

Cartographie du génome et caractères quantitatifs : L'approche biométrique et les contraintes biologiques des agrumes.

X. PERRIER, P. OLLITRAULT et Cécile DUBOIS*

GENOMIC MAPPING AND QUANTITATIVE TRAITS :
THE BIOMETRIC APPROACH AND THE CITRUS BIOLOGY
CONSTRAINTS.

X. PERRIER, P. OLLITRAULT and Cécile DUBOIS.

Fruits, Numéro spécial Agrumes 1992, p. 135-144.

ABSTRACT - Analysis of quantitative trait determinism by the detection of concerned loci (QTL) is one of the goals assigned to gene mapping with RFLP markers. Various biometrical approaches are presented, based upon a literature review, and the constraints and limits are drawn. Deviations from theoretical models linked to aspects of citrus biology are discussed. The latter include parental homozygosity, gametophytic incompatibility, chimeras and the building up of unfavourable genes in the heterozygous state.

CARTOGRAPHIE DU GENOME ET CARACTERES
QUANTITATIFS :
L'APPROCHE BIOMETRIQUE ET LES CONTRAINTES
BIOLOGIQUES DES AGRUMES.

X. PERRIER, P. OLLITRAULT et Cécile DUBOIS.

Fruits, Numéro spécial Agrumes 1992, p. 135-144.

RESUME - L'étude du déterminisme des caractères quantitatifs par la détection de leurs loci constructeurs (QTL) grâce aux marqueurs RFLP est l'un des objectifs assignés aux programmes de cartographie. Les différentes approches biométriques sont présentées sur la base d'une synthèse bibliographique afin d'en dégager les contraintes et les limites. Sont évoqués ensuite les écarts aux modèles théoriques liés à la biologie des agrumes (structure parentale hétérozygote, incompatibilité gamétophytique, chimérisme, accumulation de gènes défavorables à l'état hétérozygote ...).

INTRODUCTION

Beaucoup de caractères agronomiques soumis à sélection chez les plantes cultivées sont de nature quantitative, résultant de l'action de nombreux facteurs génétiques et environnementaux.

Depuis longtemps les sélectionneurs cherchent à comprendre le déterminisme génétique de ces caractères. L'idée d'utiliser des critères monogéniques et mendéliens liés à ces caractères comme marqueurs a été émise dès 1923 par SAX.

Elle prend un intérêt nouveau grâce aux méthodes de biotechnologie, en particulier l'étude du polymorphisme de la longueur des fragments de restriction (RFLP), qui permettent aujourd'hui d'accéder à un grand nombre de

marqueurs répartis sur le génome. Une méthodologie d'analyse biométrique de ces données est en cours d'élaboration, elle sera présentée à partir des données bibliographiques afin d'en souligner les exigences et les limites. Les conditions d'application aux agrumes seront alors discutées.

LES ELEMENTS DE L'ANALYSE

Les caractères quantitatifs.

Les paramètres d'intérêt agronomique à déterminisme monogénique peuvent être étudiés par simple observation d'une population recombinante (F2, backcross ...). Les génotypes possibles sont limités à quelques configurations souvent facilement identifiables et peu perturbées par les effets d'environnement. On parle de caractère qualitatif.

Si plusieurs gènes interviennent, le nombre de recombinaisons possibles augmente rapidement conduisant à un nombre élevé de génotypes. Les effets extérieurs s'ajoutant,

* - PERRIER et DUBOIS - CIRAD-IRFA, Service de Biométrie - B.P. 5035 - 34032 MONTPELLIER Cedex 01 (France).
OLLITRAULT - SRA San Giuliano - CIRAD-IRFA - 20230 SAN NICOLAO (Haute Corse).

les phénotypes prennent alors une distribution continue d'où le terme de caractère quantitatif (QTL : Quantitative Trait Locus) pour désigner les gènes déterminant l'expression d'un tel caractère. L'observation des individus d'une population recombinante ne permet pas d'identifier le génotype de chaque QTL. Le recours à des marqueurs est alors nécessaire.

Déterminisme génétique d'un caractère agronomique.

Soit Y le caractère agronomique étudié, déterminé par un seul locus Q. Soient deux parents homozygotes de génotype Q_1Q_1 et Q_2Q_2 , une population recombinante (une F2 par exemple) présentera, suivant le modèle mendélien, des individus de génotype Q_1Q_1 , Q_1Q_2 et Q_2Q_2 en proportions 1/4, 1/2 et 1/4.

Suivant les notations généralement adoptées (SOLLER *et al.*, 1976 ; BECKMANN et SOLLER, 1988) le caractère agronomique prendra les valeurs phénotypiques +d, -d ou h pour les génotypes Q_1Q_1 , Q_2Q_2 ou Q_1Q_2 .

d et h sont exprimés en unité d'écart type de la variable Y, Y étant centrée et réduite.

h exprime la nature de la dominance :

$h = 0$	codominance
$h = d$ ou $-d$	dominance complète de Q_1 ou Q_2
$h = x*d$	dominance partielle ($0 < x < 1$)

Si le caractère Y est déterminé par L loci, nous supposons que les effets sont additifs :

$$Y = \sum_{i=1}^L (x_i d_i + y_i h_i)$$

où $x_i = 1$ et $y_i = 0$ si le génotype de Q_i est Q_1Q_1
 $x_i = -1$ et $y_i = 0$ si le génotype de Q_i est Q_2Q_2
 $x_i = 0$ et $y_i = 1$ si le génotype de Q_i est Q_1Q_2

L'hypothèse d'additivité et donc d'absence d'épistasie est nécessaire pour les méthodes de résolution dont nous disposons aujourd'hui.

Cette hypothèse n'est peut-être pas trop lourde si l'on s'adresse à des caractères à déterminisme génétique direct (résistance à une maladie, composition chimique d'un fruit). Les phénomènes d'épistasie sont considérés comme rares ou d'effet suffisamment faible (BECKMANN et SOLLER, 1989). Ce ne serait pas le cas de caractères complexes, comme le rendement, qui doivent être décomposés en leurs différentes composantes (nombre d'épis par plante, nombre de grains par épi, poids du grain).

On fera également l'hypothèse, habituelle en génétique quantitative, d'additivité des effets génétiques et des effets d'environnement :

$$V_p = V_g + V_e$$

où V_p = variance phénotypique
 V_g = variance génétique
 V_e = variance d'environnement

Les variances phénotypiques de populations des parents homozygotes sont des estimateurs de la variance d'environnement, la variance génétique étant alors nulle.

