

# The Selective Values of Alleles in a Molecular Network Model Are Context Dependent

Jean Peccoud,<sup>1</sup> Kent Vander Velden, Dean Podlich, Chris Winkler, Lane Arthur and Mark Cooper

*Pioneer Hi-Bred International, Johnston, Iowa 50131-0552*

Manuscript received July 11, 2003

Accepted for publication January 6, 2004

## ABSTRACT

Classical quantitative genetics has applied linear modeling to the problem of mapping genotypic to phenotypic variation. Much of this theory was developed prior to the availability of molecular biology. The current understanding of the mechanisms of gene expression indicates the importance of nonlinear effects resulting from gene interactions. We provide a bridge between genetics and gene network theories by relating key concepts from quantitative genetics to the parameters, variables, and performance functions of genetic networks. We illustrate this methodology by simulating the genetic switch controlling galactose metabolism in yeast and its response to selection for a population of individuals. Results indicate that genes have heterogeneous contributions to phenotypes and that additive and nonadditive effects are context dependent. Early cycles of selection suggest strong additive effects attributed to some genes. Later cycles suggest the presence of strong context-dependent nonadditive effects that are conditional on the outcomes of earlier selection cycles. A single favorable allele cannot be consistently identified for most loci. These results highlight the complications that can arise with the presence of nonlinear effects associated with genes acting in networks when selection is conducted on a population of individuals segregating for the genes contributing to the network.

RECENTLY there has been interest in interpreting the quantitative genetic properties of gene networks at the population level (FRANK 1999; OMHOLT *et al.* 2000). This is warranted on at least three grounds: (i) much of the molecular genetic evidence points to the roles of genes in nonlinear networks in the determination of gene-to-phenotype relationships, (ii) we have a growing body of data on the structural and functional properties of the genomes of organisms and as this pool of data continues to expand it is becoming more feasible to construct models of gene networks, and (iii) for many aspects of basic and applied genetics it is necessary to study the properties of allelic variation for genes at the level of phenotypic effects and variation within populations. Bridging the molecular and population-level views of gene-to-phenotype relationships is a challenging area of research for quantitative genetics. At present there is no agreed-upon quantitative framework but a number of approaches are being investigated. We constructed a model of the gene network controlling the galactose metabolism pathway in yeast, using differential equations. This model has been used as a genotype-to-phenotype map with which to evaluate the performance of individuals in simulations of a mass selection process. Combining these two approaches makes it possible to

analyze the epistatic interactions between the genes controlling this pathway and their impact on the selection process.

Fundamental to genetics is the relationship between the genotype of an individual, the environment where it lives, and its resulting phenotype. This relationship is often referred to as genotype-to-phenotype (GP) mapping. Since the true mechanisms of gene expression have historically been poorly understood, geneticists have derived such mappings from the joint distributions of genotypic and phenotypic data. The simplest mapping, Mendelian genetics, considers traits that are determined completely by individual genes. Many traits, however, are more complex than that; they are quantitative in nature and are influenced by contributions from alleles at multiple loci. These multiple-gene cases have been studied using linear statistical models that allow both additive and nonadditive (dominance and epistasis) effects (FALCONER and MACKAY 1996). Complex traits are also often dependent on the environment in which a genotype is expressed. In addition to the direct effect of the environment, genotype-by-environment ( $G \times E$ ) interactions can have important effects on complex traits. Traditionally, genotype-to-phenotype mappings have predominantly been linear combinations of terms representing dominance, epistasis,  $G \times E$  interactions, and genotype-by-genotype ( $G \times G$ ) interactions (COOPER and PODLICH 2002).

Even though these linear statistical relationships al-

<sup>1</sup>Corresponding author: Pioneer Hi-Bred International, Bioinformatics & Discovery Research, P.O. Box 552, 7250 NW 62nd Ave., Johnston, IA 50131-0552. E-mail: jean.peccoud@pioneer.com

lowed geneticists to represent the phenotypic variability of a large number of simple traits, working beyond the limitations of linear mappings is one of the main challenges faced by geneticists today. Interactions between genes contribute to complex phenotypes in plants (ESHED and ZAMIR 1996; LI *et al.* 1997, 2001; LUO *et al.* 2001), mice (LEAMY *et al.* 2002), and microorganisms (ELENA and LENSKI 2001; STEINMETZ *et al.* 2002). Genetic factors that contribute to many pathologies do not have any direct effect on the phenotypes that are essentially determined by  $G \times E$  and  $G \times G$  interactions (PERERA 1997; WEATHERALL 2001). These observations are interpreted as nonlinear effects of gene interactions and are usually referred to collectively as epistatic effects (RUTHERFORD 2000).

De Jong has recently reviewed various families of models that have been used to represent genetic networks (DE JONG 2002). Considering the small copy number of the molecules involved in gene expression mechanisms, Markovian models (PECCOUD and YCART 1995; GOSS and PECCOUD 1998) based on a stochastic version of the mass action law are an appealing representation of gene network dynamics. However, the cost of computing Monte Carlo simulations limits their use to only those pathways having a well-documented stochastic outcome at the cellular level (ARKIN *et al.* 1998; MCADAMS and ARKIN 1999). Approximating the network dynamics by a system of differential equations provides a useful compromise between a realistic representation, speed of simulation, and a wealth of theoretical properties and analysis techniques that can complement numerical simulations.

## MATERIALS AND METHODS

**Modeling the galactose genetic switch:** The galactose pathway is an attractive system for dynamic modeling since it integrates a gene network, a metabolic pathway, and a response to environmental perturbations. In a first approximation, it is possible to associate the phenotype to the activity of the metabolic pathway and the genotype to the genes in this pathway. Our model of the galactose pathway (GAL) system (Figure 1 and Table 1) is a simplistic representation of the complex mechanisms of gene expression. It is representative of the common understanding of the molecular mechanisms responsible for the response of yeast to the presence of galactose and glucose in its environment.

**Biology:** To date, the effect of the environment has often been ignored in models of gene networks. Alternatively, it is possible to consider the environment as a set of external parameters, where simulation runs with various parameter values can be compared to evaluate the impact of the environment on the model dynamics (VENKATESH *et al.* 1999). For many situations it seems that this approach is able to capture the biological logic of the network. In the case of the galactose pathway of yeast, the environment can change the state of the genetic switch by inducing or repressing the expression of the GAL genes. However, the relationship between the network and its environment is not one-way. The induction of the GAL genes by galactose results in the transformation of galactose into glucose. This transformation introduces a feedback loop

by which the induced state of the GAL system leads to a modification of the environmental conditions that lead to this induction. In an effort to capture this behavior, we introduced in the model GalExt and GluExt, which can be regarded as external pools of molecules not affected by the dynamics of intracellular reactions. Passive diffusion or active transport of these molecules into the cell can be represented by chemical equations transforming these molecules into their intracellular counterparts, Gal and Glu, respectively (Table 1, reactions R01, R02, and R05). Gal and Glu can be regarded as the variables indicative of the intracellular environment. The values of the two control variables GalExt and GluExt indicate the presence of sugars in the environment. Absence and presence were indicated by 0 and 10, respectively. The combination of GalExt and GluExt values defines an environment.

**Dynamics:** The time evolution of the model is represented by mass-action differential equations. The set of coupled differential equations can automatically be derived from the chemical equations of Table 1 (ERDI and TOTH 1989). Specifically, the matrices of stoichiometric coefficients for the reactants  $\alpha_{ir}$  and products  $\beta_{ir}$  of the reactions can be used to represent the generic form of a chemical equation:

$$R_r = \sum_{i=1}^M \alpha_{ir} X_i \xrightarrow{k_r} \sum_{i=1}^M \beta_{ir} X_i. \quad (1)$$

The rate  $v_r$  of each reaction depends on the concentration of its reactants:

$$v_r = k_r \prod_{i=1}^M [X_i]^{\alpha_{ir}}. \quad (2)$$

The time evolution of a molecule concentration is ruled by the balance between the rates of the reactions producing this molecule and the ones using it as a reactant:

$$\frac{d[X_i]}{dt} = \sum_r \beta_{ir} v_r - \sum_r \alpha_{ir} v_r. \quad (3)$$

The complete set of differential equations is given in the APPENDIX in MATLAB format.

