

## REVIEW PAPER

# Yield–trait performance landscapes: from theory to application in breeding maize for drought tolerance

Carlos D. Messina\*, Dean Podlich, Zhanshan Dong, Mitch Samples and Mark Cooper

Pioneer Hi-Bred, A DuPont Business, 7250 NW 62nd Avenue, Johnston, IA 50131, USA

\* To whom correspondence should be addressed: E-mail: charlie.messina@pioneer.com

Received 30 March 2010; Revised 8 September 2010; Accepted 24 September 2010

## Abstract

The effectiveness of breeding strategies to increase drought resistance in crops could be increased further if some of the complexities in gene-to-phenotype ( $G \rightarrow P$ ) relations associated with epistasis, pleiotropy, and genotype-by-environment interactions could be captured in realistic  $G \rightarrow P$  models, and represented in a quantitative manner useful for selection. This paper outlines a promising methodology. First, the concept of landscapes was extended from the study of fitness landscapes used in evolutionary genetics to the characterization of yield–trait-performance landscapes for agricultural environments and applications in plant breeding. Second, the E(NK) model of trait genetic architecture was extended to incorporate biophysical, physiological, and statistical components. Third, a graphical representation is proposed to visualize the yield–trait performance landscape concept for use in selection decisions. The methodology was demonstrated at a particular stage of a maize breeding programme with the objective of improving the drought tolerance of maize hybrids for the US Western Corn-Belt. The application of the framework to the genetic improvement of drought tolerance in maize supported selection of Doubled Haploid (DH) lines with improved levels of drought tolerance based on physiological genetic knowledge, prediction of test-cross yield within the target population of environments, and their predicted potential to sustain further genetic progress with additional cycles of selection. The existence of rugged yield-performance landscapes with multiple peaks and intervening valleys of lower performance, as shown in this study, supports the proposition that phenotyping strategies, and the directions emphasized in genomic selection can be improved by creating knowledge of the topology of yield–trait performance landscapes.

**Key words:** Complex traits, fitness landscape, gene-to-phenotype, maize, maize physiology, plant breeding, yield trait performance landscape.

## Introduction

The formulation of the theory of selection, genetics, and quantitative genetics and further systematic application by plant breeders provided a framework for the sustained genetic improvement of yield and agronomic traits of crop plants for most of the 20th century (Duvick *et al.*, 2004; Janick, 2004). Yield improvement for drought tolerance in modern maize has resulted from cycles of selection among genotypes sampled from elite germplasm and evaluated for performance in a sample of environments taken to represent the target population of environments (TPE) (Bolaños and Edmeades, 1993a; Bruce *et al.*, 2002; Duvick *et al.*, 2004). Although applied breeding has proved to be effective for

yield improvement, scientists continually seek opportunities to improve the efficiency and effectiveness of breeding methodologies (Bolaños and Edmeades, 1996; Cooper and Hammer, 1996; Podlich *et al.*, 2004; Jannink *et al.*, 2010; van Eeuwijk *et al.*, 2010b). Recent investigations have provided an extension of the theoretical framework for quantitative genetics that is built on allelic variation for genes functioning within networks (Cooper *et al.*, 2005) and gene-to-phenotype ( $G \rightarrow P$ ) models structured around crop growth and development principles (Hammer *et al.*, 2006). In this paper, we report on an implementation of this framework within a maize breeding programme with the

objective of improving the drought tolerance of maize hybrids for the US Western Corn-Belt.

### Gene-to-phenotype modelling

Advances in information technologies and prediction methods based on a mixed-model framework (Podlich *et al.*, 2004; van Eeuwijk *et al.*, 2005, 2010a, b; Boer *et al.*, 2007) contributed to the realization of the potential use of DNA markers to predict the performance of new genotypes from founder genotype information. A toolkit of methodologies that quantifies trait genetic architecture based on mapping studies is now available to help breeders confront the problem of predicting the performance of new genotypes (Podlich *et al.*, 2004; Cooper *et al.*, 2009; van Eeuwijk *et al.*, 2010a, b). Theoretical studies indicate that the effectiveness of breeding strategies could be increased further if complexities in  $G \rightarrow P$  relations associated with epistasis, pleiotropy, and genotype  $\times$  environment interactions (GEI) could be captured in realistic  $G \rightarrow P$  models (Cooper *et al.*, 2005, 2009). A generalized quantitative approach to consider  $G \rightarrow P$  relations was proposed by Cooper *et al.* (2005) on the basis of the NK gene network model (Kauffman, 1993), extensions to consider the environmental context E(NK) (Cooper and Podlich, 2002), and limitations in our capacity to detect the genetic signal of complex quantitative traits fully (Podlich *et al.*, 2004),

$$P_{ijk} = [E_j(NK)_i]_D + [E_j(NK)_i]_U + [E_j(NK)_i]_I + \epsilon_{ijk} \quad (1)$$

where  $P_{ijk}$  denotes phenotype for observation  $k$  on genotype  $i$  in environment type  $j$ ,  $E_j$  identifies different environment type  $j$  within a reference TPE,  $N$  identifies the different genes influencing a trait within a given reference population of genotypes,  $K$  identifies the level of epistatic interaction between subsets of the total  $N$  genes that makeup genotype  $i$  in the form of a network graph where the edges of the graph identify the genes that interact, and  $\epsilon_{ijk}$  is the residual source of variation. The ( ) parenthesis notation indicates that the number of genes and their interactions are conditioned or can vary with the environment. The [ ] bracket notation is used here to identify subsets of the total  $N$  genes and interactions that are detected ( $D$ ), undetected ( $U$ ), and interactions arising from the intersection ( $I$ ) between subsets of detected and undetected genes. Components including undetected genes can be considered to represent the genetic background effect. This framework incorporates explicitly epistasis and GEI within the genetic model and it can be applied to the modelling of trait genetic architecture for the continuum from simple to complex trait genetics. A unique feature of this framework is the explicit consideration of unknown components of the genetic architecture of the trait, which has important implications for prediction of response to selection within the context of a breeding programme (Cooper *et al.*, 2005, 2009). The assumption about there being ‘no unknown QTL’ makes for overly-optimistic predictions of the value of known QTL and of genetic progress of QTL-based selection strategies.

The  $G \rightarrow P$  prediction problem remains a major challenge for most important traits in plant breeding (Cooper *et al.*, 2002). Different concepts have been proposed to tackle this problem (White and Hoogenboom, 1996; Chapman *et al.*, 2003; Reymond *et al.*, 2003; Tardieu, 2003; Hoogenboom *et al.*, 2004; Peccoud *et al.*, 2004; Podlich *et al.*, 2004; Yin *et al.*, 2005; Hammer *et al.*, 2005; Welch *et al.*, 2005; Messina *et al.*, 2006; Welcker *et al.*, 2007; Chenu *et al.*, 2008; Bertin *et al.*, 2010). A synthesis of these methods suggests that a generalized framework can be proposed in which the genetic architecture of the trait is represented by specific parameterizations of the NK model, and phenotypes are predicted by applying a function ( $\Gamma$ ) with parameters determined by NK

$$P_{ijk} = \Gamma[NK_i] + \epsilon_{ijk} \quad (2)$$

The nature of the function  $\Gamma$  is determined by the biophysical properties of the trait under consideration, the degree of model simplification chosen to represent the trait, and by the number and connectivity between traits and the environment in multi-trait modelling ( $\Gamma$ ). For example,  $\Gamma$  can represent a leaf elongation rate ( $dL/dt$ ) model for maize (Salah and Tardieu, 1997; Reymond *et al.*, 2003)

$$dL/dt = (T - T_0)(a + bVPD + c\psi) \quad (3)$$

where  $T$  is meristem temperature,  $VPD$  is vapour pressure deficit,  $\psi$  is soil water potential;  $b$  and  $c$  are constants coding for the response of leaf elongation rate to  $VPD$  and soil water potential after correction for  $T$  effects;  $a$  and  $T_0$  are the slope and  $x$ -intercept of the leaf elongation rate response to meristem temperature. This model can be rewritten using the NK framework (equation 2) as

$$dL/dt = (T - NK_{ip})(NK_{ip} + NK_{ip}VPD + NK_{ip}\psi) \quad (4)$$

where now the phenotype ( $dL/dt$  or its integral form) is parameterized using the NK model, which is allowed to vary among parameters  $p$ . That is, the genetic networks associated with each of the model parameters could have common components. Therefore, pleiotropic effects are formally incorporated *via* shared nodes among gene networks, epistasis is determined by  $K$  as describe before and GEI is an emergent property of the model within different environmental conditions.

