

ANNALES DE L'I. H. P., SECTION B

MICHEL GILLOIS

La relation de dépendance en génétique

Annales de l'I. H. P., section B, tome 2, n° 3 (1965-1966), p. 261-278

http://www.numdam.org/item?id=AIHPB_1966__2_3_261_0

© Gauthier-Villars, 1965-1966, tous droits réservés.

L'accès aux archives de la revue « Annales de l'I. H. P., section B » (<http://www.elsevier.com/locate/anihpb>) implique l'accord avec les conditions générales d'utilisation (<http://www.numdam.org/conditions>). Toute utilisation commerciale ou impression systématique est constitutive d'une infraction pénale. Toute copie ou impression de ce fichier doit contenir la présente mention de copyright.

NUMDAM

*Article numérisé dans le cadre du programme
Numérisation de documents anciens mathématiques
http://www.numdam.org/*

La relation de dépendance en génétique ⁽¹⁾

par

Michel GILLOIS

Station centrale de génétique animale,
Jouy-en-Josas (Seine-et-Oise), France.

SOMMAIRE. — Le but de ce travail est l'introduction logique d'une nouvelle relation binaire entre deux gènes : la relation de dépendance, définie pour ses trois états : l'absolue dépendance isoactive ou hétéroactive, la dépendance en probabilité, l'indépendance. Une analyse de la notation classique de mutation et de ses rapports avec la relation de dépendance permet d'en montrer l'ambiguïté. Des calculs simples permettent d'atteindre l'expression de la variance *a priori* d'un zygote quelconque en fonction des coefficients de dépendance. Le mode de calcul des coefficients de dépendance dans le cas des pedigrees est présenté.

I. — INTRODUCTION

Nous avons présenté (Gillois, 1964 *a-b*) une étude logique des objets de l'interprétation génique et chromosomique de l'hérédité et des relations susceptibles d'exister entre ces objets. Le gène est une quantité d'information quasi stable ayant un support biochimique chromosomique dont l'hypothèse d'existence est nécessaire et suffisante pour expliquer la ségrégation mendélienne. Soit \mathfrak{G} l'ensemble, supposé fini, des gènes de tous les zygotes d'un groupe. Soit G l'ensemble, supposé fini de tous les gènes d'un zygote.

⁽¹⁾ Communication faite aux Journées d'Études de la Commission de Génétique de la Fédération européenne de Zootechnie, La Haye, juin 1965.

Le premier ensemble est le génotype général, le second le génotype particulier. Quatre relations binaires peuvent être définies entre deux gènes appartenant à l'un de ces ensembles.

1 . La relation d'homologie.

« Deux gènes sont homologues si, et seulement si, leurs descendants s'excluent dans les quatre produits haploïdes d'une méiose ». Cette relation binaire est réflexive, symétrique et transitive. C'est donc une relation d'équivalence; elle induit sur l'ensemble du génotype général une partition en sous-ensembles disjoints appelés : *classes d'homologues*.

2 . La relation d'isoaction.

« Deux gènes sont isoactifs si, et seulement si, ils sont homologues et ont la même action sur un même caractère ». Cette relation d'équivalence induit sur chaque classe d'homologues une partition en sous-ensembles disjoints appelés : *classes d'isoaction*. La série obtenue pour une classe d'homologues, en prenant un représentant dans chacune des sous-classes d'isoaction, est une série allélique. L'expression « fréquence d'un allèle » recouvre en fait la fréquence d'une classe d'isoaction.

3 . La relation d'allélisme.

« Deux gènes sont allèles si, et seulement si, ils sont homologues et ont une action différente sur un même caractère ». Cette relation n'est pas une relation d'équivalence.

4 . La relation d'identité.

« Deux gènes sont identiques si, et seulement si, ils dérivent par descendance mendélienne, en l'absence de mutation, d'un même gène ancêtre ». Cette relation binaire d'équivalence induit sur tout ensemble de gènes une partition en sous-ensembles disjoints appelés *classes d'identité*. Deux gènes appartenant à une même classe sont identiques; la connaissance de la nature d'un seul permet de déterminer la nature de tous les autres, car il existe entre eux une relation matérielle; ils sont la copie non modifiée par les

mutations du gène ancêtre. Deux gènes appartenant à deux classes distinctes sont indépendants, car la connaissance de la nature de l'un ne permet pas de déterminer avec certitude la nature de l'autre.

Il n'est pas logique d'accepter sans réserve l'implication réciproque que nous avons définie entre la non-identité et l'indépendance.

L'analyse du rôle des mutations permet de montrer qu'il existe des situations pour lesquelles l'implication réciproque entre non-identité et indépendance n'existe pas.

Considérons une population dans laquelle existe deux classes d'isoaction A1 et A2 de représentants A et *a*. Soient *p* et *q* les probabilités attachées respectivement aux classes d'isoaction A1 et A2. Soit par ailleurs *v* la probabilité qu'un gène quelconque mute. Soit v_{12} la probabilité qu'un gène A mute en un gène *a*, soit v_{21} la probabilité de la mutation inverse. Soient deux gènes G1 et G2 à l'instant *t*₁. G1 et G2 peuvent être identiques avec la probabilité *f* et non identiques avec la probabilité $1 - f$; nous supposons que les deux gènes non identiques sont indépendants. Donc $1 - f$ est aussi la probabilité de leur indépendance. A l'instant *t*₂ les deux gènes peuvent n'avoir muté ni l'un ni l'autre avec la probabilité $(1 - v)^2$, l'un des deux gènes peut avoir muté et pas l'autre avec la probabilité $2(1 - v)v$, les deux gènes peuvent avoir muté avec la probabilité v^2 .