**Genotypes, phenotypes, and traits:** To analyze the response of a gene network to selective pressure, it is necessary to establish a correspondence between the basic properties of genetics at a population level and the characteristics of genetic networks. Our analysis relies on the following:

**Segregating loci as model parameters:** The reaction rates are genetically determined. It is well established that directed mutations of promoters or protein domains can affect the rates of protein-DNA interactions, protein-protein interactions, and gene expression or even affect the catalytic properties of an enzyme. Hence, each parameter is determined by a number of segregating loci. The precise mapping of the genetic space onto the parameter space depends on the number of genes involved ( $N$ ) and the extent of genetic polymorphism. In the case of a bimolecular reaction like R09 (Table 1, Figure 1), the rate of the binding of the Gal4p protein on the GAL promoter can be determined by the sequence coding for Gal4p and by the regulatory sequence of GAL. Potentially, two loci could determine the rate of this reaction but if only one of them is polymorphic, it is not necessary to consider in the model the locus corresponding to the conserved sequence. In the context of this article, a single locus was associated with each parameter (*i.e.*,  $N = 27$ ).

**Alleles as discrete parameter values:** The association between loci and parameters makes it natural to associate allelic polymorphism with variation of specific parameter values. Each polymorphic locus is assumed to have two alleles in this article (larger numbers can be considered). A null allele translates into a zero value of the corresponding parameter. Alleles

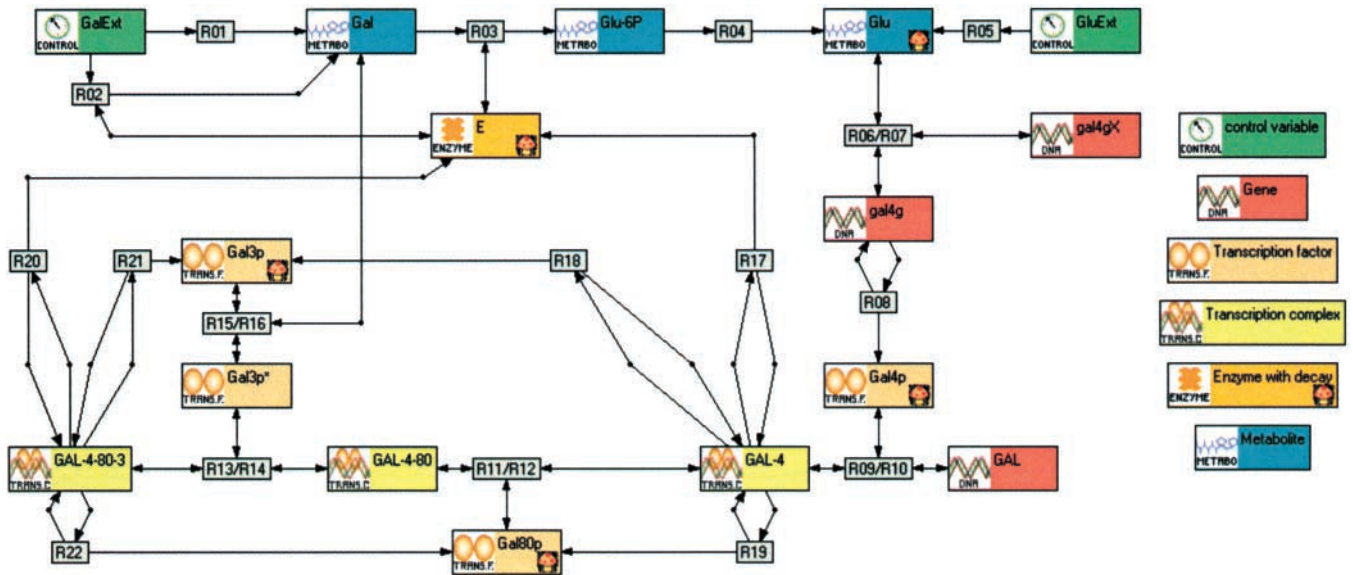


FIGURE 1.—Diagram of the galactose switch. Recent overviews of the GAL switch have been provided by IDEKER *et al.* (2001) and OSTERGAARD *et al.* (2000). GalExt and GluExt are the two environmental variables of the system. Galactose is transported into the cell primarily by Gal2p, using an ATP-dependent mechanism. It is necessary to take into consideration a small passive diffusion of galactose into the cell to trigger the induction of the GAL genes by galactose. Although there are a number of well-characterized metabolites between galactose and glucose 6-phosphate, we represent the whole pathway by a single step catalyzed by a hypothetical enzyme labeled E. Since the glucose-6-phosphatase catalyzing the transformation of Glu-6P into glucose is not part of the GAL network, it was omitted from the model. The gene coding for Gal4p, gal4g, can be in a repressed form, gal4gX, when complexed by Mig1 in the presence of glucose (RONNE 1995; OSTLING and RONNE 1998). For simplicity we considered a single enzyme in the pathway coded by a single gene noted GAL. The expression of GAL is induced by Gal4p. When in the induced state GAL-4, it expresses the E enzyme along with the Gal3p and Gal80p transcription factors. Gal80p represses this expression by binding to the GAL-4 complex. Gal3p is the galactose sensor of the GAL system. Galactose binds Gal3p through an ATP-dependent mechanism. The resulting complex Gal3p\* binds to the GAL-4-80 complex and induces the expression of the GAL genes (YANO and FUKASAWA 1997).

having a less dramatic effect result in parameters having an  $x$ -fold higher or smaller value than the wild type. The within-locus parameter values are assumed to be additive so that the heterozygous genotype is given the average parameter value of the homozygous genotypes for the two alleles. Different levels of dominance at the individual parameter level can be allowed but are not considered here. In the context of this article, we do not consider the possibility of introducing mutations. The genetic space is thus finite and discrete. Its  $3^N$  genotypes are the  $3^N$  parameter combinations resulting from the selection of one of the three possible parameter values (columns) in each of the 27 lines of Table 1.

*Phenotypes as vectors of traits:* Traditionally the phenotype of an individual is defined by the value of the biometric data that can be measured at some point in time (*e.g.*, grain yield of crops, the number of bristles on a segment of *Drosophila* spp.). These biometric data rarely translate directly into molecular variables but they are indicative of the performance of the individual. To relate a model to experimental observations, it is necessary to derive trait values from the model itself.

