

Le déséquilibre de liaison

On dit qu'il y a déséquilibre de liaison si la fréquence des gamètes porteurs de deux gènes A et B non allèles est différente du produit des fréquences de ces deux derniers ; ce produit définit l'équilibre. De nombreux déséquilibres de liaison ont été décrits, notamment dans les régions HLA de la β -globine, des gènes des apolipoprotéines A1-C3-A4 et dans les régions proches de gènes à l'origine des maladies récessives autosomiques. Les recombinaisons génétiques réduisent à chaque génération la valeur du déséquilibre de liaison, mais si les gènes sont étroitement liés, l'évolution vers l'équilibre est très lente. De nombreux mécanismes peuvent être à l'origine du déséquilibre de liaison, les plus importants étant les mélanges de populations, la dérive génétique et l'effet fondateur, la sélection naturelle.

Josué Feingold

ADRESSE

J. Feingold : directeur de recherche à l'Inserm. Unité de recherches épidémiologiques, Inserm U. 155, château de Longchamp, Bois de Boulogne, 75016 Paris, France.

m/s n° 2, vol. 7, février 91

En 1908, indépendamment, Hardy — un mathématicien anglais — et Weinberg — un médecin allemand — démontraient la loi qui porte leur nom, loi qui constitue la base de toute la génétique des populations. Cette loi dans sa forme la plus simple concerne un *locus* autosomique diallélique : p_A étant la fréquence du gène A, q_a ($q_a = 1 - p_A$) celle du gène allèle a, on montre que sous certaines conditions les fréquences des trois génotypes AA, Aa, aa sont respectivement p_A^2 , $2p_Aq_a$, et q_a^2 . Celles-ci sont atteintes en une génération quelles que soient les fréquences initiales et demeurent constantes d'une génération à l'autre. Ces fréquences d'équilibre ne sont observées que si la population est panmictique (mariage au hasard, du

moins pour le caractère considéré), de grande taille, sans migration et qu'au *locus* considéré, il n'y a ni sélection naturelle ni mutation.

Une dizaine d'années après la démonstration de la loi de Hardy et Weinberg, Jennings en 1917 [1] d'une part, Robbins en 1918 [2] d'autre part, étudièrent l'évolution de deux *loci* simultanément à l'échelle de la population. Ces auteurs ont montré que si pour chacun des *loci*, l'équilibre défini par la loi de Hardy et Weinberg est atteint en une génération, il n'en est pas de même si l'on considère les deux *loci*. Il faut, pour étudier l'évolution simultanée de deux *loci*, introduire un nouveau concept, celui de déséquilibre de liaison (DL), appelé parfois déséquilibre de *linkage*, voire association gamétique ou association allélique. Ces

RÉFÉRENCES

- Jennings HS. The numerical result of diverse systems of breedings with respect to two pairs of characters. *Genetics* 1917 ; 2 : 97-106.
- Robbins RB. Some applications of mathematics to breeding problems III. *Genetics* 1918 ; 3 : 375-89.
- Feingold N. Phénomènes de liaisons et multilocus. In : *Génétique médicale, acquisitions et perspectives*. Paris : Flammarion, 1981 : 219-27.
- Lewontin RC. The interaction of selection and linkage. I general considerations : heterotic models. *Genetics* 1964 ; 49 : 49-67.
- Hill WG, Robertson A. Linkage disequilibrium in finite populations. *Theor Appl Genet* 1968 ; 31 : 226-31.
- Feingold N. Phénomènes d'association dans une population. *Nouv Rev Franc Hemat* 1972 ; 12 : 471-5.
- Stern C. *Principles of Human Genetics* (3rd ed). San Francisco : Freeman, 1973.
- Chakraborty R, Lidsky AS, Daiger SP, et al. Polymorphic DNA haplotypes at the human phenylalanine hydroxylase locus and their relationship with phenylketonuria. *Hum Genet* 1987 ; 76 : 40-6.
- Antonarakis SE, Oetgen P, Chakravarti A, et al. DNA polymorphism haplotypes of the human apolipoprotein APOA1-APOC3-APOA4 gene cluster. *Hum Genet* 1988 ; 80 : 265-73.
- Chakravarti A, Buetow AH, Antonarakis SE, Waber PG, Boehm D, Kazazian HH. Nonuniform recombination within the human betaglobin gene cluster. *Am J Hum Genet* 1984 ; 36 : 1239-58.

