 8 (3), 337-341.

Yin, X., Struik, P.C. and Kropff, M.J., 2004. Role of crop physiology in predicting gene-to-phenotype relationships. *Trends in Plant Science,* 9 (9), 426-432.