Fluxes and Metabolic Pools as Model Traits for Quantitative Genetics. I. The L-Shaped Distribution of Gene Effects

Bruno Bost,* Christine Dillmann*,† and Dominique de Vienne*

**Station de Génétique Végétale, INRA/UPS/INAPG, Ferme du Moulon, 91190 Gif-sur-Yvette, France and †Institut National Agronomique Paris-Grignon, 75005 Paris, France*

Manuscript received June 7, 1999
Accepted for publication August 25, 1999

ABSTRACT

The fluxes through metabolic pathways can be considered as model quantitative traits, whose QTL are the polymorphic loci controlling the activity or quantity of the enzymes. Relying on metabolic control theory, we investigated the relationships between the variations of enzyme activity along metabolic pathways and the variations of the flux in a population with biallelic QTL. Two kinds of variations were taken into account, the variation of the average enzyme activity across the loci, and the variation of the activity of each enzyme of the pathway among the individuals of the population. We proposed analytical approximations for the flux mean and variance in the population as well as for the additive and dominance variances of the individual QTL. Monte Carlo simulations based on these approximations showed that an L-shaped distribution of the contributions of individual QTL to the flux variance (R^2) is consistently expected in an F_2 progeny. This result could partly account for the classically observed L-shaped distribution of QTL effects for quantitative traits. The high correlation we found between R^2 value and flux control coefficients variance suggests that such a distribution is an intrinsic property of metabolic pathways due to the summation property of control coefficients.

THE pioneering work of Kacser and Burns (1981) illustrated the power of the metabolic control theory (MCT) in accounting for fundamental genetic phenomena such as recessivity of deleterious alleles, epistasis, or selection of selective neutrality. The MCT describes how the properties of individual enzymes of a pathway influence the flux through the pathway, and thus provides a biochemical link between the genetically determined enzyme activities/concentrations and the flux, which is a global property of the pathway. This mechanistic model of a quantitative phenotype has been successfully used in quantitative and population genetics. The variability of the flux was theoretically analyzed as a function of the effect and frequency of mutations in populations (Keightley 1989), or within sibship when parental genotypic values are known (Ward 1990). Developments of this model shed light on the variability of enzyme activities in populations under mutation-selection balance (Clark 1991; Hastings 1992). The relationship between metabolic flux and fitness was explored in *Escherichia coli* (Dykhuizen *et al.* 1987), leading to the concept of natural selection of selective neutrality (Hartl *et al.* 1985). Beaumont (1988) pointed out that stabilizing selection arises as a consequence of the structure of metabolic pathways; and Keightley (1996) showed how dominance and directional epistasis, which are automatically generated in metabolic pathways, lead to an asymmetrical pattern of response to directional selection. Finally Szathmary (1993) showed that epistasis between deleterious mutations for enzyme activity is synergistic in most kinds of selection, except for selection for maximizing the flux, where epistasis is antagonistic.

In the terminology of modern quantitative genetics, the enzymatic loci can be regarded as putative quantitative trait loci (QTL) of the flux, characterized by their contribution to the flux variance in a population. Assuming that macroscopic and quantitative traits are proportional to metabolic fluxes in the cell, we considered the fluxes as model traits to analyze the quantitative genetic variation. In this work, the MCT was used to predict the shape of the distribution of flux QTL effects in a segregating population derived from the cross between two individuals drawn at random in a species. We considered both the variation of the average enzyme activities across the metabolic pathway and the variation of activity of single enzymes between individuals of the population. Using analytical developments and simulations, we showed that an L-shaped distribution of flux QTL effects is consistently observed. This distribution is related to the L-shaped distribution of flux control coefficients, which is a consequence of the summation theorem (Kacser and Burns 1973), and is also observed experimentally.

THEORETICAL BACKGROUND

Metabolic flux as a function of enzyme activity: The flux through a linear metabolic pathway is described as a hyperbolic function of the activity of each enzyme involved in the pathway (Kacser and Burns 1973). Con-

Corresponding author: Bruno Bost, Station de Génétique Végétale (INRA-UPS-INAPG), Ferme du Moulon, 91190 Gif-sur-Yvette, France. E-mail: bost@moulon.inra.fr

sider a linear metabolic pathway, with n enzymes (E_1, E_2. . . , E_i, . . . , E_n), converting a substrate (S_1) into a final product (S_{n+1}),

$$S_1 \xrightarrow{E_1} S_2 \rightarrow \ldots \rightarrow S_i \xrightarrow{E_i} S_{i+1} \rightarrow \ldots \rightarrow S_n \xrightarrow{E_n} S_{n+1}$$

and define E_i, which for simplicity will be called the *activity* of enzyme i, as

$$E_i = \frac{V_i}{M_i} K_{1,i},$$

where V_i is the maximum velocity of enzyme E_i, M_i is its Michaelis constant, and $K_{1,i} = \Pi_{j=1}^{i-1} K_{j,j+1}$ is the product of equilibrium constants of reactions 1, 2, . . . , i. At the steady state, and assuming that all enzymes are far from saturation, the flux through the pathway is

$$J = \frac{[S_1] - [S_{n+1}]/K_{1,n+1}}{\sum_{i=1}^{n}(1/E_i)}, \qquad (1)$$

where $[S_1]$ and $[S_{n+1}]$ are the concentrations of the substrate S_1 and product S_{n+1}, respectively. $[S_1]$ and $[S_{n+1}]$ are fixed parameters of the system, while the intermediate metabolite concentrations ($[S_i]$ for $i = 2$ to n) are variables.

Control coefficient of the flux: To quantify how the flux reacts when an infinitesimal change occurs in the activity of a given enzyme, Kacser and Burns (1973) defined the control coefficient C_i^J of the flux J, with respect to activity E_i of enzyme E_i, as the ratio of an infinitesimal relative variation of the flux to an infinitesimal relative variation of an enzyme:

$$C_i^J = \frac{\partial J}{J} \Big/ \frac{\partial E_i}{E_i}.$$

Under the assumptions mentioned above, we have

$$C_i^J = \frac{1/E_i}{\sum_{j=1}^{n}(1/E_j)}$$

and hence $\sum_{i=1}^{n} C_i^J = 1$.

This summation theorem (Kacser and Burns 1973) applies more generally than to linear pathways, for example in branched pathways, pathways with feedback regulation (Kacser and Burns 1973), pathways where some metabolites are involved in a moiety-conserved cycle (Hofmeyr *et al.* 1986), or pathways with two steps catalyzed by the same protein molecule (Cascante *et al.* 1990). The most important consequence of the theorem is that the control of the metabolic system may be shared among all the enzymes, a view quite different from that of "rate-limiting" or "bottleneck" concepts. In the context of the MCT, the rate-limiting steps are steps with control coefficients close to one, and they are highly dependent on the genetic background. Experimental data are rather consistent with a distribution of the control across the metabolic pathway (see results).

METHODS

To study the flux QTL distribution, we considered the populations resulting from a cross between two diploid parents drawn at random. In those populations, we varied both the average activities among loci and the extent of the genetic variability of activity of the enzymes that control the metabolic pathway. The resulting variations observed at the flux level were analytically studied, and a set of relevant variables to describe QTL effects and metabolic control was defined. Then, we used Monte Carlo simulations to analyze the distribution of these variables and their relationships.

Variation at the enzyme level: We defined a given individual k of the species by the vector $\mathbf{E_k} = \{E_{k1}, E_{k2}, \ldots, E_{ki}, \ldots\}$ of enzyme activities of the biallelic loci governing the metabolic pathway. Without any knowledge of the distribution of enzyme activities among the loci of actual metabolic pathways, we supposed that the E_{ki} are random variables, independently and identically distributed according to a given law $L(\theta_k)$, where θ_k is the vector of the parameters of L for individual k. We considered the population resulting from the cross between two individuals k and h and supposed that the loci governing enzyme activities are independent, and without linkage disequilibrium in the population. In this case, the distribution of the flux is determined by the enzyme activities $\mathbf{E_k}$ and $\mathbf{E_h}$ for each parent, and by the matrix of allelic frequencies $\{p_{ij}\}$, where p_{ij} is the frequency of allele j for enzyme i in the resulting population.

In particular, we considered the F_1 hybrid resulting from the cross between two inbred lines and the F_2 population obtained by selfing the F_1 hybrid. In case of independent loci and with no dominance at the enzyme level, enzyme activity at locus i is defined by the average activity $m_i = (E_{ki} + E_{hi})/2$ and the additive allelic effect $a_i = (E_{ki} - E_{hi})/2$. Note that m_i and a_i are not independent, because for all i, $|a_i| \leq m_i$. In an F_2 population, it is easily shown (appendix a) that the coefficient of variation (cv_i) of the activity of enzyme i is

$$cv_i = \frac{a_i}{m_i\sqrt{2}}. \qquad (2)$$

Hence, the F_2 population can alternatively be described by the distribution law $L(\theta_m)$ of the m_i and the distribution law $L(\theta_{cv})$ of the cv_i. The former describes the distribution of the average enzyme activity across the loci, while the latter describes the distribution of the differences between the parents k and h, because for a given m_i value, the a_i value can be deduced from the cv_i value.