Crop growth models that are structured to capture dynamic interactions of the physiological determinants of crop growth and development can be used as a framework for multi-trait integration ( $\Gamma$ ) and to predict consequences of genotype-by-environment-by-management interactions (Chapman *et al.*, 2003; Hammer *et al.*, 2006; Messina *et al.*, 2009). Crop models estimate growth and development using environmental resource capture and conversion efficiency concepts for radiation and water, while allowing for influences of major nutrients such as nitrogen (Muchow *et al.*, 1990; Boote *et al.*, 1999; Keating *et al.*, 2003). The value of the framework to assist plant breeding depends on the extent to which the algorithms included in the model adequately capture the physiological determinants of genetic variation

for adaptive traits of interest to the breeder (Tardieu, 2003; Cooper *et al.*, 2009; Messina *et al.*, 2009; Tardieu and Tuberosa, 2010). Fundamental physiological and genetic studies are often necessary to improve the model architecture (Tardieu, 2003; Messina *et al.*, 2006; Chenu *et al.*, 2008; Hammer *et al.*, 2009; Messina *et al.*, 2009; Bertin *et al.*, 2010; Sinclair *et al.*, 2010; Yin and Struik, 2010). The level of detail needed for crop growth models to be able to integrate processes across levels of organization while predicting emergent functional consequences for the organism is under debate (Hammer *et al.*, 2006; Bertin *et al.*, 2010). The combination of equations 1 and 2 embedded within the crop growth model provide a complete and continuous framework to model  $G \rightarrow P$  relations for multiple traits that can be used to evaluate breeding strategies

$$P_{ijk} = \Gamma \{ [NK]_{iD} + [NK]_{iU} + [NK]_{iI} + \varepsilon_{ijk} \} + [E_j(NK)]_{iD} + [E_j(NK)]_{iU} + [E_j(NK)]_{iI} + \varepsilon_{ijk} \quad (5)$$

A fundamental attribute of this framework is that it explicitly incorporates sources of uncertainty and error in the prediction. This feature has largely been ignored in process-based  $G \rightarrow P$  frameworks (Bertin *et al.*, 2010). A simplified framework based on the concepts represented in equation 4 was applied in investigations seeking to assess the value of molecular breeding and physiological knowledge to improve the rate of genetic gain relative to conventional breeding strategies (Hammer *et al.*, 2005; Cooper *et al.*, 2009).

### Fitness landscapes

The E(NK) model, as described above, has an associated trait performance landscape with a topography and ruggedness that varies in response to the complexity at the trait genetic model *via* combinations of E, N, and K (Kauffman, 1993; Cooper and Podlich, 2002; Cooper *et al.*, 2005). The concept of a fitness landscape, which originated in evolutionary biology, introduces the notion of a ‘potential surface’ or function underlying the dynamics of evolution. This landscape metaphor was first proposed by Sewall Wright (1932) as a framework to advance the theoretical grounds for evolution when assumptions about non-additive variance, pleiotropy, and epistasis are included in the genetic model. The assumptions of additivity were fundamental to the paradigm at the time, exemplified by the infinitesimal and geometric models proposed by Fisher (Orr, 2005; Gavrillets, 2004), which are determinants of a smooth and single peak landscape. Wright, contrary to Fisher, envisioned that pleiotropy and epistasis would generate design constraints leading to rugged fitness landscapes, which, in turn, would determine the phases of adaptation and speciation as populations moved across the landscapes and reached different peaks in a complex surface (Gavrillets, 2004). The implications of rugged landscapes for plant breeding are consequential as they suggest trajectories of genetic improvement conditioned by the presence of hills (high performance) and valleys (low performance). In rugged trait performance landscapes, the

response to selection is variable and conditioned upon population size, position of breeding populations in the landscape and genetic variation for adaptive traits that determine the number and height of peaks accessible to individuals in the population subjected to cycles of selection by the plant breeder (Cooper *et al.*, 2002, 2005).

Despite the important implications of rugged trait performance landscapes in plant breeding and the associated potential benefits, the creation of biologically sound landscapes for use by the breeder has been elusive. Recent work with RNA folding models has provided some views of the structure of fitness landscapes for a well-defined system (Fontana, 2002). Advances in computing capacity, data management, graph theory, and  $G \rightarrow P$  models open the opportunity to generate the first partial views of trait performance landscapes for crops, as discussed in this paper. This information offers the breeder the opportunity to examine the positions of the genotypes that comprise the reference germplasm pool under consideration in the breeding programme relative to the accessible peaks in performance, and to gain insights into the topography of the adjacent but unexplored  $G \rightarrow P$  space.

The complexity of trait performance landscapes and the inability of breeders to explore fully a vast genotype space, that remains mostly unobservable to the breeder, has been identified as a major constraint to the design of breeding programme strategies for complex traits to achieve genetic progress (Cooper and Hammer, 1996). The problem of a limited capacity empirically to explore a vast and unobserved  $G \rightarrow P$  space will not be resolved in the foreseeable future. However, recent theoretical work (Chapman *et al.*, 2003; Podlich *et al.*, 2004; Hammer *et al.*, 2005; Chenu *et al.*, 2009; Messina *et al.*, 2009) suggests that breeding simulation (Podlich and Cooper, 1998) applied to trait performance landscapes can enable breeders to explore the unknown  $G \rightarrow P$  space *in silico*. This  $G \rightarrow P$  space can only be created by means of  $G \rightarrow P$  models grounded on physiological and genetic principles. The predicted breeding programme trajectories become testable genetic hypotheses. Breeders can create genotypes to test these hypotheses by means of creating genotypic novelty through strategic sampling of germplasm, recombination and segregation, evaluation in environments representative of the TPE, and by genotypic and phenotypic selection. This paper documents the first steps towards generating the evidence to answer the question, ‘Does a knowledge-based approach that creates greater  $G \rightarrow P$  knowledge than we currently have help manage resources and accelerate genetic progress relative to what we can achieve using current technologies?’

### From fitness landscapes to yield–trait performance landscapes for plant breeding applications

The objective of a modelling framework developed and structured around  $G \rightarrow P$  models, landscapes, and breeding simulation is to enable plant breeders to achieve greater

genetic progress relative to what could be achieved using molecular breeding and phenotypic selection approaches. Considering this context, two necessary conditions are immediate and common to other breeding technologies: (i) the framework should be relevant to the germplasm developed by the breeder; and (ii) be operational in the context an ongoing breeding programme. In addition, the proposed framework must have prediction capabilities that apply beyond the set of environments and germplasm sampled in the breeding programme, a distinct characteristic from current statistically-based approaches, and should enable continuous improvements concomitant with the generation of knowledge in each cycle of selection (Podlich *et al.*, 2004).

Figure 1 is a graphical representation of the concept map after Casti (1997). The real world component represents the ongoing breeding programme with the key processes represented: (i) testing of a sample of genotypes in a sample of environments; (ii) hypothesis driven ad hoc experimentation; and (iii) selection decisions (Cooper and Hammer, 1996). The structure and operation of the breeding programme and the data produced within the breeding programme inform modelling activities. The mathematical world represents the ongoing mathematical modelling and analyses of trait genetic architecture and G→P prediction. Trait modelling informs phenotyping in the real world and ad hoc experimentation. Exercising the models, genetic, statistical, biophysical, or a combination of these, and the analyses of results inform selection decisions and put forward genetic hypotheses for testing within the breeding programme. Arrows emphasize the iterative nature of this

modelling-testing process, which is fundamental to the continuous generation of knowledge concomitant with the changes in the breeding programme, the environment, and the germplasm over cycles of improvement. Map-As-You-Go implements this iterative model building for application in molecular breeding (Podlich *et al.*, 2004).

The framework considered in this paper (herein the G→P framework) is an implementation of the iterative cycle depicted in Fig. 1 and integrates the knowledge creation capabilities of the crop model within the Map-As-You-Go methodology. The framework has five key components, which can be considered in the form of an algorithm, summarized below.

(i) *Define components for model Γ*: Understand the physiological basis of yield and adaptation in the TPE relevant to the germplasm in the breeding programme (Cooper and Hammer, 1996). Represent this knowledge as meta-processes (equation 3; see Tardieu, 2003, for concept details and an example). Design a high-throughput precision phenotyping strategy that could be executed within the breeding programme to characterize genetic variation for key components of the meta-processes (see Tardieu and Tuberosa, 2010, for a recent review on the utilization of high throughput phenotyping platforms). Integrate a suitable trait model within a crop growth and development framework (Chenu *et al.*, 2008; Hammer *et al.*, 2009; Messina *et al.*, 2009; Sinclair *et al.*, 2010; Tardieu and Tuberosa, 2010).

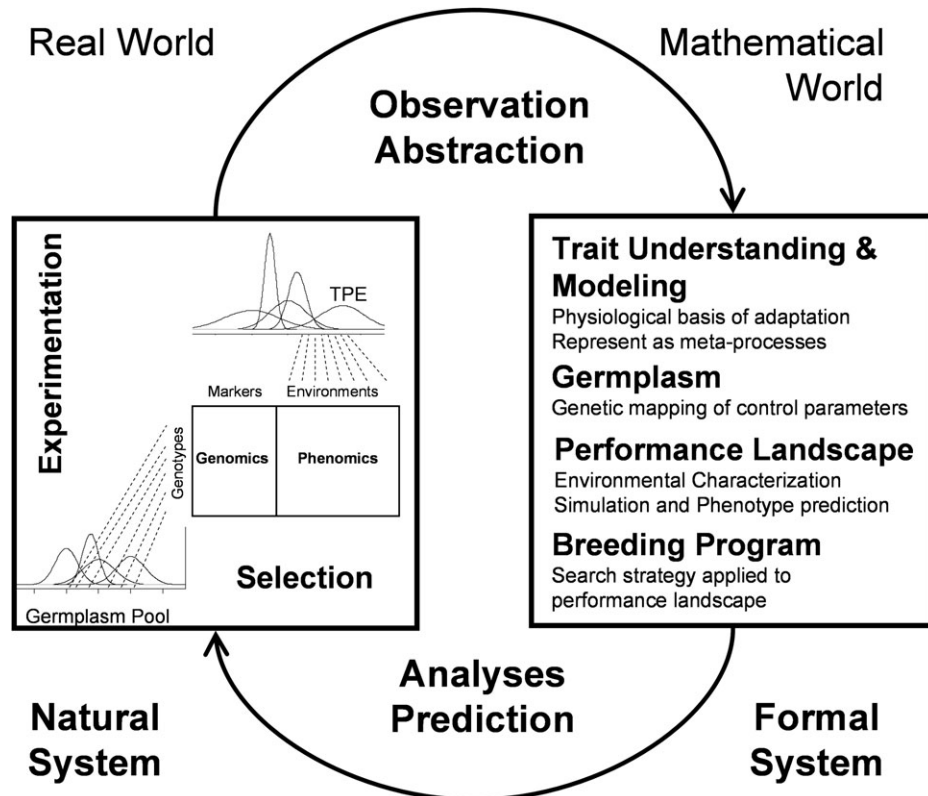


Fig. 1. Concept map of the iterative gene-to-phenotype modelling framework.