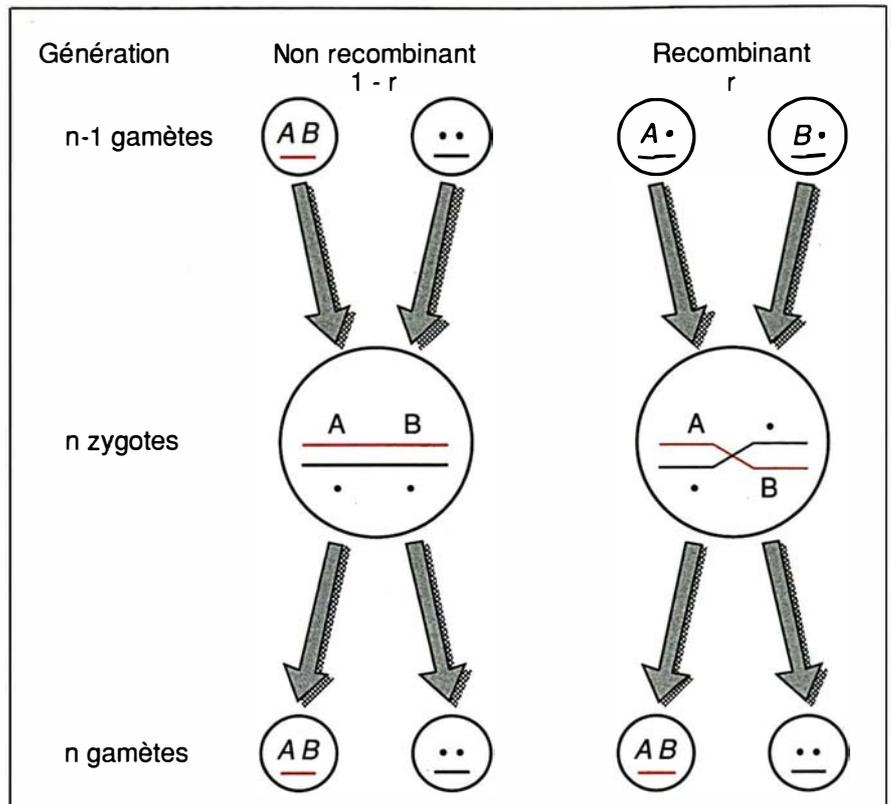


Figure 1. **Évolution du déséquilibre de liaison.** La figure indique les deux voies pour obtenir un gamète AB à la génération n : soit un des gamètes parentaux est AB et il ne subit aucune recombinaison (fréquence : $1 - r$) soit un gamète parental porte A et l'autre B et une recombinaison survient (fréquence : r). On peut déduire de ce schéma que : $P_{ABn} = (1 - r) P_{ABn-1} + r P_A P_B$. En soustrayant $p_A p_B$ de chaque membre de l'égalité on obtient $D_n = (1 - r) D_{n-1}$.

deux derniers termes sont plus corrects, mais n'ont pas été consacrés par l'usage.

Définition du déséquilibre de liaison

Nous considérerons dans un premier temps que les deux *loci* sont dialléliques : A et a sont les gènes allèles du premier locus, p_A et q_a étant leurs fréquences géniques ; B et b les gènes du deuxième locus, p_B et q_b leur fréquences géniques. Si P_{AB} est la fréquence des gamètes porteurs des gènes A et B, on définit le déséquilibre de liaison D par la relation.

$$D = P_{AB} - p_A \times p_B \quad (1)$$

Si $D = 0$, $P_{AB} = p_A \times p_B$, il y a absence de déséquilibre de liaison entre les gènes A et B. Cette dernière relation définit l'indépendance des

événements « présence de A et B dans le même gamète ». Il y a, dans ce cas, équilibre entre les deux *loci*. Si $D \neq 0$, on dit qu'il y a déséquilibre de liaison. Considérons à titre d'exemple les gènes A_1 et B_8 du système HLA, leur fréquence est d'environ 0,142 et 0,096 dans la population française, celle de leur combinaison est d'environ 0,0672 d'où : $D = 0,0672 - 0,142 \times 0,096 = 0,054$. Il y a donc déséquilibre de liaison, car 0,054 diffère statistiquement de zéro.

La relation (1) peut également prendre la forme

$D = P_{AB} \times P_{ab} - P_{Ab} \times P_{aB}$ (2) qui est historiquement la première formule décrite. P_{AB} , P_{ab} , P_{Ab} et P_{aB} sont les fréquences des 4 combinaisons gamétiques.

D peut être positif ou négatif, son intervalle de variation est :

$$-0,25 \leq D \leq 0,25.$$

La valeur minimale ou maximale de D dépend de la fréquence des gènes ; si $D > 0$, la valeur maximale de D est : $p_B (1 - p_A)$ si $p_A > p_B$, ou : $p_A (1 - p_B)$ si $p_A < p_B$; si $D < 0$, la valeur maximale en valeur absolue est la plus petite des deux valeurs $p_A p_B$ et $q_A q_B$.

Notons que si $D = -p_A p_B$, $P_{AB} = 0$, A et B sont allèles. L'allélisme peut être considéré comme une forme extrême de déséquilibre de liaison. Sur un plan historique, Bernstein, en utilisant un raisonnement équivalent, a montré que le système ABO était triallélique et non formé de deux systèmes dialléliques.