Variation at the flux level: In a segregating population, each enzyme whose activity is genetically variable explains a part of the genetic variance of the flux. In other words, the polymorphic loci responsible for enzyme variations are QTL of the flux and of any trait proportional to the flux. However, there is no simple

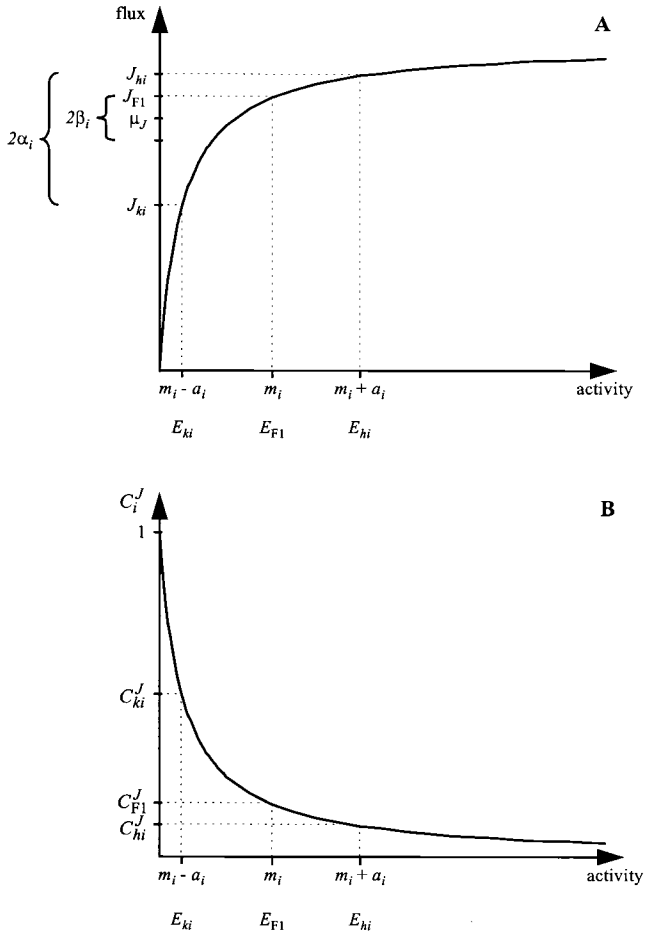

Figure 1.—Relationship between the parameters that describe the variability at the flux level, and the corresponding parameters at the enzyme level, in a simple situation where all enzymes of the pathway, except one (enzyme *i*), have the same activity for the cross between two inbred lines *k* and *h*. (A) E_{ki}, E_{hi}, and E_{F_1}, are the enzyme activities of parent *k*, parent *h* and F_1 hybrid, respectively. m_i is the average activity and a_i is the additive allelic effect of the locus *i*. J_{ki}, J_{hi}, and J_{F_1} are the flux values of parent *k*, parent *h*, and F_1 hybrid, respectively. μ_J is the average flux in an F_2 population resulting from the selfing of the F_1 hybrid. α_i and β_i are, respectively, the additive and dominance effect of the locus *i*: $2\alpha_i = J_{hi} - J_{ki}$ and $2\beta_i = J_{F_1} - (J_{hi} - J_{ki})/2$. (B) C_{ki}^J, C_{hi}^J, and $C_{F_1}^J$, are the flux control coefficients of the enzyme *i*, in parent *k*, parent *h*, and the F_1 hybrid, respectively. Parent *k* has a control coefficient higher than parent *h*, which is nearly on the plateau (C_{ki}^J is five times higher than C_{hi}^J). This situation leads to a high additive effect with an intermediary degree of positive dominance.

relationship between the variation of enzyme activity at a locus *i* and its additive and dominance effect (α_i and β_i, respectively) upon the flux.

The simplest case of an F_2 population where all enzymes but one have the same activity in both parents is represented in Figure 1. The hyperbolic relationship between flux and enzyme activity leads to a saturation in the flux when the activity increases (Figure 1A; Equation 1). The additive effect α_i of QTL *i* upon the flux is equal to half of the difference between the parents, and the degree of dominance β_i is one-half the difference between the flux of the F_1 hybrid and the mean flux of the parents. Figure 1B shows that the control coefficient increases as the enzyme activity decreases, and is close to zero for values of activity corresponding to the plateau in the flux curve. Due to this hyperbolic relationship between flux and activity, a relative change in activity leads to a smaller relative change of the flux, and the low-effect alleles are recessive (Kacser and Burns 1981). If both parents exhibit low control coefficients (*i.e.*, high values of activity), the flux values will be close to each other, while if both parents have high control values, the heterozygote will exhibit additivity, a prediction consistent with known cases of heterozygotes between sets of "lower" alleles in a series, as for pigmentation in guinea pig (Wright 1960) or mouse (Grünenberg 1952).

In the general case, all enzyme activities may differ between the parents. Hence, the flux of each individual depends on its genotype at the different enzymatic loci. The average flux μ_J in a population can be approximated by developing the function (1) expressing the flux into a second-order Taylor series. We chose the second order as a good compromise between precision and heaviness of the calculations. Provided there is no linkage disequilibrium, and taking the derivatives of μ_J with respect to allelic frequencies, the additive and dominance effects (α_i and β_i) of QTL *i*, and the epistatic (additive × additive) effect ($\alpha\alpha_{ij}$) of a pair (*i*, *j*) of QTL, could be calculated (appendix b; Kojima 1959), as well as the contributions of the QTL to the components of the flux variance: additive ($\sigma^2_{A_i}$) and dominance ($\sigma^2_{D_i}$) variances at QTL *i*, and epistatic variance (σ^2_{AAij}) for the pair of QTL (*i*, *j*). The QTL *i* contributes for a fraction R_i^2 of the total variance of the flux,

$$R_i^2 \approx \frac{\sigma^2_{A_i} + \sigma^2_{D_i}}{\sigma^2_A + \sigma^2_D + \sigma^2_{AA}}, \qquad (3)$$

where σ^2_A, σ^2_D, and σ^2_{AA}, are the additive, dominance, and epistatic (additive × additive) variances of the flux in the population, respectively. The total genetic variance of the flux also comprises other components, like the (additive × dominance) and (dominance × dominance) epistatic variances, as well as higher-order variance components, which were neglected here. In this article, the "additive allelic effect" of enzyme locus *i* refers to a_i, while the "flux QTL effect" of QTL *i* refers to R_i^2.

The sharing out of the control between the enzymes of the metabolic pathway is different for each individual of the population. To characterize each F_2 population, we computed, for each enzyme *i*, the average flux control coefficient, and its variance, $\mathrm{Var}[C_i^J]$, and we defined the concept of "populational control coefficient" for enzyme *i* as

$$\overline{C}_i^J = \frac{1/\overline{E}_i}{\sum_{j=1}^{n}(1/\overline{E}_j)}, \qquad (4)$$

where $\overline{E}_i$ (respectively $\overline{E}_j$) is the average activity of enzyme i (respectively j) in the population.

The populational control coefficient of enzyme E_i is not equal to its average control coefficient, but corresponds to the control coefficient of an "average" individual, *i.e.*, an individual displaying the average activities for all enzymes. It does not depend on the additive allelic effect a_i of the QTL, unlike the R_i^2. In F_2 populations without dominance at the enzyme level, $\overline{E}_i = m_i$ so that the populational flux control coefficient is also the control coefficient of the F_1 hybrid for enzyme *i*.

Simulation of flux QTL effects and control coefficient distributions: To simulate the distributions of flux QTL effects or control coefficients we considered 50 independent enzymatic loci in F_2 populations. A four-step procedure was used:

1. Draw of the m_i values. We considered several distribution laws for m_i, corresponding to different degrees of dispersion of the average enzyme activity across the metabolic pathway. Those distributions were constant ($m_i = 10$, $\forall i$ − reference case), uniform (in the range [0, 30]), normal ($\mu = 10$, $\sigma = 2.5$), or exponential ($\theta = 16.2$, $\sigma = 1.2$). The value of θ was chosen so that all the distributions have roughly the same range of variation, and the probability density function of the exponential law is $f(x) = (1/\sigma)\exp(-(\theta - x)/\sigma)$.
2. Draw of the cv_i values. We have chosen to consider the distribution of cv_i rather than the distribution of a_i for two reasons. First, it made it easier to take the constraint $|a_i| \leq m_i$, $\forall i$ into account. Second, we observed that our approximations for the average flux and its variance were better for cv_i values ≤ 0.3. Three contrasted distributions were considered: (i) $cv_i = 0.2$, $\forall i$, *i.e.*, there is a strict positive relationship between mean and additive effect of enzyme activity; (ii) normal, with an average of 0.35 (middle of the range for the possible values of cv_i, given m_i; see appendix a) and standard deviation fitted to get all values within the range of possible cv values; (iii) gamma, fitted to get 95% of the values between 0 and 0.3.
3. Computation of the flux QTL effects. For each pair of $\{m_i\}$ and $\{cv_i\}$ vectors, we used our approximations to compute the flux of the F_1 hybrid and the parameters of the F_2 population: populational flux control coefficient $\overline{C}^J_i$, average flux μ_J, total genetic variance, and the flux QTL effects R_i^2 (see appendix b).
4. Distribution of the flux control coefficients. For an F_2 population, 10,000 individuals were randomly generated, according to the parental genotypes at each locus. For each individual and each locus, we computed the flux control coefficient and inferred the corresponding variance $\mathrm{Var}[C^J_i]$ for each enzyme.