L'efficacité de la méthode est directement liée à la part de l'effet génétique dans la variance phénotypique. Il est donc préférable d'utiliser des parents présentant des phénotypes très différents pour le caractère étudié. Il importe également de minimiser les effets d'environnement en éliminant tout facteur limitant même si cela doit conduire à des conditions de culture très différentes de la pratique habituelle. Pour les espèces où la multiplication végétative est possible, des dispositifs comportant des répétitions de chaque génotype pourront être mis en place pour évaluer avec précision le caractère quantitatif.

Le modèle de construction d'un caractère quantitatif peut varier d'un croisement à un autre. Ce, en fonction de la présence ou de l'absence de polymorphisme des différents loci constructeurs du caractère pour le croisement considéré.

Les marqueurs moléculaires.

Les premières études utilisaient des marqueurs morphologiques monogéniques (gènes de nanisme, de pigmentation) ; mises à part les études de laboratoire sur *Drosophile* ou chez les végétaux sur *Arabidopsis*, leur utilisation a en pratique été limitée par leur très faible nombre.

Récemment une nouvelle classe de marqueurs génétiques, dits marqueurs RFLP (Restriction Fragment Length Polymorphism), donne une dimension nouvelle à ce type d'approche en fournissant des cartes du génome extrêmement denses où la distance entre marqueurs est faible ; plus de 400 marqueurs RFLP sur les 12 chromosomes de la tomate (PATERSON *et al.*, 1988), plus de 900 sur maïs, 250 sur chou, 200 sur pomme de terre (BECKMANN et SOLLER, 1989).

Il est donc possible pour tous les individus d'une population recombinante de connaître la configuration allélique d'un grand nombre de points sur le génome.

Outre leur nombre élevé, ces marqueurs présentent des avantages génétiques certains (TANKSLEY *et al.*, 1989 ; STUBER, 1989) :

- appréhendés au niveau de l'ADN, ils ne présentent pas de rapport de dominance ou de récessivité et peuvent donc être identifiés quel que soit le schéma de croisement ;

- le niveau de polymorphisme dans une population est certainement beaucoup plus élevé que pour des marqueurs morphologiques ;

- ils sont en général neutres vis-à-vis du phénotype ;

- ils ne sont apparemment pas soumis à des effets d'épistasie ;

- il n'est pas nécessaire, comme pour les marqueurs morphologiques, de disposer de plantes entières et souvent adultes ;

- le prélèvement de matériel pour analyse ne nécessite pas la destruction de l'individu.

Les marqueurs enzymatiques présentent des caractéristiques comparables (TANKSLEY *et al.*, 1982). Ils ont été utilisés antérieurement mais restaient en nombre limité. Ils sont souvent associés aux marqueurs RFLP dans la constitution de cartes de marqueurs dits moléculaires.

Tous les marqueurs disponibles pour une espèce ne sont pas polymorphes pour un croisement donné. La précision des cartes que l'on obtient est propre au couple de parents considéré et conditionne la détection des QTL pour ce croisement. Ce n'est qu'en comparant un grand nombre de cartes que l'on pourra approcher l'organisation complète du génome d'une espèce.

De telles cartes ont été établies pour des espèces cultivées comme tomate, maïs, pomme de terre, laitue. De nombreuses équipes travaillent à la recherche de marqueurs sur d'autres espèces et dans un futur proche beaucoup d'autres cartes seront établies, en particulier sur agrumes (LURO *et al.*, 1990).

LES METHODES BIOMETRIQUES D'ANALYSE.

Cartographie des marqueurs.

La première phase de l'analyse est le positionnement sur le génome des marqueurs polymorphes. Leur observation sur tous les individus d'une population recombinante donne une estimation des fréquences de recombinaison entre marqueurs deux à deux. Cette fréquence est utilisée comme indice de distance exprimée en centiMorgans (1 p. 100 de recombinaison = 1 cM), la fréquence de recombinaison entre deux loci étant d'autant plus élevée qu'ils sont éloignés sur le génome. Une distance de 50 cM qui correspond à une recombinaison aléatoire signifie l'indépendance entre les loci. C'est en particulier le cas lorsqu'ils sont situés sur des chromosomes différents.

On identifie tout d'abord les groupes de liaison : tous les marqueurs d'un groupe sont à plus de 50 cM de tous les marqueurs d'un autre groupe. Le nombre de groupes est supérieur ou égal au nombre de chromosomes, un chromosome peut être séparé en plusieurs groupes de liaison s'il existe une portion du génome de plus de 50 cM sans marqueur. On cherche ensuite, par groupe de liaison, une ordonnance des marqueurs compatible avec les fréquences observées puis on précise les distances entre marqueurs consécutifs.

LANDER *et al.* (1987 a et b) proposent pour cette dernière étape une analyse multi-points en traitant simultanément un groupe de marqueurs consécutifs qui améliore la précision de l'ordonnance et des distances. Cette méthode est utilisée dans le logiciel MAPMAKER qui permet de traiter les populations recombinantes F2 et BC.

Les solutions obtenues sont statistiques et dépendent de l'échantillon analysé. Leur précision est fonction de la taille de la population analysée et du type de population recombinante utilisée. A effectif égal, une F2 est plus informative qu'un BC.

Relation entre QTL et marqueurs.

Le problème biométrique considéré peut être résumé ainsi : soit une population de n individus (F2, backcross...) sur lesquels on mesure un caractère quantitatif et dont on sait reconnaître la forme allélique d'une série de marqueurs de position connue sur le génome, quels sont les loci intervenant dans la formation du caractère quantitatif, quel est le modèle génétique sous-jacent, quelle est la position de ces loci et peut-on quantifier leur action ?

La réponse est de nature statistique puisque seul le phénotype est observé, un effet aléatoire d'environnement venant masquer en partie l'action du génotype. Par ailleurs, comme pour la cartographie, seul un échantillon d'une population théorique est analysé.

• Modèles AQ.

Soit une portion de chromosome comprenant un marqueur A, à deux allèles codominants A_1 et A_2 , et un QTL Q, à deux allèles Q_1 et Q_2 , espacés d'une distance de r centiMorgans (cM) qui est la fréquence de recombinaison entre ces deux loci.

Soit une F1 issue du croisement de deux parents homozygotes P1 et P2 de génotype $A_1A_1Q_1Q_1$ et $A_2A_2Q_2Q_2$.

Toute population recombinée à partir de cette F1 (F2, backcross simple ou multiple sur un parent ...) est un mélange de populations de génotypes différents. On sait reconnaître les populations correspondant aux différents génotypes du marqueur et donc calculer la moyenne du caractère quantitatif pour ces populations.

Par exemple, pour une population F2, on aura 25 p. 100 de A_1A_1 de génotype moyen observé M_{11} , 50 p. 100 de A_1A_2 de moyenne M_{12} et 25 p. 100 de A_2A_2 de moyenne M_{22} .