(ii) *Develop the model  $\Gamma$* : Undertake genetic modelling to define the genetic architecture of adaptive traits. We refer here to parameters in physical or control equations (Tardieu, 2003) in the context of the NK model (e.g. equation 4; Cooper *et al.*, 2005). The recent advances in genotyping and phenotyping, together with suitable mixed model methodology to map traits, enables the estimation of allele effects for components of trait genetic architecture accessible through mapping in suitable populations (Boer *et al.*, 2007; Cooper *et al.*, 2009).

(iii) *Create performance landscapes for the defined model  $\Gamma$* : Conduct environmental characterization and define the TPE for the geography and management of interest (Chapman *et al.*, 2000; Löffler *et al.*, 2004). Using models developed in (ii) define the range of values for adaptive traits. Define the factorial experiment for all combinations of trait levels (G), environments (E), and management (M). Conduct *in silico* experiments using a model  $\Gamma$  as represented by equation 2. Represent the results in a graphical form to enable decision-making for the selection and creation of new populations of genotypes. Graphical representation of yield–trait performance landscapes to support decisions in the plant breeding programme cycle is a focus of this paper.

(iv) *Breeding modelling and simulation for  $\Gamma$* : Represent the target breeding programme in an abstract form suitable for simulation (Podlich and Cooper, 1998). Define the TPE as in (iii), germplasm as gene–genotype frequencies and organization of alleles within a reference population structure, genetic architecture of each trait as in (ii), and the breeding strategy (i.e. number and size of populations, testers, etc.). Undertake breeding simulation utilizing a specific parameterization of equation 5 for modelled and detected components, and an ensemble approach to accommodate stochastic components associated with undetected components and sources of error. The latter two stochastic terms affect  $G \rightarrow P$  prediction, and the trajectories that are realized exploring the adjacent possible genetic space *in silico*, in a manner that can be anticipated from the behaviour of dynamic complex systems (Kauffman, 1993; Hammer *et al.*, 2005; Cooper *et al.*, 2005).

(v) *Predict and evaluate*: Predict the performance of genotypes that can be created within the context of the reference germplasm, create new genotypes to test genetic hypotheses, evaluate predictions, advance the cycle of the breeding programme, and refine biophysical and genetic models [i.e. return to step (i) above] when necessary.

## Application to yield improvement of maize under drought stress

The  $G \rightarrow P$  framework described in section (ii) was applied within an operational breeding programme, which has as an objective to improve the levels of drought tolerance in

maize beyond the levels currently existing in Pioneer elite germplasm deployed in the Western Corn-Belt. This breeding programme seeks genetic improvement in two complementary heterotic groups, hereafter referred to as heterotic groups 1 and 2 (Duvick *et al.*, 2004). Doubled haploid (DH) lines in test-cross combination are evaluated in managed drought environments under stress and in a sample of environments in the Western Corn-Belt taken to represent the TPE (Cooper, 2009). Irrigation management is designed at key locations to expose genetic variation for target adaptive traits; irrigation management is informed by computer modelling and simulation as part of the iterative modelling and testing cycle (Fig. 1). Newly developed lines are evaluated and selected on performance across multiple locations in one or more years, after which selected lines are used to create new hybrid combinations and cycled back to the germplasm pool to be used as parents for future breeding cycles. The breeding programme implements reciprocal selection with molecular enhanced pedigree selection operating within the two heterotic groups. Selection decisions are informed by predictions based on  $G \rightarrow P$  statistical (equation 1) and biophysical models (equation 4) combined with yield–trait performance landscapes and breeding simulation (Podlich *et al.*, 2004; Cooper *et al.*, 2005; Messina *et al.*, 2009). Further evaluation and selection takes place in an increasing number of locations and hybrid combinations. The breeding programme was formalized within the QU-GENE software (Podlich and Cooper, 1998) to enable breeding simulation. Environmental characterization for drought stress is conducted using the EnClass® system (Löffler *et al.*, 2004) based on patterns of water deficit characterized by the ratio between water supply and demand (Messina *et al.*, 2009). For demonstration purposes, illustrative results will be presented for combinations of well-watered irrigated (ETWW) and drought-stress (ETDR) conditions.

### Trait understanding and modelling

The APSIM crop growth modelling framework (Keating *et al.*, 2003) provided the organizing principle to assess the needs for trait modelling, and the simulation platform for trait integration that enables the mapping of trait phenotypes to yield performance (used here in place of fitness as applied in the evolutionary context) in the TPE. Trait modules were developed to enable APSIM-maize to simulate the effects of leaf angle, root angle, ear growth, and competition among sinks on light use efficiency, light interception, water uptake, biomass allocation, secondary traits, and yield (Hammer *et al.*, 2009; Messina *et al.*, 2009). Root angle (RA) and leaf angle (LA) were considered as quantitative overall descriptors of canopy and root system architecture. The effects of variation in canopy architecture were modelled by means of a layered canopy photosynthesis model (Hammer and Wright, 1994). Leaf angle effects on light interception and distribution among canopy layers were modelled after Duncan *et al.* (1967). Potential water uptake rate in each soil layer was modelled using an

exponential decay function (Passioura, 1983). Root angle was used to represent root exploration of each soil layer that modulates the extent of potential water uptake that is realized (Hammer *et al.*, 2009). The root angle parameter determines the shape of the cross-section plane perpendicular to the planting row (ranging from semi-circular to semi-elliptical) that it is then compared with the area of the soil layer. Algorithms were incorporated to simulate kernel numbers in relation to ear growth and sink dynamics (Messina *et al.*, 2009). A threshold in ear mass (EBT), which determines biomass allocation to the ear, ear growth, and the number of fertilized spikelets at flowering time (Uhart and Andrade, 1995; Borrás *et al.*, 2007, 2009), connects development to growth in such a manner that changes in ear growth can affect the time to silking, the anthesis–silking interval (discussed in Hammer *et al.*, 2009), and rooting depth (Keating *et al.*, 2003). Hammer *et al.* (2009) describe the model parameterization process and verified the model to simulate faithfully diurnal patterns of photosynthesis and response to leaf angle and leaf area index. Further testing consisted in contrasting simulated and observed yield responses to irrigation applied at different times around flowering time and during grain fill (Messina *et al.*, 2009).

#### *Genetic architecture of adaptive traits*

Phenotyping strategies were designed and executed using Pioneer proprietary technologies to quantify the parameters in the process equations. However, protocols for model parameterization for the processes considered in this work have been developed and described elsewhere (Muchow and Carberry, 1989; Edmeades *et al.*, 1993; Uhart and Andrade, 1995; Birch *et al.*, 1998*a, b, c*; Vega *et al.*, 2001; Echarte *et al.*, 2004; Monneveux *et al.*, 2006; Padilla and Otegui, 2005; Borrás *et al.*, 2007, 2009). Three-hundred-and-fifty DH lines from a single cross representing one heterotic group were genotyped and phenotyped. Phenotyping was conducted to characterize the DH lines in a test-cross combination for thermal time from emergence to tassel initiation (ETI), leaf appearance rates (LAR), canopy leaf area distribution (LAD), maximum number of ovules in the apical ear (ME), and ear biomass threshold (EBT). All DH lines were evaluated for the traits as test-cross combinations with a single tester selected from the complementary heterotic group. Genome scans were applied to trait data using QTL mapping methodology within a mixed model framework following the methodology described by Boer *et al.* (2007). The results were used to describe the genetic architecture of the traits and to parameterize the NK model. The simplest genetic architecture suggested by the data, used here for the purpose of demonstrating the framework, was an additive genetic model with three QTL with equal effect sizes for each trait and two alleles at each locus (NK=3:0). This genetic architecture is assumed for all traits and does not consider epistasis or pleiotropic effects at the individual trait QTL level. It is noted, however, that GEI, trait interactions and pleiotropic effects on yield for

the trait QTL are possible and emergent properties that result from the interaction of the adaptive traits as they influence crop growth, development, and yield within the environmental context. The additive three QTL models for each trait define 27 genotypes but only seven expression states for each trait. Given the additive genetic model within a trait, different combinations of alleles for the three QTL correspond to the same trait expression state. If one considers for a given QTL *A* that allele *A* results in increased expression (+) relative to the allele *a* (–), then the expression state for the genotype is defined by the sum across the QTL of the + alleles. For example, the allelic combinations *AAbbcc*, *aaBBcc*, *aabbCC*, *AaBbcc*, *aaBbCc*, and *AabbCc* have the same expression state. As a convention for the additive QTL model, the highest expression state value 1 was defined to correspond to the genotype *AABBCC* and the lowest value 0 to *aabbcc*. Extensions to the additive genetic model considered here are possible and their main impact is on the computational load and the outcomes of the simulation strategies as discussed below.