*Traits as functions of a model:* The biometric data collected to score a trait are static, time-independent observations. Even though life is a dynamic process that develops in time, phenotypes are observed in standard conditions that remove time from the observation. Even traits tightly associated with the timing of development are considered static in genetics. The transition from vegetative growth to reproduction or flowering time provides a good illustration of this point. The whole developmental process is reduced to a single datum, the time

of the transition to flowering. The genetic analysis of this trait relies primarily on this single observation of individuals in a population. Traits are a means to score the various characteristics of genotypes. In the case of the GAL system, the most obvious trait is the capability to process galactose when it is the only source of carbon available. How does this translate in the context of our model of the galactose switch? There are several possible interpretations of this trait. The variable representing the enzymes or the variables representing metabolites can be used as indicators. In this case we elected to use Glu-6P as an indicator of the state of the galactose pathway. To quantify the trait, we assigned target values for Glu-6P in the three environments (we ignore the trivial case where no sugar is present Gal- Glu-). Arbitrarily, we decided that Glu-6P should be 0 in the two environments where the pathway should not work (Gal- Glu+, Gal+ Glu+) and 2 in Gal+ Glu-. The system of differential equations was integrated between  $t = 0$  and  $t = 10^4$  where it is assumed to reach steady state. By noting  $X_{-+}(10^4)$ , the value of Glu-6P at time  $10^4$  in the Glu+ Gal- environment, this first trait is

$$T_1(k_1, \dots, k_{27}) = \sqrt{(X_{-+}(10^4) - 2)^2 + (X_{++}(10^4) - 0)^2 + (X_{-+}(10^4) - 0)^2}. \quad (4)$$

A second trait was also defined for this model. Comparable levels of external galactose and glucose are expected to lead to comparable levels of internal glucose. By noting  $Y_{-+}(10^4)$ ,

**TABLE 1**  
**Chemical equations and parameters**

	Reaction equation	$A_1A_1$	$A_1A_2$	$A_2A_2$
D01	Gal80p $\rightarrow$ 0	<u>30</u>	50	70
D02	Gal4p $\rightarrow$ 0	16	36	<b>56</b>
D03	Gal3p $\rightarrow$ 0	22	40	<u>58</u>
D04	Glu $\rightarrow$ 0	50	50	<u>50</u>
D05	E $\rightarrow$ 0	66	82	<u>98</u>
R01	GalExt $\rightarrow$ Gal	1	<u>1</u>	1
R02	E + GalExt $\rightarrow$ Gal + E	5	6	<u>7</u>
R03	E + Gal $\rightarrow$ Glu-6P + E	<u>5</u>	12	19
R04	Glu-6P $\rightarrow$ Glu	<u>100</u>	100	100
R05	GluExt $\rightarrow$ Glu	<u>10</u>	10	10
R06	Glu + gal4g $\rightarrow$ gal4gX	7	10	<u>13</u>
R07	Glu + gal4g $\leftarrow$ gal4gX	<u>1</u>	2	3
R08	gal4g $\rightarrow$ gal4g + Gal4p	4	23	<b>42</b>
R09	GAL + Gal4p $\rightarrow$ GAL-4	<u>3</u>	7	11
R10	GAL + Gal4p $\leftarrow$ GAL-4	<u>8</u>	9	<u>10</u>
R11	GAL-4 + Gal80p $\rightarrow$ GAL-4-80	2	3.5	<u>5</u>
R12	GAL-4 + Gal80p $\leftarrow$ GAL-4-80	<u>3</u>	5	7
R13	Gal3p* + GAL-4-80 $\rightarrow$ GAL-4-80-3	6	8	<u>10</u>
R14	Gal3p* + GAL-4-80 $\leftarrow$ GAL-4-80-3	1	10	<b>19</b>
R15	Gal + Gal3p $\rightarrow$ Gal3p*	1194	<u>1320</u>	1446
R16	Gal + Gal3p $\leftarrow$ Gal3p*	700	809	<u>918</u>
R17	GAL-4 $\rightarrow$ GAL-4 + E	<u>10</u>	19	28
R18	GAL-4 $\rightarrow$ GAL-4 + Gal3p	<u>1</u>	2	3
R19	GAL-4 $\rightarrow$ GAL-4 + Gal80p	<b>15</b>	<u>101</u>	187
R20	GAL-4-80-3 $\rightarrow$ GAL-4-80-3 + E	<u>330</u>	336	342
R21	GAL-4-80-3 $\rightarrow$ GAL-4-80-3 + Gal3p	178	309	<b>440</b>
R22	GAL-4-80-3 $\rightarrow$ GAL-4-80-3 + Gal80p	294	<u>338</u>	382

Reactions are labeled in the first column. The chemical equation of the reaction is given in column 2. Each parameter has two allelic values  $A_1$  and  $A_2$ . The columns  $A_1A_1$ ,  $A_1A_2$ , and  $A_2A_2$  indicate the parameter values used when genotypes are homozygous ( $A_1A_1$  and  $A_2A_2$ ) or heterozygous ( $A_1A_2$ ). Parameters that are underlined indicate the genotype of the individual with the highest performance that was generated at the 35th generation of the 34th run. Parameters in boldface type correspond to the favorable alleles that were consistently fixed in >95% of the 1000 runs. Lines (reactions) in italics are nonsegregating in experiment 1 and experiment 2 because they correspond to interactions outside of the GAL system. Parameter values in boldface type were made nonsegregating in experiment 2.

the value of Glu at time  $10^4$  in the Gal- Glu+ environment, this second trait is

$$T_2(k_1, \dots, k_{27}) = \sqrt{(Y_{+-}(10^4) - 2)^2 + (Y_{++}(10^4) - 2)^2 + (Y_{-+}(10^4) - 2)^2}. \quad (5)$$

A trait value can be computed for each of the genotypes of the genetic spaces considered in this article. So, for instance,  $T_1(30, 36, 22, 50, 66, 1, 7, 5, 100, 10, 7, 1, 4, 3, 8, 2, 3, 6, 1, 1194, 700, 10, 1, 15, 330, 178, 294)$  is the trait value of the genotype where all loci are  $A_1A_1$  except D02 ( $A_1A_2$ ) and R02 ( $A_2A_2$ ).

*Performance as a function of traits:* A numerical performance function is computed for selection purpose. This summarizes results from a number of elementary traits that determine how well an individual performs in a given environment. There are multiple ways of combining several trait values in a performance function. In the context of this work, we considered

$$\Phi(k_1, \dots, k_{27}) = T_1(k_1, \dots, k_{27})T_2(k_1, \dots, k_{27}). \quad (6)$$

**Simulation of selection:** To simulate effects of selection operating on the model of the galactose pathway, we devel-

oped a simple genetic algorithm application that was interfaced with a gene network simulator utilizing CVODE (COHEN and HINDMARSH 1996). For this article we have limited ourselves to a mass selection strategy where the phenotype of an individual is the only criterion used to evaluate the performance of a genotype.

The initial population (500 individuals) contained equal numbers of each allele at all segregating loci in the galactose pathway model. A constant selection pressure of 20% was applied to all cycles of selection across all simulations. We simulated a case where there was sustained directional selection for smaller values of the performance function over 100 cycles of selection. One thousand replicates of the simulation were conducted.

## RESULTS

**Model:** The model of the molecular network described in this article has two specific features not commonly found in the literature on gene networks: (1) control variables are used to represent the dynamic in-

**TABLE 2**  
**Performance computation**

Environment	Glu-6P	Glu
Gal- Glu+	0.000000	2.000000
Gal+ Glu-	1.994018	3.988036
Gal+ Glu+	0.011255	2.022515
Performance = 0.025341	$T_1 = 0.012746$	$T_2 = 1.988163$

To illustrate how performance is computed, the performance of the best-performing individual generated across the entire simulation is computed. This individual was found in the 35th generation of the 34th run of the simulation. Simulations were run in the three different environments containing sugars and the values of Glu-6P and Glu at  $t = 10^4$  are reported. The two traits can be derived from these data by using Equations 4 and 5. The performance score is the product of the two trait values.

teraction of the model with the environment, and (2) trait and performance functions are defined to evaluate the performance of a model parameterization.

*Control variables:* For the sake of reproducibility, the simulations described in this article do not take full advantage of the possible time evolution of control variables. Instead of assigning a constant value to environmental factors such as the sugar concentrations, it is possible to specify the variation of these concentrations in time. This feature makes it possible to evaluate other traits of the model. For instance, it is possible to quantify the ability of the network to react to changes of the environment. The trait functions described in this work do not distinguish the networks that will quickly adapt to new conditions from the ones that will need more time to turn the galactose switch on and off. In models of other regulatory networks, control variables have also been used to represent the effect of physical parameters of the environment such as temperature, volume, or light.