De la même façon, le QTL donne 3 génotypes différents mais que l'on ne sait pas reconnaître : Q_1Q_1 de moyenne inconnue Y_{11} (= +d), Q_1Q_2 de moyenne Y_{12} (= h) et Q_2Q_2 de moyenne Y_{22} (= -d).

Le principe général d'estimation des effets des QTL est de chercher à exprimer les valeurs M_{11} , M_{12} , M_{22} en fonction de Y_{11} , Y_{12} , Y_{22} .

Les populations A_iA_j comprennent une certaine proportion de génotypes Q_{11} , Q_{12} , Q_{22} . Ces proportions sont inconnues mais on peut en calculer les espérances en fonction de r (figure 1).

On peut donc écrire les M_{ij} en fonction de Y_{11} , Y_{12} , Y_{22} :

$$\begin{aligned} M_{11} &= (1-r)^2 Y_{11} + 2r(1-r) Y_{12} + r^2 Y_{22} \\ M_{22} &= r^2 Y_{11} + 2r(1-r) Y_{12} + (1-r)^2 Y_{22} \\ M_{12} &= r(1-r) Y_{11} + [(1-r)^2 + r^2] Y_{12} + r(1-r) Y_{22} \end{aligned}$$

Remarque : il n'a pas été pris en compte les possibilités de double recombinaison. Leur probabilité variant avec r^2 , on peut les négliger dès lors que r est suffisamment faible.

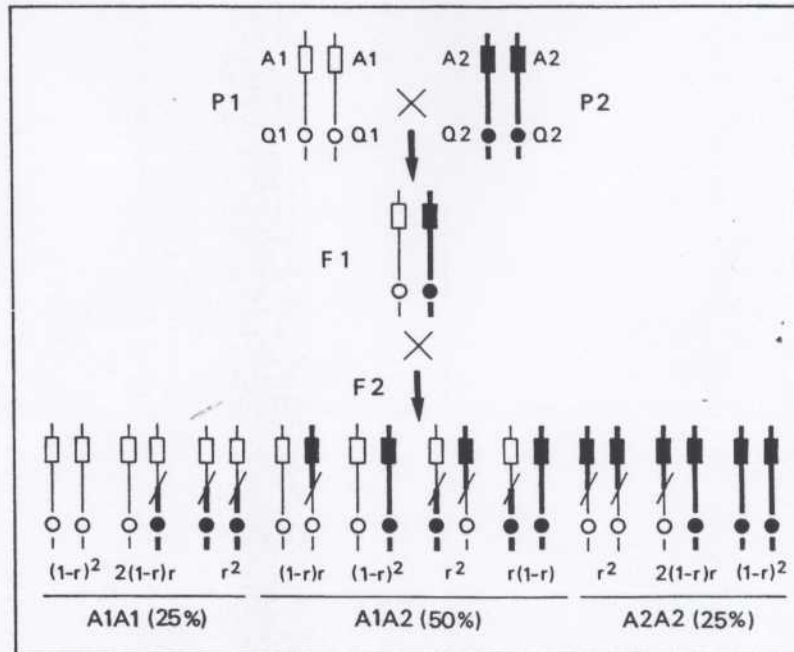


FIGURE 1 - Fréquence des différents génotypes possibles de la F2 issue de deux parents homozygotes, A₁A₁Q₁Q₁ et A₂A₂Q₂Q₂, pour le marqueur A et le QTL Q, distants de r cM.

● Modèle AQB.

Les modèles dits AQ (un marqueur, un QTL) ont été les plus étudiés. Ils correspondaient à des situations où les marqueurs, en général morphologiques, étaient rares.

Quand on dispose d'un nombre beaucoup plus élevé de marqueurs, les 2 marqueurs A et B encadrant un QTL (flanking markers) sont tous les deux suffisamment proches pour que l'on souhaite utiliser simultanément l'information apportée par chacun d'eux. Ce modèle est dit de type AQB.

En conservant l'exemple de la F2, si r_a et r_b sont les distances AQ et QB (r = r_a + r_b : distance AB), on peut estimer les fréquences des huit types de gamètes de la F1 pour ce triplet de loci. Ils génèrent 64 génotypes différents en F2 ; on sait reconnaître, par les marqueurs, 9 types différents dont les moyennes du caractère quantitatif seront notées :

	B ₁ B ₁	B ₁ B ₂	B ₂ B ₂
A ₁ A ₁	M ₁	M ₂	M ₃
A ₁ A ₂	M ₄	M ₅	M ₆
A ₂ A ₂	M ₇	M ₈	M ₉

Il est possible comme pour le modèle AQ d'écrire les M_i en fonction de Y₁₁, Y₁₂ et Y₂₂ :

$$\begin{aligned}
 M_1 &= Y_{11} \\
 M_2 &= (r_b Y_{11} + r_a Y_{12})/r \\
 M_3 &= (r_b^2 Y_{11} + 2 r_a r_b Y_{12} + r_a^2 Y_{22})/r^2 \\
 M_4 &= (r_a Y_{11} + r_b Y_{12})/r \\
 M_5 &= (r_a r_b Y_{11} + r_a r_b Y_{22} + [(1-r)^2 + r_a^2 + r_b^2] Y_{12}) / (1-2r + 2r^2)
 \end{aligned}$$

$$\begin{aligned}
 M_6 &= (r_b Y_{12} + r_a Y_{22})/r \\
 M_7 &= (r_a^2 Y_{11} + 2 r_a r_b Y_{12} + r_b^2 Y_{22})/r^2 \\
 M_8 &= (r_a Y_{12} + r_b Y_{22})/r \\
 M_9 &= Y_{22}
 \end{aligned}$$

(r étant connu, il suffira de connaître r_a pour déduire r_b puisque r = r_a + r_b).

Ce deuxième modèle est bien sûr beaucoup plus riche et sera à privilégier dans les études utilisant les marqueurs moléculaires.

Analyse des différents QTL déterminant le caractère quantitatif.

Les modèles AQ ou AQB décrivent les liens entre un seul QTL et les marqueurs voisins. En pratique le caractère quantitatif est déterminé par plusieurs QTL dont le nombre et la position ne sont pas connus.

La démarche généralement adoptée est de supposer la présence d'un QTL entre chaque marqueur et de traiter successivement chaque couple de marqueurs. Les effets de ces QTL sont comparés, à l'aide d'un test statistique approprié, aux effets d'un QTL nul (équivalent à l'absence de QTL). On ne retient que les loci pour lesquels le test dépasse un seuil prédéterminé.

