

Detection, number and effects of QTLs for a complex character

A Gallais^{1, 2}, M Rives³

¹ Station de Génétique Végétale, Ferme du Moulon, F91190 Gif-sur-Yvette;

² INA-PG, 16, rue Claude-Bernard, F75231 Paris Cedex 05;

³ INRA, Station d'Amélioration des Plantes Maraîchères, BP 94,
F84143, Montfavet Cedex, France

(Received 25 October 1992; accepted 23 June 1993)

Summary — Factors influencing the detection of quantitative trait loci (QTLs) *via* molecular markers are discussed in the case of recombinant inbred lines. The importance of the residual genetic variance present is emphasized and an expression for the minimum detectable effect is presented, which only depends on the design (trait heritability and number of lines studied); for relatively realistic situations, this minimum effect may be as high as 20 to 30% of the genetic SD. Therefore only QTLs with relatively large effects, and that are sufficiently closely linked to the marker, are likely to be detected. Increasing the genetic distance between parents can increase the probability of finding crosses in which significant (*ie* large-effect) QTLs are found, at the expense of decreasing the probability of detecting existing small-effect QTLs. Epistasis can diminish QTL detectability. Its possible importance has been discussed, and appears to be difficult to demonstrate. It has been shown that for a complex character, detected QTLs can at most only constitute a small subset of the existing QTLs. Furthermore, the occurrence of numerous small-effect QTLs regularly distributed over a chromosomal fraction in a directional manner has been shown to lead to the artefactual detection of a QTL by a centrally located marker.

quantitative trait / QTL / effective factor number / genetic effects

Résumé — **Mise en évidence, nombre et effets de QTLs pour un caractère complexe.** *Les facteurs affectant la mise en évidence des loci à effets quantitatifs (QTLs) par l'utilisation des marqueurs moléculaires sont discutés dans le cas des lignées recombinantes. L'importance de la variance résiduelle est soulignée et l'effet minimal d'un QTL pour qu'il soit révélé est déterminé : il ne dépend que du dispositif (héritabilité du caractère et nombre de lignées étudiées) ; pour des situations assez réalistes, cet effet peut être assez élevé : 20 à 30% de l'écart-type génétique. Il n'est donc possible de mettre en évidence que les QTLs à effet assez fort, ce qui implique aussi que les marqueurs soient assez fortement liés au QTL. La distance génétique entre parents peut augmenter la probabilité d'avoir des croisements avec des QTLs significatifs (à forts effets), mais elle ne peut que diminuer la probabilité de mise en évidence de QTLs à effets faibles. La présence d'épistasie affecte aussi la détectabilité des QTLs. Son importance possible est discutée ; elle paraît difficile à mettre en évidence. Il semble que, pour un caractère complexe, les QTLs révélés ne puissent être qu'un petit sous-ensemble des QTLs en cause. De plus, des QTLs nombreux, à faible effet, régulièrement répartis sur un segment chromosomique, mais de façon directionnelle, peuvent conduire à des effets détectables, comme un seul QTL par un marqueur situé au centre de ce segment.*

caractère quantitatif / QTLs / nombre de facteurs / effets génétiques

INTRODUCTION

Determining the number of loci involved in a quantitative character (QTLs) constituted one of the first developments in quantitative genetics, with approaches examining diallelic populations (F_2) (Castle, 1921; Wright in Castle, 1921; Mather, 1949; Wright, 1968, 1977; Mather and Jinks, 1982). However, hypotheses put forward to determine the number of genes, prudently called "effective factors", were too stringent to lead to realistic estimations for it is necessary to assume gene effects to be nearly constant for different loci so that:

$$(\sum_i a_i)^2 / \sum_i a_i^2 = n / (1 + cv_a^2) \sim n \quad [1]$$

where a_i represents the half-difference between the value of 2 homozygous genotypes at locus i , n is the number of loci at which the parent differ, and cv_a is the coefficient of variation of the distribution of the effects of the QTLs in the genome. $\sum_i a_i^2$ can be estimated via the breakdown of the genetic variance, but $\sum_i a_i$ may be estimated from a study of the means only when one of the parents has all favorable alleles, and as a consequence, the other all the unfavorable alleles; in the general case where there is a distribution of favorable alleles among both parents, $\sum_i a_i$ will be underestimated. A similar expression, established for the dominance effects, has the advantage of not being biased by the distribution of the favorable alleles between the parents if dominance is in 1 direction only.

Therefore the effective number of genes can only be an underestimation of their true number. As a matter of fact, in the experiments in which this number has been estimated, it has always been found to be rather small, even for complex characters such as yield. Adopting the terminology of Mather and Jinks, one deals with "effective factors", *ie* in fact, segments of chromosomes. Linkage interferes strongly with the demonstration of effective factors (Zeng *et al*, 1990); at the limit, their number is of the same order as the number of chromosomes. Darlington (in Colombo, 1992) has proposed the index $N = nc + x$ that relates N , the number of chromosome segments cut following meiosis, to the number of chiasmata observed in the anaphase, x , nc being the number of chromosomes; as a matter of fact, this index gives the number of effective factors as detected by quantitative genetics techniques. Colombo (1992) proposed a refinement of this

index to take into account the variation in chromosome length. The distribution of gene effects also influences this number. Thus Lande and Thompson (1990) have shown that this low estimated number could result from the L-shape of this distribution. In fact, Castle's formula and its derivatives (Park, 1977; Lande, 1981) may well determine only the number of "major" factors.

Nowadays, with the possibility of tagging genes through molecular markers, determining gene number involved in a quantitative character is a current issue, with an experimental approach that appears to be feasible. Results are accumulating that demonstrate that it suffices to look for QTLs to find some of them: in other words, one only needs enough markers to locate genes with quantitative effects for any complex character. Does this mean that there are a large number of QTLs? Preliminary results seem to show that indeed there are few QTLs with strong effects, but many medium- or even small-effect QTLs, thus confirming the hypothesis of an L-shaped distribution of locus effects, as already assumed by Castle (1921) and Sax (1923). The problem is that it is difficult, perhaps even impossible, to detect small-effect QTLs notably because of the limits to experimental precision. This raises the question of whether they are a reality or only a figment of the imagination.

Finally, the breeder would find it invaluable to know what the magnitude of the epistatic effects is between the detected or experimentally detectable QTLs. The absence of epistatic effects would make his work easier, as QTLs that are favorable in a given genetic background would also be favorable in another background, in such a way that he could accumulate favorable QTLs within one genotype through several cycles of crossing followed by selection. Various results suggest that epistasis may be limited (Paterson *et al*, 1991). Does this mean that there is no epistasis or that it is small with regard to the individual effects of the QTLs? If this is the case, then why, without undue restrictive hypotheses, do quantitative genetics methods demonstrate fairly often that there is epistasis for *per se* values in populations of the F_2 type, *ie* in populations of the same type as those used to localize QTLs (Bauman, 1959)?

The present paper reconsiders and discusses various elements associated with the detection, the number and the effects of detectable or detected QTLs in an effort to answer some of the questions addressed in this introduction, based

both on the test principle for the presence of QTLs *via* the use of markers, and on the experimental results published on the subject.