A l'instant *t*₂ la probabilité que les deux gènes soient identiques est $f(1 - v)^2$, et la probabilité que les deux gènes soient non identiques est :

$$f[2(1 - v)v + v^2] + (1 - f).$$

Mais la probabilité qu'à l'instant *t*₂ les deux gènes soient non identiques n'est pas égale à la probabilité de leur indépendance. Dans l'expression :

$$f[2(1 - v)v + v^2] + (1 - f)$$

$(1 - f)$ est la probabilité que les deux gènes soient indépendants,

fv^2 est la probabilité de leur absolue dépendance isoactive puisque si les deux gènes étaient AA avec la probabilité conditionnelle *p*, à l'instant *t*₁, ils sont après mutation aa avec la probabilité conditionnelle 1 à l'instant *t*₂,

$f2(1 - v)v$ est la probabilité de l'absolue hétéroactive, puisque si les deux gènes étaient AA avec la probabilité conditionnelle *p* à l'instant *t*₁, ils sont Aa avec la probabilité conditionnelle 1 à l'instant *t*₂.

Ainsi les notions de dépendance et d'identité ne se recouvrent pas.

Ce travail étudie la relation de dépendance entre gènes sans référence à leurs états d'identité ou de non-identité. Une cinquième relation binaire entre deux gènes est ainsi définie.

II. — LES RELATIONS DE DÉPENDANCE

Deux gènes sont *absolument dépendants* si la connaissance de la nature de l'un nous apporte toute l'information nécessaire pour déterminer la nature de l'autre. Cette relation binaire est une relation d'équivalence symétrique, réflexive, transitive, qui réalise sur un ensemble de gènes une partition en sous-ensembles disjoints de gènes absolument dépendants.

Une classe de gènes *absolument dépendants* n'est pas forcément un ensemble de gènes isoactifs, deux gènes hétéroactifs pouvant être absolument dépendants. Il est donc nécessaire de concevoir deux types d'absolue dépendance, une *absolue dépendance isoactive* et une *absolue dépendance hétéroactive*. L'emploi de ces deux nouvelles notions ne peut pas se faire sans préciser le nombre de classes d'isoaction de la population. Si deux gènes ne sont pas absolument dépendants, ils peuvent être dépendants en probabilité ou indépendants.

Deux gènes sont *indépendants* si la connaissance de la nature de l'un ne nous apporte aucune information modifiant notre degré d'incertitude quant à la nature de l'autre. Deux gènes sont *dépendants en probabilité* si la connaissance de la nature de l'un nous permet de modifier notre degré d'incertitude quant à la nature de l'autre.

Il s'agit de la notion mathématique de l'information. L'événement aléatoire est la détermination de la classe d'isoaction à laquelle appartient le gène. Cette détermination peut être faite *a priori*, si p_k est la probabilité attachée à la classe d'isoaction A_k , la quantité d'incertitude *a priori* est

$$H = \sum_k p_k \text{colog } p_k.$$

Cette détermination peut être faite en connaissant la classe d'isoaction d'un autre gène. Si les deux gènes sont absolument dépendants la quantité d'incertitude conditionnelle devient $H_c = 0$, si les deux gènes sont dépendants $H_c \neq H$, si les deux gènes sont indépendants $H_c = H$.

III. — COEFFICIENTS DE DÉPENDANCE DANS LE CAS DE BIALLÉLISME

Considérons une population dans laquelle nous avons deux classes d'isoaction (A_1) et (A_2) représentant A et a . Dans cette population désignons au hasard un zygote H . Soient G_H et G_H^* deux gènes homologues de ce zygote.

1^o Ces deux gènes homologues peuvent être en absolue dépendance isoactive avec la probabilité F_1 .

2^o Ces deux gènes peuvent être en absolue dépendance hétéroactive avec la probabilité F_2 . Ce qui revient à dire que nous avons une probabilité conditionnelle égale à 1 d'avoir un hétérozygote (Aa).

3^o Ces deux gènes peuvent être indépendants avec la probabilité F_3 .

Le zygote H peut être homozygote (AA) ou (aa) et hétérozygote (Aa) avec les probabilités respectives suivantes :

$$P(AA) = F_1 p + F_3 p^2$$

$$P(Aa) = F_2 + F_3 2pq$$

$$P(aa) = F_1 q + F_3 q^2.$$

Est-ce que dans le cas d'une population ne présentant que deux classes d'isoaction, on peut toujours s'affranchir de la dépendance en probabilité ? Jusqu'à présent nous n'avons pas rencontré de schémas de population ayant deux classes d'isoaction, nécessitant l'usage de la dépendance en probabilité.

F_1 , F_2 , F_3 sont appelés les *coefficients zygotiques de dépendance*.

Si les deux gènes homologues sont pris au hasard chacun dans un zygote différent, nous définissons de même les *coefficients parentaux de dépendance* Φ_1 , Φ_2 , Φ_3 .

IV. — CALCUL DES COEFFICIENTS DE DÉPENDANCE D'APRÈS LES ASCENDANCES DANS LE CAS DE BIALLÉLISME

1. Coefficients zygotiques et parentaux de dépendance.

Soit un zygote H et ses parents I et J.

Soit un taux de mutation moyen égal à u .