#### *Simulating performance for quantitative traits*

Two *in silico* factorial experiments were conducted in a high performance computing environment. For both experiments there was one reference genotype population for each of two heterotic groups, referred to hereafter as heterotic groups 1 and 2. The simulation experiment for heterotic group 1 was described by Messina *et al.* (2009). The simulation experiment for heterotic group 2 was based on five adaptive traits: ETI, LAR, LAD, ME, and EBT. The genetic models for the five traits define  $1.4 \times 10^7$  genotypes. Because of the use of the approximation of equal additive QTL effects within each trait with three QTL per trait, only  $1.6 \times 10^4$  expression states were defined. Maize phenotypes were simulated using APSIM-Maize for each expression state, 59 years of weather data for Belgrade and Holdrege, Nebraska, US. Four soil water initial conditions for a loam soil with a soil water-holding capacity of 508 mm were estimated using a long-term simulation following a procedure outlined by Hammer *et al.*, 2009. The planting date was set to 5 May for all years and plant population to 8.6 plants  $m^{-2}$ . Model parameters not phenotyped for DH lines in a test-cross combination in the mapping population were set to the values estimated for two reference hybrids for both heterotic groups.

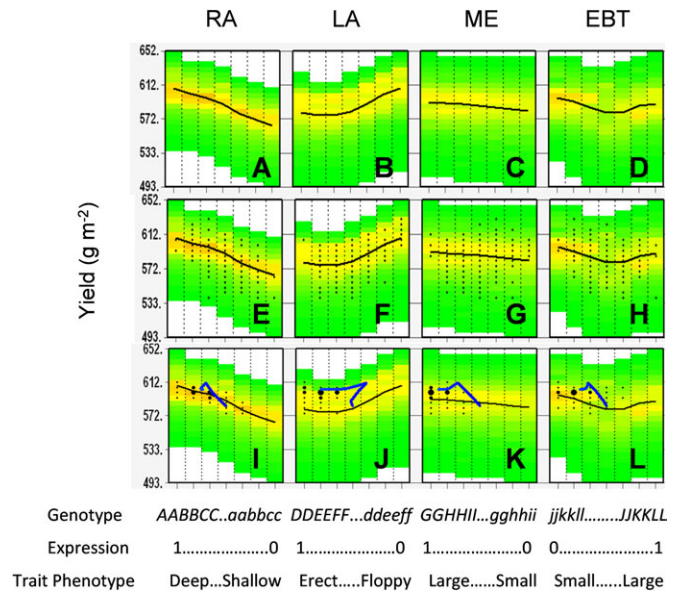
#### *Representing landscapes for plant breeding applications*

Different graphical and numerical representations have been proposed to represent trait landscapes (Wright, 1932; Kauffman, 1993; Cooper and Podlich, 2002; Fontana, 2002; Gavrillets, 2004; Cooper *et al.*, 2009). Gavrillets (2004) provides a clear synthesis in which he classifies these into three landscape types. A first class is the fitness of gene combinations, the original concept proposed by Wright (1932), in which each genotype is assigned a value of fitness.

For the application of the landscape concept in breeding, this class of landscape is referred to as the trait performance of the genotypes defined by the QTL combinations. A second type of landscape that relates the gene (QTL) allele frequencies to the mean fitness (i.e. yield performance) of the reference population can be constructed if the individual fitness (i.e. yield) values for each genotype and the population structure are known. A third type of landscape is for quantitative traits in which trait values are related to fitness (here yield performance). This third type forms the basis for representing the performance landscapes for plant breeding applications discussed in this paper. Herein this yield–trait performance landscape view is referred to as the gene-to-phenotype plot, or GP plot for shorthand, since there is a  $G \rightarrow P$  mapping of genetic variation for genes (QTL) to yield. The connection between genes (QTL) and genes is *via* their modelled influence on the physiological traits defined by the function  $\Gamma$ .

The GP plot represents a landscape as a set of conditional cross-sections for grain yield in the  $G$  and  $P$  dimensions (Fig. 2A–D). Each cross-section is a projection  $\mathbb{R}^n \rightarrow \mathbb{R}^2$  of  $n-1$  traits onto a two-dimensional space determined by the trait  $t$  and yield (Fig. 2A–D). Grain yield distributions are thus conditional on the expression states for a defined trait. Taking the landscape created for root angle (Fig. 2A) in environment ETDR as an example, seven expression states were observed for root angle, arranged on the  $x$ -axis of the GP plots, and therefore seven yield distributions, depicted on the  $y$ -axis of the GP plot. The yield distribution is determined by the simulated yield variation created by all other traits conditional on each expression level for root angle. Genotype relative frequencies for yield are shown as a heat map to visualize the grain yield distributions associated with variation for the other traits (Fig. 2).

Because the model presented in equation 2 is deterministic, a given genotype map to yield ( $y$ -axis) through the value determined by the genetic model considered for the trait phenotype ( $x$ -axis). The landscape represented by the GP plot is thus a projection of the yield–trait performance landscape for gene combinations, as conceived for fitness by Wright (1932), in two dimensions. Individuals or clusters of individuals in a breeding population can be represented as points (here shown as black dots) in the landscape with the size of the dot proportional to the genotype frequency (Fig. 2E–L). The dots in Fig. 2E–H refer to individuals in the initial population, while those in Fig. 2I–L represent groups of individuals at the end of selection. By representing in the GP plot both the frequency of the set of genotypes in a breeding population at any given stage in the breeding programme cycle and the breeding trajectory (blue line traces the average for the population at each cycle of selection for a single breeding simulation run), the GP plot is a variant of the performance (fitness) landscape as mean performance (fitness) of populations (Fig. 2I–L). The GP plot views are constructed for each environment type in the TPE to describe how the landscape shape changes in response to environment change (e.g. Fig. 3A compared with E).



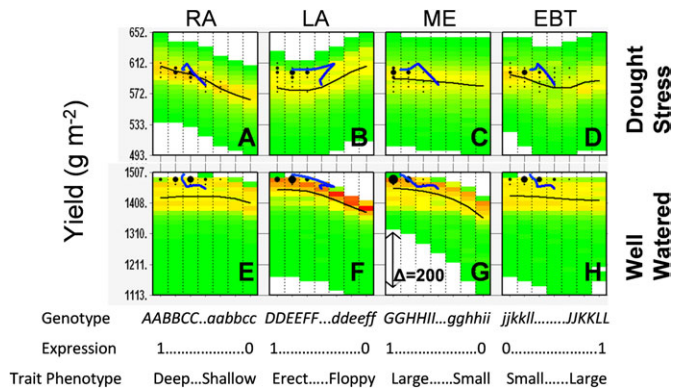
**Fig. 2.** Yield–trait performance landscapes for a population from heterotic group 1, simulated for drought stress (ETDR) environments. Cross-sections correspond to traits: root angle (RA) (A, E, I), leaf angle (LA) (B, F, J), potential number of reproductive sinks (ME) (C, G, K), and ear mass threshold (EBT) (D, H, L). Frequencies of genotypes with a yield level increases from green to red. The central tendency in the landscape is shown as a black line. Individual or clusters of genotypes are shown as black dots with sizes proportional to the frequency of the genotype in the population (E–L). The population mean breeding trajectory over cycles of selection is shown in blue (I–L).

### Analyses of performance landscapes

Two approaches were used to study the properties of the yield–trait performance landscapes using the GP plots. The first approach focuses on measuring static properties of the landscape for a given environment and population and seeks to describe the geometric properties of the landscape. In the second approach, dynamic properties were studied by simulating the trajectories of breeding populations over cycles of selection.

The geometric properties considered here include local and global maxima and minima, trends in central tendency, trends in maxima and minima, ridges, valleys, and plateaus. The global maximum is identified as the maximum value in the  $y$ -axis across all frequency distributions for a given trait conditioned to  $G$ . The global maximum has coordinates in  $n$  dimensions, one for each trait. For a given genetic and environment context the global maximum could be conceived as an ideotype conditional to the germplasm and the TPE. In the example presented in Fig. 3A–D for drought-stress environments, this maximum corresponds to genotype QTL *AABBCC ddeeff GGHHII JJKKLL*.

Local maxima (minima) correspond to each maximum (minimum) value for yield conditional on each expression state and  $G$ . Local maxima provide information about the potential for yield improvement for a population under selection with access to a subset of the peaks on the



**Fig. 3.** Yield–trait performance landscapes for a population from heterotic group 1, simulated for drought-stress (ETDR) and well-watered (ETWW) environments. Cross-sections correspond to traits: root angle (RA) (A–E), leaf angle (LA) (B–F), potential number of reproductive sinks (ME) (C–G), and ear mass threshold (EBT) (D–H). The central tendency in the landscape is shown as a black line. The population mean breeding trajectory over cycles of selection is shown in blue and individual or clusters of genotypes are shown as black dots with sizes proportional to the frequency of the genotype in the population. Frequencies of genotypes within a yield level increases from green to red.