La notion de déséquilibre de liaison a été généralisée, d'une part, aux systèmes multialléliques et, d'autre part, au cas à plus de deux *loci*. L'interprétation de ce type d'analyse est difficile.

Évolution du déséquilibre de liaison

Le déséquilibre de liaison évolue d'une génération à l'autre sous l'effet des recombinaisons génétiques (*crossing-over*). Si r est la fréquence des recombinaisons entre les deux *loci* considérés ($0 \geq r \geq 0,50$), on a (figure 1) :

$$D_n = (1 - r) D_{n-1} \quad (3)$$

D_n et D_{n-1} sont respectivement le déséquilibre de liaison aux n^e et $(n - 1)^e$ générations. On a $D_n < D_{n-1}$. La formule (3) permet d'obtenir une loi de récurrence donnant la valeur de D à la génération n en fonction de celle à la génération 0 (génération initiale) :

$$D_n = (1 - r)^n D_0.$$

Quand le nombre de générations augmente, $(1 - r)^n$ diminue puisque $(1 - r)$ est inférieur à 1 et D_n diminue donc aussi. Si n tend vers l'infini ($n \rightarrow \infty$), $(1 - r)^\infty$ tend vers zéro et donc D_∞ aussi. L'évolution se fait donc vers l'équilibre.

La loi d'atteinte de l'équilibre est fonction du temps exprimé par n , nombre de générations, et du taux de recombinaison r entre les *loci*. Si le taux de recombinaison est de 0,50 (*loci* génétiquement indépendants), l'équilibre est atteint rapidement ; en revanche, si les deux *loci* sont étroitement liés ($r < 0,01$), l'équilibre n'est atteint qu'au bout de très nombreuses générations (figure 2). Donc, en cas de liaison étroite entre les *loci*, il est possible d'observer des déséquilibres de liaison au niveau de la population. En effet, les fréquences des différentes associations alléliques

sont certes gamétiques mais également chromosomiques et quasi stables d'une génération à l'autre, d'où une estimation possible du déséquilibre de liaison par des études familiales ou de populations, comme on le verra. Rappelons qu'on appelle haplotypes ces combinaisons alléliques qui se transmettent d'une génération à l'autre, les recombinaisons étant rares entre les *loci* considérés. Il faut noter que si les *loci* sont très rapprochés ($r < 0,001$) les mutations peuvent jouer un rôle dans l'évolution vers l'équilibre.

Causes du déséquilibre de liaison

Elles sont de deux ordres, les unes sont en rapport avec l'évolution démographique de la population, les autres avec la sélection naturelle [3].

- Une fusion entre deux populations peut créer un déséquilibre de liaison si les fréquences géniques sont différentes pour les deux *loci* concernés. Un déséquilibre de liaison significatif peut se maintenir, rappelons-le, si la distance génétique entre les deux *loci* est faible.

Un déséquilibre de liaison peut également être créé par dérive génétique ou par effet fondateur. Ces deux phénomènes — non exclusifs — ne peuvent concerner au début que des populations isolées et de taille réduite.

- La sélection naturelle peut être à l'origine d'un déséquilibre de liaison si elle favorise une certaine combinaison chromosomique de gènes ; il s'agit donc d'un phénomène qui favorise un haplotype donné. Le déséquilibre de liaison ne peut persister que si la distance entre les *loci* est inférieure à un ou deux centimorgans.

Parfois la sélection naturelle ne favorise qu'un gène, celui-ci entraînant les gènes neutres qui l'entourent. C'est l'effet auto-stop (*hitchhiking effect*). Celui-ci peut donc être à l'origine d'un déséquilibre de liaison entre deux gènes neutres situés au voisinage d'un gène « favorisé » ou entre un gène neutre et un gène « favorisé ». L'effet auto-stop explique vraisemblablement les déséquilibres de liaison notés entre les gènes de la drépanocytose, de la mucovis-