Si deux gènes homologues pris au hasard l'un dans I et l'autre dans J sont indépendants, qu'il y ait eu mutation ou non, ils restent indépendants lorsqu'ils sont réunis dans le zygote H :

$$F_3 = \Phi_3.$$

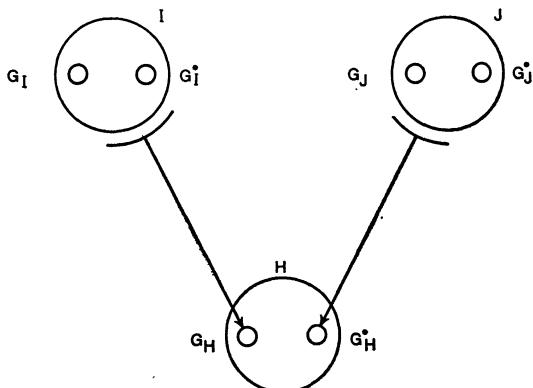


FIG. 1.

Les deux gènes homologues de H peuvent être absolument dépendants isoactifs (isocontraints) :

— soit parce que les deux gènes homologues pris au hasard l'un dans I et l'autre dans J sont absolument dépendants isoactifs (isocontraints) et n'ont pas muté, ou ont muté l'un et l'autre au cours de leur transfert à H,

— soit parce que les deux gènes homologues pris au hasard l'un dans I et l'autre dans H sont absolument dépendants hétéroactifs (hétérocontraints) et que l'un d'entre eux, seul, a muté au cours de leur transfert à H :

$$F_1 = \Phi_1[(1-u)^2 + u^2] + \Phi_2(1-u)u.$$

Les deux gènes homologues de H peuvent être absolument dépendants hétéroactifs (hétérocontraints) :

— soit parce que les deux gènes homologues pris au hasard l'un dans I l'autre dans J sont absolument dépendants isoactifs (isocontraints) et que l'un d'entre eux, seul a muté,

— soit parce que les deux gènes homologues pris au hasard l'un dans I et l'autre dans J sont absolument dépendants hétéroactifs (hétérocontraints) et n'ont pas muté, ou ont muté tous les deux au cours de leur transfert à H :

$$F_2 = \Phi_1 2(1-u)u + \Phi_2[(1-u)^2 + u^2].$$

2. Calcul des coefficients parentaux de dépendance.

Considérons le pedigree suivant pour lequel il existe un seul parent commun A.

Soient n_1 le nombre de chaînons entre I et A, et n_2 le nombre de chaînons

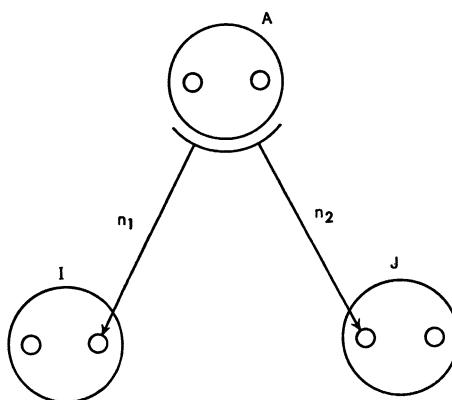


FIG. 2.

entre J et A. Soit \mathfrak{C}_0 la probabilité qu'aucun des deux gènes ne mutent au cours de leurs $(n_1 + n_2)$ transferts. Soit \mathfrak{C}_1 la probabilité qu'au cours du transfert des deux gènes les mutations se soient produites en nombre impair. Soit \mathfrak{C}_2 la probabilité qu'au cours du transfert des deux gènes les mutations se soient produites en nombre pair.

$$\mathfrak{C}_0 = (1 - u)^{n_1 + n_2}$$

$$\mathfrak{C}_1 = 1 - \mathfrak{C}_0 - \mathfrak{C}_2$$

$$\mathfrak{C}_2 = \sum_{n_i} u^{2n_i} (1 - u)^{n_1 + n_2 - 2n_i} \cdot C_{n_1 - n_2}^{2n_i}$$

$$1 \leq 2n_i \leq n_1 + n_2.$$

Les expressions des coefficients parentaux de dépendance sont les suivants :

$$\Phi_{1,U} = (1/2)^{n_1 + n_2} [(1/2)(\mathfrak{C}_0 + \mathfrak{C}_2) + (1/2)F_{1,A}(\mathfrak{C}_0 + \mathfrak{C}_2) + (1/2)F_{2,A}\mathfrak{C}_1]$$

$$\Phi_{2,U} = (1/2)^{n_1 + n_2} [(1/2)\mathfrak{C}_1 + (1/2)F_{1,A}\mathfrak{C}_1 + (1/2)F_{2,A}(\mathfrak{C}_0 + \mathfrak{C}_2)]$$

$$\Phi_{3,U} = 1 - \Phi_{1,U} - \Phi_{2,U}$$

$$\Phi_{3,U} = (1/2)^{n_1 + n_2} (1/2)F_{3,A} + 1 - (1/2)^{n_1 + n_2}$$

$$\Phi_{3,U} = 1 - (1/2)^{n_1 + n_2} [1 - (1/2)F_{3,A}]$$

Si le pedigree est quelconque, les expressions ci-dessus des coefficients parentaux de dépendance sont des sommes d'expression semblables à celles écrites ci-dessus, la sommation étant étendue à tous les parents communs de I et de J, reliés à I et à J par une chaîne de parenté n'ayant que ce parent comme point commun.