Those steps were iterated 100 times to simulate 100 different F_2 populations. Hence, we computed 100 μ_J values and a total of $100 \times 50 = 5000$ different values for m_i, cv_i, $\overline{C}^J_i$, R_i^2, and $\mathrm{Var}[C^J_i]$. The expected distributions of those parameters were obtained by pooling the 5000 resulting values. The populational control coefficient and R_i^2 distributions were characterized by the following parameters: mean, skewness (Sokal and Rohlf 1995), $\sum_{i=1}^{n} R_i^2$ averaged over populations, and percentage of values <0.02, which is the value expected for 50 equivalent QTL (1/50).

RESULTS

Relationships between enzymatic allelic effects and flux QTL effects: As the relationship between flux and enzyme activities is nonlinear, the average flux μ_J in a population depends not only on the average enzyme activities, $\overline{E}_i$, but also on the variances, $\sigma^2_{E_i}$, of enzyme activities, with a negative relationship between average flux and activity variances. As shown in appendix b,

$$\mu_J \approx f(\overline{E}_1, \ldots, \overline{E}_n) - \sum_{i=1}^{n}\left[\sigma^2_{E_i} - \frac{2K}{\overline{E}_i^3}\frac{\sum_{j\neq i}(1/\overline{E}_j)}{(\sum_{j=1}^{n}(1/\overline{E}_j))^3}\right]. \qquad (5)$$

For the same reason, the flux variance is related not only to the variances of enzyme activity at each QTL but also to their average activities. These features clearly differ from the classical additive models used in quantitative genetics. Moreover, the formulas show that it is not the average activity of the enzymatic locus that directly influences the flux variance, but the relative weight of the enzyme in the pathway, expressed as the "populational control coefficient" ($\overline{C}^J_i$), or control coefficient of the "average" individual (Equations B13–B15). The additive contribution of QTL *i* to the flux variance ($\sigma^2_{A_i}$) is also affected by the other QTL through their variability (contribution of a QTL is reduced by an increase of the variability of the other enzymes) and through their populational control coefficients, which are related by the summation property,

$$\sigma^2_{A_i} \approx \sigma^2_{E_i}(\overline{C}^J_i)^4 K^2[1 + 3cv_i^2(\overline{C}^J_i - 1)^2 + \sum_{j\neq i}[cv_j^2\overline{C}^J_j(3\overline{C}^J_j - 2)]]^2, \qquad (6)$$

where $cv_i = a_i/m_i\sqrt{2}$, with m_i the average activity of enzyme i and a_i its additive allelic effect.

It is worth noting that the relationship between those factors is not tight: an enzymatic locus with a large additive allelic effect may have a small effect upon the flux variance if its control on the pathway is weak in both parents (Figure 1).

The flux QTL effects are L-shaped distributed: Figure 2 compares the distributions of the average enzyme activity, m_i, across the loci, to the corresponding distributions of flux QTL effects, R_i^2, for a gamma distribution of cv_i. When all enzymes have identical m_i and a_i, the QTL have the same R_i^2 (Figure 2A). Our simulations show that, as soon as there is any difference in their average activity or variability, the distribution of R_i^2 exhibits an L shape: few steps have high R_i^2 values, more

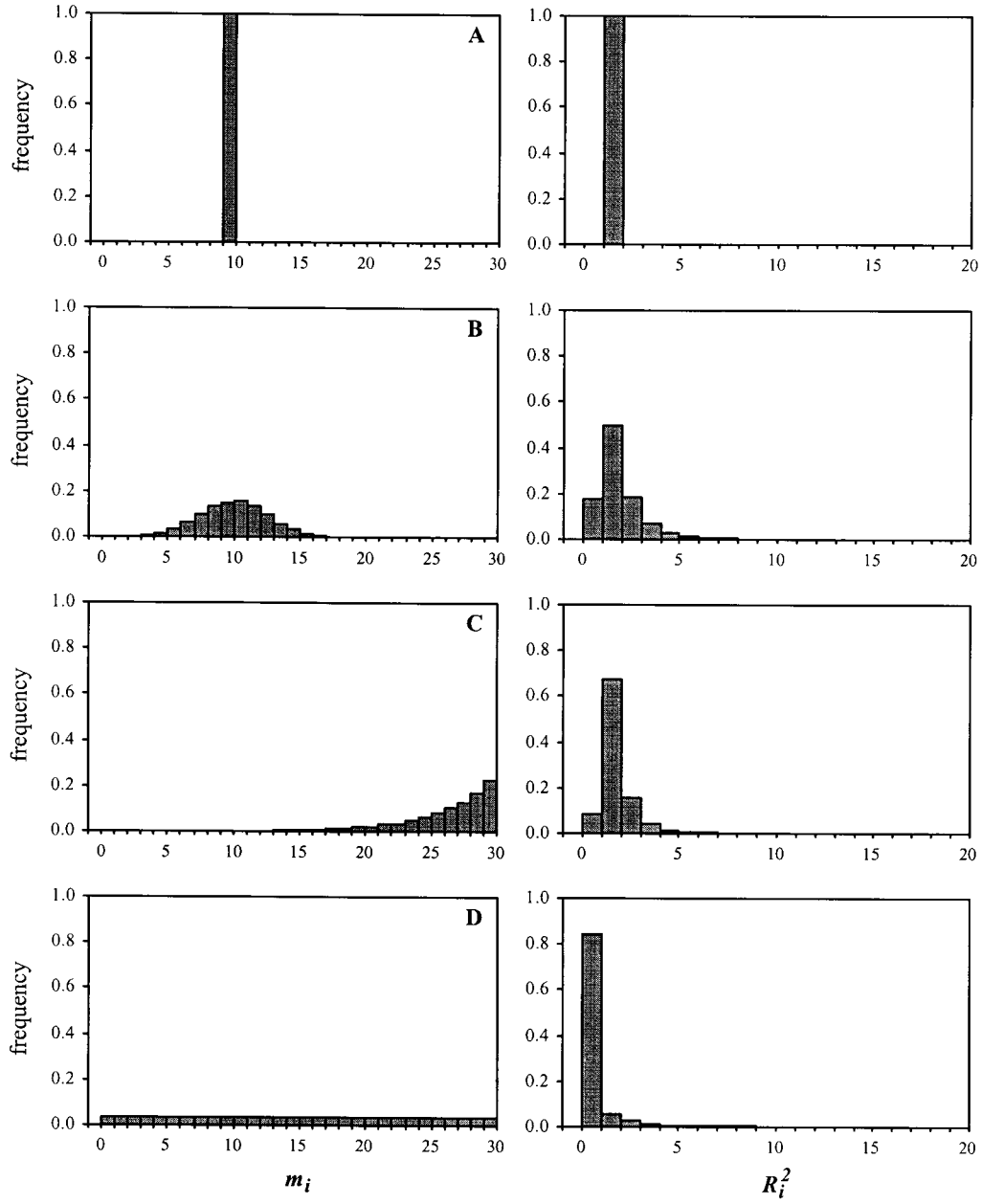

Figure 2.—Distributions of flux QTL R^2 (percentages) for a 50-enzyme pathway in 100 F_2 populations, computed for four distributions of average activity (m_i): (A) Reference (same m_i for all the QTL); (B) normal; (C) exponential; and (D) uniform. The coefficients of genetic variation of the QTL (cv_i) are randomly drawn in a gamma distribution.

have moderate values, and a large number have small or very small values. This shape is more pronounced with the uniform distribution of m_i, which leads to more contrasted m_i values, but remains with the exponential, which still displays a J-shaped distribution of m_i, or with the normal distribution (Figure 2). Numerical characteristics of the R_i^2 distributions confirm these observations: all skewness values are significantly positive ($P = 0.001$). As shown Table 1, 61.4–91.5% of the QTL have an R_i^2 value below 1/50, depending on the m_i and cv_i distributions, and do not really contribute to the variance of the flux in the population. However, this skewness is more affected by unequal m_i values across loci rather than by unequal cv_i values: for a given distribution of the m_i values, it is about the same whatever the distribution of the cv_i.

Another feature of flux QTL is that they behave as if they are nearly additive. Without epistasis, the R_i^2 should sum up to 100% in a given population (Equation 3). With the parameters chosen in our simulations, the average R_i^2 value is just below 1/50, and their average sum over the flux QTL ranges from 95.1 to 99.9%. Of course we did not consider all the epistatic terms in the denominator of Equation 3. However, we checked, by calculating the difference between the total genetic variance and the denominator, that the epistatic terms of higher order are negligible so this approximation does not significantly modify the results. The more pronounced

TABLE 1

Distributions of flux QTL R^2 for a 50-enzyme pathway in an F_2 population

cv_i and m_i distributions	QTL R^2 distribution			
	Average (%)	Sum (%)	Skewness	Percentage under 2%
$\forall i$, $cv_i = 0.2$				
$\forall i$, $m_i = m$ (reference)	1.99	99.5	—	100.0
m_i: normal	1.99	99.3	4.86	67.4
m_i: exponential	1.99	99.3	21.02	76.2
m_i: uniform	1.97	98.4	7.12	89.0
cv_i: normal distribution				
$\forall i$, $m_i = m$ (reference)	1.96	98.2	3.46	68.4
m_i: normal	1.96	98.1	9.25	71.8
m_i: exponential	1.96	98.2	13.66	74.1
m_i: uniform	1.90	95.1	6.90	90.6
cv_i: gamma distribution				
$\forall i$, $m_i = m$ (reference)	1.99	99.7	6.95	76.2
m_i: normal	1.99	99.7	8.30	78.0
m_i: exponential	1.99	99.7	9.21	77.4
m_i: uniform	1.99	99.4	7.07	90.9

The R^2 distributions were computed for a 50-enzyme pathway in 100 F_2 populations, with four distributions of average activity (m_i), reference (same m_i for all the QTL), normal, exponential, and uniform; and three distributions of coefficients of genetic variation (cv_i), same cv_i for all the QTL, normal, and gamma. Mean R^2 value of the distribution and average sum of the 50 R^2 are in percentage of total flux variance.

the L-shaped distribution across loci, the higher the variance that results from additive × additive interactions: it is equal to 0.6% with constant m_i and cv_i, and rises up to 4.9% with uniform m_i and normal cv_i. Moreover, when the number of QTL decreases, the additive × additive epistasis increases—for example, from 0.6 to 2.8% for 50 and 10 QTL, respectively (constant m_i and cv_i, not shown).