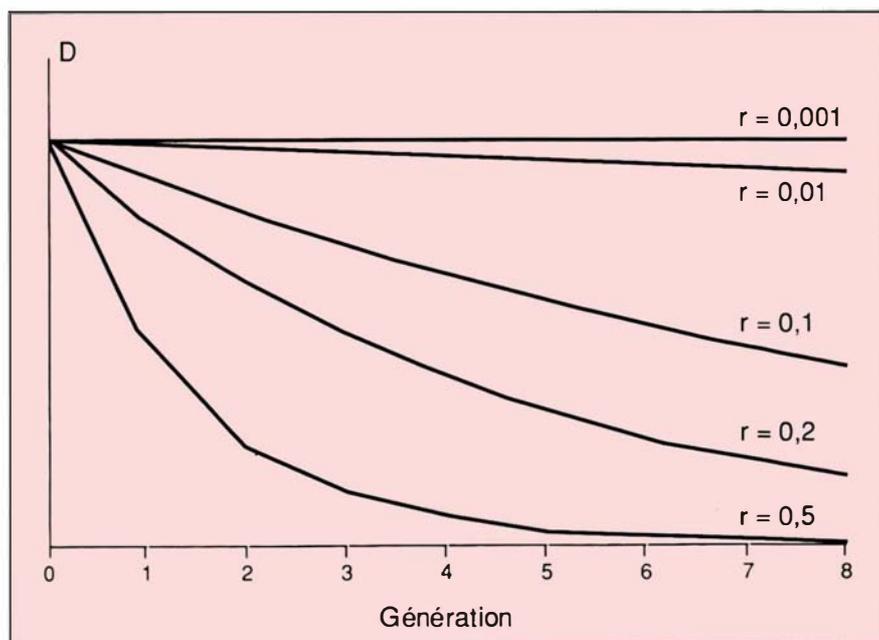


Figure 2. Évolution du déséquilibre de liaison (D) en fonction du taux de recombinaison (r), au cours des générations.

RÉFÉRENCES

11. Chakravarti A, Phillips AJ, Melits KH, Buetow KH, Kazazian HH, Seeburg PH. Patterns of polymorphisms and linkage disequilibrium suggest independent origin of the human growth hormone gene cluster. *Proc Natl Acad Sci USA* 1984 ; 81 : 6085-9.
12. Tzall S, Ellenbogen A, Eng F, Hirschhorn R. Identification and characterization of nine RFLPs at the adenosine deaminase (ADA) locus. *Am J Hum Genet* 1989 ; 44 : 864-75.
13. Ferec C, Verlingue C, Parent P. Déséquilibre de liaison et marqueurs de l'ADN associés au gène de la fibrose kystique. *J Genet Hum* 1989 ; 37 : 407-23.
14. Cox NJ, Bell IG, Xiang KS. Linkage disequilibrium in the human insulin/insulin like growth factor II region of human chromosome 11. *Am J Hum Genet* 1988 ; 43 : 495-501.
15. Lobos EA, Rudnick CH, Watson MS, Isenberg KE. Linkage disequilibrium study of RFLPs detected at the human muscle nicotinic acetylcholine receptor subunit genes. *Am J Hum Genet* 1989 ; 44 : 522-33.
16. Hegele RA, Plaetke R, Lalouel JM. Linkage disequilibrium between DNA markers at the low-density lipoprotein receptor gene. *Genet Epidemiol* 1990 ; 7 : 69-81.
17. Litt M, Jorde LB. Linkage disequilibrium between pairs of loci in a highly polymorphic region on chromosome 2q. *Am J Hum Genet* 1986 ; 39 : 166-78.

cidose, par exemple, et certains polymorphismes de restriction. (Un gène « favorisé » dans le cas d'une maladie récessive autosomique est un gène qui, à l'état hétérozygote, entraîne un avantage sélectif).

- Ces mécanismes très simples ne peuvent expliquer les nombreux déséquilibres de liaison observés dans certaines régions chromosomiques telle la région HLA.

Mesure du déséquilibre de liaison

Les formules (1) et (2) permettent de mesurer le déséquilibre de liaison. Cependant la valeur du déséquilibre de liaison dépend en partie, comme on l'a dit, de la fréquence des gènes. Pour essayer d'avoir des mesures indépendantes de ces fréquences, deux coefficients ont été décrits.

- Le coefficient D' (Lewontin) [4] est égal au rapport $\frac{D}{D_{max}}$. Dmax est

égal à la valeur maximale (en valeur absolue) que prend D pour les fréquences observées des gènes (voir plus haut). D' peut varier entre - 1 et + 1. Par exemple, dans le système HLA, les « D » des deux couples de gènes A₁ et B₈ d'une part, Bw₆₂ et Cw₃ d'autre part, sont proches (0,054 et 0,053) ; tandis que les D' sont différents : 0,64 et 0,83 respectivement.

- Le coefficient Dst (D standardisé, ou coefficient de corrélation gamétique) a été décrit par Hill et Robertson [5]. La fréquence des quatre types de gamètes (ou de chromosomes) peut être présentée sous la forme d'un tableau carré (ci-dessous).

		gènes		
		A	a	
gènes	B	P _{AB}	P _{aB}	P _{AB} + P _{aB} + P _{Ab} + P _{ab} = 1
	b	P _{Ab}	P _{ab}	

Si l'on donne la valeur 1 à la présence de A ou B et 0 à celle de a ou b, on peut calculer un coefficient de corrélation qui est le Dst. Dst varie entre - 1 et + 1, mais Dst et

D', bien qu'ayant les mêmes variations, sont deux mesures différentes. Si l'on considère les effectifs des quatre types de gamètes (ou chromosomes) on peut tester l'association par un test du khi deux.

On montre que $Dst = \sqrt{\chi^2/N}$, N étant le nombre de gamètes étudiés.