*Performance function:* To assess the way individuals are scored by the performance value, we looked for the individual with the lowest performance value that was generated across the entire experiment. This individual was found at the 35th cycle of run 34. It is interesting that this individual was not found in the population generated at the end of the selection process (cycle 100). The performance value of the best individual is  $\sim 0.025$  (see Table 2). The values achieved at the end of the selection process are typically close to 0.20. This eightfold difference tends to indicate that a dramatic loss of performance occurred during the selection process. That is when it becomes necessary to examine how these performance values are achieved, *i.e.*, the property of the trait and performance functions used in the simulation. The values in Table 2 show that the target values for Glu-6P are reached in the three environmental conditions and  $T_1$  can reach a very low value. This is not the case for  $T_2$ . The target values for Glu are reached

in Gal- Glu+ and Gal+ Glu+ conditions but we cannot get close to the target value of 2 in the Gal+ Glu- condition. When the best individual is compared to the best individuals typically found in the last cycle of selection, it turns out that their behavior is very comparable. Minimal changes in Glu-6P values result in a significant difference in the  $T_1$  value, which propagates to the performance value. Even more interesting is the examination of the time evolution of Glu-6P and Glu when the molecular network is integrated. Asymptotic values are quickly reached in Gal- Glu+ and Gal+ Glu- but the system oscillates when placed in Gal+ Glu+ conditions. The amplitudes of the oscillations are significant (0.5 for Glu and 0.3 for Glu-6P) but the values are close to the target values at  $t = 10^4$ . This shows how dependent the outcome of the selection process is on the trait and performance functions.

**Simulations:** Running such an experiment is a significant computational challenge. There are very significant differences of simulation time between runs since some simulations can be achieved in 1.7 hr while others would take up to 4.4 hr on a processor running at 2.8 MHz. Most of the time is spent evaluating the performance of 50,000 individuals generated during 100 populations of 500 individuals. Since the model needs to be simulated in three different environmental conditions, the differential equations are integrated 150,000 times in each run. To speed up the genetic algorithm, previously evaluated phenotypes are recorded in a cache. Some simulations are more likely than others to explore larger regions of the genetic space. Even though the total number of individuals evaluated is the same for all simulations, some will evaluate more genotypes than others, which explains the differences of simulation time. To complete the 1000 runs in an acceptable time ( $\sim 15$  hr), the simulations were distributed over the 56 nodes of a Linux cluster, each node having two processors. The selection process simulated in this experiment is extremely basic. Its implementation did not require much programming. To analyze the response of regulatory networks to actual breeding programs, we also interfaced the molecular network simulation environment with QU-GENE, an environment for simulating breeding strategies (PODLICH and COOPER 1998; MICALLEF *et al.* 2001).

**Response to selection:** There are two ways to analyze the network response to selection. The time evolution of performance is indicative of the effect of selection while the time evolution of allele frequencies tells us how this effect is achieved. Two experiments (series of 1000 simulations with identical parameters and initial conditions) were conducted. In experiment 1, the only nonsegregating loci were those corresponding to interactions that are often considered “outside” the galactose switch. In a second experiment, experiment 2, we also fixed the favorable alleles of loci having an additive

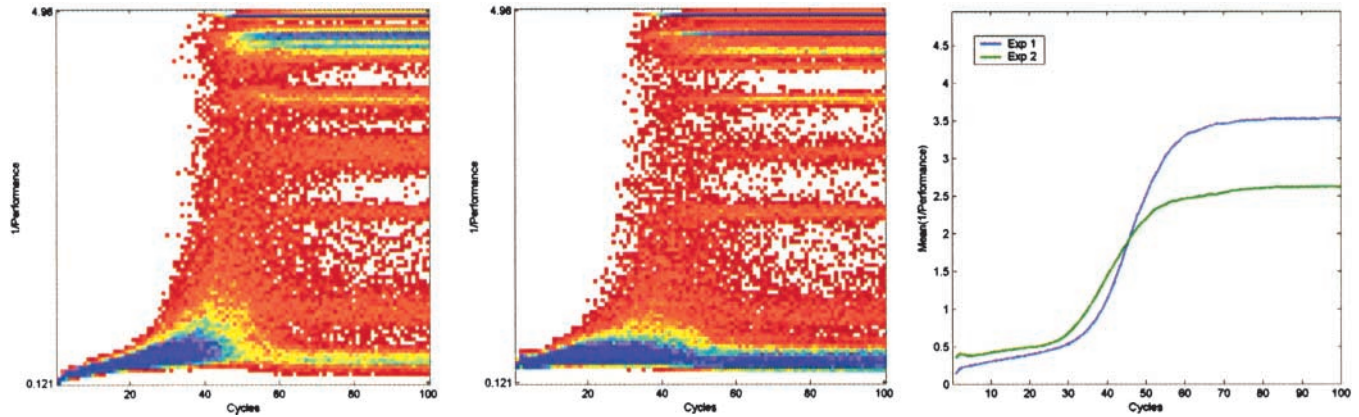


FIGURE 2.—Time evolution of the population average performance distribution. Data were recorded during experiment 1 and experiment 2, each consisting of 1000 simulations. Histograms of the inverse of the population average performance values were computed for each of the 100 cycles of selection. The frequency of each performance category was color coded. Results from experiment 1 (left) show that the distribution is clearly nonnormal since it exhibits at least eight modes. Beyond cycle 80, the selection process has reached its asymptotic distribution. The distribution observed in experiment 2 (center) is fairly similar to results of experiment 1. The main difference is the weight of the bottom mode (blue peak), indicating that a large fraction of the simulations never achieved good performance values. To better compare these two distributions, the time evolutions of their mean values were plotted on the third graph (right). It shows that better performance is achieved in experiment 2 (green line) for the early phases of the selection process. However, the long-term response to selection in experiment 2 is not as good as that in experiment 1.

effect in the results of experiment 1 (see Table 1 for details).

The response to selection of this genetic system can be illustrated by graphing the evolution over cycles of selection of the mean performance value of the population. However, in the case of this experiment this graph did not appear to be the most appropriate. The best performers in our experiment have the lowest performance value. As a result the selection results in a reduction of performance values over time. The other problem is that the performance function has an absolute lower bound. So the plot of the mean performance values over cycles of selection is difficult to read since all the runs tend to accumulate toward 0. To overcome these difficulties, the statistical distribution of the inverse of the mean performance value was plotted (Figure 2).

*Experiment 1:* The statistical distribution of mean performance values is initially unimodal (Figure 2, left). Beyond cycle 50 or so up to eight modes can be identified. Interestingly, there is a mode corresponding to poor levels of performance. There are also two major modes corresponding to good performance and a few minor modes of intermediate values. Beyond cycle 50, the selection process appears to have reached its asymptotic distribution. However, the observation of individual trajectories indicates that despite a constant selection pressure, the populations can move from one mode to the other, resulting in quick gains or losses of performance even in the stationary regime. This pattern indicates that the performance landscape is complex with multiple local maxima and that the fluctuations of the

selection process are large enough to move the population from one peak to the next.

Allele frequencies exhibit a fairly complex behavior at most loci (Figure 3, top). Fixation of one of the two alleles in  $>95\%$  of the runs is observed for seven loci (D02, D05, R08, R14, R18, R19, and R21). In the other cases the final allele frequencies are variable and are distributed between 0 and 1 with peaks at 0, 50, and 100%. Thus, either one of the homozygotes or the heterozygote could be favored depending on the replicate. So for most loci it is not possible to clearly identify a consistently favorable allele; the favorable allele is highly context dependent. Also, since a small percentage of the runs lead to retaining the heterozygous state, both alleles could be retained. Also included in Figure 3 is the time evolution of the four parameters that are not polymorphic. They are the only ones exhibiting a random drift behavior. These loci can thus be considered as negative controls. All the polymorphic loci have some selective value in this experiment since none of them drift as the nonpolymorphic ones do.