As seen in Figure 1 and methods, the additive and dominance effects of an enzymatic locus upon the flux are related to the difference in the control coefficients between the parental genotypes. On the other hand the control coefficients are linked through the summation property. Thus, looking for a possible intrinsic relationship between flux QTL R^2 and summation property, we analyzed the sharing out of the control in the parents, and we studied the relationship between the R_i^2 and the variance of the control coefficients in the population.

The parental flux control coefficients are L-shaped distributed: In linear pathways of unsaturated Michaelian enzymes, the flux control coefficients are all positive. With an *n*-enzymes pathway, the summation theorem implies that, for a given individual, when one or a few steps have control coefficients greater than $1/n$, the other steps will necessarily have coefficients below $1/n$. Hence, the average value of the control coefficients is expected to be $1/n$. If a mutation decreases one enzyme activity close to 0, its control coefficient will rise up to a value close to 1, and the other coefficients will become negligible. Thus, as soon as there is some variation for enzyme activity across loci in large metabolic systems, there would be many steps exhibiting small or very small control coefficients and a few steps with a large control; *i.e.*, the distribution of control coefficients across loci is expected to exhibit an L shape.

Experimental data: We analyzed three experimental or modeling studies by pooling for each one all the control coefficients estimated under various conditions (Groen *et al.* 1986; Albe and Wright 1992; Hill *et al.* 1993). Groen *et al.* (1986) have estimated the control coefficients of eight steps of gluconeogenesis in isolated rat liver cells under various experimental conditions (each "step" actually included several reactions, the whole pathway being considered). About 70% of the control coefficients were under the average value (0.125), with 50% under 0.05. The shape of the distribution of the coefficients was skewed to the right, with a skewness value of 20.42 (significant with $P = 0.001$; Figure 3A). Similar results were obtained for five steps of succinate oxidation in cucumber cotyledon mitochondria under various experimental conditions (Hill *et al.* 1993; skewness 27.00, $P = 0.001$, Figure 3B). Finally, a steady-state model for the tricarboxylic acid cycle has been established from experimental data in *Dictyostelium discoideum* (Albe and Wright 1992). The control coefficients for the CO_2 flux produced by the cycle and the total CO_2 production were estimated for each of the 26 steps, with six different ranges of variation of enzyme activities. The distribution of the coefficients was skewed toward weak control (skewness 12.29, $P = 0.001$, Figure 3C), and 66% of the values were below the expected average of 1/26. These data are consistent with numerous partial characterizations of other metabolic path-

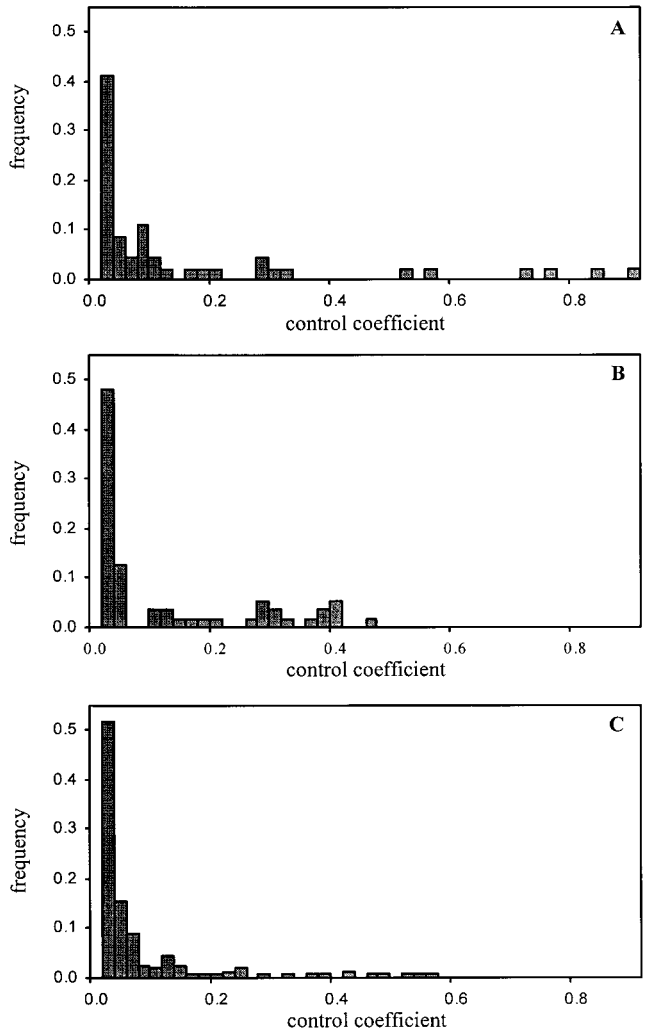

Figure 3.—(A) Distribution of flux control coefficients for 8 steps of the gluconeogenesis pathway, from lactate to glucose, in isolated rat liver cells in various experimental conditions. Data from Groen *et al.* (1986). (B) Distribution of flux control coefficients for 5 steps of the succinate oxidation pathway in isolated cucumber cotyledon mitochondria. The flux is the O_2 consumption. Data from Hill *et al.* (1993). (C) Distribution of CO_2 flux control coefficients for 26 steps of a computer model of the tricarboxylic acid cycle in *Dictyostelium discoideum.* Data from Albe and Wright (1992).

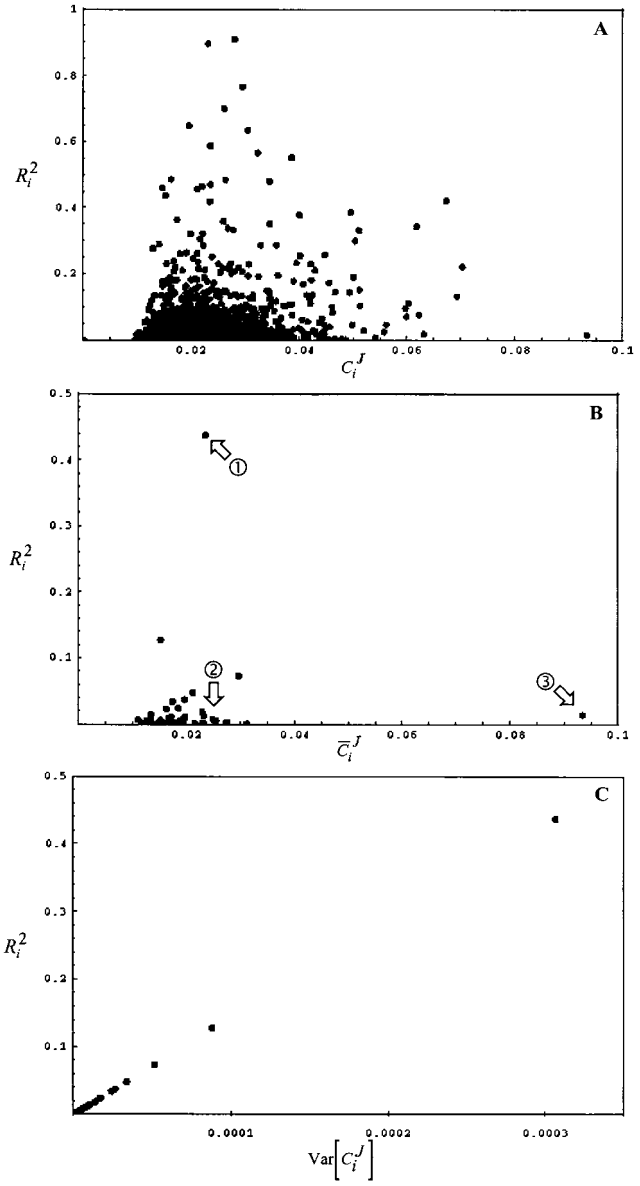

Figure 4.—Relationship between flux QTL R^2 and populational control coefficients (A and B) or variance of the control coefficient (C) for a 50-enzyme pathway in 100 F_2 populations. Average activity (m_i) is normally distributed (see Figure 2B). (A) The coefficient of genetic variation of each QTL (cv_i) is randomly drawn in a gamma distribution. (B) As for A, but only 1 of the 100 populations is represented. For the three marked QTL, see results. (C) Relationship between R^2 and control coefficient variances in the same population as for B.

ways, which show that the control is not equally shared between the different steps of metabolic chains.