DETECTING QTLs

For a given cross between 2 lines, and to uncover QTLs using markers, the first condition is that polymorphism exists both for the marker locus and for the QTL. Once this is the case, the problem can be stated in statistical terms. To make the presentation simpler, this will be made through a comparison of the means for an experiment with recombinant inbred lines. This is equivalent to the 1-factor analysis of variance (ANOVA) approach. Lander and Botstein's (1989) method or related methods simultaneously using several markers will not be adopted here. This is justified because experimentally, Stuber *et al* (1992) have shown that both methods provide the same result and theoretically, and Rebai *et al* (1993) have shown that interval mapping is only slightly more powerful than ANOVA for intervals < 30 cM. However, the use of several markers ANOVA or multiple regression will be discussed briefly later. In the case under consideration, for a polymorphic locus there will be only 2 genotypes, M₁M₁ and M₂M₂. There will be one or several QTLs detected in the vicinity of the marker locus if the means Y₁₁ and Y₂₂ of M₁M₁ and M₂M₂ are significantly different, that is in statistical terms if:

$$(Y_{11} - Y_{22}) / \sqrt{[1/n_1 + 1/n_2]\sigma_R^2} > t_s \quad [2]$$

or:

$$(Y_{11} - Y_{22})^2 / (1/n_1 + 1/n_2)\sigma_R^2 > t_s^2$$

or in expectation, considering the situation with random effects (Gallais and Charcosset 1993):

$$a^{*2} / (\sigma_R^2 / N) > t_s^2 - 1 = k_s^2 \quad [3]$$

where *t_s* is the value of Student's *t*-test for the level of significance chosen; *n₁* and *n₂* represent the numbers of M₁M₁ and M₂M₂ lines, with *n₁* + *n₂* = *N*, the total number of recombinant lines, σ_R^2 is the (intra-class) variance of the distribution of M₁M₁ and M₂M₂ individuals (we assume these 2 variances to be equal, which is true in the absence of epistasis and if there is independence between the various QTLs; see Annex). *a** is the half-difference between the true values of the M₁M₁ and M₂M₂ classes; the variance explained

by the marker locus is directly written as a function of this parameter: $\sigma_g^2 = a^{*2}$.

The intra-class variance σ_R^2 is the sum of a genetic intra-class variance, that due to the other QTLs ($\sigma_G^2 - \sigma_g^2$), ($\sigma_G^2 = \sum_i a_i^2$ being the total genetic variance), and of the environmental variance (σ_e^2).

Expression [3] exhibits the various factors that influence the detection of the marker-QTL association:

- the value of the residual variance that depends both on the environmental variance (σ_e^2) and on the residual genetic variance ($\sigma_G^2 - \sigma_g^2$), which depends on the genetic distance between the parents;
- the number of lines under study: increasing this increases the power of the test;
- the actual value of the difference between the means of the marker genotypes, that is a direct function of the distance between the marker and the QTL, and of the effect of the QTL;
- the level of significance. Which level to choose? Taking into account the large number of markers, some associations will be found significant by chance (type I risk); it is preferable to deal with the risk at the level of the whole genome and therefore to choose a per locus risk threshold that is more stringent than usually selected: this, however, strongly increases the risk to conclude that there is no difference (no QTL) while indeed there is some (type II risk). This problem will not be discussed here.

The experimental limit

For a complex character, owing to its polygenic control and to its sensitivity to the environment, the residual variance σ_e^2 is expected to be very high. Therefore it is obvious that the probability of detecting an effect of a given QTL which is assumed to be polymorphic will be higher the smaller σ_R^2 (*ie* the smaller σ_e^2 , or $\sigma_G^2 - \sigma_g^2$, or, even better, if both are smaller at the same time), and the larger *N*. As a matter of fact, introducing $m^2 = \sigma_g^2 / \sigma_G^2$ for the proportion of the genetic variance explained by the marker, $q^2 = \sigma_g^2 / \sigma_G^2$ for that due to the QTL (with $\sigma_G^2 = \sum_i a_i^2$), and with $h^2 = \sigma_G^2 / (\sigma_G^2 + \sigma_e^2)$ for the heritability, expression [3] becomes:

$$m^2 h^2 N / (1 - m^2 h^2) > k_s^2 \text{ or } m^2 h^2 > k_s^2 / (N + k_s^2) \sim k_s^2 / N \quad [4]$$

(when N is sufficiently large > 40) or, with $m^2 = \lambda^2 q^2$ (Soller and Beckman, 1983; Cowen, 1988), λ being the linkage parameter of Schnell (1961), $\lambda = 1 - 2C$, and C is the recombination rate at the level of the lines, with $C = c$, or $C = 2c / (1 + 2c)$, according to whether the lines are doubled haploid lines or recombinant inbred lines obtained by SSD (Haldane and Waddington, 1937):

$$k^2 q^2 h^2 N / (1 - m^2 h^2) > k_S^2 \quad [5]$$

When all QTLs have the same effect, which is artificial, expression [5] becomes:

$$N \lambda^2 / h^2 (1 - h^2) \approx N k^2 h^2 / l > k_S^2 \quad [6]$$

where l is the number of polymorphic QTLs (m^2 becoming equal λ^2/l).

These expressions make it possible to illustrate the limit of the QTL detection tool (tables I, II). Suppose $N = 200$, which is already high. With $h^2 = 0.50$, which is relatively realistic for a character such as yield in maize, it is then possible at the 0.05 level to detect a QTL that explains at least 4% of the variance, *ie* 20% of the genetic SD (see tables I, II). With $\lambda = 1$ (absolute linkage), formula [6] shows that it would be necessary that there be no more than 25 QTLs for all of them to be revealed (and only 12 if $\lambda = 0.70$). With only 100 recombinant lines, only QTLs explaining at least 8% of the variance, *ie* 28% of the SD will be detected.

Thus, even with important means it will be difficult to detect QTLs with effects on the SD $< 20\%$. That is to say that it is possible to detect only QTLs with a rather strong effect on the SD. Indeed, for the plant breeder it is much more with regard to the SD that one must judge the impact of the QTL than *vis à vis* the variance. Moreover, the number of detected QTLs will be limited. In order to detect a QTL that explains 10% of the SD, *ie* 1% of the variance ($m^2 = 0.01$), and again with $h^2 = 0.50$, N ought to be equal to 800, which is simply impossible with family evaluation. The same type of conclusion would be reached for other types of progenies (F_2 s, etc). (In these types of progenies, it is more usual to express the results through the square of the correlation between marker and phenotypic value; $r^2 = \sigma_G^2 / \sigma_P^2$; with $r^2 = 0.30$, $h^2 = 0.50$, this means that $m^2 = 0.60$). Therefore failure to detect QTLs explaining < 10 to 20% of the genetic SD cannot be taken as evidence against their existence; this may simply be due to the extreme difficulty in uncovering them. In other

Table I. Minimum number of lines that ought to be studied in order to detect the effect of a QTL (at the 0.05 level), as a function of m , m^2 , h^2 , $m^2 h^2$ (m^2 is the proportion of the genetic variance explained by the marker). In fact, $m^2 h^2$, *ie* the coefficient of phenotypic correlation between marker and QTL, is the relevant criterium to be considered.

m	m^2	h^2	$m^2 h^2$	N
0.70	0.50	0.20	0.10	40
0.70	0.50	0.10	0.05	80
0.50	0.25	0.10	0.025	160
0.50	0.25	0.05	0.0125	320
0.35	0.125	0.05	0.0062	640
0.25	0.0625	0.05	0.0031	1 280

words, and whatever the situation, the precision of the experiment, as measured by h^2 , is a first major factor in determining the size of the mesh in the net used to fish for QTLs.

When testing families (here recombinant inbred lines), under fixed experimental means in terms of the number of plants that it is possible to study, the question of optimizing the design can be raised. Is it better to study many families, and few plants per family, or fewer families with more plants and therefore a better precision in the evaluation of family values? Is there an optimum? Let N_t be the total number of plants that can be studied, n the number of plants per line, and N_l the number of lines: $N_t = n * N_l$. Assuming full randomization of the plants of a line, which ensures maximal efficiency for a fixed n , condition [3] will be written thus:

Table II. Minimum value of m^2 (or m), *ie* the proportion of the genetic variance (or of the genetic SD) explained by the marker in order for its effect to be detectable (at the 0.05 level) for fixed N and as a function of h^2 the heritability.

h^2	$N = 200$		$N = 100$	
	m^2 (%)	m (%)	m^2 (%)	m (%)
0.10	20.0	44.0	40.0	63.2
0.20	10.0	31.6	20.0	44.7
0.30	6.6	28.8	13.3	36.5
0.40	5.0	22.4	10.0	31.6
0.50	4.0	20.0	8.0	28.3
0.60	3.3	18.2	6.6	25.8
0.70	2.8	16.9	5.6	23.7
0.80	2.5	15.8	5.6	22.4
0.90	2.2	14.9	4.4	21.0
1.00	2.0	14.0	4.0	20.0

$$N_t m^2 / [n(1 - m^2) + (1 - h^2) / h^2] > k_S^2$$

For fixed values of N_t , h^2 , m^2 , the calculated value of k_S^2 is maximal for $n = 1$, ie one plant per line. For instance, with $N_t = 1\ 000$, $h^2 = 0.10$, and 20 plants per family, it will be possible to detect only those QTLs that explain at least 10.7% of the variance, while with only 1 plant, this threshold will be 3.98%. The problem remains that with only 1 plant per family, one has to study the markers on n times the number of families.