*Experiment 2:* In experiment 2, the seven loci that had a favorable allele in experiment 1 were fixed and thus made nonpolymorphic (see Table 1). By fixing the favorable allele in the parameter file, it was anticipated that the transient phase of the selection process would be shortened. It turns out that the initial mean performance values are actually better (Figure 2, center), as anticipated. However, the asymptotic distributions are significantly different. The heavily loaded mode at the bottom of the plot indicates that a large fraction of the simulations never manage to achieve good levels of

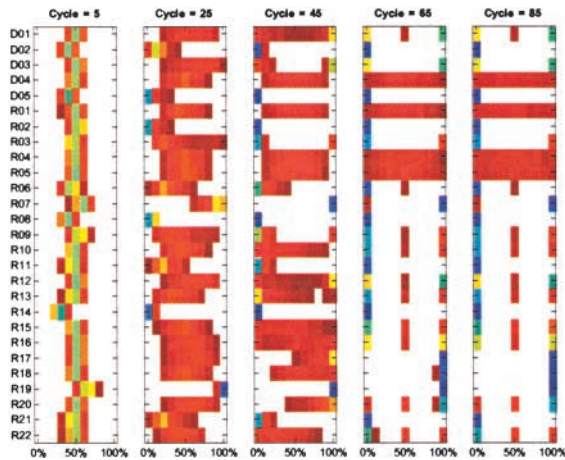


FIGURE 3.—Evolution of allele frequencies under selection. During the Monte Carlo simulations corresponding to Figure 2, the frequencies of allele  $A_1$  at each of the 27 loci were recorded. Histograms of these frequencies were color coded as in Figure 2. To illustrate the effect of the selection process on the genetic make-up of the population, five histograms corresponding to the selection cycles 5, 25, 45, 65, and 85 are displayed. In experiment 1 (D02, D05, R08, R14, R18, R19, and R21) one of the two alleles is consistently fixed in  $>95\%$  of the simulations. For most loci, however, no allele is fixed. Frequency distribution is multimodal with peaks at 0, 100, and often 50%. Nonpolymorphic loci (D04, R01, R04, and R05) exhibit a pattern indicative of genetic drift. Indirectly, they indicate that all other loci have some selective value. Plots at cycles 65 and 85 are very similar, indicating that the asymptotic distribution of the process has been reached. Results of experiment 2 are generally similar (not shown).

performance. This is confirmed by comparing the time evolution of the mean of these two distributions (Figure 2, left). The mean for experiment 2 (green line) is initially higher than the mean of experiment 1 (blue line). Since, initially, the performance response is slow, this results in almost a 10-cycle advantage provided by the fixation of favorable alleles. However, there is a long-time cost to this more limited genetic variability since the long-term response of experiment 2 is not as good as that in experiment 1. The response of allele frequencies to selection is very similar in experiment 1 and experiment 2 even though some minor quantitative difference can be observed.

Future work will relate the peaks of the performance distribution (Figure 3) with the distributions of the allele frequencies. It appears that the context-dependent combinations of alleles emphasized by the results of the different replicates of the selection process correspond to different peaks of performance on a moderately rugged landscape (data not shown).

## DISCUSSION

**Molecular networks as GP maps:** GP maps have traditionally been based on statistical models. In some cases we now have enough understanding of the molecular

mechanisms to capture their dynamics into mathematical models. There are some indications in the recent literature that we now have models with some predictive power of the phenotype (ELOWITZ and LEIBLER 2000; HASTY *et al.* 2002; HOUCHEMANDZADEH *et al.* 2002; KAERN *et al.* 2003). Analyzing the genetic properties of regulatory networks raises a number of theoretical and technical problems, which explains the limited numbers of articles dealing with this problem.

*Nonlinear GP maps:* Introduction of nonlinear terms in genotype-to-phenotype mappings leads to considerable theoretical difficulties that prevent any closed-form expression of the model properties. As suggested by KEMPTHORNE (1988), the development of software to simulate genotype-environment systems (*e.g.*, plant breeding programs) has enabled geneticists to explore the genetic consequences of nonlinear mappings *in silico* (PODLICH and COOPER 1998; MICALLEF *et al.* 2001) without the need for an analytic result. The  $E(NK)$  framework provides a foundation for an *in silico* approach to genetic analysis of the properties of linear and nonlinear gene-to-phenotype mappings at the individual and population levels (COOPER and PODLICH 2002). It is specified as a generalization of Kauffman's  $NK$  gene network model (KAUFFMAN 1993), where a set of  $N$  genes is assumed to be under the control of, on average,  $K$  other genes in the network. The  $E(NK)$  framework incorporates  $G \times E$  interactions through allowing a series of  $NK$  genotype-to-phenotype relationships corresponding to different environment types for a given target population of environment types. Here the target population of environments is defined as a mixture model of different environment types. Within this generic modeling framework various types of genotype-to-phenotype mappings can be implemented (PODLICH and COOPER 1999; COOPER *et al.* 2002). So far we have examined a wide range of artificial gene networks, results from molecular map-based genetic mapping of traits, and a combination of genetic analysis and crop growth models (COOPER *et al.* 2002). In this article, we describe a way to build a genotype-to-phenotype map within the  $E(NK)$  framework that relies on our understanding of the molecular mechanisms of gene expression.

It is interesting to relate the results presented in this article to previous work based on the  $E(NK)$  framework. In a broad perspective, molecular networks can be considered as  $E(NK)$  models. In the context of this article we have  $N = 27$  loci and  $E = 3$  environments. Even though the loci in our model interact, quantifying the level of connections between genes,  $K$ , proves difficult. In molecular networks, interactions between genes often involved more than one reaction. Hence there is no straightforward way of computing  $K$ . This limitation does not really matter since it is often used as a summary statistic in experiments based on an ensemble approach to gene networks. Since in this article the network we

analyze is not random, the actual topology of the network is more meaningful than the parameter  $K$ .

**Computational challenges:** Simulating the evolution of a population of network models requires solving the model with a large number of different parameterizations (size of the population  $\times$  number of generations). To estimate the fluctuations of the selection process, it is necessary to repeat the simulation of the network evolution a large number of times. Since dynamic models are orders of magnitude more expensive to simulate than a static model of a GP map, running an experiment such as the one described in this article is a significant technical challenge.

**Multiscale models:** A major challenge in using regulatory gene networks or metabolic pathways as genotype-to-phenotype mappings is that gene networks are dynamical systems and consequently their properties are defined by reference to their time evolution. In contrast, the common genetics view is a more static vision of the relationship between the genotype of an individual and its phenotype. Time is included to describe the evolution of populations of individuals across generations. Analysis of the genetics of gene networks requires introducing a different timescale. By introducing a correspondence between genetic loci and the parameter space of a gene network on the one hand and by defining trait functions to quantify the performance of a model parameterization on the other hand, we reconcile a theoretical framework that assumes a static relationship between phenotype and genotype with dynamical models of gene expression.

An important step of this approach is to reduce the time evolution of the gene network into a set of static gene-to-phenotype relationships. So far, the performance of gene networks has been reduced to the asymptotic level of expression of one or a few genes in one particular set of simulation conditions (FRANK 1999; OMHOLT *et al.* 2000). In this article we have formalized and generalized the notion of trait and performance functions applied to models of molecular interactions. Instead of focusing on the level of expression of specific genes, the traits considered in this article are derived from metabolite concentrations. These indicators integrate the effect of all genes in the system along with the effects of environmental parameters. This approach makes it possible to integrate the environment in the GP maps derived from molecular networks. In other simulations, we have defined on the same model trait functions to quantify the ability of the model to react quickly to environmental perturbations or to quantify the stability and robustness of a network (not shown).