Les expressions deviennent alors :

$$\begin{aligned}\Phi_{1,u} &= \sum_i (1/2)^{n_1+n_i} [(1/2)(\mathcal{C}_0 + \mathcal{C}_2) + F_{1,Ai}(\mathcal{C}_0 + \mathcal{C}_2) + (1/2)F_{2,Ai}\mathcal{C}_1] \\ \Phi_{2,u} &= \sum_i (1/2)^{n_1+n_i} [1/2]\mathcal{C}_1 + (1/2)F_{1,Ai}\mathcal{C}_1 + (1/2)F_{2,Ai}(\mathcal{C}_0 + \mathcal{C}_2)] \\ \Phi_{3,u} &= 1 - \Phi_{1,u} - \Phi_{2,u} = 1 - \sum_i (1/2)^{n_1+n_i} [1 - (1/2)\delta_{Ai}].\end{aligned}$$

V. — VARIANCE GÉNOTYPIQUE A PRIORI D'UN ZYGOTE DANS LE CAS DE BIALLÉLISME

Soit un zygote H tiré au hasard dans une population. Les deux gènes homologues G_H et G_H^* de ce zygote peuvent être :

- soit absolument dépendants isoactifs (isocontraints) avec la probabilité F_1 ;
- soit absolument dépendants hétéroactifs (hétérocontraints) avec la probabilité F_2 ;
- soit indépendants (non contraints) avec la probabilité F_3 .

Nous ne considérons que les variables aléatoires génotypique Z , génique X et X^* , de dominance D et zygotique Y , telles que :

$$Z = X + X^* + D$$

$$Y = X + X^*.$$

Les grandeurs s et t de l'activité des gènes A et a sont déterminées par la méthode des moindres carrés en rendant minimum $E_p(D^2)$, espérance mathématique du carré de la dominance et en posant $E_p(Z)$ nulle; ces espérances mathématiques étant calculées sous l'hypothèse de l'indépendance réalisée entre les gènes homologues G_H et G_H^* .

$$\text{VAR } (Z) = E(Z^2) - [E(Z)]^2.$$

A. CALCUL DE $E(Z^2)$.

$$E(Z^2) = E[(Y + D)]^2 = E(Y^2) + EE(YD) + E(D^2).$$

a) Calcul de $E(Y^2)$.

$$E(Y^2) = F_1E_1(Y^2) + F_2E_2(Y^2) + F_3E_3(Y^2)$$

$$E_1(Y^2) = 4E(X^2) = 2E_p(Y^2)$$

$$E_2(Y^2) = (s + t)^2 = E_h(Y^2)$$

h est pris comme le symbole de la condition d'absolue dépendance hétéroactive.

$$\begin{aligned} E_s(Y^2) &= E_p(Y^2) \\ E(Y^2) &= (2F_1 + F_3)E_p(Y^2) + F_2E_h(Y^2). \end{aligned}$$

b) *Calcul de $E(YD)$.*

$$\begin{aligned} E(YD) &= F_1E_1(YD) + F_2E_2(YD) + F_3E_3(YD) \\ E_1(YD) &= p2s(i - 2s) + q2t(k - 2t) = E_c(YD) \\ E_2(YD) &= (s + t)(j - s - t) = E_h(YD) \\ E_3(YD) &= p^22s(i - 2s) + 2pq(s + t)(j - s - t) + q^22t(k - 2t) = 0 \\ E(YD) &= F_1E_c(YD) + F_2E_h(YD). \end{aligned}$$

c) *Calcul de $E(D^2)$.*

$$\begin{aligned} E'(D^2) &= F_1E_1(D^2) + F_2E_2(D^2) + F_3E_3(D^3) \\ E_1(D^2) &= p(i - 2s)^2 + q(k - 2t)^2 = E_c(D^2) \\ E_2(D^2) &= (j - s - t)^2 = (pi + qk)^2 \\ &= [E_c(D)]^2 = E_p(D^2) \\ E_3(D^2) &= F_1E_c(D^2) + (F_1 + F_3)E_p(D^2). \end{aligned}$$

d) *Expression de $E(Z^2)$.*

$$\begin{aligned} E(Z^2) &= (2F_1 + F_3)E_p(Y^2) + F_2E_h(Y^2) + 2F_1E_c(YD) + 2F_2E_h(YD) \\ &\quad + F_1E_c(D^2) = (F_2 + F_3)E_p(D^2). \end{aligned}$$

B. CALCUL DE $[E(Z^2)]$.

$$\begin{aligned} E(Z) &= F_1E_1(Z) + F_2E_2(Z) + F_3E_3(Z) \\ * \quad E_1(Z) &= E_1(Y + D) = E_1(Y) + E_1(D) \\ \text{or} \quad & \end{aligned}$$

$$\begin{aligned} E_1(Y) &= E_c(Y) = 0 \\ E_1(D) &= E_c(D) \end{aligned}$$

donc

$$\begin{aligned} E_1(Z) &= E_c(D) \\ * \quad E_2(Z) &= E_2(Y + D) = E_2(Y) + E_2(D) \\ \text{or} \quad & \end{aligned}$$

$$\begin{aligned} E_2(Y) &= (s + t) = E_h(Y) \\ E_2(D) &= (j - s - t) = pi + qk = E_e(D) \\ * \quad E_3(Z) &= 0 \text{ par convention,} \\ E(Z) &= F_2E_h(Y) + (F_1 + F_2)E_c(D) \end{aligned}$$