Effect of the choice of parents and the size of the QTL

The percentage of genetic variance explained by the markers is a second important factor in the detectability of QTLs. Actually other QTLs with effect on the quantitative character that are not linked to the marker decrease the probability of detecting the QTLs under consideration. To detect a particular QTL, it is better to choose 2 parents that are not too different from each other, provided that there is polymorphism between these 2 parents, for this QTL or else they would also differ for many other QTLs.

Thus, in fact, 2 factors are to be taken into consideration:

- the within-progeny variance that increases with the distance between the parents, thus diminishing the power of the test;
- the difference between the means of the 2 marker genotypes that is likely to be more significant when it is large, even with a large within-progeny variance. With distant parents, it is more likely to detect QTLs with large effects in each cross, simply due to the number of sampled QTLs being larger than in crosses between close parents. This can compensate for the unfavorable effect of the increase in the residual variance.

The masking effect of the QTLs with large effect

Through their large contribution to the residual genetic variance, large-effect QTLs can “mask” QTLs with lesser effects. Thus let us assume that detectable QTLs explain p^2 % of the genetic variance σ_G^2 of a cross. Let us imagine that it is possible to conceive another cross in which these QTLs would no longer be polymorphic, as they would be homozygous: the genetical variance

would then become $(1-p^2) \sigma_G^2$, and the residual variance would be smaller. Expression [3] then becomes:

$$Nm^2h^2 > k_S^2[1 - h^2 (m^2 + p^2)] \tag{7}$$

heritability being that defined for the large base cross, in order to make comparisons possible. It becomes now possible to bring to light QTLs that were previously masked.

Table III shows that the gain in sharpness of the analysis once some QTLs with strong effects are fixed as highly significant for high heritabilities. Thus, with $h^2 = 0.80$, which is realistic, and still with $N = 200$, it is possible to pass from a threshold m^2 or 2.5%, ie 15.8% of the SD, to 0.6%, ie 7.7% of the SD, when 95% of the variance has been eliminated (as in the case of a cross between related lines). Thus a marker that explains only 0.5% of the variance of a trait from a large base-cross will be detectable if it is the sole source of variation, making up 100% of the variance for the narrow base cross under consideration.

This example shows that percentages of variation explained by a marker do not mean anything if one does not specify the genetic base over which they are defined. It is impossible to conclude that there is a large effect before this has been referred to a common larger genetic base. 100% explained variance can be associated with a small effect. Another possible expression for the effects is by reference to the mean instead of to the genetic variance or SD, the connection between the 2 being made through the genetic or the phenotypic coefficient of variation. It can be shown that the effect in terms of the mean, a^*/μ is equal to the product of the ge-

Table III. Effect of p^2 , the percentage of variance explained by markers for “strong” QTLs on the detection thresholds of QTLs with smaller effects (undetectable in a large genetic base).

$p^2(\%)$	$h^2 = 0.50$		$h^2 = 0.80$	
	$m_c^2(\%)$	$m_c(\%)$	$m_c^2(\%)$	$m_c(\%)$
0	4.0	20.0	2.5	15.8
50	3.0	17.3	1.5	12.2
80	2.4	15.5	0.9	9.5
90	2.2	14.8	0.7	8.4
95	2.1	14.5	0.6	7.7
1 QTL only	2.1	14.1	0.5	7.0

netic coefficient of variation, cv_G by $m = \sigma_{G^*} / \sigma_G$. Table IV shows the correspondence between the 3 possible expressions of the effect of a QTL for a given value of the genetic coefficient of variation. With $cv_G = 0.15$, a QTL explaining 4% of the genetic variance corresponds to 20% of the SD and to only 3% of the mean. A more logical expression of the effect of a QTL at the homozygous level would be the ratio $a^*/\sum_i a_i$. The problem is that $\sum_i a_i$, the difference between the line with all favorable alleles, and that with all the unfavorable alleles, is difficult not to say impossible to estimate.

The previous argument also allows it to be shown that for a given cross, with relatively distant parents, it is possible to detect only "large" QTLs. To reveal "small" QTLs it would be advisable to study them in narrow base-crosses between related parents, the ideal situation being when parents differ only by that QTL under consideration. In any case, an experimental limit to the detection will always remain.

Minimum detectable effect and always detectable effect

It is thus possible to look for the minimum detectable effect, when it is known that this QTL is the only source of variation. In this case, assuming to simplify that there is total linkage between the marker and the QTL, which will approximately be the case with a very dense marking of the genome, it is sufficient that:

Table IV. Correspondence between various expressions of the effect of one QTL as a function of the genetic coefficient of variation cv_G .

% Variance	% SD	% Mean (a^* / μ)		
		$cv_G: 0.10$	$c_G: 0.15$	$c_G: 0.30$
0.2	4.4	0.4	0.7	1.3
0.5	7.0	0.7	1.1	2.1
1	10	1.0	1.5	3.0
2	14	1.4	2.1	4.2
4	20	2.0	3.0	6.0
6	24.5	2.4	3.7	7.4
10	31.6	3.1	4.7	9.4
20	44.7	4.4	6.7	13.4
30	54.8	5.4	8.2	16.4
40	63	6.3	9.4	18.8
50	70.7	7.0	10.6	21.2
70	83.6	8.3	12.5	25.0
90	94.9	9.5	14.2	28.4

$$a^2 / (\sigma_e^2 / N) > k_s^2$$

or:

$$m^2 > m_{inf}^2 \text{ with } m_{inf}^2 = (1 - h^2) / Nh^2 \quad [8]$$

h^2 being the heritability σ_G^2 / σ_P^2 , at the level of the large genetic base. This threshold also corresponds to values given by [7] with $1 - p^2 = m^2$. In addition, it is possible to determine the effect of a QTL that will always be detectable whatever the remaining QTLs in the genetic base. This corresponds to values given by [7] for $p^2 = 0$; then:

$$m_{sup}^2 = k_s^2 / h^2 (N + k_s^2) \quad [9]$$

For example, with $h^2 = 0.50$, $N = 100$, $k_s^2 = 4$, $m_{inf}^2 = 0.040$, and $m_{sup}^2 = 0.077$. This means that in all crosses from the genetic base that involve a QTL the m^2 of which is larger than 0.077, one will conclude that there are indeed QTLs. Let us now consider crosses involving various QTLs drawn from the genetic background. If the effect of the strongest QTL is inferior to m_{inf}^2 , no QTL will be detected; if it is comprised between m_{inf}^2 and m_{sup}^2 , then the results will depend on the sample of QTLs drawn, and if the sample contains at least 1 QTL the effect of which is superior to m_{sup}^2 , then the test for the presence of QTLs will be positive, whatever the sample of remaining QTLs drawn. In such a situation, which supposes only the presence of QTLs with diverse effects, if there are QTLs with effects inferior to m_{inf}^2 that are relatively numerous, but at the same time QTLs with effects superior of the m_{sup}^2 threshold, a wide genetic base will be likely to rather systematically include these latter, and will more often lead to significant results than a narrow genetic base. On the average, the degree of significance of the "positive" crosses will go decreasing, while the number of positive crosses will go on increasing.