Mise en évidence du déséquilibre de liaison

Deux types d'étude sont possibles — selon que l'échantillon est composé de familles ou d'individus —, et dans chacun, le déséquilibre de liaison concerne dans la quasi-totalité des cas des loci étroitement liés, du moins en génétique humaine.

- **Les enquêtes familiales.** Les familles étudiées peuvent être, soit représentatives de celles de la population générale, soit au contraire avoir été recensées du fait de la présence d'un ou de plusieurs sujets atteints d'une maladie héréditaire. Habituellement, l'échantillon concerne des familles nucléaires, c'est-à-dire composées des parents et de leurs enfants ; leur étude permet de déduire les quatre haplotypes parentaux (figure 3).

- Dans les études des familles représentatives, on peut, à partir des haplotypes parentaux, calculer la fréquence des différents allèles, celle des combinaisons chromosomiques qui sont équivalentes aux combinaisons gamétiques lorsque les loci sont étroitement liés. A partir de ces différents éléments on peut calculer D, D' et Dst. La signification de l'association est donnée par le test du khi deux. Ce type d'analyse a longtemps été

utilisé pour l'étude du système HLA ; actuellement il l'est également pour la recherche de déséquilibre de liaison entre polymorphismes de restriction.

- L'étude des familles des sujets atteints concerne dans la majorité des cas une maladie récessive autosomique. L'examen des parents et des enfants malades, et si nécessaire des enfants sains, permet de déterminer la fréquence d'un marqueur M parmi les chromosomes porteurs du gène normal A [f(M si A)] et parmi ceux porteurs du gène pathologique a [f(M si a)]. Une différence significative entre les deux fréquences témoigne de l'existence d'un déséquilibre de liaison. Ce type d'étude ne permet de calculer que Dst, mais il faut noter qu'il s'agit d'un cas particulier où la fréquence du gène normal et celle de l'allèle pathologique sont égales à 0,50 (parents de malades atteints par une maladie récessive autosomique).

Dans la mucoviscidose, par exemple, la sonde KM-19 avec l'enzyme de restriction Pst I révèle un polymorphisme diallélique. L'analyse des haplotypes dans les familles françaises a montré que la fréquence de l'allèle 1 est de 0,732 si le chromosome porte le gène normal et de 0,095 s'il porte le gène de la mucoviscidose. La valeur de Dst est de - 0,65.

• **Les études de population.** Deux types d'enquêtes sont possibles selon que l'échantillon est représentatif de la population générale ou au con-

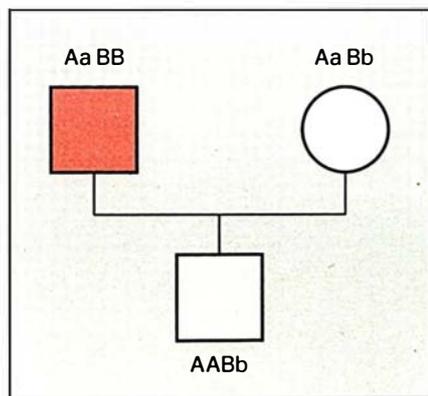


Figure 3. **Déduction des haplotypes à partir de l'étude d'une famille nucléaire.** Les haplotypes de l'enfant sont AB, Ab ; ceux du père sont AB, aB et ceux de la mère Ab, aB (les deux loci considérés sont génétiquement liés).

traire est constitué de malades et de témoins.

- Dans le premier type d'étude on cherche une association entre deux caractères héréditaires [A] et [B]. L'association entre les phénotypes est mise en évidence par le coefficient de corrélation calculé à partir d'un tableau de contingence. A partir de la fréquence des caractères et de l'association phénotypique on peut calculer p_A , p_B et P_{AB} , et par conséquent le coefficient D [6]. Prenons à titre d'exemple l'association qui existe à l'échelle des populations entre les antigènes HLA-A1, et B8. L'encadré ci-dessous donne les répartitions phénotypiques.

		B8	
		+	-
A1	+	a = 376	b = 235
	-	c = 91	d = 1 265

L'association est hautement significative ($\chi^2 = 699,4$). On peut, à partir de ces données, calculer la fréquence p_A et p_B des gènes A1 et B8, et la fréquence du chromosome non porteur de A1 et B8. La valeur

de D est égale à $\sqrt{\frac{d}{N}} - (1 - p_A)(1 - p_B)$ soit 0,077 ($N = a + b + c + d$), $\sqrt{\frac{d}{N}}$ est la fréquence

des chromosomes non porteurs, ni de A1, ni de B8. On a utilisé, pour le calcul de D, la fréquence des gènes « non A1 », « non B8 » et de l'haplotype « non A1 », car ce sont les paramètres qu'on peut estimer en premier.