Simulations: The same kind of L-shaped distribution was found in our simulations, considering the 2×100 parents of the F_2 populations and a 50-step metabolic pathway (results not shown). Whatever the distribution of enzyme activity, skewness values are positive and highly significant ($P = 0.001$) and there are much more control values below 1/50 (51–81% of the values) than over.

Relationship between control coefficients and flux QTL effects: From Equations 5 and B20, it appears that the contributions of QTL *i* to the flux additive and dominance variances are related to the populational control coefficient and to the variance of the activity of the enzyme *i*. However, the relationship between the R_i^2 and the populational control coefficient in a given population is complex and appears to be very loose (Figure 4A). QTL with similar populational control coefficients may have quite different R_i^2, depending on a_i values, while QTL may exhibit low R_i^2 even though the control coefficient is high, if a_i is low (compare QTL 1, 2, and 3 in Figure 4B).

On one hand, a given allelic additive effect (a_i) is expected to affect the flux all the more if the difference

between parental flux control coefficients of enzyme *i* is high (Figure 1). On the other hand, a large difference between parental flux control coefficients of enzyme *i* would result in a high variance (Var[C_i^j]) of the flux control coefficient in the resulting population. We therefore expect a positive relationship between R_i^2 and Var[C_i^j]. Such a high positive correlation was actually found in our simulations for each F_2 population as illustrated in Figure 4C. This striking result implies that the major flux QTL are those for which the parents are the most contrasted for the flux control coefficients. We can relate this result to the L-shaped distribution of flux control coefficients in the parents. As shown previously, the summation theorem implies that any variation of enzyme activity across loci would result in a few enzymes with a high control on the metabolic pathway. If the parents of the cross are nonrelated, those enzymes with a high control on the flux are not expected to be the same in both parents. They will therefore appear as major QTL for the flux. Hence, the L-shaped distribution of flux QTL effects is simply a consequence of intrinsic properties of metabolic pathways, through the summation property of flux control coefficients.

DISCUSSION

Understanding the relationship between gene polymorphism and quantitative trait variability is one of the main goals of quantitative genetics. The MCT provides a theoretical framework to analyze the consequences of the polymorphism of the genes controlling the enzymes concentration / activity in a linear pathway on the steady-state flux through this pathway, or on any trait proportional to this flux. We developed approximations for the flux variance components in any population without linkage disequilibrium and for any number of biallelic enzymatic QTL. These approximations lose precision for high coefficients of variation of enzymatic activities (roughly >0.3). Other methods (*e.g.*, Keightley 1989) allow us to take into account deleterious variants but are analytically restricted to models with one or two QTL.

Simulations based on these formulas have shown that the L-shaped distribution of flux QTL R^2 in a segregating population is inevitable for a flux through a linear pathway at the steady state: L-shaped distributions are generated as soon as there is any difference between the activities of the enzymes across the pathway. We have shown that such distributions arise as an indirect consequence of the summation theorem for the flux control coefficients, through the sharing out of the control in the parents. Flux QTL with major effect should correspond to enzymes that exhibit a great difference between parental flux control coefficients, namely enzymes that have a high control, *i.e.*, a low activity, in one parent only. Thus the measurement of parental control coefficients, through metabolic control analysis methods (see Fell 1992 for a review), allows us to identify which enzymes could provide good candidate loci—regulatory or structural loci of these enzymes—to explain the quantitative trait variation in the segregating population studied.

In the framework of the metabolic control theory, the low-activity alleles appear to be recessive at the flux level. Unrelated parents with different evolutionary histories are not expected to exhibit the same deficient enzymes. As a consequence, their hybrid will exhibit heterosis for the flux due to positive dominance at different loci (B. Bost, C. Dillmann and D. de Vienne, unpublished results). This result generalizes the classical result from Kacser and Burns (1981) on the biochemical basis of dominance to a multilocus situation.

For simplicity, we took into account a 50-enzyme pathway with only one structural or regulatory polymorphic locus per enzyme. It is now well documented that the amounts / activities of enzymes are themselves polygenic traits (Laurie-Ahlberg *et al.* 1982; Clark and Keith 1988; Damerval *et al.* 1994; Causse *et al.* 1995; Mitchell-Olds and Pedersen 1998). However, this does not modify our conclusions. In fact, considering polygenic enzyme activities will probably only result in partitioning the control of a given enzyme between its different loci. In the simulations, we also considered that the enzymatic loci have additive effects. Even though there are some exceptions (Clark and Wang 1997), this is consistent with most experimental studies on enzyme activity (Kacser and Burns 1981). In maize, most of the proteins revealed by 2D-PAGE displayed additive inheritance for their quantity (Leonardi *et al.* 1988). The majority were found to be enzymes (Touzet *et al.* 1996).

In maize, tomato, rice, or Drosophila, where numerous QTL have been mapped for various complex traits, compilations consistently revealed extremely skewed distributions of QTL effects, with few QTL having large effects, more QTL having moderate effects, and likely a lot having small effects (depending on the power of detection methods), resulting in a typical L-shaped distribution. For example in Drosophila, many loci have small effects on abdominal and sternopleural bristle number, but few loci cause most of the genetic variation (Mackay 1996). Edwards *et al.* (1987) searched for associations between ~20 marker loci and 82 traits in two F_2 populations of maize, each of ~1900 individuals. With a type I error of 0.05 they found 2460 significant associations, with a typically L-shaped distribution of the R^2, and 94.5% of the associations exhibiting R^2 values smaller than 5%. Other examples can be found in the literature (*e.g.*, Sing and Boerwinkle 1987; Shrimpton and Robertson 1988; Paterson *et al.* 1991; Zehr *et al.* 1992; Schön *et al.* 1994; Grandillo and Tanksley 1996; Lee *et al.* 1996).

Statistical artifacts can contribute to that distribution (Carbonell *et al.* 1992, 1993; Beavis 1994). Moreover, factors such as linkage between QTL, phase (coupling /

repulsion), or heritability, may produce L-shaped distributions for traits following the classical additive model of quantitative genetics, even though the skewness is consistently higher with the metabolic model (B. Bost, C. Dillmann and D. de Vienne, unpublished results). However, it is highly likely that the L-shaped distribution also has biological bases. First, there is no genetic reason for a discontinuity between all-or-null (wild-type/mutant) variation and quantitative variation. Thus in pea, a major gene for Ascochyta blight resistance was mapped on chromosome 4 using both a QTL detection approach and a Mendelian analysis after partitioning the distribution of the resistance in the progeny into two classes (Dirlewanger *et al.* 1994). Second, a survey of the literature shows that for various traits, the same major QTL may be found in different populations or environments, which seems quite unlikely in case of erroneous estimates of the QTL effects. For example, in two different F_2 populations derived from crosses between maize and teosinte, Doebley and Stec (1993) found similar suites of QTL for architectural traits, with the same order of QTL effects for several of them. In three tomato populations derived from the same parents (one F_2 and two $F_{2/3}$) and grown in different environments, Paterson *et al.* (1991) found QTL for fruit traits common to two or even three conditions. Also significant are the QTL apparently commonly found in different species of Poaceae (Paterson *et al.* 1995) or Fabaceae (Fatokun *et al.* 1992; Maughan *et al.* 1996; Timmerman-Vaughan *et al.* 1996). Finally, in a noninfinitesimal model, equal QTL effects would imply that for any given trait, there would be a given, limited number of detectable QTL, which seems to be biologically irrelevant.

Our biochemical modeling is relevant to account for quantitative traits that are themselves components of metabolic pathways (starch concentration in grain, angiotensinogen concentration in blood, maysin content in maize silks, anthocyanin content, etc.) or that are proportional to such components. For example the weight of maize kernel is correlated to starch amount, so that the study of the variability of kernel weight could come within the framework of the MCT, relying on the well-known steps of the starch biosynthesic pathway. On the other hand, the analysis is valid for any gene, structural or regulatory, controlling the activity of an enzyme. Some identified QTL proved to be genes coding for enzymes, such as angiotensin-converting enzyme (ACE; Schächter *et al.* 1994), inducible nitric oxide synthase (Deng and Rapp 1995), or renin (Rapp *et al.* 1989) in rat. Transcription factors controlling enzyme amounts also obey the same theoretical framework, such as locus *p1* in maize, which is a transcription activator of a portion of the biosynthesis pathway of flavonoids (Byrne *et al.* 1996). But there are examples of QTL that do not control, at least in a direct way, enzyme activities such as angiotensinogen-coding gene (Jeunemaître *et al.* 1992), apolipoprotein gene (Corder *et al.* 1993), or HLA loci (Thomson 1988; Tomlinson and Bodmer 1995) in humans. Actually, the proportion of polymorphic genes encoding or controlling enzymes is not known, and the extent to which our analysis applies to gene products involved in the timing or tissue specificity of gene expression, or to hormone/receptor systems, is difficult to appreciate.

In conclusion, the hyperbolic relation between enzyme activity and metabolic flux accounts for the L-shaped distribution of control coefficients and hence could be one of the factors explaining the L-shaped distribution of gene effects in quantitative genetics.

Thanks go to F. Hospital for useful discussions and reading the manuscript, and also to C. Damerval and A. Leonardi for reading the manuscript. We thank G. de Jong, who brought to our attention the Kojima approach on gene effect decomposition. B. Bost was supported by a Ph.D. grant from the French Ministry of National Education, Research and Technology (MENRT).

LITERATURE CITED

Albe, K. R., and B. E. Wright, 1992 Systems analysis of the tricarboxylic acid cycle in *Dictyostelium discoideum*. II. Control analysis. J. Biol. Chem. **267:** 3106–3114.