$$[E(Z)]^2 = (F_1^2 + F_2^2 + 2F_1F_2)E_p(D^2) + F_2^2E_h(Y^2) + (2F_1F_2 + 2F_2^2)E_h(YD).$$

Ce résultat est obtenu en utilisant les relations :

$$[E_c(D)]^2 = E_p(D^2)$$

$$E_h(Y)E_c(D) = E(s+t)(pi+qk) = E_h(YD).$$

C. CALCUL DE VAR (Z).

$$\begin{aligned} \text{VAR } (Z) &= E(Z^2) - [E(Z)]^2 \\ &= (2F_1 + F_3)E_p(Y^2) + (F_2 - F_1^2)E_h(Y^2) \\ &\quad + 2F_1E_c(YD) + (2F_2 - 2F_1F_2 - 2F_1^2)E_h(YD) \\ &\quad + F_1E_c(D^2) + (F_2 + F_3 - F_1^2 - F_2^2 - 2F_1F_2)E_p(D^2). \end{aligned}$$

D. ÉTUDE DE QUELQUES CAS LIMITES.

1^o Posons :

$$F_1 = f; \quad F_2 = 0, \quad F_3 = (1-f)$$

f ayant le sens d'un coefficient de consanguinité, c'est-à-dire que les deux gènes homologues G_H et G_H^* sont en état d'identité avec une probabilité f , et que cet état entraîne leur absolue dépendance isoactive (isocontrainte).

Il vient :

$$\text{VAR } (Z) = 2(1+f)E(X^2) + 4fE_c(XD) + fE_c(D^2)(1-f-f^2)E_p(D^2).$$

Expression que nous avions déjà obtenue dans le cas où nous supposions une implication réciproque entre non-identité et indépendance des gènes.

2^o Posons :

$$F_1 = F_2 = 0 \quad \text{et} \quad F_3 = 1$$

$$\text{VAR } (Z) = 2E(X^2) + E_p(D^2) = E_p(Y^2) + E_p(D^2).$$

3^o Posons :

$$F_1 = 1 \quad \text{et} \quad F_2 = F_3 = 0.$$

$$\text{VAR } (Z) = 4E(X^2) + 4E_c(XD) + E_c(D^2) - E_p(D^2)$$

$$\text{VAR } (Z) = 2E_p(Y^2) + 2E_c(YD) + E_c(D^2) - E_p(D^2).$$

4^o Posons :

$$F_2 = 1 \quad \text{et} \quad F_1 = F_3 = 0.$$

$$\text{VAR } (Z) = 0.$$

Ce résultat s'explique puisqu'il n'y a qu'un seul état possible.

VI. — AMBIGUITÉ DE LA NOTION CLASSIQUE DE MUTATION

La notion de mutation est expérimentale; c'est le changement d'activité d'un gène. Dans le cadre de l'analyse logique des objets de la théorie génique et chromosomique de l'hérédité et des relations susceptibles d'exister entre

ces objets, la mutation d'un gène est son changement de classe d'isoaction. Tout changement de classe d'isoaction implique pour un gène la rupture de sa relation d'identité avec tous les gènes de la classe d'identité à laquelle il appartenait. Si entre deux gènes la relation d'identité est rompue, la relation d'absolue dépendance isoactive qui les liait disparaît. Quelle nouvelle relation de dépendance ou d'indépendance s'établit alors entre les deux gènes devenus non identiques ? La mutation — cause de la rupture de l'identité entre deux gènes — nous intéresse, du point de vue logique, précisément parce qu'elle rompt l'identité de deux gènes. Nous considérerons deux types de mutations, les alpha-mutations correspondant aux changements de classe d'isoaction et les bêta-mutations correspondant aux changements de classes d'identité. Si toute alpha-mutation implique une bêta-mutation, la réciproque n'est pas nécessairement vraie. Une bêta-mutation rompt une relation d'identité sans imposer comme conséquence inéluctable un changement de classe d'isoaction.

Toute mutation du gène, qui modifie le site actif de la protéine que ce gène détermine, est une alpha-mutation, si cette modification entraîne un changement d'activité. Toute mutation du gène qui modifie la protéine que ce gène détermine, en dehors du site actif et sans modification obligatoire de sa spécificité fonctionnelle, est une bêta-mutation. La connaissance des séquences d'acides aminés d'une protéine permettrait de différencier ces deux types de mutations.

Soit p_i la probabilité *a priori* qu'un gène soit a_i , c'est-à-dire appartienne à la classe d'isoaction AI. Soit u la probabilité de bêta-mutation d'un gène quelconque. Soit u_{ij} la probabilité qu'un gène a_i mute en un gène a_j .

$$u = \sum_i \left[\sum_j p_i u_{ij} \right] = \sum_{ij} p_i u_{ij}.$$

La probabilité qu'un gène bêta-muté appartienne à la classe d'isoaction AK est :

$$\frac{\sum_i p_i u_{ik}}{\sum_i \left(\sum_j p_i u_{ij} \right)} = \varepsilon_k.$$

Si et seulement si $\varepsilon_k = p_k$ (pour tout k) la bêta-mutation entraîne l'indépendance entre deux gènes qui étaient identiques. Dans tous les autres cas pour lesquels $\varepsilon_k \neq p_k$ il y a dépendance en probabilité, car la bêta-mutation modifie notre degré d'incertitude quant à la nature du gène muté.