Cette technique de balayage du génome demande une quantité de calcul importante, les marqueurs étant par principe nombreux. Elle nécessite des outils informatiques puissants.

Elle fait l'hypothèse implicite d'indépendance, au sens génétique, entre les QTL : ils sont sur des chromosomes

différents ou à plus de 50 cM sur un même chromosome. En effet, dans le modèle AQB par exemple, les fonctions donnant les M_i en fonction de Y_{11} , Y_{12} et Y_{22} ne sont vraies que si le caractère n'est déterminé que par ce QTL. Si d'autres gènes interviennent, ils agissent sur le phénotype M_i :

$$M_i = f(Y_{11}, Y_{12}, Y_{22}) + K_i$$

L'estimation n'a de sens que si les K_i sont tous égaux, ceci ne peut être vrai que si la configuration allélique des autres QTL est indépendante de celle du locus étudié.

Solutions mathématiques et exemple d'application.

Dans les modèles AQ comme AQB, l'objet est l'estimation des paramètres Y_{11} , Y_{22} et Y_{12} et r (ou r_a) et de leur intervalle de confiance.

Trois méthodes d'estimation des paramètres sont classiquement utilisées : la méthode des moments, la méthode des moindres carrés et celle du maximum de vraisemblance.

Quelques travaux font référence. On peut citer pour son antériorité WELLER (1986) bien qu'il ne travaille que sur des modèles AQ. Une judicieuse combinaison des méthodes des moments et du maximum de vraisemblance permet d'estimer les divers paramètres et leur variance.

Sur des modèles AQB deux approches légèrement différentes sont aujourd'hui retenues.

KNAPP (1989, 1990) utilise des méthodes de régression non linéaire à partir de la procédure NLIN du logiciel SAS. Le calcul est répété sur chaque QTL hypothétique placé entre chaque couple de marqueurs, un test statistique permet de retenir les QTL ayant une signification.

LANDER et BOTSEIN (1989) utilisent des méthodes de maximum de vraisemblance développées pour la cartographie du génome humain. L'originalité de leur démarche réside dans l'approche d'«interval mapping» : l'intervalle entre deux marqueurs est balayé et le test statistique réalisé à chaque cM entre les marqueurs.

PATERSON *et al.* (1988) appliquent cette technique à un croisement entre la tomate domestique (*Lycopersicon esculentum*) à gros fruits et faible teneur en sucres et une tomate sauvage à très petits fruits verts et très sucrés (*L. chmielecvskaïi*). 71 marqueurs polymorphes sont disponibles, espacés en moyenne de 14 cM. Une population backcross de 237 individus est observée : poids, teneur en sucre, pH. La figure 2 montre les résultats obtenus pour 4 chromosomes. On note sur le chromosome 6 un locus contrôlant fortement les 3 caractères et qui explique les corrélations que l'on peut observer sur ces paramètres. L'effet sur le poids et la teneur en sucre sont opposés, la sélection progressive de tomate à gros fruits a donc progressivement diminué la teneur en sucre. Il apparaît plus intéressant pour augmenter la taille des fruits de faire porter la sélection sur le QTL du chromosome 11 qui est neutre vis-à-vis du pH ou de la teneur en sucre.

L'efficacité des méthodes proposées dépend de la taille de la population étudiée, qui dans les exemples rencontrés

dans la littérature, est de plusieurs centaines. Elles permettent en général de détecter les QTL à effet fort (gènes majeurs) qui, pour les caractères agronomiques étudiés, sont rarement supérieurs à 3 ou 4 (BECKMANN et SOLLER, 1983). Les gènes à effet plus faible sont plus difficiles à reconnaître et nécessitent des dispositifs puissants.

La précision des estimations est meilleure sur les effets des QTL (Y), que sur leur position (r). Ceci tient en partie aux effectifs servant à les établir. Seuls les génotypes recombinés, beaucoup moins nombreux, servent à estimer r . Il est donc préférable de disposer de cartes très denses où la distance, entre deux marqueurs, est faible.

LES POPULATIONS ANALYSABLES

Toute population recombinante, issue de parents homozygotes P_1 et P_2 , peut être utilisée pour rechercher les liens entre QTL et marqueurs (figure 3).

Les **backcross** (BC) sur un des parents ont souvent servi de modèle aux biométriciens car les différents génotypes possibles sont en plus faible nombre. Ils ont cependant l'inconvénient majeur de ne pas permettre la mise en évidence d'effets de dominance.

Les populations **F2** nécessitent un modèle d'analyse plus complexe puisque les génotypes possibles sont plus nombreux mais n'exigent pas l'hypothèse de codominance.

Les **haploïdes doublés** (DH) issus des gamètes de la F1 simplifient fortement la cartographie des marqueurs : ils ne sont pas toujours faciles à obtenir et l'on soupçonne des perturbations importantes de la régulation du génome, pour l'expression des caractères quantitatifs, lors de l'haploïdisation.

Les populations **S1** sont des F3 issues de l'autofécondation des individus de la F2. L'information est comparable à celle de la F2. L'intérêt est de disposer, à partir d'un individu de la F2 dont le génotype peut être analysé pour les marqueurs, d'une population proche sur laquelle le caractère quantitatif peut être observé de façon plus fiable : génétiquement la dérive est faible, par contre le fait d'observer un ensemble d'individus peut permettre pour certains caractères de réduire très sensiblement les effets d'environnement.

La même démarche peut être appliquée, quelle que soit la population recombinante, aux espèces à multiplication végétative, la population observée étant alors génétiquement identique à l'individu initial.

Les **lignées améliorées** (RI : recombinant inbred lines) sont obtenues par un certain nombre d'autofécondations à partir des individus de la F2 -au minimum six- qui assurent la fixation des caractères. Elles permettent la détection des effets principaux mais pas des effets de dominance.