- Dans les études « malades-témoins », on compare la fréquence d'un marqueur génétique chez les malades et les témoins. Nous ne considérerons ici que les maladies mendéliennes ; en effet, c'est uniquement dans ce cas qu'on peut affirmer qu'il y a un déséquilibre de liaison entre le gène marqueur et le gène responsable de la maladie. L'exemple le plus classique est l'association entre l'hémochromatose idiopathique,

maladie récessive autosomique, et l'antigène HLA-A3. La fréquence de ce dernier est de 76 % chez les malades et de 28 % chez les témoins. Ce type d'étude permet de mettre en évidence un déséquilibre de liaison mais ne permet de calculer aucun des coefficients mesurant le déséquilibre de liaison.

Biais dans la mise en évidence d'un déséquilibre de liaison

Une association entre deux caractères génétiques à l'échelle des populations n'est pas toujours due à un

DL. Le Tableau I, page suivante, décrit les principales situations pouvant conduire ou non à une association, ou à une liaison génétique (*linkage*) à l'échelle des familles.

Deux biais sont à souligner [3] :

- la stratification de la population : l'existence de deux groupes avec peu d'échanges génétiques peut être à l'origine d'une association entre deux caractères si la fréquence des gènes n'est pas la même dans les deux groupes. Cette situation peut être la première étape précédant un mélange ;
- une interaction entre gènes non alléliques ou épistasie peut être à l'origine d'une association significative. Le groupe sanguin Lewis et le caractère sécréteur sont un exemple de ce type de situation ; les quatre combinaisons phénotypiques sont : [Se, Lea +] [se, Lea +], [Se Lea -], [se Lea -] dont les fréquences respectives sont, dans la population européenne, 0 ; 0,187 ; 0,797 et 0,015. Le coefficient de corrélation est : - 0,95 [6].

Déséquilibre de liaison et génome

Il est très difficile actuellement d'esti-

RÉFÉRENCES

18. Barker D, Holm T, White R. A locus on chromosome 11p with multiple restriction site polymorphisms. *Am J Hum Genet* 1984 ; 36 : 1159-71.
19. Thompson EA, Deeb S, Motulsky AG. The detection of linkage disequilibrium between closely linked markers : RPLPs at the A1-CIII apolipoprotein genes. *Am J Hum Genet* 1988 ; 35 : 133-54.
20. Simon M, Pawlowsky Y, Bourel M, Fauchet R, Genetet B. Hémochromatose idiopathique : maladie associée à l'antigène tissulaire HLA-A3. *Nouv Presse Méd* 1974 ; 4 : 1432.
21. Kan YW, Dozy AM. Polymorphism of DNA sequence adjacent to human beta globin structural gene : relationship to sickle mutation. *Proc Natl Acad Sci USA* 1978 ; 75 : 5631-5.
22. Theilmann J, Kanani S, Shiang R, et al. Non-random association between alleles detected at D4S95 and D4S98 and the Huntington's disease gene. *J Med Genet* 1989 ; 26 : 676-81.
23. Degos L, Dausset J. Human migrations and linkage disequilibrium of HLA system. *Immunogenetics* 1974 ; 3 : 195-210.
24. Pagnier J, Mears JG, Dunda-Belkhodja O, et al. Evidence for the multicentric origin of the sickle cell hemoglobin gene in Africa. *Proc Natl Acad Sci USA* 1984 ; 81 : 1771-3.
25. Serre JL, Simon-Bouy B, Mornet E, et al. Studies of RFLP closely linked to the cystic fibrosis locus throughout Europe lead to new considerations in population genetics. *Hum Genet* 1990 ; 84 : 449-54.

mer même grossièrement l'importance des déséquilibres de liaison dans les populations humaines, mais il ne s'agit pas d'un phénomène exceptionnel, comme on l'a longtemps cru. Cette évolution des connaissances est facile à comprendre, et liée au fait que, aujourd'hui, ne sont pas seulement étudiés des *loci* éloignés. Il faut analyser séparément les déséquilibres de liaison observés apparemment hors de tout contexte pathologique et ceux observés dans la région d'un gène muté. Dans les deux cas, les déséquilibres de liaison concernent des *loci* étroitement liés. C'est, rappelons-le, l'étude de la région HLA qui a montré que les déséquilibres de liaison pouvaient être observés dans les populations humaines. Hors de la région HLA et de la région codant pour les chaînes lourdes des immunoglobulines, un certain nombre de déséquilibres de liaison non liés à des gènes pathologiques ont été rapportés. Il peut s'agir de déséquilibres de liaison intragéniques, de déséquilibres de liaison entre polymorphismes proches d'un gène, d'une combinaison des deux (un polymorphisme intragénique, l'autre extragénique) et enfin de déséquilibres de liaison dans des régions anonymes. Le *Tableau II*, qui n'est pas exhaustif, donne quelques résultats publiés. Il appelle quelques commentaires :

- il est vraisemblable que les résultats négatifs (absence de déséquilibres de liaison) sont moins souvent publiés que les positifs ;