Detecting QTLs with small effects through analysis of variance or regression

As a consequence of the aforementioned, 1-way ANOVA (the classification according to the marker genotype for the locus under consideration) will detect only QTLs with a sufficiently strong effect, *ie* for which $m^2 > m_{inf}^2$. But if several marker loci are considered at the same time, one of which is closely linked to a QTL with a strong effect, that will mask the other associations in 1-way ANOVA, and then multiple-way ANOVA will

permit detection of some of these associations. Assuming that statistical independence exists between the marked QTLs, the residual from the ANOVA will be decreased from the variation that is explained by each marker locus that is introduced into the analysis. The non-significant 1-factor test will then eventually become significant. This approach, which is equivalent to a multiple regression of the character on the marker loci, is all the more interesting as it permits testing whether there are epistatic effects among the marked QTLs taken 2 at a time.

Effect of the dispersion and the linkage among QTLs

What is marked is not a locus, but a chromosome segment (QTS, for quantitative trait segment; Rives, 1991). In such a segment, there may be loci with favorable or unfavorable alleles: the results of Paterson *et al* (1990), with "fine mapping" clearly show that on 3cM, this may be the case. Such a situation may lead to a zero resultant effect of the segment, or to identical values in the 2 parents, whereas the genotypes at the individual loci are different. Thus, on 1 segment, the difference between homozygous markers can be written as: $a^* = \sum_i a_i \lambda_i$ with $\lambda_i = 1 - 2c_i$, where c_i is the percentage recombination (a_i can be positive or negative). With dispersion of the favorable genes, the existence of QTSs will decrease the probability of detecting elementary QTLs, the more so the looser the marking.

An interesting situation to consider is that of a large number of QTLs with small effects, uniformly distributed throughout the genome. In this case, one can show (see Annex 2) that if for 1 chromosome all the favorable alleles come from the same parent, then the effects of 1 marker on that chromosome will possibly represent 3% (if it is distal) to 5% (if central) of the mean (computation made for a 150 cM chromosome from a 1 500 cM genome). With a genetic coefficient of variation of 20%, this corresponds to explained percentages of variance of 2.25 and 6.25 (see table IV): this will be experimentally detectable, at least for the markers with a central position on the chromosome. There is therefore a risk to conclude on the presence of QTLs with significant effects, while there is no gene with a detectable effect. This risk increases in importance the more the values of the parents differ.

As a consequence, if QTLs with small effects, that are individually impossible to detect are nu-

merous and regularly dispersed on 1 chromosome or distributed in "segments", it will be possible to find associations with the markers due to these QTLs; in this case, one should find a deficit of QTLs at the chromosome extremities. One way to attempt to differentiate them from strong-effect QTLs is to study the associations with different sufficiently closely spaced markers. If there is a QTL with a sufficiently strong effect, its correlation with the marker should decrease with its distance to the marker. If there are, however, several markers with a relatively strong effect on the same chromosome, and these are evenly distributed, then it will be difficult to sort out the 2 situations, except at the price of very expensive experiments, such as those of Paterson *et al* (1990) "fine mapping". Then QTL or QTS? The debate is still open. And apart the fact that it risks being lengthy, the search for a gene that is supposed to lie in the neighborhood of a marker through "chromosome walking" will not always be successful.

The existence of QTSs could well be at the origin of apparent superdominance observed in maize by Edwards *et al* (1987, 1992) for kernel weight and various characters that are associated with it, as has already been demonstrated by Comstock and Robinson (1952) using the "quantitative genetics" approach.

Effect of the distance between marker and QTL

As the difference between means of 2 homozygous marker genotypes can be written $a_i \lambda_i$, it is obvious that if λ_i is small, ie if c is close to 1/2, then it will not be possible to detect the QTL, even if its effect is large. Figure 1 shows that experiments will detect only QTLs that are sufficiently closely linked, ie $c < 0.30$, and with a sufficiently strong effect.

Effect of epistasis

Epistasis between the QTLs or QTS can reinforce the similarity of action of homologous blocks and thus decrease the probability of detecting the effects of such blocks. However, the main effect of the presence of epistasis will be to decrease the significance of principal effects: with strong epistasis, these effects no longer have meaning. As a matter of fact, with epistasis

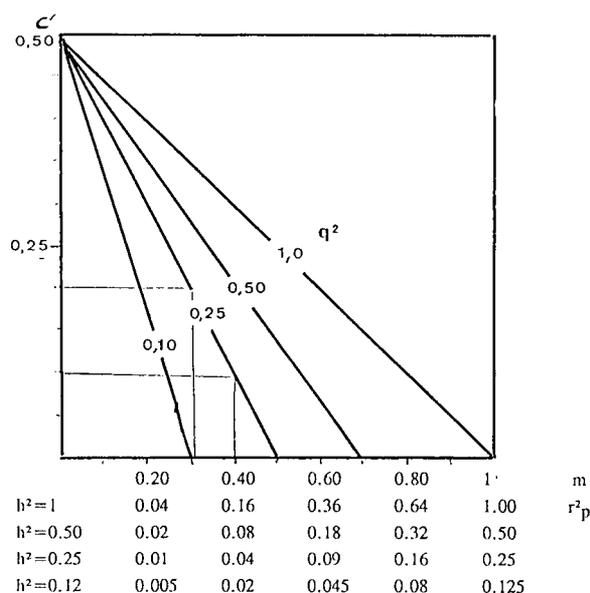


Fig 1. Possible variation range of the recombination rate (c) as a function of the percent phenotypic variation (r_p^2) explained by the marker and of the part of genetic variance due to the QTL. As for a given design one can detect only those QTLs having a r_p^2 superior to a certain threshold, it is possible to detect the QTL only when the linkage is strong enough and when the effect of the QTL is sufficiently strong as well. With $h^2 = 0.25$ and $q^2 = 0.25$ (which is high) and the threshold $r_p^2 = 0.04$ ($-k^2 / N$), it will be possible to detect only QTLs located at < 10 cM; with $q^2 = 0.10$, nothing will be detected; doubling the number of units studied, the r_p^2 threshold will become 0.02 and the recombination rate for detection can then vary between 0 and 20 cM.

and with a random effect model that can be shown to be preferable (Charcosset and Gallais, 1993), the ANOVA test of QTL effects should not be made with the residual but with the interaction mean square. Some significance of the main effects of QTLs must therefore have been overestimated.

ON THE PRESENCE OF EPISTASIS

Absence of epistasis between marked QTLs

Various results (Paterson *et al*, 1991) tend to show that there is little epistasis between marked QTLs, *ie* with strong effects. Given what has been said, this result should not be surprising: epistasis when present can decrease the probability of detecting individual effects, or then they lose any meaning. Thus, in most cases, detected QTLs with the strongest effect are likely to be those that show the least epistasis. To check

on this, it is necessary to study epistasis between markers which individually taken detect nothing significant. In addition, to obtain a sufficiently precise test of the existence of epistasis between 2 QTLs requires more means, *ie* more experimental recombinant lines than for the test of additive QTLs. For a given size of experiment, however, the test for epistasis will always be less precise than that of the individual effects of the QTLs. Finally, a hypothesis that is difficult to reject is that of epistasis integrated within some QTLs.

However in maize, Zehr *et al* (1992) found a relatively strong epistasis for QTLs affecting grain yield in top-cross. Epistasis will be more difficult to detect relative to top-cross values than to *per se* values, as a consequence of the "dilution" of effects by the tester, their variance being divided by 16 (Gallais, 1991).