Considérons deux gènes homologues identiques G_H et G_H^* . La probabilité que l'un d'eux mute et que l'autre ne mute pas est $2u(1 - u)$. La probabilité que ce couple de gènes soit $a_i a_j$ sachant qu'ils sont identiques l'un d'eux seul a muté est $2p_i p_j$, si l'hypothèse $\varepsilon_k = p_k$ (pour tout k) est satisfaite.

La bêta-mutation, utilisée avec l'hypothèse $\varepsilon_k = p_k$ (pour tout k), hypothèse réalisée asymptotiquement, est implicitement employée en génétique de population (Malécot, 1948). Cette notion de mutation utilisée en génétique de population n'est pas conforme à son modèle expérimental.

L'alpha-mutation conduit toujours à une dépendance en probabilité puisqu'un gène qui subit une alpha-mutation ne peut rester dans la classe d'isoaction qui était la sienne avant mutation. Soit v la probabilité d'une alpha-mutation. Soit v_{ij} la probabilité qu'un gène a_i mute en un gène a_j . Il est immédiat que la probabilité v_{ii} est nulle pour tout i .

$$v = \sum_i \left(\sum_j p_i v_{ij} \right) = \sum_{ij} p_i v_{ij}$$

avec $v_{ii} = 0$ pour tout i .

La probabilité qu'un gène alpha-muté appartienne à la classe d'isoaction AK est :

$$\frac{\sum_i p_i v_{ik}}{\sum_{ij} p_i v_{ij}} = \pi_k.$$

Considérons deux gènes homologues identiques G_H et G_H^* . La probabilité que l'un d'eux ne mute pas est $2u(1 - u)$. La probabilité que ce couple de gènes soit $a_i a_j$ sachant qu'ils sont identiques un seul d'entre eux a muté est :

$$p_i \frac{\pi_j}{1 - \pi_i} + p_j \frac{\pi_i}{1 - \pi_j}$$

π_i étant la probabilité que parmi tous les gènes mutés, un gène pris au hasard soit a_i .

Si l'hypothèse $\pi_k = p_k$ (pour tout k), hypothèse réalisée asymptotiquement, alors l'expression ci-dessus devient :

$$p_i \frac{p_j}{1 - p_i} + p_j \frac{p_i}{1 - p_j}.$$

L'alpha-mutation ne conduit jamais à l'indépendance entre gènes identiques dont un seul mute. Si nous supposons que le nombre de classes d'isoaction est très grand, et que les probabilités qui leur sont attachées sont du même ordre de grandeur alors :

$$p_i \frac{p_j}{1 - p_i} \text{ est peu différent de } p^i p_j.$$

Mais même dans ce cas il n'y a pas d'indépendance puisque ce couple de gènes ne peut être isoactif.

Remarquons que dans le cas où il n'y a que deux classes d'isoaction :

$$\frac{p_j}{1 - p_i} = \frac{p_i}{1 - p_j} = 1.$$

VII. — LES RELATIONS DE DÉPENDANCE DANS LES CAS DE MULTIALLÉLISME

Les précédentes définitions de l'absolue dépendance isoactive (isocontrainte) et de l'indépendance restent valables dans le cas de multiallélisme. Par contre la définition de l'absolue dépendance hétéroactive (hétérocontrainte) ne peut être transposée de biallélisme au multiallélisme sans modification.

Deux voies de généralisation sont possibles. La première consiste à définir des probabilités, de symbole $F_{2,ij}$, attachées à tout couple de gènes hétéroactifs $a_i a_j$ qui sont de plus dépendants. La somme des probabilités $F_{2,ij}$ est égale à F_2 , probabilité que deux gènes soient absolument dépendants hétéroactifs. S'il n'existe que deux classes d'isoaction $F_{2,ij} = F_2$.

La seconde généralisation définit « l'absolue dépendance hétéroactive » comme la négation simultanée de l'absolue dépendance isoactive et de l'indépendance. Il s'agit en fait d'une dépendance en probabilité.

Nous appellerons une telle situation une « dépendance hétéroactive en probabilité ». Soit F_2 la probabilité que deux gènes soient en dépendance hétéroactive en probabilité, la probabilité qu'alors ce couple de gènes homologues soit $a_i a_j$ est :

$$p^i \frac{p_j}{1 - p_j} + p_j \frac{p_i}{1 - p_i}.$$

S'il n'existe que deux classes d'isoaction cette probabilité conditionnelle est égale à 1. C'est cette deuxième généralisation que nous utiliserons dans la suite de ce travail.

**VIII. — CALCUL DES COEFFICIENTS DE DÉPENDANCE
D'APRÈS LES ASCENDANCES
DANS LE CAS DE MULTIALLÉLISME**

1 . Les coefficients de dépendance zygotiques et parentaux.

Soit un zygote H et soient I et J ses parents.

1^o Soit v la probabilité d'une alpha-mutation.