Sur un plan pratique, les lignées améliorées présentent l'avantage considérable d'être génétiquement stables et de pouvoir être propagées par graines alors qu'une population recombinante comme une F2 est par essence éphémère (TANKSLEY, 1989). A partir d'une unique détermination du génotype des marqueurs, il est possible de répéter l'ob-

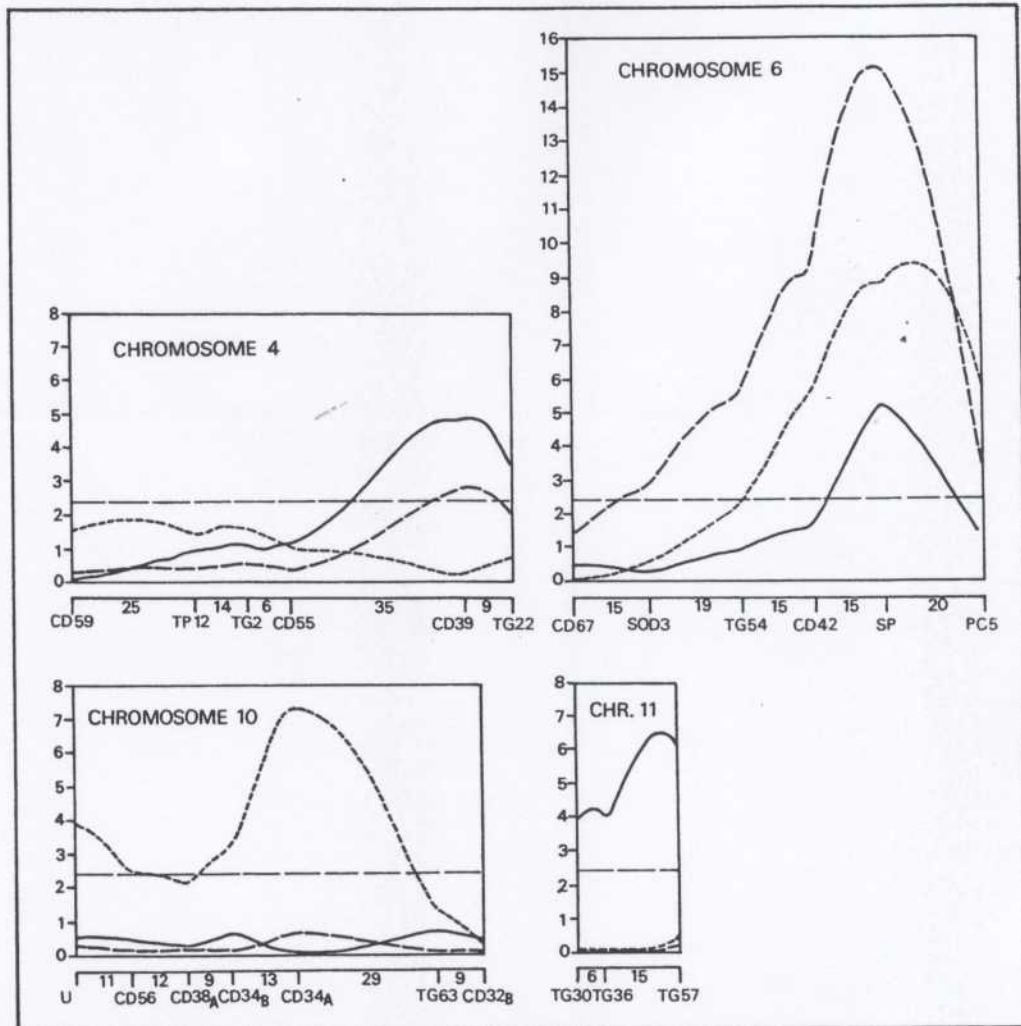
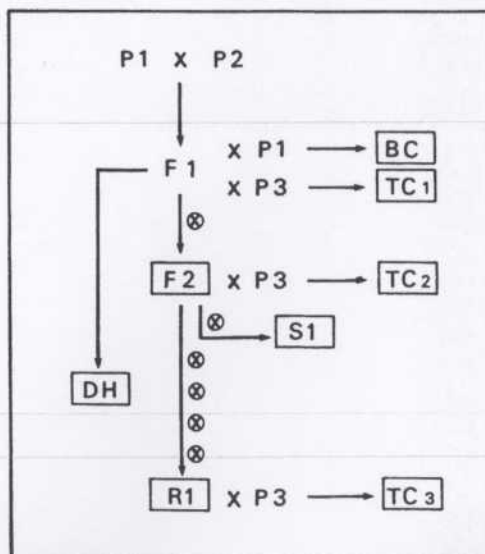


FIGURE 2 - Carte de 4 chromosomes de la tomate (voir texte)(d'après PATERSON, 1989).
 En abscisse, position sur le chromosome.
 En ordonnée, indice de vraisemblance de la présence d'un QTL codant pour le poids (ligne continue),
 le pH (tirets) et la teneur en sucres (points).



servation d'un caractère quantitatif soit dans l'espace pour prendre en compte la variabilité d'environnement, soit dans le temps pour étudier ultérieurement d'autres caractères quantitatifs. L'inconvénient le plus sérieux est le temps nécessaire à l'obtention de ces lignées qui, même pour des espèces à cycle court, peut durer plusieurs années.

Les différents testcross (TC) issus du croisement d'un parent homozygote P₃ avec une F₁, une F₂ ou même une lignée améliorée peuvent également être utilisés. Là encore, seuls les effets principaux sont détectables.

Sur le plan de l'analyse, ces différentes populations peuvent être regroupées en deux types (KNAPP, 1989) :

- les BC, DH, TC et RI qui présentent 4 génotypes possibles du couple de marqueurs A et B, ils n'autorisent la mise en évidence que des effets principaux ;

FIGURE 3 - Les différentes populations recombinantes utilisables (voir texte) (d'après COWEN, 1988).

- les F2, F3 ou S1 pour qui 9 génotypes différents sont observés et permettent d'appréhender les effets de dominance.

Il apparaît donc que le choix de la population à analyser dépend de considérations diverses et est principalement dicté par les caractéristiques de l'espèce étudiée :

- longueur du cycle.
- possibilité d'autofécondation,
- possibilité de multiplication végétative,
- disponibilité de techniques d'obtention d'haploïdes doublés.

APPLICATION A LA CONNAISSANCE DU GENOME DES CITRUS

Il a été souligné les différentes hypothèses nécessaires aux méthodes proposées :

- additivité des effets des QTL, sans phénomène d'épistasie.
- respect des règles du mendélisme.
- indépendance des QTL constructeurs.
- additivité des effets génétiques et d'environnement.
- homozygotie des parents.

Ces méthodes ont donné des résultats intéressants sur certaines espèces comme la tomate (PATERSON *et al.*, 1988 ; WELLER *et al.*, 1988) ou le maïs (FREI *et al.*, 1986). On manque cependant de recul pour juger de la pertinence des résultats. Il faut s'inquiéter en particulier du degré de généralité des résultats obtenus qui pourraient être fortement perturbés dans un contexte génétique différent (autres parents) ou dans un environnement différent (interaction génotype-milieu) (STUBER, 1989).

Par ailleurs les particularités des espèces étudiées génèrent des contraintes qui doivent être intégrées dans les modèles d'analyse. Elles sont de deux ordres chez les Citrus : il est en effet possible de distinguer les contraintes pratiques qui vont rendre difficile l'obtention de certains types de population et l'analyse d'effectifs importants, des contraintes qui entraînent des écarts aux modèles théoriques dans les populations analysées.