It is to be noted that: a) Paterson *et al*'s (1991) experiment deals with interspecific crosses (wild x cultivated tomato); therefore it is likely that those QTLs that were detected are QTLs of domestication, *ie* the QTLs with sufficiently strong action that differ between wild and cultivated genotypes. As a matter of fact, it is known that such QTLs with relatively strong effects generally influence morpho-physiological traits with a relatively simple genetic control, hence exhibit little epistasis. In this interspecific experiment the sampling of the QTLs was therefore likely to be biased and it is difficult to draw general conclusions relevant to the intraspecific situation from this experiment. The case of the maize x teosinte cross (Doebley *et al*, 1990, 1991; Szabo, 1991) was better analyzed from the point of view of evolution and is in favor of this interpretation; b) Paterson *et al* (1991) considered their results to be coherent with those obtained through the quantitative genetics approaches. It is true that investigations of variances in populations rarely conclude on an important epistasis; but, with a few exceptions (Gallais, 1990, 1992), their power to detect epistasis is very low. On the contrary, some rare investigations, *eg* those of Bauman (1959) or Mather and Jinks (1982), were well adapted to detect epistasis and indeed they rather frequently detected some interaction; c) the test of epistasis between effective (*ie* linked to QTLs) linked markers can be misleading. Indeed, it can be shown that when there is only 1 QTL between 2 markers, this generates apparent epistasis (see Annex 3).

Epistasis between unmarked QTLs

The preceding discussion, which indicates that epistasis could prevail for QTLs undetected by their direct effects, is essentially statistical. From the biological point of view, however, the more a gene, even with a very important role, is involved in a network of actions and interactions, the more difficult it will be to detect. This does not mean that all QTLs with small effects must include epistasis: they can have small additive effects, or small strong interacting effects.

It is probably this type of epistasis that is revealed by quantitative genetics methods, which could explain the apparent disagreement between the 2 types of analysis. These studies therefore retain their interest, especially when they make use of a design that is well adapted to detect epistasis: this is the case with the lines obtained by haplodiploidization (Choo *et al.*, 1980; Gallais, 1990), or the study of top-cross progenies (Gallais, 1991). It would be especially interesting to perform both types of analysis on an F_2 . This requires only at least 2 HD lines or 2 F_3 plants (to be crossed to the tester) to be obtained per F_2 plant. It is also possible to study the structure of the total genetic variance with other designs such as Bauman's (1959), or the "triple test cross" as presented by Mather and Jinks (1982).

Occurrence of epistasis between QTLs and the genetic background

If there are interactions between the variable QTLs and the QTLs that are fixed (genetic background), and this is difficult to exclude, then they will not be detected. Hence, not to find epistasis in a given cross cannot mean that there is no epistasis among the QTLs for the trait under consideration: epistatic effects may appear when certain QTLs will be introduced into genetic backgrounds in which they have not been studied. A functional epistasis could well exist: QTLs being additive when far from the saturation of the functions and interactive, *ie* non-additive, when close to saturation of functions (Dillmann, 1992).

Experimental proof of epistasis vis-à-vis the genetic background is shown when QTLs detected for the value in top-cross are not always a subset of those detected for own-value (Pioneer HY, 1989; Stuber, 1989; Baud, 1992). If there were only additivity, QTLs detected in top-crosses should be the same as for own-value, or

be a subset of these. With dominance but without epistasis, they would necessarily be a subset of these. But this is not the case, as top-cross specific QTLs have been observed. Charcosset and Causse (1993), studying recombinant inbred lines from 3 connected crosses, have also observed interactions between QTLs and genetic background in spite of the absence of interactions between detected QTLs.

In conclusion, some experimental results already show the occurrence of epistasis involving the QTLs, and in any case, for different reasons, its importance is bound to be underestimated. The genome cannot be reduced to a string of genes; this is all the more so the more closely associated with selective value the traits are.

ON NUMBER, NATURE AND DISTRIBUTION OF DETECTED QTLs

Restriction in causes of variation permits the detection of QTLs

The genetic base

The fact of finding for a given cross in a given location only a limited number of QTLs is evidence that in this location there are some important limiting factors: it is obvious that there cannot be a large number of factors with relatively strong effects. In order to detect them, it may be necessary to diminish the influence of all that can blur or dilute their effects. Now why should these limiting factors always have strong effect? Why cannot situations exist with a large number of "limiting" factors with small effects, which would hence be still more difficult to detect?

The study of various crosses reveals common and specific QTLs: according to the cross, these can be very different if the crosses are made between distant parents. It is necessary to note at this point that if only a very limited number of QTLs appear in the wild x cultivated tomato crosses (Paterson *et al.*, 1990), this may be due to the fact that "big" detected QTLs (QTLs of domestication) could have masked QTLs with smaller effects, not specifically involved in domestication. The existence of "hot points" in the genome (Helentjaris, 1991), *ie* of points in the genome that are frequently involved in QTLs, is not surprising and is in good agreement with a kind of L-shaped distribution of effects, the most

readily detected QTLs being those with strong effects, hence often the same ones. It can however, also indicate problems associated with the frequency of the alleles at the QTL; it is sufficient that in the population, 2 relatively different alleles be equally frequent to give the illusion of hot points. On a set of crosses, one will repeatedly find the QTL corresponding to the highest frequency among the heterozygous genotypes. If there are QTLs in which multiallelism is more frequent, then these QTLs will appear more often, for in a population of crosses, the probability of heterozygosity will be higher for those QTLs than for the other in which there is only diallelism. The "polyallelism" that is obvious in QTSs with rather strong effects associated with each cross can also generate this situation which will be associated, however, with an apparent "superdominance".

Environment of the test and nature of the traits

Changing the environment of the test reveals common QTLs but also specific ones (Paterson *et al.*, 1991). This indicates that in a given environment, there cannot be many limiting factors, but, insofar as the detected QTLs change with the environment, hence with the nature of the limiting factor, this indeed means that there are numerous QTLs. These will be easier to detect if one studies for each of them the component elementary trait which they control.

The probability of detecting a QTL is higher in a well-controlled environment, and at the level of elementary characters, components or factors of the complex trait. Environment control ought to eliminate random variations that increase the experimental error; observing elementary characters amounts to reducing the number of loci involved, hence to obtaining a better precision.

An important problem is whether QTLs detected by the markers are dependent on the environment. In contrast to Paterson *et al.* (1991) results, Stuber *et al.* results (1992) show only common QTLs. The problem is whether the genotype x environment interaction is important. If it is, specific QTLs can be expected, provided that there are only some genes controlling adaptation to particular environments. If specific QTLs cannot be detected this could mean that their effects are too small, due to their number. However another explanation is possible; i) it has been observed

that genotype x environment interaction is more controlled by specific than general effects (Rojas and Sprague, 1951); ii) specific effects can be due to epistasis; and iii) the presence of epistasis can mask the detection of the QTLs. The detected QTLs could then be non-epistatic and non-interactive with environments in spite of the genotype x environment interaction.

QTLs for a complex trait such as yield are likely to be numerous

Evidence from long-term selection experiments

The simplest hypothesis for explaining the response to 70 generations of selection for protein or oil content of the maize kernel is indeed that it is controlled by a large number of QTLs (Dudley, 1974). Other hypotheses exist such as transpositions, mutations etc, but they seem to lack foundation. It is particularly difficult to explain the immediate response to reverse selection after 45 generations by anything other than the occurrence of many QTLs in the starting material, the favorable and unfavorable alleles of which are relatively dispersed. This implies a "reserve variability" through linkage and linkage disequilibrium created by selection; this is "Bulmer's effect" (Bulmer, 1976, 1980) that can be easily explained with a large number of loci.

Evidence from simulations

Simulation experiments of Zhang and Smith (1992) show that response to long-term selection can only be explained through the occurrence of a large number of independent QTLs, this number being smaller if there is interaction: a simple model with a limited number of non-interactive QTLs leads relatively quickly to a selection plateau through the exhaustion of the available variability.

Evidence from changing the test circumstances

The fact that detected QTLs change with the genetic base, with the environment, or with the trait speaks in favor of a large number of detectable QTLs: many physiological traits must have an impact on a character such as yield.