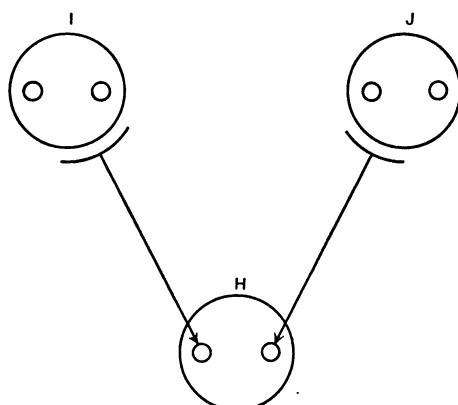


FIG. 3.

$$F_{1,H} = \Phi_{1,H}(1-v)^2 + \Phi_{2,H}X_1$$

$$F_{2,H} = \Phi_{2,H}[(1-v)^2 + X_2] + \Phi_{1,H}[2(1-v)v + u^2]$$

$$F_{3,H} = \Phi_{3,H}$$

$$X_1 = \sum_{ij} p_i \left(\frac{p_j}{1-p_i} \right) \left[v_{ij} \left(1 - \sum_{l \neq j} v_{il} \right) + v_{ji} \left(1 - \sum_{l \neq i} v_{il} \right) + v_{ij}v_{jl} \right]$$

$$X_2 = \sum_{ik} p_i \left(\frac{p_k}{1-p_i} \right) \left[v_{jk} \left(1 - \sum_{l \neq k} v_{il} \right) + v_{ik} \left(1 - \sum_{l \neq k} v_{il} \right) + v_{ik}v_{jl} \right]$$

par convention, puisqu'il s'agit d'alpha-mutation $l \neq j$, $l \neq i$, $i \neq j$.

Les expressions de ces coefficients de dépendance zygotiques en fonction des coefficients de dépendance parentaux ne sont pas utilisables puisqu'il

est toujours difficile d'avoir des renseignements sur les probabilités des alpha-mutations. Nous allons montrer, ci-après, comment l'emploi de la notion de bêta-mutation permet d'atteindre des expressions plus simples.

2^o Soit u la probabilité d'une bêta-mutation, nous considérons le cas où une bêta-mutation entraîne l'indépendance, hypothèse simplificatrice qu'il est impossible de faire dans le cas des alpha-mutations.

$$F_{1,H} = \Phi_{1,U}(1-u)^2$$

$$F_{2,H} = \Phi_{2,U}(1-u)^2$$

$$F_{3,H} = \Phi_{1,U}[u^2 + 2(1-u)u] + \Phi_{2,U}[u^2 + 2(1-u)u] + \Phi_{3,U}.$$

CALCUL DES COEFFICIENTS PARENTAUX DE DÉPENDANCE D'APRÈS LES ASCENDANCES. — Ce calcul n'est fait que dans le cas des bêta-mutations, dont la probabilité est :

$$\Phi_{1,U} = \sum_i \left[\left(\frac{1-u}{2} \right)^{n_1+n_2} \left(\frac{1+F_{1,i}}{2} \right) \right]$$

$$\Phi_{2,U} = \sum_i \left[\left(\frac{1-u}{2} \right)^{n_1+n_2} \left(\frac{F_{2,i}}{2} \right) \right]$$

$$\Phi_{3,U} = \sum_i \left[1 - \left(\frac{1-u}{2} \right)^{n_1+n_2} \left(\frac{1+F_{1,i}+F_{2,i}}{2} \right) \right]$$

$$\Phi_{3,U} = 1 - \Phi_{1,U} - \Phi_{2,U}.$$

IX. — VARIANCE GÉNOTYPIQUE *A PRIORI* D'UN ZYGOTE DANS LE CAS DU MULTIALLÉLISME

$$\text{VAR}(Z) = E(Z^2) - [E(Z)]^2.$$

A. CALCUL DE $E(Z^2)$.

$$E(Z^2) = F_1 E_1(Z^2) + F_2 E_2(Z^2) + E_3 E_3(Z^2)$$

$$E_1(Z^2) = 2E_p(Y^2) + 2E_c(YD) + E_c(D^2)$$

$$E_3(Z^2) = E_p(Y^2) + E_p(D^2).$$

Seul le calcul du terme $E_2(Z^2)$ présente une difficulté : posons :

$$E_h(Y^2) = \sum_{kl} p_k \frac{p_l}{1-p_k} (s_k + s_l)^2$$

$$E_h(YD) = \sum_{\substack{kl \\ k \neq l}} p_k \frac{p_l}{1-p_k} (s_k + s_l)(j_{kl} - s - s)$$

$$E_h(D^2) = \sum_{\substack{kl \\ k \neq l}} p_k \frac{p_l}{1-p_k} (j_{kl} - s_k - s_l)^2$$

en conséquence :

$$E_2(Z^2) = E_h(Y^2) + E_h(YD) + E_h(D^2)$$

d'où

$$\begin{aligned} E(Z^2) &= (2F_1 + F_3)E_p(Y^2) + 2F_1E_c(YD) + F_1E_c(D^2) \\ &\quad + F_3E_p(D^2) + F_2E_h(Y^2) + F_2E_h(YD) + E_2E_h(D^2) \\ &= (2F_1 + F_3)E_p(Y^2) + F_2E_h(Y^2) \\ &\quad + 2F_1E_c(YD) + F_2E_h(YD) \\ &\quad + F_1E_c(D^2) + F_2E_h(D^2) + F_3E_p(D^2). \end{aligned}$$