Beaumont, M. A., 1988 Stabilizing selection and metabolism. Heredity **61:** 433–438.

Beavis, W. D., 1994 The power and deceit of QTL experiments: lessons from comparative QTL studies, pp. 250–266 in *49th Annual Corn and Sorghum Research Conference*. ASTA, Washington, DC.

Byrne, P. F., M. D. McMullen, M. E. Snook, T. A. Musket, J. M. Theuri *et al.*, 1996 Quantitative trait loci and metabolic pathways: genetic control of the concentration of maysin, a corn earworm resistance factor, in maize silks. Proc. Natl. Acad. Sci. USA **93:** 8820–8825.

Carbonell, E. A., T. M. Gerig, E. Balansard and M. J. Asins, 1992 Interval mapping in the analysis of nonadditive quantitative trait loci. Biometrics **48:** 305–315.

Carbonell, E. A., M. J. Asins, M. Balsega, E. Balansard and T. M. Gerig, 1993 Power studies in the estimation of genetic parameters and the localization of quantitative trait loci for backcross and doubled haploid populations. Theor. Appl. Genet. **86:** 411–416.

Cascante, M., E. I. Canela and R. Franco, 1990 Control analysis of systems having two steps catalyzed by the same protein molecule in unbranched chains. Eur. J. Biochem. **192:** 369–371.

Causse, M. A., J.-P. Rocher, A.-M. Henry, A. Charcosset, J.-L. Prioul *et al.*, 1995 QTLs for carbon metabolism and early growth in maize, with emphasis on key-enzyme loci. Mol. Breed. **1:** 259–272.

Clark, A. G., 1991 Mutation-selection balance and metabolic control theory. Genetics **129:** 909–923.

Clark, A. G., and L. E. Keith, 1988 Variation among extracted lines of *Drosophila melanogaster* in triacylglycerol and carbohydrate storage. Genetics **119:** 595–607.

Clark, A. G., and L. Wang, 1997 Epistasis in measured genotypes: Drosophila P-element insertions. Genetics **147:** 157–163.

Corder, E. H., A. M. Saunders, W. J. Strittmatter, D. E. Schmechel, P. C. Gaskell *et al.*, 1993 Gene dose of apolipoprotein E type 4 allele and the risk of Alzheimer's disease in late onset families. Science **261:** 921–923.

Damerval, C., A. Maurice, J. M. Josse and D. de Vienne, 1994 Quantitative trait loci underlying gene product variation: a novel perspective for analyzing regulation of genome expression. Genetics **137:** 289–301.

Deng, A. Y., and J. P. Rapp, 1995 Locus for the inducible, but not a constitutive, nitric oxide synthase cosegregates with blood pressure in the Dahl salt-sensitive rat. J. Clin. Invest. **95:** 2170–2177.

Dirlewanger, E., P. G. Isaac, S. Ranade, M. Belajouza, R. Cousin *et al.*, 1994 Restriction fragment length polymorphism analysis of loci associated with disease resistance genes and developmental traits in *Pisum sativum* L. Theor. Appl. Genet. **88:** 17–27.

Doebley, J., and A. Stec, 1993 Inheritance on the morphological

differences between maize and teosinte: comparison of results for two F_2 populations. Genetics **134:** 559–570.
Dykhuizen, D. E., A. M. Dean and D. L. Hartl, 1987 Metabolic flux and fitness. Genetics **115:** 25–31.
Edwards, M. D., C. W. Stuber and J. F. Wendel, 1987 Molecular-marker-facilitated investigations of quantitative-trait loci in maize. I. Numbers, genomic distribution and types of gene action. Genetics **116:** 113–125.
Fatokun, C. A., D. I. Menancio-Hautea, D. Danesh and N. D. Young, 1992 Evidence for orthologous seed weight genes in cowpea and mung bean based on RFLP mapping. Genetics **132:** 841–846.
Fell, D. A., 1992 Metabolic control analysis: a survey of its theoretical and experimental background. Biochem. J. **286:** 313–330.
Grandillo, S., and S. D. Tanksley, 1996 QTL analysis of horticultural traits differentiating the cultivated tomato from the closely related species *Lycopersicon pimpinellifolium*. Theor. Appl. Genet. **92:** 935–951.
Groen, A. K., C. W. T. Van Roermund, R. C. Vervoorn and J. M. Tager, 1986 Control of gluconeogenesis in rat liver cells. Flux control coefficients of the enzymes in the gluconeogenic pathway in the absence and presence of glucagon. Biochem. J. **237:** 379–389.
Grünenberg, H., 1952 *The Genetics of the Mouse*. Martinus Nijhoff, The Hague.
Hartl, D. L., D. E. Dykhuizen and A. M. Dean, 1985 Limits of adaptation: the evolution of selective neutrality. Genetics **111:** 655–674.
Hastings, I. M., 1992 The population genetics of alleles affecting enzyme activity. J. Theor. Biol. **157:** 305–316.
Hill, S. A., J. H. Bryce and C. J. Leaver, 1993 Control of succinate oxidation by cucumber (*Cucumis sativus* L.) cotyledon mitochondria. The role of the adenine-nucleotide translocator and extramitochondrial reactions. Planta **190:** 51–57.
Hofmeyr, J.-H., H. Kacser and K. J. van der Merwe, 1986 Metabolic control analysis of moiety-conserved cycles. Eur. J. Biochem. **155:** 631–641.
Jeunemaître, X., F. Soubrier, Y. V. Kotelevtsev, R. P. Lifton, R. R. Williams *et al.*, 1992 Molecular basis of human hypertension: role of angiotensinogen. Cell **71:** 169–180.
Kacser, H., and J. A. Burns, 1973 The control of flux. Symp. Soc. Exp. Biol. **27:** 65–104.
Kacser, H., and J. A. Burns, 1981 The molecular basis of dominance. Genetics **97:** 639–666.
Keightley, P. D., 1989 Models of quantitative variation of flux in metabolic pathways. Genetics **121:** 869–876.
Keightley, P. D., 1996 Metabolic models of selection response. J. Theor. Biol. **182:** 311–316.
Kojima, K.-I., 1959 Role of epistasis and overdominance in stability of equilibria with selection. Proc. Natl. Acad. Sci. USA **45:** 984–989.
Laurie-Ahlberg, C. C., A. N. Wilton, J. W. Curtsinger and T. H. Emigh, 1982 Naturally occurring enzyme activity variation in *Drosophila melanogaster*. I. Sources of variation for 23 enzymes. Genetics **102:** 191–206.
Lee, S. H., M. A. Bailey, M. A. R. Mian, T. E. Carter Jr., E. R. Shipe *et al.*, 1996 RFLP loci associated with soybean seed protein and oil content across populations and locations. Theor. Appl. Genet. **93:** 649–657.
Leonardi, A., C. Damerval and D. de Vienne, 1988 Organ-specific variability and inheritance of maize proteins revealed by two-dimensional electrophoresis. Genet. Res. Camb. **52:** 97–103.
Mackay, T. F., 1996 The nature of quantitative genetic variation revisited: lessons from Drosophila bristles. Bioessays **18:** 113–121.
Maughan, P. J., M. A. S. Maroof and G. R. Buss, 1996 Molecular-marker analysis of seed-weight: genomic locations, gene action, and evidence for orthologous evolution among three legume species. Theor. Appl. Genet. **93:** 574–579.
Mitchell-Olds, T., and D. Pedersen, 1998 The molecular basis of quantitative genetic variation in central and secondary metabolism in Arabidopsis. Genetics **149:** 739–747.
Paterson, A. H., S. Damon, J. D. Hewitt, D. Zamir, H. D. Rabinowitch *et al.*, 1991 Mendelian factors underlying quantitative traits in tomato: comparison across species, generations, and environments. Genetics **127:** 181–197.
Paterson, A. H., Y. R. Lin, Z. Li, K. F. Schertz, J. F. Doebley *et al.*, 1995 Convergent domestication of cereal crops by independent mutations at corresponding genetic loci. Science **269:** 1714–1718.
Rapp, J. P., S.-M. Wang and H. Deene, 1989 A genetic polymorphism in the renin gene of Dahl rats cosegregates with blood pressure. Science **243:** 542–544.
Schächter, F., L. Faure-Delanef, F. Guénot, H. Rouger, P. Froguel *et al.*, 1994 Genetic associations with human longevity at the *APOE* and *ACE* loci. Nat. Genet. **6:** 29–32.
Schön, C. C., A. E. Melchinger, J. Boppenheimer, E. Brunklaus-Jung, R. G. Herrmann *et al.*, 1994 RFLP mapping in maize: quantitative trait loci affecting testcross performance of elite european flint lines. Crop Sci. **34:** 378–388.
Shrimpton, A. E., and A. Robertson, 1988 The isolation of polygenic factors controlling bristle score in *Drosophila melanogaster*. 2. Distribution of third chromosome bristle effects within chromosome sections. Genetics **118:** 445–459.
Sing, C. F., and E. Boerwinkle, 1987 Genetic architecture of inter-individual variability in apolipoprotein, lipoprotein and lipid phenotypes, pp. 99–121 in *Molecular Approaches to Human Polygenic Diseases. CIBA Foundation Symposium No. 130*, edited by G. Bock and G. M. Collins. Wiley, Chichester, United Kingdom.
Sokal, R. R., and F. J. Rohlf, 1995 *Biometry. The Principles and Practice of Statistics in Biological Research*. W. H. Freeman and Company, New York.
Szathmáry, E., 1993 Do deleterious mutations act synergistically? Metabolic control theory provides a partial answer. Genetics **133:** 127–132.
Thomson, G., 1988 HLA disease associations: models for insulin dependent diabetes mellitus and the study of complex human disorders. Annu. Rev. Genet. **22:** 31–50.
Timmerman-Vaughan, G. M., J. A. McCallum, T. J. Frew, N. F. Weeden and A. C. Russell, 1996 Linkage mapping of quantitative trait loci controlling seed weight in pea (*Pisum sativum* L.). Theor. Appl. Genet. **93:** 431–439.
Tomlinson, I. P. M., and W. F. Bodmer, 1995 The HLA system and the analysis of multifactorial genetic disease. Trends Genet. **11:** 493–498.
Touzet, P., F. Riccardi, C. Morin, C. Damerval, J.-C. Huet *et al.*, 1996 The maize two-dimensional gel protein database: towards an integrated genome analysis program. Theor. Appl. Genet. **93:** 997–1005.
Ward, P. J., 1990 The inheritance of metabolic flux: expressions for the within-sibship mean and variance given the parental phenotype. Genetics **125:** 655–667.
Wright, S., 1960 Posnatal changes in the intensity of coat color in diverse genotypes of the guinea pig. Genetics **45:** 1503–1529.
Zehr, B. E., J. W. Dudley, J. Chojecki, M. A. Saghai Maroof and R. P. Mowers, 1992 Use of RFLP markers to search for alleles in a maize population for improvement of an elite hybrid. Theor. Appl. Genet. **83:** 903–911.