**Trait and performance functions:** We were surprised to find networks exhibiting oscillations in one environment at the end of the selection process. This observation illustrates the dependence of the selection outcome on the trait functions and performance index. By using a naïve expression of the trait that relied on a single data

point rather than on calculating a trend, the selection process led to parameterization consistent with our specification of the selection target but more complex than we anticipated. Similarly, we illustrate that finding the right expression to combine several traits into a single performance index is challenging. Again, the examination of the outcome of the selection showed that the performance function we used in this experiment is not optimal. The choice of trait and performance functions is partly subjective since there is no one single way to quantify the properties that will be maximized by the selection process. By comparing the outcomes of simulated selection using different performance functions, it might be possible to evaluate their relevance in the computer before using them in actual breeding programs.

**Genetics of molecular networks:** Even though the model of the galactose switch considered in this article has not been validated by any experimental data, the results are probably representative of the results we would get from a model derived from molecular data. It will be necessary to apply the same approach to a number of molecular network models to better understand how the model topology and regulation translate into genetic properties.

**Performance landscape:** The multiple modes of the asymptotic distribution of the average performance values demonstrate that the outcome of the selection process is highly uncertain from a common starting point. In the context of plant breeding programs where there is only a single realization of the selection process, this observation raises a number of issues for risk management and breeding program design. From an evolutionary perspective, it is striking that given a deterministic genotype-to-phenotype mapping and a stable environment, the selection process can have a large diversity of outcomes. It would be interesting to investigate the properties of the performance landscape *in vivo*. This would require conducting a large number of selection experiments in parallel, starting from identical conditions. Conducting such an experiment requires having first derived from a molecular network model a GP map explaining a large number of observed genotype-to-phenotype relationships. Such a map should also have some prediction power on the unobserved regions of the genetic space. The derivation of validated GP maps from the understanding of molecular mechanisms controlling the expression of complex traits remains a major scientific challenge (GUET *et al.* 2002).

**Exploration of the genetic space:** Assuming that a GP map with good prediction power is available, then another possible application of this type of simulation is the identification of the genotypes with outstanding levels of performance by exploring the genetic space *in silico* rather than *in vivo*. These genotypes could then be assembled by fixing alleles one locus at a time, using genotyping techniques and marker-based selection. This application



could be evaluated today by introducing a genetic variability in artificial gene networks (HASTY *et al.* 2002; KAERN *et al.* 2003).

*Molecular noise:* In the context of this article, the gene network dynamics have been represented by differential equations. It is recognized that the small copy number of some molecules involved in the mechanisms of gene expression (*e.g.*, transcription complexes, genes) can result in molecular fluctuations that are responsible for some level of phenotypic variability. This has been addressed theoretically (PECCOUD and YCART 1995), numerically (ARKIN *et al.* 1998; GOSS and PECCOUD 1998), and experimentally (ELOWITZ *et al.* 2002). Using a stochastic model of the gene network dynamics might have a significant impact on the outcome of the selection process. It is likely to smooth the performance landscape. Having nondeterministic performance values would also reduce the likelihood of the process being trapped in local performance minima. By modeling molecular interactions with mass-action equations as opposed to specialized biochemical kinetics, it is possible to simulate the fluctuations of molecular interactions without changing the model. In a follow-up article we will show how molecular noise can influence the response to selection of a molecular network. It seems likely that molecular noise influences the expression of some complex traits in higher organisms (COOK *et al.* 1998; KEMKEMER *et al.* 2002). The framework described in this article makes it possible to investigate its evolutionary consequences.

*Context-dependency of genetic effects:* For seven of the loci, one of the alleles was fixed in >95% of the runs. These alleles can be regarded as favorable within the context of this parameterization of the genotype-to-phenotype mapping of the galactose pathway. In a first approximation, these alleles have a strong additive effect on performance. However, for the remaining polymorphic loci, the contribution of each allele to performance is context dependent and it is not possible to classify either of the alleles as favorable without specifying the context. At the individual level, the context refers to the alleles present at other loci associated to the trait. At this level, the context dependency of allele values results from the nonlinearity of the model of molecular interactions. Context dependency can also be considered at the population level. The selective values of the allele at one particular locus depend on the allele frequencies of all other loci associated to the trait being selected.

Epistasis is a challenging concept with different meanings in molecular biology and genetics. At the molecular level, all the genes of the GAL system are engaged in some form of *cis* or *trans* interaction. Epistasis seems prevalent at this level. For geneticists, epistasis is associated with the limits of the additive model of gene action. If the complexity of the selection process indicates epistatic effects, it is nonetheless striking that most genetic gain takes place during the first 50 cycles of selection

in our simulation experiment. This suggests that at the population level the system is initially in a largely additive state, despite these molecular interactions. However, following cycle 50 the results of the selection process are much less predictable. This indicates that the initial cycles of selection predictably fix particular alleles at seven loci. The additive genetic variation associated with these five loci is exploited by selection. Following this additive gain, the population structure is such that the system moves into a state where there are more context-dependent, nonadditive effects exploited by selection. The consequence is the many possible selection end points by cycle 100. It may thus be necessary to refine our understanding of the consequences of molecular interactions by, for instance, relating genetic epistasis to the control properties of the regulatory circuits of the gene network model (THOMAS 1999). Further, the results we observe reinforce that views of genetic variation based on the concepts of additive and nonadditive (dominance, epistatic) components of variance for a trait are population specific and are therefore time dependent in relation to the cycles of selection (CARROLL *et al.* 2003). The work presented in this article paves the way to a more formal analysis of the genetic properties of molecular networks. In particular, it is necessary to analyze physiological and statistical genetic effects (CHEVERUD and ROUTMAN 1995; HOLLAND 2001). The techniques for analyzing genetic interactions between more than two loci raise a number of theoretical and computational problems that are beyond the scope of this article.

It is an inspirational first step to use models of molecular interactions for gene networks and their gene-to-phenotype mappings, such as our representation of the galactose pathway, to consider the complex biological processes involved in the changes brought about by plant breeding. In turn this provides a demonstration of important issues that must be considered in the design of molecular plant breeding strategies.

We thank Howie Smith and two anonymous reviewers for valuable comments and suggestions. This work would not have been possible without the support of Bob Merrill and Roy Luedtke.