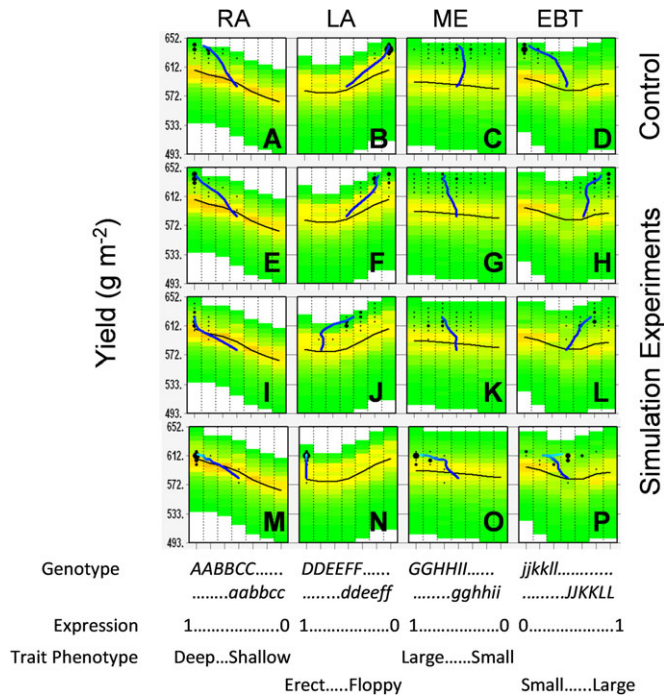
landscape. Local minima emphasize the role of the traits and the expression state on the probability of yield failure. Thus, knowledge of these features of the landscape helps assess the role of the trait in determining yield stability. For example, increasing rooting depth associated with increasing frequency of (+) alleles at loci A, B, and C increases the minimum yield under drought stress. This breeding strategy has been advocated and genetic variation for root angle has been documented (Sinclair and Muchow, 2001; Tuberosa *et al.*, 2002; Giuliani *et al.*, 2005; Manschadi *et al.*, 2006; Messina *et al.*, 2009). Variations in the location of minima and maxima for any given trait provide insights to the interpretation of the yield distributions for other traits. Under well-watered conditions, the predominantly flat landscape (green area) across expression states for RA is largely determined by the variation in minimum yields for ME (Fig. 3G). Increasing the potential number of sinks underpinned yield improvement in Argentina (Echarte *et al.*, 2004) and it has been proposed as a possible path towards yield improvement under well-watered conditions in the US Corn-Belt (Messina *et al.*, 2009). When local minima are zero the GP plot provides a view of what Gavrillets (2004) refers to as holey landscapes, i.e. landscapes with valleys or holes defined by unviable genotypes, zero yields in this case.

The yield distribution for a given expression state as shown in Fig. 3, can be viewed as analogous to a Poincare section of a multi-dimensional space within which trajectories cross the landscape at that point in  $G \rightarrow P$  space. Local maxima and minima define the size of this section. The landscape for ME (Fig. 3G) provides an example of a tight cross-section for genotype *GGHHII* that contrasts with a wide section for genotype *gghhii*. While population

trajectories are strongly constrained by the genotype *GGHHII*, they are not by the genotype *gghhii*. At the trait expression state associated with genotype *gghhii*, other traits have more influence on yield range performance than ME; this is reflected by the yield range difference between these two genotypes, mainly determined by differences among genotypes in local minima ( $\Delta=200 \text{ g m}^{-2}$ ; Fig. 3G) rather than differences in local maxima ( $\Delta=100 \text{ g m}^{-2}$ ). The size of the Poincare section is an important descriptor of the landscape. This statistic could be used to help the breeder anticipate the expected genetic variation in a breeding population and potential changes that can be expected over cycles of selection.

Trends in local maxima and central tendency statistics (Fig. 3, black line) help identify ridges and valleys, and provide a first view of gradients that can shape the path of populations under selection pressure towards local or global maxima. The identification of valleys is of particular interest to the breeder because they provide a first assessment about the possibility to attain maximum yields given the position of the population in the landscape and the need for populations to move downhill before reaching a new position that provides access to the global maximum or higher local maxima. Figure 3 provides two examples (Fig. 3B, D) that illustrate the presence of saddles that can constrain the population access to the global maxima. Figure 3B suggests that increased levels of drought tolerance are feasible by either increasing or decreasing leaf inclination angle. Both strategies lead to improved water status during reproductive stages associated with reduced growth and water conservation. But only reduced growth due to lower radiation use efficiency and lower average leaf inclination angle leads to the global maxima under drought stress. However, this strategy has the associated cost of reduced productivity under well-watered conditions (Fig. 3F). Water conservation strategies that improve the partitioning of water use between vegetative and reproductive stages (Sinclair *et al.*, 1984; Condon *et al.*, 2004) and opportunities for breeding for specific adaptation in the western Corn Belt (Messina *et al.*, 2009) have been discussed previously.

A second approach to study performance landscapes is through the study of the dynamics of populations over cycles of selection. Breeding simulation enables an estimation of the length of the adaptive walks *via* higher yielding neighbours from a starting reference population towards a local maxima, the overall complexity of the landscape through the quantification of the correlation structure of the landscape (Kaufmann, 1993; Cooper *et al.*, 2005), how this complexity affects rates of genetic gain over cycles of selection (Cooper *et al.*, 2009), and the dependency of such breeding trajectories on the initial structure of the population. Sensitive dependence on initial conditions is a fundamental characteristic of complex systems (Lorenz, 1995; Kauffman, 1993). It can be anticipated that this feature of the breeding system will have some importance for the plant breeder in the use of prediction methods to enable genetic improvement of complex traits for different crosses within



**Fig. 4.** Breeding simulation applied to the study of performance landscapes. Simulations are for a breeding population from heterotic group 1 and selection in drought-stress environments (ETDR). Results are shown for cycle 8 of selection. Cross-sections correspond to traits: root angle (RA) (A, E, I, M), leaf angle (LA) (B, F, J, N), potential number of reproductive sinks (ME) (C, G, K, O), and ear mass threshold (EBT) (D, H, L, P). The central tendency in the landscape is shown as a black line. The population mean breeding trajectory over cycles of selection is shown in blue and individual or clusters of genotypes are shown as black dots with sizes proportional to the frequency of the genotype in the population. Frequencies of genotypes with a yield level increases from green to red. Panels (A–D) correspond to a control simulation experiment where (+) allele frequencies were set at 0.5 in the cycle 0 reference population of genotypes. Three simulation experiments where the frequency of (+) alleles in the cycle 0 reference population of genotypes increased from 0.5 to 0.8 for the trait EBT (E–H), from 0.5 to 0.8 for the trait LA (I–L) and from 0.5 to 0.9 for the trait LA (M–P).

different reference populations of genotypes and different target environments.

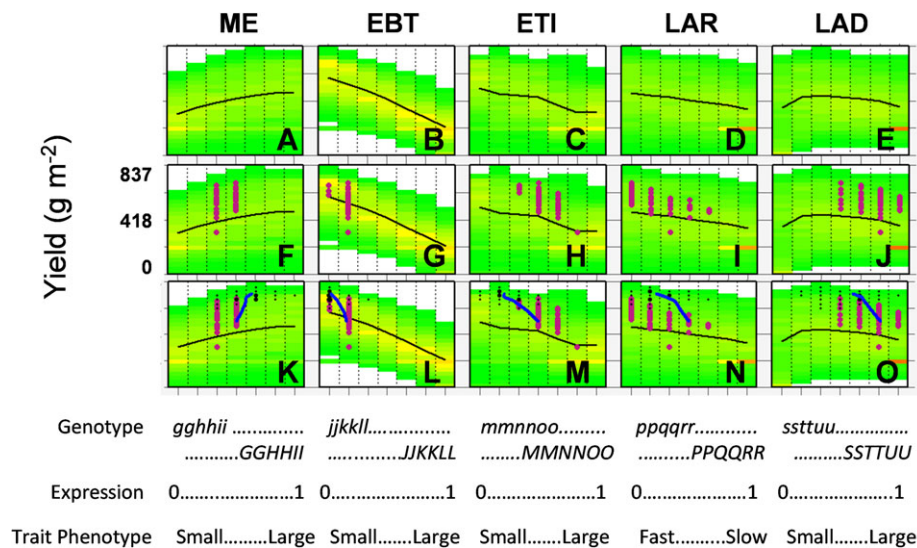
Figure 4 presents the results of a simulation experiment designed to study trajectories of maize populations and how these are affected by changes in the structure of the initial breeding populations. Results for the last run from an ensemble of 50 simulations are shown for cycle 8 of selection under drought stress conditions (ETDR). Drought environments were sampled over time from the classification of environments described in Messina *et al.* (2009). When the population structure was defined by the frequency of (+) alleles and set at 0.5 within the reference population for all traits (control simulation experiment), yield increased with increasing rooting depth (Fig. 4A), decreasing leaf angle and radiation use efficiency (Fig. 4B), and decreasing

EBT (Fig 4D). The reduction in EBT was associated with simulated lower anthesis silking interval (ASI) (data not shown), as discussed by Campos *et al.* (2004) and Borrás *et al.* (2009). This result from simulation conforms to empirical data that show a correlated reduction in ASI with increased ear growth and yield (Edmeades *et al.*, 1993; Campos *et al.*, 2004; Borrás *et al.*, 2007, 2009). Reduction in EBT and increased partitioning to the ear has been implicated in the genetic improvement for yield (Bolaños and Edmeades, 1993b; Edmeades *et al.*, 1993; Echarte *et al.*, 2004; Hammer *et al.*, 2009; Messina, 2009). Increasing the frequency of the (+) alleles in the reference population for EBT from 0.5 to 0.55 led to a bifurcation of trajectories with populations selected towards any of the two highest local maxima (data not shown). Further increases in the frequency of the (+) alleles from 0.55 to 0.8 resulted in the population following a trajectory in the opposite direction from the one shown for the control run (compare Fig. 4D with H). Higher values of EBT increased ASI but also increased rooting depth associated with delayed flowering time, resulting in a trade-off between resource use efficiency and resource capture.

A third simulation experiment was defined such that the breeding population was positioned on the landscape at the centre of the valley identified in the cross-section created for LA (Fig. 4I–L). Two key results of these two simulation experiments are the reduction in the rate of genetic gain (compare yields after 8 cycles of selection as indicated by the end of the trajectory show in blue in Fig. 4A with I) and the change in the trajectory in EBT (compare Fig. 4D with L) in response to the increase in initial frequency of (+) alleles for LA. Further increases in the frequency of (+) alleles for LA led to yield stagnation due to convergence of the population on a local maxima determined by the highest average leaf inclination angle (Fig. 4N) and a population that largely remains within the valley observed for EBT (Fig. 4P).