Les contraintes pratiques.

Les premières contraintes communes à tous les arbres sont :

- **l'encombrement des descendance**s qui restreint considérablement les effectifs analysables pour les caractères devant être évalués au stade adulte (en particulier les caractères pomologiques ou les caractères de production) : la précision des estimations, aussi bien pour l'établissement de la carte que pour la détection des QTL, sera faible. On peut cependant, pour la cartographie des marqueurs, étudier dès la phase juvénile une population beaucoup plus abondante dont une partie seulement sera conduite au stade adulte.

- **la longueur de la phase juvénile** qui retarde de manière importante l'établissement de populations de seconde gé-

nération d'hybridation. Il sera judicieux d'exploiter au mieux ces populations en observant le maximum de caractères quantitatifs.

Les autres contraintes pratiques sont celles associées au régime de reproduction des agrumes :

● La polyembryonie.

L'apomixie partielle de nombreux agrumes limite leur utilisation comme géniteur femelle ou comme parent d'autofécondation. Cette contrainte n'est toutefois pas rédhibitoire puisqu'il est possible de sauver les embryons zygotiques grâce à la culture *in vitro* et de distinguer rapidement les jeunes plants d'origine nucléaire et zygotique grâce à l'électrophorèse d'isozymes.

● le système d'autoincompatibilité gamétophytique.

L'autoincompatibilité gamétophytique a été mise en évidence chez de nombreuses variétés monoembryonnées d'agrumes (SOOST, 1969). Elle est contrôlée par un gène possédant une série d'allèles S : un grain de pollen possédant un allèle Si ne peut pas féconder une plante possédant ce même allèle.

Elle est donc un obstacle à l'obtention de population issue d'autofécondation et en particulier à l'établissement de populations de type F2 ou lignées améliorées. Par ailleurs, les BC ne peuvent être réalisés qu'en utilisant l'hybride comme pollinisateur, la moitié des grains de pollen (qui portent l'allèle Si du parent femelle) sont exclus de la fécondation.

Les agrumes présentent toutefois, comme toutes les espèces à multiplication végétative, l'avantage déterminant de pouvoir multiplier, à volonté, chaque génotype d'une descendance. Il est ainsi possible d'obtenir une estimation beaucoup plus fine de la variance génotypique des caractères quantitatifs, et également de conserver chaque génotype pour des analyses complémentaires.

Les écarts aux modèles théoriques.

● L'hétérozygotie résiduelle des génotypes parentaux.

La plupart des modèles théoriques partent de structures initiales homozygotes pour établir des familles F2, des backcross, etc. Chez les agrumes, l'hétérozygotie est modérée à élevée suivant les variétés (OLLITRAULT et FAURE, 1992). En tout état de cause, l'hétérozygotie résiduelle n'est jamais nulle. Les problèmes de cartographie à partir de parents hétérozygotes ont été étudiés par RITTER *et al.* (1990) mais ne sont pas traités par les logiciels disponibles. De nouveaux modèles sont à développer pour l'analyse des QTL.

La création de dihaploïdes pourrait constituer une solution à cette contrainte. Toutefois les données bibliographiques concernant l'haploïdisation des agrumes sont peu encourageantes en raison de la très faible vigueur des plantules obtenues (HIDAKA *et al.*, 1979 ; HIDAKA et KAJIURA, 1989).

- Les sélections gamétiques et zygotiques et les écarts au mendélisme.

La sélection gamétique traduit une différence d'efficacité (ex. : compétition pollinique) ou de viabilité de certains types de gamète tandis que la sélection zygotique est due à la non germination de certains hybrides. Si ces différences de viabilité ou de vigueur sont associées systématiquement à l'état allélique d'un locus donné, elles entraînent des ségrégations non mendéliennes pour tous les gènes liés à ce locus (à moins de 50 cM). L'écart au mendélisme est fonction du taux de recombinaison entre le locus soumis à la sélection et le gène étudié. Ceci implique l'estimation d'un paramètre supplémentaire traduisant l'effet de sélection, avec des problèmes particuliers pour l'estimation des effets et de la position des QTL. De plus la disparition partielle, ou totale, de certains génotypes diminue fortement les effectifs de certaines combinaisons et donc la précision des estimations.

Deux facteurs majeurs induisent, chez les Citrus, des sélections zygotiques ou gamétiques :

- Le système d'incompatibilité gamétophytique.

Le système d'incompatibilité gamétophytique peut conduire dans certaines hybridations, à l'exclusion de la moitié des grains de pollen du processus de reproduction. Cette situation sera plus fréquente lors des hybridations entre variétés apparentées qui ont davantage de chance de posséder un allèle S en commun. A contrario, le recours à des hybridations entre géniteurs éloignés doit permettre d'éviter ce facteur d'écart au mendélisme.

- Les mutations délétères.

Du fait de la multiplication végétative, de nombreuses mutations délétères ou défavorables se sont accumulées à l'état hétérozygote. Elles peuvent entraîner des ségrégations anormales, lors des fécondations croisées (sélection gamétique) ou des autofécondations (sélection gamétique et zygotique). L'une des solutions biologiques pour diminuer les distorsions liées à ces sélections pourrait être le recours à des hybrides récents comme géniteurs. Le premier niveau d'hybridation doit en effet constituer un filtre pour éliminer un certain nombre de mutations délétères. Il est clair toutefois, que la probabilité est très faible de pouvoir transférer, par hybridation, un gène intéressant, fortement lié à une mutation délétère. Ces mutations peuvent donc constituer un obstacle important à la recombinaison sexuée.

Chez les agrumes, des distorsions de ségrégations ont été obtenues lors de l'étude du déterminisme génétique de certains marqueurs enzymatiques pour des descendances d'hybrides pamplemousse x rough lemon (TORRES *et al.*, 1985) ainsi que des hybrides pamplemousse x lime Brasil sweet, pamplemousse x citron Meyer et pamplemousse x rough lemon (OLLITRAULT et FAURE, 1992). En revanche les ségrégations observées pour l'hybridation pamplemousse x mandarine Cléopâtre x citrange Carrizo) étaient très proches des règles du mendélisme.

- La dépression de consanguinité.

L'accumulation de mutations défavorables chez les variétés se traduit également par une forte dépression de consanguinité dans les descendances d'autofécondation (SOOST et CAMERON, 1975). Les résultats obtenus sur des semis de *C. volkameriana*, laissent supposer l'existence de nombreux gènes défavorables perturbant le métabolisme de base. En effet, la grande majorité des plants zygotiques présentent un défaut de vigueur très marqué (figure 4). Il est à craindre que ces mutations affectant le métabolisme de base, ne masquent les effets des gènes directement impliqués dans l'expression des caractères étudiés.