#### LITERATURE CITED

- ARKIN, A., J. ROSS and H. H. McADAMS, 1998 Stochastic kinetic analysis of developmental pathway bifurcation in phage  $\lambda$ -infected *Escherichia coli* cells. *Genetics* **149**: 1633–1648.
- CARROLL, S. P., H. DINGLE and T. R. FAMULA, 2003 Rapid appearance of epistasis during adaptive divergence following colonization. *Proc. R. Soc. Lond. B Biol. Sci.* **270** (Suppl 1): S80–S83.
- CHEVERUD, J. M., and E. J. ROUTMAN, 1995 Epistasis and its contribution to genetic variance components. *Genetics* **139**: 1455–1461.
- COHEN, S. D., and A. C. HINDMARSH, 1996 CVODE, a stiff/nonstiff ODE solver in C. *Comput. Phys.* **10**: 138–143.
- COOK, D. L., A. N. GERBER and S. J. TAPSCOTT, 1998 Modeling stochastic gene expression: implications for haploinsufficiency. *Proc. Natl. Acad. Sci. USA* **95**: 15641–15646.
- COOPER, M., and D. W. PODLICH, 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., S. C. CHAPMAN, D. W. PODLICH and G. L. HAMMER, 2002 The GP problem: quantifying gene to phenotype relationships. *In Silico Biol.* **2**: 151–164.
- DE JONG, H., 2002 Modeling and simulation of genetic regulatory systems: a literature review. *J. Comput. Biol.* **9**: 67–103.
- ELENA, S. F., and R. E. LENSKI, 2001 Epistasis between new mutations and genetic background and a test of genetic canalization. *Evolution* **55**: 1746–1752.
- ELOWITZ, M. B., and S. LEIBLER, 2000 A synthetic oscillatory network of transcriptional regulators. *Nature* **403**: 335–338.
- ELOWITZ, M. B., A. J. LEVINE, E. D. SIGGIA and P. S. SWAIN, 2002 Stochastic gene expression in a single cell. *Science* **297**: 1183–1186.
- ERDI, P., and J. TOTH, 1989 *Mathematical Models of the Chemical Reaction*, Vol. 1. Princeton University Press, Princeton, NJ.
- ESHED, Y., and D. ZAMIR, 1996 Less-than-additive epistatic interactions of quantitative trait loci in tomato. *Genetics* **143**: 1807–1817.
- FALCONER, D. S., and T. F. C. MACKAY, 1996 *Introduction to Quantitative Genetics*, Ed. 4. Prentice Hall, Englewood Cliffs, NJ.
- FRANK, S. A., 1999 Population and quantitative genetics of regulatory networks. *J. Theor. Biol.* **197**: 281–294.
- GOSS, P. J. E., and J. PECCOUD, 1998 Quantitative modeling of stochastic systems in molecular biology using stochastic Petri nets. *Proc. Natl. Acad. Sci. USA* **95**: 6750–6755.
- GUET, C. C., M. B. ELOWITZ, W. HSING and S. LEIBLER, 2002 Combinatorial synthesis of genetic networks. *Science* **296**: 1466–1470.
- HASTY, J., D. McMILLEN and J. J. COLLINS, 2002 Engineered gene circuits. *Nature* **420**: 224–230.
- HOLLAND, J. B., 2001 Epistasis and plant breeding. *Plant Breed. Rev.* **21**: 27–92.
- HOUGHMANDZADEH, B., E. WIESCHAUS and S. LEIBLER, 2002 Establishment of developmental precision and proportions in the early *Drosophila* embryo. *Nature* **415**: 798–802.
- IDEKER, T., V. THORSSON, J. A. RANISH, R. CHRISTMAS, J. BUHLER *et al.*, 2001 Integrated genomic and proteomic analyses of a systematically perturbed metabolic network. *Science* **292**: 929–934.
- KAERN, M., W. J. BLAKE and J. J. COLLINS, 2003 The engineering of gene regulatory networks. *Annu. Rev. Biomed. Eng.* **5**: 179–206.
- KAUFFMAN, S. A., 1993 *The Origins of Order: Self-Organization and Selection in Evolution*. Oxford University Press, London.
- KEMKEMER, R., S. SCHRANK, W. VOGEL, H. GRULER and D. KAUFMANN, 2002 Increased noise as an effect of haploinsufficiency of the tumor-suppressor gene neurofibromatosis type 1 in vitro. *Proc. Natl. Acad. Sci. USA* **99**: 13783–13788.
- KEMPTHORNE, O., 1988 An overview of the field of quantitative genetics, pp. 47–56 in *Proceedings of the Second International Conference on Quantitative Genetics*, edited by B. S. WEIR, E. J. EISEN, M. M. GOODMAN and G. NAMKOONG. Sinauer, Sunderland, MA.
- LEAMY, L. J., E. J. ROUTMAN and J. M. CHEVERUD, 2002 An epistatic genetic basis for fluctuating asymmetry of mandible size in mice. *Evolution* **56**: 642–653.
- LI, Z., S. R. PINSON, W. D. PARK, A. H. PATERSON and J. W. STANSEL, 1997 Epistasis for three grain yield components in rice (*Oryza sativa* L.). *Genetics* **145**: 453–465.
- LI, Z. K., L. J. LUO, H. W. MEI, D. L. WANG, Q. Y. SHU *et al.*, 2001 Overdominant epistatic loci are the primary genetic basis of inbreeding depression and heterosis in rice. I. Biomass and grain yield. *Genetics* **158**: 1737–1753.
- LUO, L. J., Z. K. LI, H. W. MEI, Q. Y. SHU, R. TABIEN *et al.*, 2001 Overdominant epistatic loci are the primary genetic basis of inbreeding depression and heterosis in rice. II. Grain yield components. *Genetics* **158**: 1755–1771.
- MCADAMS, H. H., and A. P. ARKIN, 1999 It's a noisy business! Genetic regulation at the nanomolar scale. *Trends Genet.* **15**: 65–69.
- MICALLEF, K. P., M. COOPER and D. W. PODLICH, 2001 Using clusters of computers for large QU-GENE simulation experiments. *Bioinformatics* **17**: 194–195.
- OMHOLT, S. W., E. PLAhte, L. OYEHAUG and K. XIANG, 2000 Gene regulatory networks generating the phenomena of additivity, dominance and epistasis. *Genetics* **155**: 969–980.
- OSTERGAARD, S., L. OLSSON, M. JOHNSTON and J. NIELSEN, 2000 Increasing galactose consumption by *Saccharomyces cerevisiae* through metabolic engineering of the GAL gene regulatory network. *Nat. Biotechnol.* **18**: 1283–1286.
- OSTLING, J., and H. RONNE, 1998 Negative control of the Mig1p repressor by Snf1p-dependent phosphorylation in the absence of glucose. *Eur. J. Biochem.* **252**: 162–168.
- PECCOUD, J., and B. YCART, 1995 Markovian modelling of gene products synthesis. *Theor. Popul. Biol.* **48**: 222–234.
- PERERA, F. P., 1997 Environment and cancer: Who are susceptible? *Science* **278**: 1068–1073.
- PODLICH, D. W., and M. COOPER, 1998 QU-GENE: a simulation platform for quantitative analysis of genetic models. *Bioinformatics* **14**: 632–653.
- PODLICH, D. W., and M. COOPER, 1999 Modelling plant breeding programs as search strategies on a complex response surface. *Lect. Notes Comput. Sci.* **1585**: 171–178.
- RONNE, H., 1995 Glucose repression in fungi. *Trends Genet.* **11**: 12–17.
- RUTHERFORD, S. L., 2000 From genotype to phenotype: buffering mechanisms and the storage of genetic information. *BioEssays* **22**: 1095–1105.
- STEINMETZ, L. M., H. SINHA, D. R. RICHARDS, J. I. SPIEGELMAN, P. J. OEFNER *et al.*, 2002 Dissecting the architecture of a quantitative trait locus in yeast. *Nature* **416**: 326–330.
- THOMAS, R., 1999 Deterministic chaos seen in terms of feedback circuits: analysis, synthesis, labyrinth chaos. *Int. J. Bifurcat. Chaos* **9**: 1889–1905.
- VENKATESH, K. V., P. J. BHAT, R. A. KUMAR and P. DOSHI, 1999 Quantitative model for Gal4p-mediated expression of the galactose/melibiose regulon in *Saccharomyces cerevisiae*. *Biotechnol. Prog.* **15**: 51–57.
- WEATHERALL, D. J., 2001 Phenotype-genotype relationships in monogenic disease: lessons from the thalassaemias. *Nat. Rev. Genet.* **2**: 245–255.
- YANO, K., and T. FUKASAWA, 1997 Galactose-dependent reversible interaction of Gal3p with Gal80p in the induction pathway of Gal4p-activated genes of *Saccharomyces cerevisiae*. *Proc. Natl. Acad. Sci. USA* **94**: 1721–1726.