## Application to selection decisions

The G→P framework was applied to enable selection decisions. For all breeding lines with acceptable agronomic phenotypes (i.e. reduced lodging), the merit of the inbred lines was determined from the lines relative position in the performance landscape, the predicted performance, and the potential to contribute to further yield improvement. Inbred lines in test-cross combinations and reference hybrids were characterized for five adaptive traits. A performance landscape was created for a range of quantitative traits that encompass both the breeding population and reference hybrids and was extended to include an adjacent possible genotype space (Fig. 5A–E). The predicted test-cross yield performance values of the inbred lines for their trait combinations were then positioned in this landscape (Fig. 5F–J) and the plausible trajectories of this population in G→P space, based on selection under drought stress, were investigated *via* simulation (Fig. 5K–O).



**Fig. 5.** Yield–trait performance landscapes applied to plant breeding. Simulated performance for a breeding population from the heterotic group 2 and representation of landscape (A–E). Phenotyping of inbred lines and projection of performance in the performance landscape (F–J). Simulation of breeding strategy for a specific breeding population and projected trajectory over cycles of selection (K–O). Selection experiments were conducted by sampling drought-stress environments (ETDR). Results are shown for cycle 4 of selection (K–O). Cross-sections correspond to traits: potential number of reproductive sinks (ME) (A, F, K), ear mass threshold (EBT) (B, G, L), thermal time from emergence to flower initiation (ETI) (C, H, M), leaf appearance rate (LAR) (D, I, N), and leaf area distribution (LAD) (E, J, O). The central tendency in the landscape is shown as a black line. The population mean breeding trajectory over simulated cycles of selection is shown in blue and individual or clusters of genotypes are shown as black dots with sizes proportional to the frequency of the genotype in the population. Frequencies of genotypes with a yield level increases from green to red. Observed phenotypes for 350 inbred lines in test-cross combination are represented as red dots.

The  $G \rightarrow P$  space of the target germplasm is characterized by a complex and holey landscape. In contrast with previous examples (Figs 3, 4), the presence of zero yields are common when these genotypes are exposed to severe drought stress and largely determined by genotypes with a large expression of EBT. Presence of two saddles (see cross-section corresponding to the thermal time from emergence to tassel initiation; Fig. 5C) and the absence of a clear single trait determinant of yield bring complexity to the landscape. However, trends could be identified and indicate yield under drought stress increasing with increasing ME, and decreasing EBT, ETI, LAR, and LAD. Together, the trends in these traits indicates that yield improvement for this germplasm could be attained by reducing canopy size (both reduction in LAD, ETI, and LAR) and time to flowering, and by increasing the potential number of reproductive sinks and the allocation of resources to the ear. The strategy has components of stress avoidance (early flowering), water conservation and optimal partitioning of water use between pre-anthesis and post-anthesis (early flowering and small canopy size), and stress resistance by means of improved kernel set. The empirical evidence indicates that there is little room for further improvement of drought resistance for the characterized breeding material by further change in the expression of EBT; individuals in this population have already explored the edge of the  $G \rightarrow P$  space (Fig. 5G). Breeding simulation suggests that opportunities for further yield improvements exist. Earliness and smaller canopy size can improve further

the water use pattern provided seed fill duration remains constant (Fig. 5M, O). Bolaños *et al.* (1993) reported similar changes in development and canopy architecture with cycles of selection in CIMMYT drought-tolerant germplasm. Increasing the potential number of reproductive sinks can contribute further to yield improvement, but only after improvement in the water use pattern and stress avoidance were in place (compare yields after 8 cycles of simulated selection indicated by the ends of the blue lines; Fig. 5K and M).

## Discussion and remarks

Plant breeders and physiologists recognize drought tolerance and yield as complex traits (Cooper and Hammer, 1996; Bruce *et al.*, 2002; Campos *et al.*, 2004; Duvick *et al.*, 2004; Ribaut *et al.*, 2004). Many studies have sought to unravel the genetic and physiological basis of drought tolerance and yield improvement in field crops (Sinclair *et al.*, 1984; Ludlow and Muchow, 1990; Condon *et al.*, 2004; Ribaut, 2006). The difficulty of studying complex systems in multiple trait dimensions with the available methodologies has forced discussions about the value of physiological traits to improved drought resistance to orthogonal contrasts. This study has implemented a theoretical framework grounded on realistic  $G \rightarrow P$  models and fitness landscapes within a breeding programme to help breeders deal with the complexity of how these traits

interact with each other during crop growth and development to determine yield within an environmental context. Previous studies advocated for the implementation of similar frameworks but these have remained on theoretical grounds (Chapman *et al.*, 2003; Cooper *et al.*, 2005; Hammer *et al.*, 2005; Chenu *et al.*, 2009; Messina *et al.*, 2009). This paper documents the first application of a G→P framework within an operational breeding programme. The application of the framework to the genetic improvement of drought tolerance in maize supported selection of DH lines with improved levels of drought tolerance based on physiological and genetic knowledge, prediction of yield within the TPE, and their predicted potential to sustain further genetic progress with additional cycles of selection.

The G→P modelling framework considered in this work opens up new opportunities for selection that are difficult to consider with the current statistical approaches to molecular breeding. A limitation of current methods in plant breeding is that they provide a static view of marker–trait associations, leaving to the breeder the subjective interpretation of the interplay and emergent trade-offs among traits and the subjective prediction of the selection trajectories in genotype and phenotype space. Because the G→P model considered here is constructed on genetic and biophysical principles, emergent behaviour resulting from the interplay between the traits is captured and enables prediction for genotypes that were not created or environments in the TPE that were inadequately sampled. The framework proved useful to enable selection decisions due to previous work conducted on model adaptive traits relevant to the genetic variation within a set of elite breeding populations. When considering applying this technology to other populations it is easy to foresee the need to advance the crop growth model framework to incorporate or improve trait models incorporated in APSIM. Some targets for improvement are routines to enable the simulation of root growth (Tuberosa *et al.*, 2002; Bengough *et al.*, 2006; Whitmore and Whalley, 2009; Dupuy *et al.*, 2010; White and Kirkegaard, 2010), the regulation of biomass and nitrogen allocation at the organ level (Horton, 2000; Echarte *et al.*, 2004; Condon *et al.*, 2004; Messina *et al.*, 2009; van Oosterom *et al.*, 2009a, b), and the mechanisms underpinning growth maintenance under drought (Tardieu, 2003; Horton, 2000) and high temperature stress (Suwa *et al.*, 2010). The proposed iterative nature of the framework emphasizes the need to conduct this work in the context of the target germplasm, management and TPE such that breeding objectives, physiological questions, phenotyping strategies, and genetic hypotheses are co-ordinated.

This paper provides a set of empirical views of relevant yield–trait performance landscapes with projections of maize elite germplasm onto the modelled G→P space (Fig. 5). This moves us beyond those views created for the folding properties of RNA molecules (Fontana, 2002). The analyses of the landscapes provided a wealth of knowledge to inform selection decisions, phenotyping, and target traits for genetic analyses. Some of the key outcomes from the simulation studies undertaken in this work to support

selection decisions are: (i) the existence of multiple paths towards yield improvement for drought tolerance within a breeding population; and (ii) the outcomes of the cycles of selection for a given breeding population are conditional on the environmental challenges and the physiological genetic background of the germplasm. Studies seeking to understand the causes of improved drought tolerance in tropical maize support this proposition and provide empirical evidence for the existence of multiple selection trajectories. Reduced canopy temperature, reduced rate of leaf senescence, and increased leaf elongation rates were shown to be correlated with improved yield under drought stress (Fischer *et al.*, 1989). These traits could be associated with improved water status, water capture, and growth maintenance under drought stress. Chenu *et al.* (2009) suggest that growth maintenance under drought can improve maize adaptation to drought stress. A body of evidence supports an alternative path towards yield improvement under drought stress associated with a reduction in the anthesis–silking interval caused by increased biomass allocation to the ear, and ear growth rate per ovule (Bolaños *et al.*, 1993; Bolaños and Edmeades, 1993b, 1996). The results obtained from the simulation experiments show the importance of both improvements in resource capture and use relative to improvements in reproductive efficiency, and how these change with cycles of selection and breeding populations (Figs 4, 5). Simulation results can thus help breeders design phenotyping strategies and prioritize trait phenotyping based on their immediate and predicted long-term effects of the trait on yield improvement.

This paper showed that small changes in initial structure of the germplasm in the reference population, and types of environmental conditions used to phenotype populations and rank genotypes for selection, can have a profound impact on the selection trajectories achieved in a breeding population and thus rate of genetic gain (Fig. 4). This result is of significance as it affects how we collectively think about the utilization of QTL information and other sources of genetic diversity in plant breeding. This study demonstrated that yield performance landscapes can be rugged and that genetic gain can be constrained by such landscape structure (Fig. 4). However, the landscapes were continuous and conform to the notion that the effects of single alterations in the physiological process were dampened so that small changes were observed in plant growth and yield (Sinclair and Purcell, 2005). The absence of discontinuities in the performance landscapes suggests that the impact of a single QTL in a complex trait would be small (Figs 3, 4, 5) supporting the arguments for the utilization of genomic selection for complex traits (Heffner *et al.*, 2009; Jannink *et al.*, 2010). The existence of bifurcation, saddles, and rugged landscapes supports the proposition that the directions emphasized in such genomic selection can be improved by the knowledge of the yield–trait performance landscapes structure as demonstrated here (Podlich *et al.*, 2004).