Comme précédemment, on peut penser que l'utilisation d'hybrides récents, comme géniteurs, permettra de limiter l'importance de ce facteur.

- Les variations de structure et de taille du génome.

Les études de cytogénétique ont mis en évidence des translocations et des inversions chez certaines variétés (IWAMASA, 1966). De plus, il semble exister des variations de structure (NAIR et RANDHAWA, 1969) et de taille (OLLITRAULT et MICHAUX-FERRIERE, 1991) du génome entre différentes espèces diploïdes du genre Citrus. La cartographie du génome établie à partir d'un génotype donné n'est donc pas transposable in extenso à l'ensemble des agrumes. On peut cependant penser que la majorité des groupes de liaison sont conservés.

- Les chimères.

Les mutations de bourgeon sont fréquentes chez les agrumes et conduisent à la formation de chimères. Certains cultivars présentant des caractères intéressants se sont ainsi avérés être des chimères périclines (ex. : certains pomelos rouges, OLSON *et al.*, 1976). Si la couche cellulaire portant le caractère muté ne participe pas à la gamétogenèse il n'y a pas transmission du caractère à la descendance. La localisation du gène portant la mutation favorable n'est donc pas réalisable. Ce facteur limitant est, toutefois, certainement marginal pour l'étude des caractères de résistance nous intéressant en priorité.

CONCLUSION

L'utilisation des marqueurs moléculaires de type RFLP pour l'étude du déterminisme génétique des caractères quantitatifs est conceptuellement très élégante. Elle ouvrira la porte dans un avenir proche aux méthodes de génie génétique d'insertion de gènes intéressants.

Dès à présent l'établissement de cartes saturées permet la description et l'étude de la diversité génétique. Elles autorisent l'étude de la recombinaison et de la transmission des caractères monogéniques, ce qui est un apport essentiel, en particulier dans le cadre d'hybridations interspécifiques où les éliminations chromosomiques et les retours vers les formes parentales sont fréquents. La localisation des loci et l'estimation de leur effet fournissent une aide fondamentale pour la compréhension des mécanismes (poids et teneur

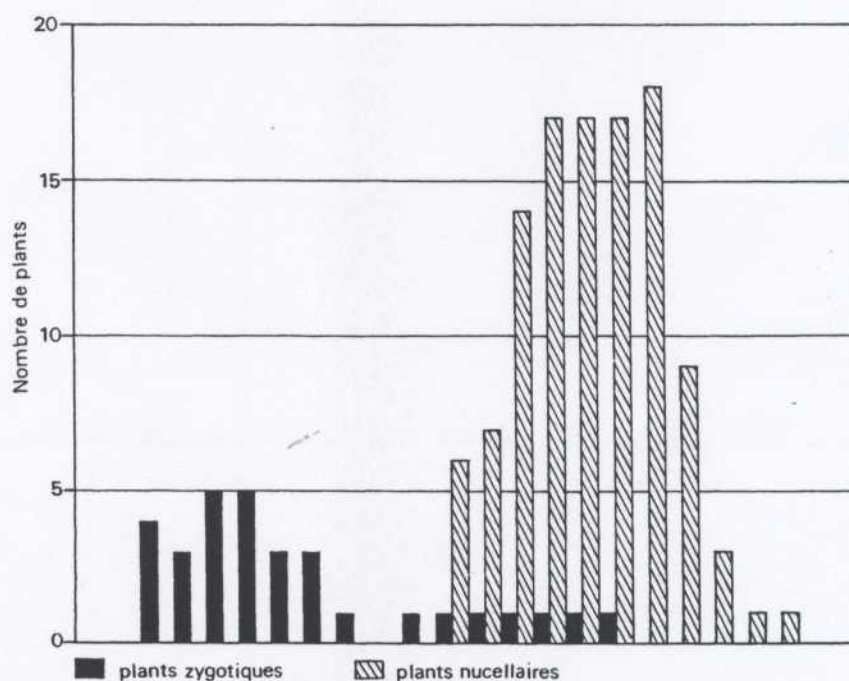


FIGURE 4 - Dépression de consanguinité dans un semis de *Volkameriana*.
 Histogramme des coordonnées sur le premier axe factoriel d'une analyse en composantes principales sur des caractères quantitatifs de vigueur.

en sucre chez la tomate) et pour l'élaboration de programmes d'amélioration pertinents.

Un intérêt annexe, mais essentiel pour les espèces à cycle très long, est l'utilisation des marqueurs comme critères précoces de sélection.

Cependant les contraintes biologiques liées au régime de reproduction et aux structures du génome, en particulier sur agrumes, sont importantes, certaines peuvent être contournées par des solutions biologiques (sauvetage d'embryons, hybridations préliminaires, haplodiploïdisation...). Toutefois des solutions biométriques spécifiques devront également être apportées :

- pour l'analyse de populations de type hybride trois voies (un génotype fortement homozygote croisé par un hybride non apparenté) qui présentent de nombreux avantages (limi-

tation des sélections gamétiques et zygotiques, peu de risque d'incompatibilité gamétophytique, limitation de la dépression de consanguinité). L'analyse de ces populations s'apparente à celle d'un backcross pour les caractères monogéniques et codominants (ex. : les marqueurs). Elle réclame en revanche un modèle d'analyse pluri-allélique pour les autres caractères, particulièrement si le génotype homozygote n'est pas totalement fixé.

- pour l'analyse de populations présentant des écarts au mendélisme sur certaines portions du génome.

Les agrumes peuvent ainsi constituer un modèle méthodologique de choix pour la cartographie des espèces ligneuses et des espèces à multiplication végétative, pour lesquelles des contraintes biologiques similaires pourraient se poser. Ils présentent, en effet, l'avantage de la diploïdie et d'un génome de petite taille.

BIBLIOGRAPHIE

BECKMANN (J.S.) and SOLLER (M.). 1983.
 Restriction fragment length polymorphisms in genetic improvement : methodologies, mapping and costs.
Theoretical and Applied Genetics, 67 : 35-43.

BECKMANN (J.S.) and SOLLER (M.). 1988.
 Detection of linkage between marker loci and loci affecting quantitative traits in crosses between segregating populations.
Theoretical and Applied Genetics, 76, 228-236.

BECKMANN (J.S.) and SOLLER (M.). 1989.
 Genomic genetics in plant breeding.
Vorträge für Pflanzenzüchtung, 16, 228-236.

COWEN (N.M.). 1988.
 The use of replicated progenies in marker-based mapping of QTL's.
Theoretical and applied genetics, 75 : 857-862.