Evidence from the effect of undetected QTLs

In addition, one must count all QTLs that are not detected due to the experimental limit or because of epistasis. Various results provide evidence of their existence. In addition to those already mentioned in this paper, 2 selection experiments with markers were especially demonstrative in this regard.

In Stuber's (1989) experiment, individual phenotypic selection and genotypic selection restricted to the markers, applied to yield, gave the same results over 1 selection cycle. But the variation in frequency of the (favorable) markers was far larger for genotypic selection than for phenotypic selection. In the latter selection, progress must have been made from other sources of variation than those associated with the marked QTLs. In this experiment, however, the marking of the genome was not very dense.

In Paterson *et al's* (1991) experiment as well, it is apparent that phenotypic selection was more efficient than marker-assisted selection for fruit weight, a relatively heritable trait; this shows that the remainder of the variation undetected through the markers had a strong impact on the response to selection.

On the importance of effects and their distribution law

The large size of the r^2 values (square of the correlation between trait and marker) should not be misleading: significance is tested vis-à-vis zero, but the confidence interval can be relatively large and asymmetrical towards high values. Moreover, r^2 is relevant to a rather narrow genetic base, and to one cross only. Thus, with only one locus, $m^2 = 1$, so r^2 is equal to the heritability of the trait under study. Hence for the same QTL, it is possible to pass from a high r^2 , equal to h^2 , to a very small r^2 , simply by increasing the number of polymorphic QTLs. As a consequence it is not possible from the results obtained on 1 cross to make an inference on the level of the population. To predict the explained variance in the population from these results it is necessary to assume that there is no epistasis and to take into account the probability of the genotypes.

On the L-shaped distribution

The fact that there ought to be few genes with strong effects and many with small effects seems

rather evident and always has been present in the mind of quantitative geneticists (Wright, 1921, in Castle, 1921; Sax, 1923; Wright, 1968, 1977). The problem is to determine:

- from which threshold do we establish that a gene has a strong effect? What do we call a major gene? A major gene has visible effects on the distribution of phenotypes; the QTLs that are under investigation in general have no visible effect on the distribution. The effect will be considered strong if it is easily detected: in the example taken above, a QTL that explained < 4% of the genetic variance, *ie* 20% of the SD appeared very difficult to detect: 4% is small, but 20% is "big"! Hence one has to pay attention to the reference used: SD or means as a reference seem biologically more correct; but then the additivity properties of the contributions are lost, whereas they exist with variance if there is no epistasis.
- what proportion of the genetic variance is explained by the detected QTLs? This has meaning only for the cross under consideration and it is difficult to extend to the population level.

The first experimental results on plants (Edwards *et al*, 1987; Paterson *et al*, 1990) and on *Drosophila* (Shrimpton and Robertson, 1988) indeed show this L-shaped distribution, however truncated for smaller effects. It is necessary to remark that its shape can have been accentuated by the occurrence of false positive; the distribution of errors on r^2 is of necessity disymmetric; as a consequence, the L-shaped distribution is expected even with small and little variable (or even uniform) effects of the QTLs.

On the nature of detected QTLs

Genes that affect a quantitative trait have no reason whatsoever to be of a different nature to those affecting a qualitative character. They can, however, be involved in interaction networks, especially for traits associated with selected value. The problem is then more that of breaking down the complex trait into elementary characters. From this point of view, it is remarkable that experimental results on maize (Edwards *et al*, 1992) show the existence of QTLs for height in zones of the genome where dwarf mutants are already known to exist. As has been proposed by Thompson (1975), detected QTLs could be alleles of major genes. While this may be a worthwhile trail to look for QTLs starting from some major genes affecting a quantitative trait

(one only has to retain markers that are close to these major genes; this is supported by studies on mice and man; Hilbert *et al*, 1991) the problem is that not all "major" genes are known (this is especially true of yield genes).

Reviewing the hypotheses of quantitative genetics

Quantitative genetics has made the hypothesis of a large number of genes with small effects only to provide a genetic basis for the theory of long-term natural or artificial selection. This was necessary for analyzing the concentration of selection cycles, or making possible a statistical approach to the decrease in genetic variance. By the same token, Mather and Jinks (1982) introduced the hypothesis of a uniform distribution of effects, without necessarily implying a large number of loci (but where does "large number" begin?) to be able to estimate the number of effective factors. This, by definition, can only be inferior to the true number of loci and could, according to Lande and Thompson (1990), be equal to the number of large-effect QTLs.

At this point it may be useful to draw a clear definition of and a distinction between the concepts that are described by the words: *ie* between QTLs, effective factors or "QTSs" (for quantitative trait segments). Some confusion arises in many papers because QTLs are alternatively and indifferently taken either as true loci (*ie*, at least implicitly; coding sequence plus regulating elements expressed as a protein-enzyme) or as chains of loci, *ie* QTSs or effective factors. A very important consequence, in addition to the operative one that QTSs offer no perspective of being able to be cloned and transferred, is that interactions can occur within QTSs, while not occurring between the QTSs, which means that recombination can alter and possibly destroy QTSs and their effects (or build new ones, which is the goal of recurrent selection), whereas QTLs would remain intact, except for mutations (and rare within locus recombination).

If one is not interested in selection, however, nor in estimating the number of genes involved, *ie* if one is interested only in gene effects and their variance, quantitative genetics does not make strong hypotheses on gene numbers and their effects: the analysis of variance is indeed little sensitive to deviations from normality. The

very difficulty in assigning a sufficiently logical and simple law to the L-shaped curve is the main reason why it has so far rarely been introduced into the models. Lande and Thompson (1990) have introduced it in a discontinuous way from a geometric series, and Zhang and Smith (1992) did consider a γ -type law. If one considers, however, that genes with the larger effects can be subject to "tagging", then the remaining genes are closer to the classical hypotheses of quantitative genetics. Combining tagging and quantitative genetics ought to give more power to the analysis of the sources of genetic variation and to their utilization in selection.

ON THE DIRECTED BUILDING OF GENOTYPES

One of the main advantages of gene tagging is the possibility of controlling recombinations to obtain more efficient transgressions. This argument is fully justified and the packing together of favorable QTLs into one and the same genotype will be the easier as epistasis will not be present. If there were epistasis, it would be necessary to test and identify the most favorable associations, which would be heavier and would require more investments, but still justify the use of tagging. The more troublesome epistasis is that resulting from the interaction of the marked QTLs with the genetic background without possibility of identifying the relevant QTL(s). This is mainly what impairs gene transfer programs. A typical example of this type of situation is afforded by the opaque-2 gene in maize: this gene has a very strong effect on the protein composition of the kernel, but unfavorable pleiotropic effects on yield, and many other important traits, when it is introduced by back-crossing into normal genotypes. It is now known that these effects of apparent pleiotropy are due to epistasis, *ie* unfavorable interactions with the remainder of the genome; modifier genes have been selected that correct the main defects of opaque-2. As a matter of fact, opaque-2 is both pleiotropic and epistatic, or rather pleiotropic through epistasis, and a truly constitutive epistasis for what matters, as opaque-2 has been shown to be a regulation gene controlling the expression of many other genes at the same time through the operation of a protein that binds to the regulatory elements upstream from their TATA box (Burr, 1990).

CONCLUSION

From this discussion, there emerge 3 important conclusions vis-à-vis the genetic study of quantitative traits with the help of markers:

– finding a marker–trait association does not necessarily mean that there is only one QTL; there may be several; the marker marks a segment (QTS) and the difference between QTS and QTL can be difficult to demonstrate;

– it is not easy to conclude on the number of QTLs for a given quantitative character. The limit in experimental precision will always make it difficult to detect small-effect QTLs or QTLs that are involved in an interaction network; the failure to detect QTLs does not mean that there are none. A solution will be to reduce the sources of variation due to the genetic background and to environment. One should not forget that 4% non-explained variance leads to differences of 20% at the level of the SD, which is considerable for the breeder.