The utilization of genome-wide prediction methods (van Eeuwijk *et al.*, 2010b) applied to the prediction of yield is one path forward. The G→P framework as represented in

equation 5 provides a prediction methodology that can help the breeder make selection decisions utilizing both the biological knowledge created for single QTL or other sources of genetic variation (set the genetic context that determines the trajectory of a breeding population), and genome-wide information (genome-wide prediction applied to non-modelled traits) to anticipate and assess outcomes of alternative selection strategies in the short and long term by means of landscape methodology. To enable such a comprehensive knowledge-based selection methodology will require implementation of high throughput phenotyping strategies, advanced information management systems, and methodological developments that further integrate crop growth modelling platforms with quantitative genetics models.

## Acknowledgements

We acknowledge the field support provided by Darren Schneider, Carla Gho, Andrea Salinas, Andres Reyes, Karen Thompson, and Neil Hausmann, and the information management support provided Zac Oler, Andy Beatty, Jason Thompson, Duhong Chen, and Tim Fast.

## References

- Bengough GA, Bransby MF, Hans J, MacKenna SJ, Roberts TJ, Valentine TA.** 2006. Root responses to soil physical conditions: growth dynamics from field to cell. *Journal of Experimental Botany* **57**, 437–447.
- Bertin N, Martre P, Genard M, Quilot B, Salon C.** 2010. Under what circumstances can process-based simulation models link genotype to phenotype for complex traits? Case study of fruit and grain quality traits. *Journal of Experimental Botany* **61**, 955–967.
- Birch CJ, Hammer GL, Rickert KG.** 1998a. Temperature and photoperiod sensitivity of development in five cultivars of maize (*Zea mays* L.) from emergence to tassel initiation. *Field Crops Research* **55**, 93–107.
- Birch CJ, Hammer GL, Rickert KG.** 1998b. Improved methods for predicting individual leaf area and leaf senescence in maize (*Zea mays* L.). *Australian Journal of Agricultural Research* **49**, 249–262.
- Birch CJ, Rickert KG, Hammer GL.** 1998c. Modelling leaf production and crop development in maize (*Zea mays* L.) after tassel initiation under diverse conditions of temperature and photoperiod. *Field Crops Research* **58**, 81–95.
- Boer MP, Wright D, Feng L, Podlich DW, Luo Lang, Cooper M, van Eeuwijk FA.** 2007. A mixed-model quantitative trait loci (QTL) analysis for multiple-environment trial data using environmental covariables for QTL-by-environment interactions, with an example in maize. *Genetics* **177**, 1801–1813.
- Bolaños J, Edmeades GO.** 1993a. Eight cycles of selection for drought tolerance in lowland tropical maize. I. Responses in grain yield, biomass, and radiation utilization. *Field Crops Research* **31**, 233–252.
- Bolaños J, Edmeades GO.** 1993b. Eight cycles of selection for drought tolerance in lowland tropical maize. II. Responses in reproductive behavior. *Field Crops Research* **31**, 253–268.
- Bolaños J, Edmeades GO.** 1996. The importance of the anthesis–silking interval in breeding for drought tolerance in tropical maize. *Field Crops Research* **48**, 65–80.
- Bolaños J, Edmeades GO, Martinez L.** 1993. Eight cycles of selection for drought tolerance in lowland tropical maize. III. Responses in drought-adaptive physiological and morphological traits. *Field Crops Research* **31**, 269–286.
- Boote KJ, Jones JW, Hoogenboom GH.** 1998. Simulation of crop growth: CROPGRO Model. In: Peart RM, Curry RB, eds. *Agricultural systems modelling and simulation*. New York: Marcel Dekker, 651–693.
- Borras L, Astini JP, Westgate ME, Severini AD.** 2009. Modeling anthesis to silking in maize using a plant biomass framework. *Crop Science* **49**, 937–948.
- Borras L, Westgate ME, Astini JP, Echarte L.** 2007. Coupling time to silking with plant growth rate in maize. *Field Crops Research* **102**, 73–85.
- Bruce WB, Edmeades GO, Barker TC.** 2002. Molecular and physiological approaches to maize improvement for drought tolerance. *Journal of Experimental Botany* **53**, 13–25.
- Campos H, Cooper M, Habben JE, Edmeades GO, Schussler JR.** 2004. Improving drought tolerance in maize: a view from industry. *Field Crops Research* **90**, 19–34.
- Casti JL, ed.** 1997. *Would-be worlds*. United States: John Wiley and Sons.
- Chapman S, Cooper M, Podlich D, Hammer G.** 2003. Evaluating plant breeding strategies by simulating gene action and dryland environment effects. *Agronomy Journal* **95**, 99–113.
- Chapman SC, Cooper M, Hammer GL, Butler D.** 2000. Genotype by environment interactions affecting grain sorghum. II. Frequencies of different seasonal patterns of drought stress are related to location effects on hybrid yields. *Australian Journal of Agricultural Research* **50**, 209–222.
- Chenu K, Chapman SC, Hammer GL, Mclean G, Ben Haj Salah, Tardieu F.** 2008. Short-term responses of leaf growth rate to water deficit scale up to whole-plant and crop levels: an integrated modelling approach in maize. *Plant, Cell and Environment* **31**, 378–391.
- Chenu K, Chapman SC, Tardieu F, McLean G, Welcker C, Hammer GL.** 2009. Simulating the yield impacts of organ-level quantitative trait loci associated with drought response in maize: a ‘Gene-to-Phenotype’ modeling approach. *Genetics* **183**, 1507–1523.
- Condon AG, Richards RA, Rebetzke GJ, Farquhar GD.** 2004. Breeding for high water-use efficiency. *Journal of Experimental Botany* **55**, 2447–2460.
- Cooper M.** 2009. *Applications of molecular breeding: drought tolerance in corn*. 64th Corn and Sorghum Seed Research Conference. 8–11 December. Chicago, Illinois.
- Cooper M, Chapman SC, Podlich DW, Hammer GL.** 2002. The GP problem: quantifying gene-to-phenotype relationships. *In Silico Biology*. (Available online at <http://www.bioinfo.de/isb/2002/02/0013/>; verified 14 August 2008).

- Cooper M, Hammer GL, eds.** 1996. *Plant adaptation and crop improvement*. Wallingford, UK: CAB International.
- Cooper M, Podlich D.** 2002. The E(NK) model: extending the NK model to incorporate gene-by-environment interactions and epistasis for diploid genomes. *Complexity* **7**, 31–47.
- Cooper M, Podlich DW, Smith OS.** 2005. Gene-to-phenotype models and complex trait genetics. *Australian Journal of Agricultural Research* **56**, 895–918.
- Cooper M, van Eeuwijk FA, Hammer GL, Podlich DW, Messina C.** 2009. Modeling QTL for complex traits: detection and context for plant breeding. *Current Opinion in Plant Biology* **12**, 231–240.
- Duncan WG, Loomis RS, Williams WA, Hanau R.** 1967. A model for simulating photosynthesis in plant communities. *Hilgardia* **38**, 181–205.
- Dupuy L, Gregory PJ, Bengough G.** 2010. Root growth models: towards a new generation of continuous approaches. *Journal of Experimental Botany* **61**, 2131–2148.
- Duvick DN, Smith JSC, Cooper M.** 2004. Long-term selection in a commercial hybrid maize breeding programme. *Plant Breeding Review* **24**, 109–151.
- Echarte L, Andrade FH, Vega CRC, Tollenaar M.** 2004. Kernel number determination in Argentinean maize hybrids released between 1965 and 1993. *Crop Science* **44**, 1654–1661.
- Edmeades GO, Bolaños J, Hernandez M, Bello S.** 1993. Causes for silk delay in lowland tropical maize population. *Crop Science* **33**, 1029–1035.
- Fischer KS, Edmeades GO, Johnson EC.** 1989. Selection for the improvement of maize under moisture deficits. *Field Crops Research* **22**, 227–243.
- Fontana W.** 2002. Modelling ‘Evo-Devo’ with RNA. *BioEssays* **24**, 1164–1177.
- Gavrilets S.** 2004. *Fitness landscapes and the origin of species*. United States: Princeton University Press.
- Giuliani S, Sanguineti MC, Tuberosa R, Bellotti M, Salvi S, Landi P.** 2005. Root-ABA1, a major constitutive QTL, affects maize root architecture and leaf ABA concentration at different water regimes. *Journal of Experimental Botany* **56**, 3061–3070.
- Hammer GL, Chapman S, van Oosterom E, Podlich DW.** 2005. Trait physiology and crop modelling as a framework to link phenotypic complexity to underlying genetic systems. *Australian Journal of Agricultural Research* **56**, 947–960.
- Hammer G, Cooper M, Tardieu F, Welch S, Walsh B, van Eeuwijk F, Chapman S, Podlich D.** 2006. Models for navigating biological complexity in breeding improved crop plants. *Trends in Plant Science* **11**, 1360–1385.
- Hammer GL, Dong Z, McLean G, Doherty A, Messina C, Schussler J, Zinselmeier C, Paszkiewicz S, Cooper M.** 2009. Can changes in canopy and/or root systems architecture explain historical maize yield trends in the U.S. Corn Belt? *Crop Science* **49**, 299–312.
- Hammer GL, Wright GC.** 1994. A theoretical analysis of nitrogen and radiation use efficiency in peanut. *Australian Journal of Agricultural Research* **45**, 575–579.
- Heffner EL, Sorrells ME, Jannink J-L.** 2009. Genomic selection for crop improvement. *Crop Science* **49**, 1–12.
- Hoogenboom G, White JW, Messina CD.** 2004. From genome to crop: integration through simulation modeling. *Field Crops Research* **90**, 145–163.
- Horton P.** 2000. Prospects for crop improvement through the genetic manipulation of photosynthesis: morphological and biochemical aspects of light capture. *Journal of Experimental Botany* **51**, 475–485.
- Janick J, ed.** 2004. *Plant Breeding Reviews 24, Part 1: Long-term selection: Maize*. New Jersey: John Wiley & Sons Inc.
- Jannink J-L, Lorenz AJ, Iwata H.** 2010. Genomic selection in plant breeding: from theory to practice. *Briefings in Functional Genomics and Proteomics* **9**, 166–177.
- Kauffman SA.** 1993. *The origins of order: self-organization and selection in evolution*. Oxford University Press.
- Keating BA, Carberry PS, Hammer GL, et al.** 2003. An overview of APSIM, a model designed for farming systems simulation. *European Journal of Agronomy* **18**, 267–288.
- Löffler CM, Wei J, Fast T, Gogerty J, Langton S, Bergman M, Merrill RE, Cooper M.** 2005. Classification of maize environments using crop simulation and geographic information systems. *Crop Science* **45**, 1708–1716.
- Lorenz EN, ed.** 1995. *The essence of chaos*. Seattle: University of Washington Press.
- Ludlow MM, Muchow RC.** 1990. A critical evaluation of traits for improving crop yields in water-limited environments. *Advances in Agronomy* **43**, 107–153.
- Manschadi AM, Christopher J, deVoil P, Hammer GL.** 2006. The role of root architectural traits in adaptation of wheat to water-limited environments. *Functional Plant Biology* **33**, 823–837.
- Messina CD.** 2009. *Understanding maize yield trends in the U.S. Corn Belt*. 64th Corn and Sorghum Seed Research Conference, 8–11 December, Chicago, Illinois.
- Messina CD, Hammer GL, Dong Z, Podlich D, Cooper M.** 2009. Modelling crop improvement in a G\*E\*M framework via gene–trait–phenotype relationships. In: Sadras V, Calderini D, eds. *Crop physiology: interfacing with genetic improvement and agronomy*. The Netherlands: Elsevier, 235–265.
- Messina CD, Jones JW, Boote KJ, Vallejos CE.** 2006. A gene-based model to simulate soybean development and yield responses to environment. *Crop Science* **46**, 456–466.
- Monneveux P, Sanchez C, Beck D, Edmeades GO.** 2006. Drought tolerance improvement in tropical maize source populations: evidence of progress. *Crop Science* **46**, 180–191.
- Muchow RC, Carberry PS.** 1989. Environmental control of phenology and leaf growth in tropically adapted maize. *Field Crops Research* **20**, 221–236.
- Muchow RC, Sinclair TR, Bennett JM.** 1990. Temperature and solar radiation effects on potential maize yield across locations. *Agronomy Journal* **82**, 338–343.
- Orr HA.** 2005. The genetic theory of adaptation: a brief history. *Nature Reviews Genetics* **6**, 119–127.