FREI (O.M.), STUBER (C.W.) and GOODMAN (M.M.). 1986.
 Use of allozymes as genetic markers for predicting performance in maize single cross hybrids.
Crop Science, 26 : 37-42.

HIDAKA (T.), YAMADA (Y.) and SHICHIJO (T.). 1979.
 In vitro differentiation of haploid plants by anther culture in *Poncirus trifoliata* (L.) Raf.
Japan J. Breed., 29 (3), 248-254.

HIDAKA (T.) and KAJIURA (I.). 1989.
 A simple method for acclimatation of *in vitro* plantlets of *Citrus*.
Bull. Fruit Tree Res. Stn. B., 16, 19-28.

IWAMASA (M.). 1966.
 Studies on the sterilities in the genus *Citrus* with special reference to the seedlessness.
Bul. Hort. Res. Sta. Japan, Ser. B 6 : 1-77.

- KNAPP (S.J.). 1989.
Quasi-Mendelian analysis of traits using molecular marker linkage maps.
Vorträge für Pflanzenzüchtung, 6, 51-67.
- KNAPP (S.J.), BRIDGES (W.C.) and BIRKES (D.). 1990.
Quasi-mendelian analyses of quantitative trait loci using molecular marker linkage map. *Genetics*.
Theoretical and Applied Genetics, 79, 583-592.
- LANDER (E.S.) and GREEN (P.). 1987 a.
Construction of multilocus genetic linkage maps in human.
Proc. Natl. Acad. Sci. USA, 84, 2363-2367.
- LANDER (E.S.), GREEN (P.), ABRAHAMSON (J.), BARLOW (A.), DALY (M.J.), LINCOLN (S.E.) and NEWBURG (L.). 1987 b.
MAPMAKER : An interactive computer package for constructing primary genetic linkage maps of experimental and natural populations.
Genomics, 1, 174-181.
- LANDER (E.S.) and BOTSTEIN (D.). 1989.
Mapping mendelian factors underlying quantitative traits using RFLP linkage maps.
Genetics, 121, 185-199.
- LURO (S.), LAIGRET (S.) et OLLITRAULT (P.). 1991.
Cartographie du génome des agrumes.
Journées Agrumes IRFA, Montpellier, 4-10 septembre 1991.
- NAIR (P.K.F.) and RANDHAWA (G.S.). 1969.
Chromosome morphology of the Pachyten stage with respect to different Citrus types.
Proc. 1st. Int. Citrus Symp., vol. 1, 215-223.
- OLLITRAULT (P.) and MICHAUX-FERRIERE (Nicole). 1992.
Application of flow cytometry for citrus genetic and breeding.
VII Citrus Cong. Acireale (Italy), 8-13 March 1992.
- OLLITRAULT (P.) et FAURE (X.). 1992.
Système de reproduction et organisation de la diversité génétique dans le genre Citrus.
Coll. int. «Complexes d'espèces, flux de gènes et ressources génétiques des plantes», Paris, 8-10 janvier 1992, 12 p. (à paraître).
- OLSON (E.O.), CAMERON (J.W.) and SOOST (R.K.). 1966.
The Burgundy sport : further evidence of the chimeral nature of pigmented grapefruits.
Hortsciences, 1, 57-59.
- PATERSON (A.H.), LANDER (E.S.), HEWITT (J.D.), PETERSON (S.) and LINCOLN (S.E.). 1988.
Resolution of quantitative traits into Mendelian factors by using a complete linkage map of restriction fragment length polymorphisms.
Nature, 335, 721-726.
- RITTER (E.), GEBHARDT (C.) and SALAMINI (F.). 1990.
Estimation of recombination frequencies and construction of RFLP linkage maps in plants from crosses between heterozygous parents.
Genetics, 125, 645-654.
- SOLLER (M.), BRODY (T.) and GENIZI (A.). 1976.
On the power of experimental designs for the detection of linkage between marker loci and quantitative loci in crosses between inbred lines.
Theoretical and Applied Genetics, 47 (1), 35-39.
- SOOST (R.K.). 1969.
The incompatibility gene system in Citrus.
Proc. 1st Int. Citrus Symposium, H. Chapman ed. University of California, Riverside, 1, 189-190.
- SOOST (R.K.) and CAMERON (J.W.). 1975.
Citrus. In advances in fruit breeding.
J. Janick and J.N. Moore eds. Purdue University Press, West Lafayette, India, 507-540.
- STUBER (C.W.). 1989.
Marker-based selection for quantitative traits.
Vorträge für Pflanzenzüchtung, 16, 31-49.
- TANKSLEY (S.D.), MEDINA-FILHO (H.) and RICK (C.M.). 1982.
Use of naturally-occurring enzyme variation to detect and map genes controlling quantitative traits in an interspecific backcross of tomato.
Heredity, 49 (1), 11-25.
- TANKSLEY (S.D.), YOUNG (N.D.), PATERSON (A.H.) and BONIERBALE (M.W.). 1989.
RFLP mapping in plant breeding : new tools for an old science.
Bio/Technology, 7, 257-264.
- TORRES (A.M.), MAU-LASTOVICKA (T.), WILLIAMS (T.E.) and SOOST (R.K.). 1985.
Segregation distortions and linkages of Citrus and Poncirus isozymes genes.
J. Hered., 76, 289-294.
- WELLER (J.I.). 1986.
Maximum likelihood techniques for the mapping and analysis of quantitative trait loci with the aid of genetic markers.
Biometrics, 42 (3), 627-640.
- WELLER (J.I.), SOLLER (M.) and BRODY (T.). 1988.
Linkage analysis of quantitative traits in a interspecific cross of tomato (*Lycopersicon esculentum* x *Lycopersicon pimpinellifolium*) by means of genetic markers.
Genetics, 118, 329-339.

CARTOGRAFIA DEL GENOMIO Y CARACTERES CUANTATIVOS : EL ENFOQUE BIOMETRICO Y LAS LIMITANTES BIOLÓGICAS DE LOS CITRICOS.

X. PERRIER, P. OLLITRAULT y Cécile DUBOIS.

Fruits, Numéro spécial Agrumes 1992, p. 135-144.

RESUMEN - El estudio del determinismo de los caracteres cuantitativos mediante la detección de los loci constructores (QTL) con los marcadores RFLP es uno de los objetivos asignado a los programas de cartografía.

Los diferentes enfoques biométricos son presentados sobre la base de una síntesis bibliográfica para destacar los apremios y las limitantes. Después se señalan las digresiones con los modelos teóricos ligados a la biología de los cítricos (estructura parental heterocigótica, incompatibilidad gametofítica, quimerismo, acumulación de genes desfavorables al estado heterocigótico ...).