– it is difficult to conclude on the importance of epistasis.

It is clear that the identification of QTLs with sufficiently strong effect and if possible independent of the genetic background, will permit a better analysis of the physiological bases of a quantitative trait and will facilitate certain “genetic manipulations”, such as transfer of favorable genes through back-crossing, or genetic engineering, increasing their frequency through crossing followed by selection, directed build-up of favorable genotypes, etc.

To use all the sources of variation, *ie* to use the variation associated with the small-effect QTLs that are undetected because undetectable, it will always be necessary to resort to classical quantitative genetics, defined as the art of performing genetics without identifying the genes. It is evident that combining the 2 approaches, tagging and biometrical genetics, will render selection more efficient. The possibility of marking chromosome fragments, even without detectable QTLs, ought to permit the development of a new quantitative genetics with a more powerful predictive value than the former and thus the construction of a new theory of selection that should lead to new, still more efficient selection methods.

ANNEX 1

Homogeneity or heterogeneity of the variance of the genotypic classe of the marker in the case of recombinant inbred lines

The case when there is epistasis between 2 QTLs

Let there be 2 QTLs with at locus 1, alleles A and a and at locus 2, B and b with a polymorphic marker (M and m) linked to A (M in coupling to A). Following Mather's notation, the coded genotypic values of the recombinant inbred lines will be:

$$\begin{aligned} AABB &= a_1 + a_2 + i_{12} \\ AAbb &= a_1 - a_2 - i_{12} \\ aaBB &= -a_1 + a_2 - i_{12} \\ aabb &= -a_1 - a_2 + i_{12} \end{aligned}$$

With linkage between loci 1 and 2, the mean of the marker genotypes MM and mm is not affected by epistasis:

$$\begin{aligned} MM &= a_1 \lambda_1 \\ mm &= -a_1 \lambda_1 \end{aligned}$$

On the contrary, the intra-class variance is affected:

$$\begin{aligned} V_{MM} &= (1 - \lambda_1^2) a_1^2 + i_{12}^2 + 2\lambda_1 a_2 i_{12} \\ V_{mm} &= (1 - \lambda_1^2) a_1^2 + i_{12}^2 - 2\lambda_1 a_2 i_{12} \end{aligned}$$

The case of linkage disequilibrium between QTLs without epistasis

In this case loci 1 and 2 and the marker locus are supposed to be on the same chromosome. To simplify, let us assume that crossing-overs are independent. In this case, 2 parameters are sufficient to describe recombination probabilities, c_1 , the recombination rate between locus 1 and the marker, and c_2 that between locus 2 and the marker.

The mean of the 2 marker genotypes will be (λ_1 being the linkage parameter between locus 1 and the marker, and λ_2 that between locus 2 and the marker):

$$\begin{aligned} MM &= a_1\lambda_1 + a_2\lambda_2 = a^* \\ mm &= -a_1\lambda_1 - a_2\lambda_2 = -a^* \end{aligned}$$

Variances are not affected by linkage disequilibrium:

$$V_{MM} = V_{mm} = (1 - \lambda_1^2)a_1^2 + (1 - \lambda_2^2)a_2^2$$

It is easily seen that this result can be extrapolated to n loci linked to the marker:

$$a^* = \sum_i \lambda_i a_i$$

and:

$$V_{\text{intra-class}} = \sum_i a_i^2 - \sum_i \lambda_i^2 a_i^2$$

The linkage disequilibrium creates a covariance between the 2 QTLs which is found again in totality at the inter-class level:

$$V_{\text{inter-class}} = \lambda_1^2 a_1^2 + \lambda_2^2 a_2^2 + 2\lambda_1\lambda_2 a_1 a_2$$

It can indeed be verified that:

$$V_{\text{total (individual)}} = V_{\text{inter-class}} + V_{\text{intra-class}}$$

As a consequence, without epistasis between QTLs, the variance homogeneity hypothesis is justified. It is to be recalled that, in F_2 or F_3 , this hypothesis will no longer be justified, because of dominance. This is a disadvantage of F_2 s as opposed to recombinant inbred lines for detecting QTLs through analysis of variance methods.

ANNEX 2

Working out the effect of a chromosome or a chromosome region detected by markers, when there are a large number of genes with very small effects

In *Annex 1*, we have shown that when the favorable genes on a chromosome are all originated from the same parent, the effect associated with the marker can be written:

$$a^* = \sum_i \lambda_i a_i$$

Let L_c be the length, in Morgans of the chromosome and L the total length of the genome. Assuming a large number of genes with small effects, all equivalent, and using a transformation introduced by Dekkers and Dentine (1991), it is possible to write:

$$a^* = \sum_i \lambda_i (S_a / L) \Delta x$$

S_a representing $\sum_i \lambda_i a_i$ over the whole of the genome, and Δx the (very small) distance between 2 consecutive QTLs; replacing c_j in $\lambda_j = 1 - c_j$ by Haldane's (1919) function, $c = 1/2 (1 - e^{-2x})$, for a marker located at L_{c1} and L_{c2} Morgans of the chromosome extremities ($L_{c1} + L_{c2} = L_c$), one obtains:

$$a^* = S_a / L \left[\int_0^{L_{c1}} e^{-2x} dx + \int_0^{L_{c2}} e^{-2x} dx \right]$$

$$\text{or } a^* / S_a = [(1 - e^{-2L_{c1}}) + (1 - e^{-2L_{c2}})] / 2L$$

This ratio can be considered as a superior limit for a^* / μ , the effect of the QTL as a function of the mean.

Example: with $L = 15$ Morgans, $L_c = 1.5$ Morgans, and $L_{c1} = L_{c2}$, $a^* / S_a = 0.051$.

ANNEX 3

Effect of linkage among markers on the detection of epistasis between QTLs linked to those markers

Consider a QTL Q, with alleles Q_1 and Q_2 between 2 linked markers M with alleles M_1 and M_2 and N with alleles N_1 and N_2 . With the assumption of independence between cross-overs, and with alleles with the same subscript originate from the same parent, they are 8 types of recombinant inbred lines, $M_1 Q_1 N_1$, $M_1 Q_1 N_2$, $M_1 Q_2 N_1$, $M_1 Q_2 N_2$, $M_2 Q_1 N_1$, $M_2 Q_1 N_2$, $M_2 Q_2 N_1$, $M_2 Q_2 N_2$ with respectivement frequencies $c_1 c_2 / 2$, $c_1 c_2 / 2$, with $c_1 = 1 - c_1$, $c_2 = 1 - c_2$, c_1 and c_2 being the recombination frequencies between M and Q and between Q and N.

At the level of markers they are only 4 classes with the following mean values around the general mean:

$$M_1 N_1 = a(\lambda_1 + \lambda_2) / (1 + \lambda_3)$$

$$M_1 N_2 = a(\lambda_1 - \lambda_2) / (1 - \lambda_3)$$

$$M_2 N_1 = -a(\lambda_1 - \lambda_2) / (1 - \lambda_3)$$

$$M_2 N_2 = -a(\lambda_1 + \lambda_2) / (1 + \lambda_3)$$

with $\lambda_1 = 1 - 2c_1$, the Schnell's linkage parameters with the property $\lambda_1\lambda_2 = \lambda_3$, c_3 being the recombination frequency between M and N, a is half of the difference between the 2 homozygous geno-

types for the QTL. The average effect of marker M is obviously $\lambda_1 a$, and that of N $\lambda_2 a$. It is clear that due to the linkage between M and N ($\lambda_3 \neq 0$) the main effects of M and N are not additive for determining the value of the 4 genotypes. This means that the linkage among markers "generates" a pseudo-epistasis.