Communicating editor: P. D. Keightley

APPENDIX A

Coefficient of genetic variation for an enzymatic locus in an F_2 population: The coefficient of variation (cv_i) for the enzymatic locus i is

$$cv_i = \frac{\sqrt{\sigma^2_{E_i}}}{\overline{E}_i}, \tag{A1}$$

where $\overline{E}_i$ and $\sigma^2_{E_i}$ are, respectively, the average activity and the variance of activity of the enzyme i in the population. In an F_2 population derived from the cross between two homozygous lines, we have, with our notations, $\overline{E}_i = m_i$ and $\sigma^2_{E_i} = \frac{1}{2}a_i^2$. Thus

$$cv_i = \frac{a_i}{m_i\sqrt{2}}. \tag{A2}$$

As there is a constraint on the possible a_i values—$\forall i$, $|a_i| \leq m_i$—, the values of cv_i are in the range

$$0 \le \mathrm{cv}_i \le \frac{1}{\sqrt{2}}. \quad \text{(A3)}$$

APPENDIX B

QTL contributions to the flux variance components: Shown are calculation of approximations of the additive and dominance effects (α_i and β_i) and additive and dominance variance ($\sigma^2_{A_i}$ and $\sigma^2_{D_i}$) at QTL *i*, and approximations of epistasis (additive × additive) effect and variance ($\alpha\alpha_{ij}$ and $\sigma^2_{AA_{ij}}$) for a pair of QTL, *i* and *j*.

The QTL are controlling the flux through a linear pathway of *n* enzymes, in a segregating population, assuming that the QTL are not linked. Each enzyme of the pathway is controlled by one biallelic QTL *i*, with a frequency p_i for the upwardly acting allele.

The flux in the pathway for an individual *j* in the population is

$$\begin{aligned} J(j) &= K \Big/ \sum_{i=1}^{n} \frac{1}{E_i(j)} \\ &= f(E_1(j), \ldots, E_i(j), \ldots, E_n(j)), \end{aligned} \quad \text{(B1)}$$

where $K = [S_1] - [S_{n+1}]/K_{1,n+1}$ (see Equation 1).

Expanding B1 into a second-order Taylor series, we have

$$\begin{aligned} &f(E_1(j), \ldots, E_i(j), \ldots, E_n(j)) \\ &\quad \approx f(\bar{E}_1, \ldots, \bar{E}_n) + \sum_{i+1}^{n}\left[(E_i - \bar{E}_i)\frac{\partial f}{\partial E_i}(\bar{E}_1, \ldots, \bar{E}_n)\right] \\ &\qquad + \frac{1}{2}\sum_{i=1}^{n}\left[(E_i - \bar{E}_i)^2\frac{\partial^2 f}{\partial E_i^2}(\bar{E}_1, \ldots, \bar{E}_n)\right] \\ &\qquad + \sum_{i=1}^{n}\sum_{j>i}\Big[(E_i - \bar{E}_i)(E_j - \bar{E}_j) \\ &\qquad\qquad \times \frac{\partial^2 f}{\partial E_i \partial E_j}(\bar{E}_1, \ldots, \bar{E}_n)\Big], \end{aligned} \quad \text{(B2)}$$

where $\bar{E}_i$ is the population average activity of the enzyme *i*.

The population average flux is

$$\mu_J = E[f(E_1(j), \ldots, E_i(j), \ldots, E_n(j))]. \quad \text{(B3)}$$

Some simplification occurs:

$$\mathbf{E}[f(\bar{E}_1, \ldots, \bar{E}_n)] = f(\bar{E}_1, \ldots, \bar{E}_n)$$
$$\mathbf{E}[E_i - \bar{E}_i] = 0.$$

And as there is no linkage disequilibrium in the population,

$$\mathbf{E}[(E_i - \bar{E}_i)(E_j - \bar{E}_j)] = 0$$
$$E[(E_i - \bar{E}_i)^2] = \sigma^2_{E_i}.$$

Combining (B2), (B3), and these simplifications, an approximation of the population average flux is

$$\mu_J \approx f(\bar{E}_1, \ldots, \bar{E}_n) + \frac{1}{2}\sum_{i=1}^{n}\left[\sigma^2_{E_i}\frac{\partial^2 f}{\partial E_i^2}(\bar{E}_1, \ldots, \bar{E}_n)\right]. \quad \text{(B4)}$$

The second partial derivative of the function *f* with respect to E_i is

$$\frac{\partial^2 f}{\partial E_i^2}(\bar{E}_1, \ldots, \bar{E}_n) = -\frac{2K}{\bar{E}_i^3}\frac{\sum_{j\neq i} 1/\bar{E}_j}{(\sum_{j=1}^{n} 1/\bar{E}_j)^3}. \quad \text{(B5)}$$

Thus, introducing (B5) into (B4), we have an approximation of the population average flux,

$$\begin{aligned} \mu_J &\approx \frac{K}{\sum_{i=1}^{n} 1/\bar{E}_i}\left[1 - \sum_{i=1}^{n}\left[\frac{\sigma^2_{E_i}}{\bar{E}_i^2}\bar{C}_i^J(1 - \bar{C}_i^J)\right]\right] \\ &= \frac{K}{\sum_{i=1}^{n} 1/\bar{E}_i}\left[1 - \sum_{i=1}^{n}[\mathrm{cv}_i^2\bar{C}_i^J(1 - \bar{C}_i^J)]\right], \end{aligned} \quad \text{(B6)}$$

where $\sigma^2_{E_i}$ and cv_i are, respectively, the variance and the genetic coefficient of variation of activity of enzyme *i* in the population, and $\bar{C}_i^J$ is the "populational" flux control coefficient of enzyme *i*:

$$\bar{C}_i^J = \frac{1/\bar{E}_i}{\sum_{j=1}^{n} 1/\bar{E}_j}.$$

Kojima (1959) showed that the additive and dominance effects of a gene *i* upon a trait are related, respectively, to first and second partial derivatives of the population mean of the trait with respect to allelic frequency (p_i). The (additive × additive) epistasis effect of a pair of QTL (*i* and *j*) is related to the mixed partial derivative of the mean with respect to both allelic frequencies (p_i and p_j). Applying these formulas to the flux, we get

$$\alpha_i = \frac{1}{2}\frac{\partial \mu_J}{\partial p_i} \quad \text{(B7)}$$

$$\Rightarrow \sigma^2_{A_i} = 2p_i(1 - p_i)(\alpha_i)^2 \quad \text{(B8)}$$

$$\beta_i = \frac{1}{2}\frac{\partial^2 \mu_J}{\partial p_i^2} \quad \text{(B9)}$$

$$\Rightarrow \sigma^2_{D_i} = p_i^2(1 - p_i)^2(\beta_i)^2 \quad \text{(B10)}$$

$$\alpha\alpha_{ij} = \frac{1}{4}\frac{\partial^2 \mu_J}{\partial p_i \partial p_j} \quad \text{(B11)}$$

$$\Rightarrow \sigma^2_{AA_{ij}} = 4p_ip_j(1 - p_i)(1 - p_j)(\alpha\alpha_{ij})^2. \quad \text{(B12)}$$

So following Equation B7, we have, for the flux additive effect of the locus *i*,

$$\alpha_i \approx \frac{K}{2}(\bar{C}_i^J)^2\left[\begin{array}{l}\left[\frac{\partial \bar{E}_i}{\partial p_i}\right] \times \left[1 - \sum_{j=1}^{n}[\mathrm{cv}_j^2\bar{C}_j^J(3\bar{C}_j^J - 2)] + \mathrm{cv}_i^2(3 - 4\bar{C}_i^J)\right] \\ \quad - \left[\frac{\partial \sigma^2_{E_i}}{\partial p_i}\right] \times \frac{1 - \bar{C}_i^J}{\bar{E}_i}\end{array}\right]. \quad \text{(B13)}$$