Communicating editor: D. VOYTAS

## APPENDIX

The model of the GAL switch is given in a format suitable for use with the MATLAB functions for numerically integrating differential equations:

---

Y0 = zeros(16, 1);      % initial condition

---

Y0(1) = 0.000000;      % GalExt  
 Y0(2) = 0.000000;      % GluExt  
 Y0(3) = 0.000000;      % Gal  
 Y0(4) = 0.000000;      % Glu-6P  
 Y0(5) = 0.000000;      % Gal80p  
 Y0(6) = 0.000000;      % Gal4p  
 Y0(7) = 0.000000;      % Gal3p  
 Y0(8) = 0.000000;      % GAL-4-80  
 Y0(9) = 0.000000;      % GAL-4-80-3  
 Y0(10) = 0.000000;      % Glu  
 Y0(11) = 0.000000;      % gal4g  
 Y0(12) = 1.000000;      % gal4gX  
 Y0(13) = 0.000000;      % E  
 Y0(14) = 1.000000;      % GAL  
 Y0(15) = 0.000000;      % GAL-4  
 Y0(16) = 0.000000;      % Gal3p\*

---

KC = zeros(27, 1);      % vector of kinetic parameters

---

KC(1) = 50.000000;      % Gal80p → 0  
 KC(2) = 36.000000;      % Gal4p → 0  
 KC(3) = 40.000000;      % Gal3p → 0  
 KC(4) = 50.000000;      % Glu → 0

KC(5) = 82.000000;	% $E \rightarrow 0$	r18 = KC(18) * Y(9);	% Gal3p* + GAL-480 $\leftarrow$ GAL-480-3
KC(6) = 1.000000;	% GalExt $\rightarrow$ Gal	r19 = KC(19) * Y(15);	% GAL-4 $\rightarrow$ GAL-4 + Gal80p
KC(7) = 6.000000;	% $E +$ GalExt $\rightarrow$ Gal + $E$	r20 = KC(20) * Y(15);	% GAL-4 $\rightarrow$ GAL-4 + Gal3p
KC(8) = 12.000000;	% $E +$ Gal $\rightarrow$ Glu-6P + $E$	r21 = KC(21) * Y(9);	% GAL-480-3 $\rightarrow$ GAL-480-3 + Gal80p
KC(9) = 100.000000;	% Glu-6P $\rightarrow$ Glu	r22 = KC(22) * Y(9);	% GAL-480-3 $\rightarrow$ GAL-480-3 + Gal3p
KC(10) = 10.000000;	% Glu + gal4g $\rightarrow$ gal4gX	r23 = KC(23) * Y(9);	% GAL-480-3 $\rightarrow$ GAL-480-3 + $E$
KC(11) = 1.000000;	% Glu + gal4g $\leftarrow$ gal4gX	r24 = KC(24) * Y(15);	% GAL-4 $\rightarrow$ GAL-4 + $E$
KC(12) = 23.000000;	% gal4g $\rightarrow$ gal4g + Gal4p	r25 = KC(25) * Y(2);	% GalExt $\rightarrow$ Glu
KC(13) = 7.000000;	% GAL + Gal4p $\rightarrow$ GAL-4	r26 = KC(26) * Y(3) * Y(7);	% Gal + Gal3p $\rightarrow$ Gal3p*
KC(14) = 9.000000;	% GAL + Gal4p $\leftarrow$ GAL-4	r27 = KC(27) * Y(16);	% Gal + Gal3p $\leftarrow$ Gal3p*
KC(15) = 3.500000;	% GAL-4 + Gal80p $\leftrightarrow$ GAL-480	Y2 = zeros(16, 1);	% state of the system
KC(16) = 5.000000;	% GAL-4 + Gal80p $\leftarrow$ GAL-480	<hr/>	
KC(17) = 8.000000;	% Gal3p* + GAL-480 $\rightarrow$ GAL-480-3	Y2(3) = 1 * r6 + 1 * r7 + -1 * r8 + -1 * r26 + 1 * r27;	% GalExt
KC(18) = 10.000000;	% Gal3p* + GAL-480 $\leftarrow$ GAL-480-3	Y2(4) = 1 * r8 + -1 * r9;	% GluExt
KC(19) = 101.000000;	% GAL-4 $\rightarrow$ GAL-4 + Gal80p	Y2(5) = -1 * r1 + -1 * r15 + 1 * r16 + 1 * r19 + 1 * r21;	% Gal
KC(20) = 2.000000;	% GAL-4 $\rightarrow$ GAL-4 + Gal3p	Y2(6) = -1 * r2 + 1 * r12 + -1 * r13 + 1 * r14;	% Glu-6P
KC(21) = 338.000000;	% GAL-480-3 $\rightarrow$ GAL-480-3 + Gal80p	Y2(7) = -1 * r3 + 1 * r20 + 1 * r22 + -1 * r26 + 1 * r27;	% Gal80p
KC(22) = 309.000000;	% GAL-480-3 $\rightarrow$ GAL-480-3 + Gal3p	Y2(8) = 1 * r15 + -1 * r16 + -1 * r17 + 1 * r18;	% Gal4p
KC(23) = 336.000000;	% GAL-480-3 $\rightarrow$ GAL-480-3 + $E$	Y2(9) = 1 * r17 + -1 * r18;	% Gal3p
KC(24) = 19.000000;	% GAL-4 $\rightarrow$ GAL-4 + $E$	Y2(10) = -1 * r4 + 1 * r9 + -1 * r10 + 1 * r11 + 1 * r25;	% GAL-480
KC(25) = 10.000000;	% GluExt $\rightarrow$ Glu	Y2(11) = -1 * r10 + 1 * r11;	% GAL-480-3
KC(26) = 1320.000000;	% Gal + Gal3p $\rightarrow$ Gal3p*	Y2(12) = 1 * r10 + -1 * r11;	% Glu
KC(27) = 809.000000;	% Gal + Gal3p $\leftarrow$ Gal3p*	Y2(13) = -1 * r5 + 1 * r23 + 1 * r24;	% gal4g
% reaction rates		Y2(14) = -1 * r13 + 1 * r14;	% gal4gX
<hr/>		Y2(15) = 1 * r13 + -1 * r14 + -1 * r15 + 1 * r16;	% $E$
r1 = KC(1) * Y(5);	% Gal80p $\rightarrow$ 0	Y2(16) = -1 * r17 + 1 * r18 + 1 * r26 + -1 * r27;	% GAL
r2 = KC(2) * Y(6);	% Gal4p $\rightarrow$ 0	<hr/>	
r3 = KC(3) * Y(7);	% Gal3p $\rightarrow$ 0		
r4 = KC(4) * Y(10);	% Glu $\rightarrow$ 0		
r5 = KC(5) * Y(13);	% $E \rightarrow$ 0		
r6 = KC(6) * Y(1);	% GalExt $\rightarrow$ Gal		
r7 = KC(7) * Y(1) * Y(13);	% $E +$ GalExt $\rightarrow$ Gal + $E$		
r8 = KC(8) * Y(3) * Y(13);	% $E +$ Gal $\rightarrow$ Glu-6P + $E$		
r9 = KC(9) * Y(4);	% Glu-6P $\rightarrow$ Glu		
r10 = KC(10) * Y(10) * Y(11);	% Glu + gal4g $\rightarrow$ gal4gX		
r11 = KC(11) * Y(12);	% Glu + gal4g $\leftarrow$ gal4gX		
r12 = KC(12) * Y(11);	% gal4g $\rightarrow$ gal4g + Gal4p		
r13 = KC(13) * Y(6) * Y(14);	% GAL + Gal4p $\rightarrow$ GAL-4		
r14 = KC(14) * Y(15);	% GAL + Gal4p $\leftarrow$ GAL-4		
r15 = KC(15) * Y(5) * Y(15);	% GAL-4 + Gal80p $\leftrightarrow$ GAL-480		
r16 = KC(16) * Y(8);	% GAL-4 + Gal80p $\leftarrow$ GAL-480		
r17 = KC(17) * Y(8) * Y(16);	% Gal3p* + GAL-480 $\rightarrow$ GAL-480-3		