- les déséquilibres de liaison concernent surtout des polymorphismes de restriction qui ne représentent qu'une faible part des polymorphismes de l'ADN ;

- quelques travaux ont montré qu'il n'y avait pas de relation monotone entre déséquilibre de liaison et distance physique [17, 19] : un déséquilibre de liaison peut exister entre deux gènes et ne pas être observé entre ces deux gènes et un troisième situé entre eux. Un exemple de ce type a été décrit au niveau du complexe génique apo A1-C3-A4 (voir *m/s* n° 5, vol. 5, p. 385). On ne peut donc ordonner les *loci* à partir de l'étude des déséquilibres de liaison. Cependant il faut noter que la non-mise en évidence d'un déséquilibre de liaison peut être due à la petite taille relative de l'échantillon. Il a été montré [19] qu'il faut des échantillons relativement importants soit si le déséquilibre de liaison est négatif et les fréquences géniques faibles, soit si le déséquilibre de liaison est positif et petit.

Les déséquilibres de liaison non associés à un gène pathologique peuvent avoir plusieurs origines. La plupart sont vraisemblablement dus à un mélange des populations ou à la

Tableau I		
ASSOCIATION ENTRE DEUX CARACTÈRES HÉRÉDITAIRES (D'APRÈS C. STERN MODIFIÉ) [7]. LES ÉTUDES FAMILIALES (FRATRIE) PERMETTENT DE METTRE EN ÉVIDENCE UNE LIAISON GÉNÉTIQUE (LINKAGE)		
Base génétique	Association	
	Population générale	Fratrises
• Effet pléiotropique d'un gène	+	+
• Effet de stratification		
• gènes non liés	+	-
• gènes liés	+	+ ou - *
• Population panmictique		
• gènes non liés	-	-
• gènes liés DL ≠ 0	+	+ ou - *
DL = 0	-	+ ou - *
• Interaction entre gènes	+	+

* Selon la distance génétique. Les associations positives ou négatives sont marquées +. L'absence d'association est marquée -. DL : déséquilibres de liaison.

Tableau II

DÉSÉQUILIBRES DE LIAISON OBSERVÉS DANS CERTAINES RÉGIONS CHROMOSOMIQUES EN DEHORS DE TOUT CONTEXTE PATHOLOGIQUE (LISTE NON EXHAUSTIVE). (LES POPULATIONS EUROPÉENNES ÉTUDIÉES SONT AMÉRICAINES DANS LA MAJORITÉ DES CAS)

Locus ou région chromosomique (Réf)	Populations	Nombre de polymorphismes étudiés	Nombre de chromosomes étudiés	Nombre de comparaisons 2 à 2	Seuil de signification	Nombre de DL (%)
Phénylalanine hydroxylase (8)	Danemark	8	66	28	0,01	17 (61)
Apo A1-C3-A4 (9)	Méditerranée	11	129	55	0,0009	14 (26)
	Noirs améric.	10	67	45	0,001	4 (9)
β-globine (10)	Grèce	13	18 à 70	78	0,05	24 (31)
	Italie	13	28 à 85	78	0,05	22 (28)
	Inde	12	9 à 46	66	0,05	27 (41)
	Noirs améric.	14	14 à 108	91	0,05	5 (5)
Hormone de croissance (11)	Europe (Nord)	4	34 à 42	6	0,01	6 (100)
	Méditerranée	4	16 à 22	6	0,01	6 (100)
	Noirs améric.	4	24 à 42	6	0,01	2 (33)
Adénosine déaminase (12)	Européennes	7	90	21	p < 0,005	13 (62)
Région du locus de la mucoviscidose (13)	Bretagne	7	110	21	p < 0,01	5 (24)
Insuline-Facteur de croissance II (14)	Chinoise	5	≥ 144	10	p < 0,01	5 (50)
	Européennes	4	≥ 72	6	p < 0,01	5 (83)
Récepteur de l'acétylcholine musculaire (15)	Européennes	Locus β : 2	96	1	p < 0,01	1 (100)
		Locus γδ : 5	106	10	p < 0,01	6 (60)
Récepteur des LDL (16)	Européennes	12	88 à 184	66	p < 0,001	5 (8)
Région anonyme en 2q (17)	Européennes	6	67 à 100	15	p < 0,01	3 (20)
Région anonyme 11 p (18)	Européennes	4	200	6	p < 0,01	4 (67)

DL : déséquilibre de liaison.