From Equation B9 we get the flux dominance effect of the locus *i*,

$$\beta_i \approx \frac{K}{2}(\overline{C}_i^{\,J})^2 \left[\begin{array}{l} \left[\frac{\partial \overline{E}_i}{\partial p_i}\right]^2 \times \left[\begin{array}{l} 3\,(\overline{C}_i^{\,J})^2 \sum_{j=1}^{n}\left[\frac{\mathrm{cv}_j^2}{\overline{E}_j}(2\overline{C}_j^{\,J}-1)\right] \\ +\,\overline{C}_i^{\,J}\sum_{j=1}^{n}\left[\frac{\mathrm{cv}_j^2}{\overline{E}_j}(2-3\overline{C}_j^{\,J})\right] \\ -\,2\frac{\mathrm{cv}_i^2}{\overline{E}_i}[6(\overline{C}_i^{\,J})^2 - 8\overline{C}_i^{\,J}+3] - \frac{1-\overline{C}_i^{\,J}}{\overline{E}_i} \end{array} \right] \\ -\left[\frac{\partial^2\sigma^2_{\overline{E}_i}}{\partial p_i^2}\right] \times \frac{1-\overline{C}_i^{\,J}}{\overline{E}_i} \\ +\left[\frac{\partial \overline{E}_i}{\partial p_i}\frac{\partial \sigma^2_{\overline{E}_i}}{\partial p_i}\right] \times 6\left(\frac{1-\overline{C}_i^{\,J}}{\overline{E}_i}\right)^2 \\ +\left[\frac{\partial^2 \overline{E}_i}{\partial p_i^2}\right] \times [1 - \sum_{j=1}^{n}[\mathrm{cv}_j^2\overline{C}_j^{\,J}(3\overline{C}_j^{\,J}-2)] + \mathrm{cv}_i^2(3-4\overline{C}_i^{\,J})] \end{array} \right]. \tag{B14}$$

And from Equation B11 we get the epistatic effect of the pair of loci i, j,

$$\alpha\alpha_{ji} \approx \frac{K}{4}(\overline{C}_i^{\,J}\overline{C}_j^{\,J})^2 \left[\begin{array}{l} \left[\frac{\partial \overline{E}_i}{\partial p_i}\frac{\partial \sigma^2_{\overline{E}_j}}{\partial p_j}\right] \times \frac{3\overline{C}_j^{\,J}-1}{\overline{C}_j^{\,J}(\overline{E}_j)^2} \\ +\left[\frac{\partial \overline{E}_i}{\partial p_i}\frac{\partial \overline{E}_j}{\partial p_j}\right]\left(\sum_{k=1}^{n}\frac{1}{\overline{E}_k}\right)\left[\begin{array}{l} 6\mathrm{cv}_i^2(1-2\overline{C}_i^{\,J}) \\ +\,2 + 6\sum_{k=1}^{n}[\mathrm{cv}_k^2\overline{C}_k^{\,J}(2\overline{C}_k^{\,J}-1)] \\ +\,6\mathrm{cv}_j^2(1-2\overline{C}_j^{\,J}) \end{array} \right] \\ +\left[\frac{\partial \overline{E}_j}{\partial p_j}\frac{\partial \sigma^2_{\overline{E}_i}}{\partial p_i}\right] \times \frac{3\overline{C}_i^{\,J}-1}{\overline{C}_i^{\,J}(\overline{E}_i)^2} \end{array} \right] \tag{B15}$$

In an F_2 population with additive enzymatic loci, there are some simplifications in Equations B13 to B15:

$$\overline{E}_i = m_i \qquad \frac{\partial \overline{E}_i}{\partial p_i} = 2a_i \qquad \frac{\partial^2 \overline{E}_i}{\partial p_i^2} = 0 \qquad \left| \begin{array}{l} \overline{C}_i^{\,J} = \frac{1/m_i}{\sum_{j=1}^{n}[1/m_j]} \\ \mathrm{cv}_i = \frac{a_i}{m_i\sqrt{2}}. \end{array} \right.$$

$$\sigma^2_{\overline{E}_i} = \frac{1}{2}a_i^2 \qquad \frac{\partial \sigma^2_{\overline{E}_i}}{\partial p_i} = 0 \qquad \frac{\partial^2 \sigma^2_{\overline{E}_i}}{\partial p_i^2} = -4a_i^2$$

Hence an approximation of the average flux in an F_2 population is

$$\mu_J \approx \frac{K}{\sum_{i=1}^{n} 1/m_i}\left[1 - \sum_{i=1}^{n}[\mathrm{cv}_i^2\overline{C}_i^{\,J}(1-\overline{C}_i^{\,J})]\right]. \tag{B16}$$

An approximation of the additive effect of the QTL i upon the flux is

$$\alpha_i \approx Ka_i(\overline{C}_i^{\,J})^2\left[1 + 3\mathrm{cv}_i^2(\overline{C}_i^{\,J}-1)^2 + \sum_{j\neq i}[\mathrm{cv}_j^2\overline{C}_j^{\,J}(3\overline{C}_j^{\,J}-2)]\right] \tag{B17}$$

and the contribution of the QTL i to the flux additive variance is

$$\sigma^2_{A_i} \approx \sigma^2_{E_i}(\overline{C}_i^{\,J})^4K^2\left[1 + 3\mathrm{cv}_i^2(\overline{C}_i^{\,J}-1)^2 + \sum_{j\neq i}[\mathrm{cv}_j^2\overline{C}_j^{\,J}(3\overline{C}_j^{\,J}-2)]\right]^2. \tag{B18}$$

The dominance effect of the QTL i is

$$\beta_i \approx 2Ka_i^2(\overline{C}_i^{\,J})^2\left[\begin{array}{l} 3(\overline{C}_i^{\,J})^2\sum_{j\neq i}\left[\frac{\mathrm{cv}_j^2}{m_j}(2\overline{C}_j^{\,J}-1)\right] \\ +\,\overline{C}_i^{\,J}\sum_{j\neq i}\left[\frac{\mathrm{cv}_j^2}{m_j}(2-3\overline{C}_j^{\,J})\right] \\ +\,\frac{1}{m_i}[6\mathrm{cv}_i^2(\overline{C}_i^{\,J}-1)^3 + \overline{C}_i^{\,J}-1] \end{array} \right] \tag{B19}$$

and the contribution of the QTL i to the flux dominance variance is

$$\sigma^2_{D_i} \approx (\sigma^2_{E_i})^2(\overline{C}_i^{\,J})^4K^2\left[\begin{array}{l} 3(\overline{C}_i^{\,J})^2\sum_{j\neq i}\left[\frac{\mathrm{cv}_j^2}{m_j}(2\overline{C}_j^{\,J}-1)\right] \\ +\,\overline{C}_i^{\,J}\sum_{j\neq i}\left[\frac{\mathrm{cv}_j^2}{m_j}(2-3\overline{C}_j^{\,J})\right] \\ +\,\frac{1}{m_i}[6\mathrm{cv}_i^2(\overline{C}_i^{\,J}-1)^3 + \overline{C}_i^{\,J}-1] \end{array} \right]^2. \tag{B20}$$

The epistatic effect of the pair of loci i, j ($i \neq j$) is

$$\alpha\alpha_{ji} \approx Ka_ia_j(\overline{C}_i^{\,J}\overline{C}_j^{\,J})^2\left(\sum_{k=1}^{n}\frac{1}{m_k}\right)\left[\begin{array}{l} 3\mathrm{cv}_i^2(1-2\overline{C}_i^{\,J})(1-\overline{C}_i^{\,J}) \\ +\,2 + 3\sum_{k\neq i,j}[\mathrm{cv}_k^2\overline{C}_k^{\,J}(2\overline{C}_k^{\,J}-1)] \\ +\,3\mathrm{cv}_j^2(1-2\overline{C}_j^{\,J})(1-\overline{C}_j^{\,J}) \end{array} \right] \tag{B21}$$

and the contribution of the QTL i and j ($i \neq j$) to the flux epistatic variance is

$$\sigma^2_{AA_{ij}} \approx \sigma^2_{E_i}\sigma^2_{E_j}(\overline{C}_i^{\,J}\overline{C}_j^{\,J})^4K^2\left[\begin{array}{l} 3\mathrm{cv}_i^2(1-2\overline{C}_i^{\,J})(1-\overline{C}_i^{\,J}) \\ +\,2 + 3\sum_{k\neq i,j}[\mathrm{cv}_k^2\overline{C}_k^{\,J}(2\overline{C}_k^{\,J}-1)] \\ +\,3\mathrm{cv}_j^2(1-2\overline{C}_j^{\,J})(1-\overline{C}_j^{\,J}) \end{array} \right]^2 \tag{B22}$$

with $\forall i$, $\sigma^2_{AA_{ii}} = 0$.

The R^2 of the QTL i is calculated as described in Equation 3 (methods), with

$$\sigma^2_A = \sum_{i=1}^{n}\sigma^2_{A_i}, \quad \sigma^2_D = \sum_{i=1}^{n}\sigma^2_{D_i}, \quad \sigma^2_{AA} = \sum_{i=1}^{n}\sum_{j\geq i}\sigma^2_{AA_{ij}}.$$