REFERENCES

- Baud S (1992) Recherche de corrélations entre marqueurs moléculaires de type RFLP et caractères agronomiques chez *Zea mays*. Thèse Doct Fac Sci Lyon 1
- Bauman LF (1959) Evidence of non-allelic gene interaction in determining yield, ear height and kernel row number in corn. *Agron J* 51, 531-534
- Bulmer MG (1976) The effect of selection on genetic variability: a simulation study. *Genet Res* 28, 101-117
- Bulmer MG (1980) *The Mathematical Theory of Quantitative Genetics*. Clarendon Press, Oxford, 225 p
- Burr B (1990) Maize regulatory gene opaque-2 encodes a protein with a "leucine zipper" motif that binds to zein DNA. *Proc Natl Acad Sci USA* 87, 46-50
- Castle WE (1921) An improved method of estimating the number of genetic factors concerned in cases of blending inheritance. *Science* 54, 223
- Charcosset A, Causse M (1993) QTL detection in three connected populations of maize recombinant inbred lines: the case of earliness. In: *Gordon Conference*. Ventura, USA, February 1993
- Choo TM (1980) Doubled haploids for estimating additive epistatic variances in self-pollinated crops. *Can J Genet Cytol* 22, 125-127
- Colombo PC (1992) A new index for estimating genetic recombination from chiasma distribution data. *Heredity* 69, 412-415
- Comstock RE, Robinson HF (1952) *Estimation of Average Dominance of Genes in Heterosis* (Gowen JW, ed) Iowa State Coll Press, 494-516
- Cowen NM (1988) The use of replicated progenies in marker-based mapping of QTLs. *Theor Appl Genet* 75, 857-862
- Dekkers JCM, Dentine MR (1991) Quantitative genetic variance associated with chromosomal markers in segregating populations. *Theor Appl Genet* 81, 212-220
- Dillmann C (1993) Organisation de la variabilité génétique chez les plantes : modélisation des effets d'interaction, conséquences sur la réponse à la sélection. Thèse Institut national agronomique Paris-Grignon
- Doebley J, Stec A, Wendel J, Edwards M (1990) Genetic and morphological analysis of a maize-teosinte F_2 population: implication for the origin of maize. *Proc Natl Acad Sci USA* 87, 9888-9892
- Doebley J, Stec A (1991) Genetic analysis of the morphological differences between maize and teosinte. *Genetics* 129, 285-295
- Dudley JW (1974) *Seventy Generations of Selection for Oil and Protein in Maize*. *Crop Sci Soc Am*, Madison, WI, USA
- Edwards MD, Stuber CW, Wendel JF (1987) Molecular marker facilitated investigations of quantitative trait loci in maize. 1. Numbers, genome distribution and types of gene action. *Genetics* 115, 113-125
- Edwards MD, Helentjaris T, Wright S, Stuber CW (1992) Molecular marker facilitated investigations of quantitative trait loci in maize. 4 Analysis based on genome saturation with isozyme and RFLP markers. *Theor Appl Genet* 83, 765-774
- Gallais A (1990) Quantitative genetics of doubled haploid populations and application to the theory of line improvement. *Genetics* 124, 199-206
- Gallais A (1992) A general approach of the study of a population of testcross progenies and consequences for recurrent selection. *Theor Appl Genet* 81, 493-503
- Haldane JBS (1919) The combination of linkage values and the calculation of distances between loci of linked factors. *J Genet* 8, 299-309
- Haldane JBS, Waddington CJ (1937) Inbreeding and linkage. *Genetics* 16, 192-195
- Helentjaris TG (1991) RFLP analysis for manipulating agronomic traits in plants. In: *Plant Breeding in the 1990* (Stalker HT, Murphy JP, eds) North Carolina Univ Press, Raleigh, NC, USA
- Hilbert P, Lindpaintner K, Beckmann JS, Serikawa T, Soubrier F, Dubay C, Cartwright P, De Gouyon B, Julier C, Takahasi S, Vincent M, Ganten D, Georges M, Lathrop GM (1991) Chromosomal mapping of two genetic loci associated with blood pressure regulation in hereditary hypertensive rats. *Nature* 353, 521-529
- Lande (1981) The minimum number of genes contributing to quantitative genetic variation between and within populations. *Genetics* 99, 541-553
- Lande R, Thompson R (1990) Efficiency of marker-assisted selection in the improvement of quantitative traits. *Genetics* 124, 743-756
- Lander ES, Botstein D (1989) Mapping mendelian factors underlying quantitative traits using RFLP linkage maps. *Genetics* 121, 185-199
- Mather K (1949) *Biometrical genetics*. Methuen, London
- Mather K, Jinks JL (1982) *Biometrical Genetics: The Study of Continuous Variation*. Chapman and Hall, NY, 3rd edn
- Park YV (1977a) Theory for the number of genes affecting quantitative characters. I. Estimation of an variance of the estimation of gene number for quantitative traits controlled by additive genes having equal effects. *Theor Appl Genet* 50, 153-161
- Park YV (1977b) Theory for the number of genes affecting quantitative characters. II Biases from drift,

- dominance, inequality of gene effects, linkage disequilibrium and epistasis. *Theor Appl Genet* 50, 163-172
- Paterson AH, Damon S, Hewitt JD, Zamir D, Rabinovitch HD, Lincoln SDE, Lander ES, Tanksley SD (1991) Mendelian factors underlying quantitative traits in tomato: comparison across species, generations and environment. *Genetics* 181-197
- Pioneer Hy Biol Intern (1989) *Identification and Genetic Localization of Loci affecting Yield and Other Agronomic Traits in Maize*. Poster, XII Congr EUCARPIA, Göttingen, Germany
- Rebaï A, Goffinet V, Mangin B (1993) Comparing power of different methods for QTL detection. *Biometrics* (in press)
- Rives M (1991) La recombinaison génétique. In: *Méribel 91. Marquage Moléculaire et Sélection* (M Lefort-Buson, ed) Méribel-les-Allues, INRA, Dept Amélior Plantes
- Sax K (1923) The association of size differences with seed-coat pattern and pigmentation in *Phaseolus vulgaris*. *Genetics* 8, 552-560
- Schnell FW (1961) Some general formulations of linkage effects in inbreeding. *Genetics* 46, 947-952
- Shrimpton AE, Robertson A (1988) The isolation of polygenic factors controlling bristle score in *Drosophila melanogaster*. II. Distribution of third chromosome bristle effects within chromosome sections. *Genetics* 118, 457-459
- Soller M, Beckman JS (1983) Genetic polymorphism in varietal identification and genetic improvement. *Theor Appl Genet* 67, 25-33
- Stuber CW (1989) Marker-based selection for quantitative traits. In: *Science for Plant Breeding: Proc XII Congr EUCARPIA* (Robbelen G, ed), Paul Parey Scientific, Berlin, 31-49
- Stuber CW, Lincoln SE, Wolff DW, Helentjaris T, Lander ES (1992) Identification of genetic factors contributing to heterosis in a hybrid from two elite maize inbred lines using molecular markers. *Genetics* 132, 823-839
- Szabo V (1991) Cartographie des gènes qui différencient maïs et téosine à l'aide de marqueurs. Thèse Fac Sci, Orsay, France
- Thompson JN Jr (1975) Quantitative variation and gene number. *Nature* 258, 665-668
- Wright S (1968) *Evolution and the Genetics of Populations. I. Genetic and Biometric Foundations*. Univ Chicago Press, Chicago, USA
- Wright S (1977) *Evolution and the Genetics of Populations. 3. Experimental Results and Evolutionary Characters*. Univ Chicago Press, Chicago, USA
- Zehr BE, Dudley JW, Chojecki J, Sghai Maroof MA, Mowers RP (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
- Zeng ZB, Houle H, Cockerham CC (1990) How informative is Wright's estimator of the number of genes affecting a quantitative character. *Genetics* 126, 235-247
- Zhang W, Smith C (1992) Computer simulation of marker assisted selection utilizing linkage disequilibrium. *Theor Appl Genet* 83, 813-820