dérive génétique. Cependant certains doivent avoir un intérêt sur le plan biologique et par conséquent avoir été favorisés par la sélection naturelle. Il s'agit de déséquilibres de liaison situés dans le voisinage d'un gène, voire intragéniques ; ils peuvent donc jouer un rôle dans le fonctionnement (terme volontairement vague) de ce dernier. Si cette hypothèse est correcte, on devrait trouver le même déséquilibre de liaison en étudiant différentes populations. C'est le cas du déséquilibre de liaison noté dans la région de la β-globine entre le polymorphisme Hinc II situé en 5' du gène et le polymorphisme Hind III dans le gène Gγ ; il a été en effet retrouvé dans les populations noire, indienne, grecque et italienne [10]. Depuis la découverte du déséquilibre de liaison entre le gène de l'hémochromatose et le gène HLA-A3 [20], le

gène de la drépanocytose et le polymorphisme de restriction situé en 3' du gène β-globine et révélé par l'enzyme Hpa I (fragment de 13 Kb) [21], de nombreuses autres associations concernant les maladies récessives autosomiques ont été décrites. Il peut s'agir de déséquilibre de liaison entre le gène pathologique et aussi bien un polymorphisme de restriction qu'un haplotype défini par plusieurs polymorphismes.

Les déséquilibres de liaison sont systématiquement recherchés par le biologiste moléculaire lors de la « marche » vers le gène. La découverte d'une association suggère que l'on est proche du gène ; en effet, comme nous l'avons dit, on ne met en évidence que des déséquilibres de liaison concernant des loci proches. Les déséquilibres de liaison observés dans les maladies récessives auto-

miques sont vraisemblablement dus à la sélection naturelle : l'effet auto-stop est le plus probable, mais on ne peut exclure les autres mécanismes. Il faut noter que si une maladie se maintient par un équilibre entre mutation et sélection, on ne doit pas observer de déséquilibre de liaison.

Dans les maladies dominantes autosomiques, les déséquilibres de liaison sont rarement rapportés, ces affections se maintenant par mutations récurrentes. Cependant on en a récemment décrit un concernant deux polymorphismes de restriction et le gène de la chorée de Huntington [22]. Ce déséquilibre de liaison, concernant des malades appartenant à des populations différentes, est un argument en faveur de l'origine quasi unique de la mutation responsable de cette maladie. Rappelons que la chorée de Huntington est une maladie

grave, mais qui a été peu soumise à la sélection naturelle du fait de son apparition relativement tardive dans la vie. Ce dernier fait montre l'intérêt des études du déséquilibre de liaison en génétique des populations. Rappelons que l'analyse conjointe des fréquences géniques et des déséquilibres de liaison observés dans la région HLA a permis de retracer en partie l'histoire des migrations des populations humaines [23].

Dans ce domaine, d'autres faits intéressants ont été rapportés.

- L'origine multicentrique du gène de la drépanocytose a été démontrée par la mise en évidence de son association avec des haplotypes différents dans les populations du Sénégal, du Bénin, d'Afrique centrale et d'Asie [24].

- Dans certaines maladies telles que la phénylcétonurie ou la β -thalassémie, les différents types de mutations sont associés à des haplotypes particuliers, ces associations n'étant pas absolues du fait des recombinaisons génétiques et des mutations survenant au niveau des sites polymorphes. Elles sont vraisemblablement en rapport avec un avantage sélectif des hétérozygotes.

- La mucoviscidose fournit un autre exemple. En effet, avant le clonage du gène et la découverte de la muta-

tion prédominante (ΔF 508), on avait montré que le gène muté était associé à un haplotype appelé B. Cet haplotype est défini par les sondes XV2C (allèle 1) et KM-19 (allèle 2). Dans la population française, le gène de la mucoviscidose (gène CF) est associé dans plus de 80 % des cas à l'haplotype B, tandis que le gène normal ne l'est que dans moins de 20 %. Depuis, on a montré que le gène muté ΔF 508 est très fortement associé à l'haplotype B. Mais si l'on considère les haplotypes porteurs d'une mutation CF mais non ΔF 508, on trouve également un excès d'haplotype B. Ce fait pourrait être dû à une coïncidence, ou bien c'est la combinaison haplotype B (ou un gène qui lui est associé) - gène CF qui est favorisée à l'état hétérozygote. En outre, l'étude du déséquilibre de liaison dans différentes populations a permis d'émettre quelques hypothèses concernant l'origine géographique et la date d'apparition de la mutation CF [25].

Ces exemples montrent qu'une meilleure connaissance du génome a donné un nouvel élan à la génétique des populations qui, à son tour, permettra d'expliquer en partie les découvertes de la génétique moléculaire ■

Summary

Linkage disequilibrium

Linkage disequilibrium can be defined as a deviation from gametic equilibrium: at equilibrium the frequency of the AB gamete is the product of the frequencies of the component alleles. It has been described in a number of regions of the human genome in particular in the HLA complex, the β -globin region, the apolipoprotein genes A1-C3-A4 region and around the genes for many recessive diseases. Recombination tends to reduce any initial linkage disequilibrium but genes very closely linked remain far from equilibrium for hundreds of generations. A number of mechanisms could produce linkage disequilibrium, the most important are: population admixture, random drift and founder effect, selection effect.

TIRÉS A PART

J. Feingold.