- Padilla JM, Otegui ME.** 2005. Co-ordination between leaf initiation and leaf appearance in field-grown maize (*Zea mays*): gGenotypic differences in response of rates to temperature. *Annals of Botany* **96**, 997–1007.
- Passioura JB.** 1983. Roots and drought resistance. *Agricultural Water Management* **7**, 265–280.
- Peccoud J, Vander Velden K, Podlich DW, Winkler C, Arthur L, Cooper M.** 2004. The selective values of alleles in a molecular network model are context-dependent. *Genetics* **166**, 1715–1725.
- Podlich DW, Cooper M.** 1998. QU-GENE: a platform for quantitative analysis of genetic models. *Bioinformatics* **14**, 632–653.
- Podlich DW, Winkler CR, Cooper M.** 2004. Mapping as you go: an effective approach for marker-assisted selection of complex traits. *Crop Science* **44**, 1560–1571.
- Reymond M, Muller B, Leonardi A, Charcosset A, 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.
- Ribaut J-M, ed.** 2006. *Drought adaptation in cereals*. United States: The Hawthorn Press Inc.
- Ribaut J-M, Hoisington D, Bänziger M, Setter TL, Edmeades GO.** 2004. Genetic dissection of drought tolerance in maize: a case study. In: Nguyen HT, Blum A, eds. *Physiology and biotechnology integration for plant breeding*. New York: Marcel Dekker Inc, 571–609.
- Salah HBH, Tardieu F.** 1997. Control of leaf expansion rate of droughted maize plants under fluctuating evaporative demand. A superposition of hydraulic and chemical messages? *Plant Physiology* **114**, 893–900.
- Sinclair T, Messina CD, Beatty A, Samples M.** 2010. Assessment across the United States of the benefits of altered soybean drought traits. *Agronomy Journal* **102**, 475–482.
- Sinclair TR, Muchow RC.** 2001. System analysis of plant traits to increase grain yield on limited water supplies. *Agronomy Journal* **93**, 263–270.
- Sinclair TR, Purcell LC.** 2005. Is a physiological perspective relevant in a 'genocentric' age? *Journal of Experimental Botany* **421**, 2777–2782.
- Sinclair TR, Tanner CB, Bennett JM.** 1984. Water-use efficiency in crop production. *Bioscience* **34**, 36–40.
- Suwa R, Hakata H, Hara H, El-Shemy HA, Adu-Gyamfi JJ, Nguyen NT, Kanai S, Lightfoot DA, Mohapatra PK, Fujita K.** 2010. High temperature effects on photosynthate partitioning and sugar metabolism during ear expansion in maize (*Zea mays* L.) genotypes. *Plant Physiology and Biochemistry* **48**, 124–130.
- Tardieu F.** 2003. Virtual plants: modelling as a tool for the genomics of tolerance to water deficit. *Trends in Plant Science* **8**, 9–14.
- Tardieu F, Tuberosa R.** 2010. Dissection and modelling of abiotic stress tolerance in plants. *Current Opinion in Plant Biology* **13**, 206–212.
- Tuberosa R, Salvi S, Sanguineti MC, Landi P, Maccaferri M, Conti S.** 2002. Mapping QTL regulating morpho-physiological traits and yield: case studies, shortcomings and perspectives in drought-stressed maize. *Annals of Botany* **89**, 941–963.
- Uhart SA, Andrade FH.** 1995. Nitrogen deficiency in maize. II. Carbon–nitrogen interaction effects on kernel number and yield. *Crop Science* **35**, 1384–1389.
- van Eeuwijk FA, Bink MCAM, Chenu K, Chapman SC.** 2010a. Detection and use of QTL for complex traits in multiple environments. *Current Opinion in Plant Biology* **13**, 1–13.
- van Eeuwijk FA, Boer M, Totir LR, et al.** 2010b. Mixed model approaches for the identification of QTLs within a maize hybrid breeding programme. *Theoretical and Applied Genetics* **120**, 429–440.
- van Eeuwijk FA, Malosetti M, Yin X, Struik PC, Stam P.** 2005. Statistical models for genotype by environment data: from conventional ANOVA models to ecophysiological QTL models. *Australian Journal of Agricultural Research* **56**, 883–894.
- van Oosterom EJ, Borrell AK, Chapman SC, Broad IJ, Hammer GL.** 2009a. Functional dynamics of the nitrogen balance of sorghum. I. N demand of vegetative plant parts. *Field Crops Research* **115**, 19–28.
- van Oosterom EJ, Chapman SC, Borrell AK, Broad IJ, Hammer GL.** 2009b. Dynamics of the nitrogen balance of sorghum. II. Grain filling period. *Field Crops Research* **115**, 29–38.
- Vega CRC, Andrade FH, Sadras VO, Uhart SA, Valentinuz OR.** 2001. Seed number as a function of growth. A comparative study in soybean, sunflower, and maize. *Crop Science* **41**, 748–754.
- Welch SM, Dong Z, Roe JL, Das S.** 2005. Flowering time control: gene network modelling and the link to quantitative genetics. *Australian Journal of Agricultural Research* **56**, 919–936.
- Welcker C, Bousuge1 B, Bencivenni C, Ribaut J-M, Tardieu F.** 2007. Are source and sink strengths genetically linked in maize plants subjected to water deficit? A QTL study of the responses of leaf growth and of anthesis–silking interval to water deficit. *Journal of Experimental Botany* **58**, 339–349.
- White RG, Kirkegaard JA.** 2010. The distribution and abundance of wheat roots in a dense, structured subsoil: implications for water uptake. *Plant, Cell and Environment* **33**, 133–148.
- White JW, Hoogenboom G.** 1996. Simulating effects of genes for physiological traits in a process-oriented crop model. *Agronomy Journal* **88**, 416–422.
- Whitmore AP, Whalley WR.** 2009. Physical effects of soil drying on roots and crop growth. *Journal of Experimental Botany* **60**, 2845–2857.
- Wright S.** 1932. The roles of mutation, inbreeding, crossbreeding and selection in evolution. *Proceedings of the 6th International Congress of Genetics* Ithaca, NY, 356–366.
- Yin X, Struik PC, van Eeuwijk FA, Stam P, Tang J.** 2005. QTL analysis and QTL-based prediction of flowering phenology in recombinant inbred lines of barley. *Journal of Experimental Botany* **56**, 967–976.
- Yin X, Struik PC.** 2010. Modelling the crop: from system dynamics to systems biology. *Journal of Experimental Botany* **61**, 2171–2183.