B. CALCUL DE $[E(Z)]^2$, TERME CORRECTEUR.

$$E(Z) = F_1E_c(Z) + F_2E_h(Z) + F_3E_p(Z)$$

$$E_c(Z) = F_c(Y + D) = E_c(D)$$

$$E_h(Z) = E_h(Y + D) = E_h(Y) + E_h(D)$$

$$E_h(Y) = \sum_{\substack{kl \\ k \neq l}} p_k \frac{p_l}{1-p_k} (s_k + s_l)$$

$$E_h(D) = \sum_{\substack{kl \\ k \neq l}} p_k \frac{p_l}{1-p_k} (j_{kl} - s_k - s_l)$$

$$E_p(Z) = 0$$

donc :

$$E(Z) = F_1E_c(D) + F_2E_h(Y) + F_2E_h(D)$$

d'où

$$\begin{aligned} [E(Z)]^2 &= F_1^2[E_c(D)]^2 + F_2^2[E_h(Y)]^2 + F_2^2[E_h(D)]^2 \\ &\quad + 2F_1F_2E_c(D)E_h(Y) + 2F_1F_2E_c(D)E_h(D) \\ &\quad + 2F_2^2E_h(Y)E_h(D). \end{aligned}$$

C. ÉTUDE DE QUELQUES CAS LIMITES.

1°

$$F_1 = F_2 = 0 \quad F_3 = 1$$

$$\text{VAR}(Z) = E_p(Y^2) + E_p(D^2).$$

$$2^o \quad F_1 = 1 \quad F_2 = F_3 = 0$$

$$\text{VAR}(Z) = 2E_p(Y^2) + 2E_c(YD) + E_c(D^2) + E_p(D^2).$$

$$3^o \quad F_2 = 1 \quad F_1 = F_2 = 0$$

$$\text{VAR}(Z) = E_h(Y^2) + E_h(YD) + E_h(D^2) + E_h(Y)^2 + E_h(D^2) + 2E_h(Y)E_h(D).$$

Ce résultat ne s'identifie pas avec celui du point D3 du paragraphe V.

X. — CONCLUSION

La connaissance de la nature exacte des relations de dépendance et d'indépendance qui existent entre les variables aléatoires génétiques est absolument nécessaire pour pouvoir établir un modèle mathématique susceptible de décrire l'évolution probable d'une population.

C'est dans ce sens qu'est utilisée la relation d'identité. Si deux gènes homologues sont identiques, les variables aléatoires génétiques qui leur sont attachées suivent la même loi et sont absolument dépendantes. Malheureusement, la non-identité de deux gènes ne permet pas de déduire quelles sont les relations de dépendance ou d'indépendance des variables aléatoires génétiques qui leur sont attachées. Supposer une implication réciproque entre la non-identité des gènes et leur indépendance est une faute logique qui restreint particulièrement la portée du nouveau langage génétique que nous avions introduit (Gillois M., 1964).

Les relations de dépendance que nous avons définies complètent et achèvent cette étude logique. Elles permettent l'étude simple des populations homogames et d'effectif limité.

RÉSUMÉ

L'étude logique des objets de l'interprétation génique et chromosomique de l'hérédité et des relations susceptibles d'exister entre ces objets (Gillois M., 1964) était perfectible, cette précédente étude ne permettant pas de traiter de façon cohérente le cas des populations homogames. Une analyse plus poussée montre l'ambiguité de la relation d'identité. Le fait que la non-identité de deux gènes entraîne leur indépendance n'est pas une propriété intrinsèque de la relation d'identité. Cette propriété est relative à la structure de la population. Pour atteindre plus de généralité, nous introduisons la relation de dépendance sous ses trois aspects (absolue dépendance

isoactive ou isocontrainte, dépendance en probabilité hétéroactive ou hétéro-contrainte, indépendance ou non-contrainte).

Comme application de ces nouvelles définitions, nous donnons l'expression de la variance *a priori* d'un zygote quelconque, le mode de calcul des coefficients de dépendance d'après les ascendances, dans le cas de biallélisme, puis de multiallélisme.

Pour pouvoir atteindre ces résultats, une analyse très poussée du rôle des mutations en génétique des populations a permis de distinguer les changements de classes d'isoaction (alpha-mutations), des changements de classes d'identité (bêta-mutations).

REMERCIEMENTS

Nous adressons à M. Malécot G., Professeur de Mathématiques à la Faculté des Sciences de Lyon, France, notre gratitude respectueuse pour l'aide attentive qu'il nous a toujours apportée au cours de nos travaux.

M. Malécot G. a établi à propos de ce travail que les hypothèses $\pi_k = p_k$ (dans le cas des alpha-mutations) et que $\varepsilon_k = p_k$ (dans le cas des bêta-mutations) sont asymptotiquement réalisées, car

$$p_{k(n)} = p_{k(n+1)} = \sum_i p_i v_{ij} = \pi_k$$

avec

$$v_{ii} = 1 - \sum_{i \neq k} p_i v_{ij}$$

et

$$\sum_{ij} p_i v_{ij} = 1.$$

BIBLIOGRAPHIE

- GILLOIS M., 1964 *a*, La relation d'identité en génétique. *Thèse Fac. Sciences*, Paris, 294 pages.
- GILLOIS M., 1964 *b*, Logique d'un langage génétique : la relation d'identité. *Journées d'Études de la Commission de Génétique de la Fédération européenne de Zootechnie*, Lisbonne, octobre 1964.
- GILLOIS M., 1964 *c*, Corrélation génotypique entre apparentés. *Journées d'Études de la Commission de Génétique de la Fédération européenne de Zootechnie*, Lisbonne, octobre 1964.
- GILLOIS M., 1964 *d*, Calcul des corrélations d'identité d'après les liens de parenté. *Journées d'Études de la Commission de Génétique de la Fédération européenne de Zootechnie*, Lisbonne, octobre 1964.

Déposé le 22 octobre 1965.