

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

M. GILLOIS

Relation d'identité en génétique I. - Postulats et axiomes mendéliens

Annales de l'I. H. P., section B, tome 2, n° 1 (1965-1966), p. 1-94

<http://www.numdam.org/item?id=AIHPB_1965__2_1_1_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/>

Relation d'identité en génétique

I. — Postulats et axiomes mendéliens

par

M. GILLOIS

Station Centrale de Génétique animale,
Centre National de Recherches zootechniques,
Jouy-en-Josas (S.-et-O.),
I. N. R. A.

Échange Annales

SOMMAIRE

I. Introduction	1
II. L'analyse de la méiose chez un pyrénomycète	2
III. Les objets de la théorie de l'hérédité et leurs relations	5
IV. Les différents aspects de la relation d'identité	16
V. Les situations et les coefficients d'identité restreinte	20
VI. Les situations et les coefficients d'identité généralisée d'ordre 2.	32
VII. Les tirages gamétiques multiples dans un même zygote	33
VIII. Conclusion.	35
IX. Résumé.	36
X. Bibliographie	37

I. — INTRODUCTION

Depuis que des études très poussées ont abordé soit la structure fine du gène, soit les conditions permettant l'expression des potentialités d'un gène, il n'existe plus de « bons gènes ». En fait, la notion de gène qui a été dégagée des résultats des expériences d'hybridation sexuelle est à la fois précise et floue. C'est, en génétique classique, une unité discontinue de fonc-

tion, de duplication, de mutation, de recombinaison. Le mot fonction est pris dans un sens assez vague, il s'agit de l'ensemble des phénomènes aboutissant à l'expression d'un caractère donné (ailes longues de la drosophile). Au cours des divisions cellulaires indirectes (mitose et méiose), le gène, unité discontinue, donne naissance par « duplication » à d'autres gènes généralement identiques au gène ancêtre. Mais parfois la duplication est incorrecte puisqu'un gène descendant n'a plus la même fonction que le gène ancêtre, un tel phénomène est une mutation. Le gène est enfin une unité de recombinaison. Les segments des chromosomes peuvent être séparés par le crossing-over, alors que les gènes eux-mêmes ne sont pas susceptibles d'être scindés.

Mais l'augmentation du pouvoir de résolution des expériences d'hybridation sexuelle (et parasexuelle des microorganismes), après avoir rendu une image floue du gène, a fait éclater la notion même du gène dont les débris épars sont les cistrons, les recons, les mutons, les polarons et les opérans.

Pour un critère qualitatif (morphologique ou biochimique) la notion de « bon gène » dépend en fait du pourcentage minimum de recombinaisons que les tests statistiques permettent de détecter au cours des expériences d'hybridation.

Les modèles mathématiques que nous allons exposer reposent exclusivement sur l'analyse des conclusions des expériences d'hybridation sexuelle qui ont permis de dégager la notion de gène.

II. — L'ANALYSE DE LA MÉIOSE CHEZ UN PYRÉNOMYCÈTE

Soient deux souches d'un champignon pyrénomycète *Bombardia lunata* dont l'une sauvage (S) a les ascospores noires à maturité et un mycélium très pigmenté, et dont l'autre dite *lactea* (L) a les ascospores blanches et un mycélium non pigmenté. Il s'agit de deux « races » pures ne produisant chacune que des gamétophytes d'un seul type. Les asques des périthèces d'un croisement d'une souche sauvage et d'une souche *lactea*, les deux souches ayant des signes sexuels complémentaires, appartiennent à six types, schématisés ci-contre (fig. 1).

Chacun de ces asques est le résultat d'une seule méiose et les généticiens présentent de la méiose l'analyse génétique suivante, que nous illustrons dans le cas de l'asque n° 3 (tableau I).

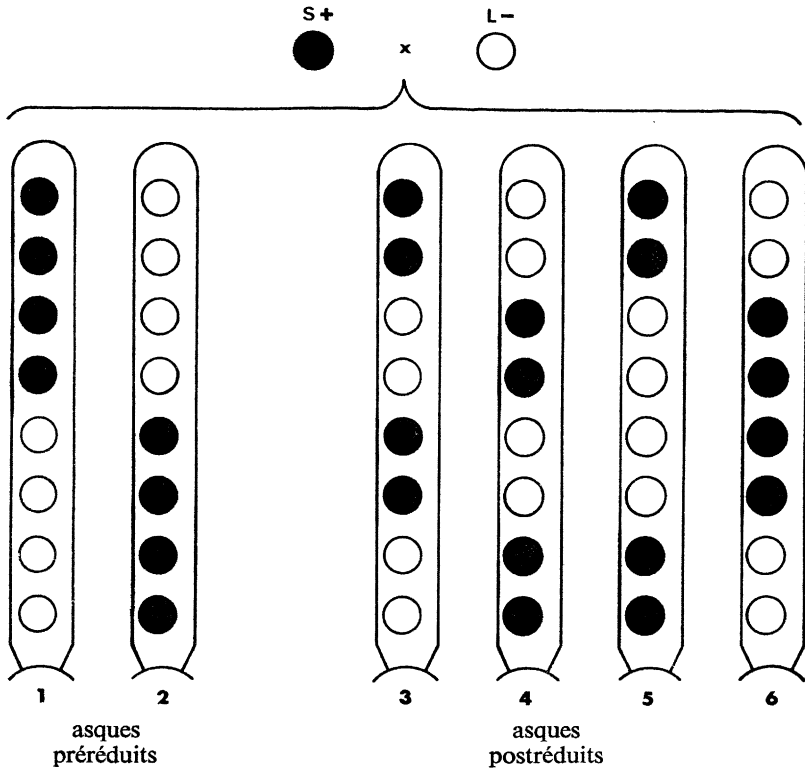


FIG. 1.

Pour expliquer la transmission potentielle des caractères sauvage et *lactea* au travers de la méiose nous supposons l'existence d'un objet appelé gène, support de l'information nécessaire à l'expression des caractères. Cet objet est discontinu et quasi stable et se repère sur un point du chromosome, considéré comme un segment de droite orientée appelé carte chromosomique.

Numérotons de 1 à 8 les huit gènes de l'asque final de notre exemple.

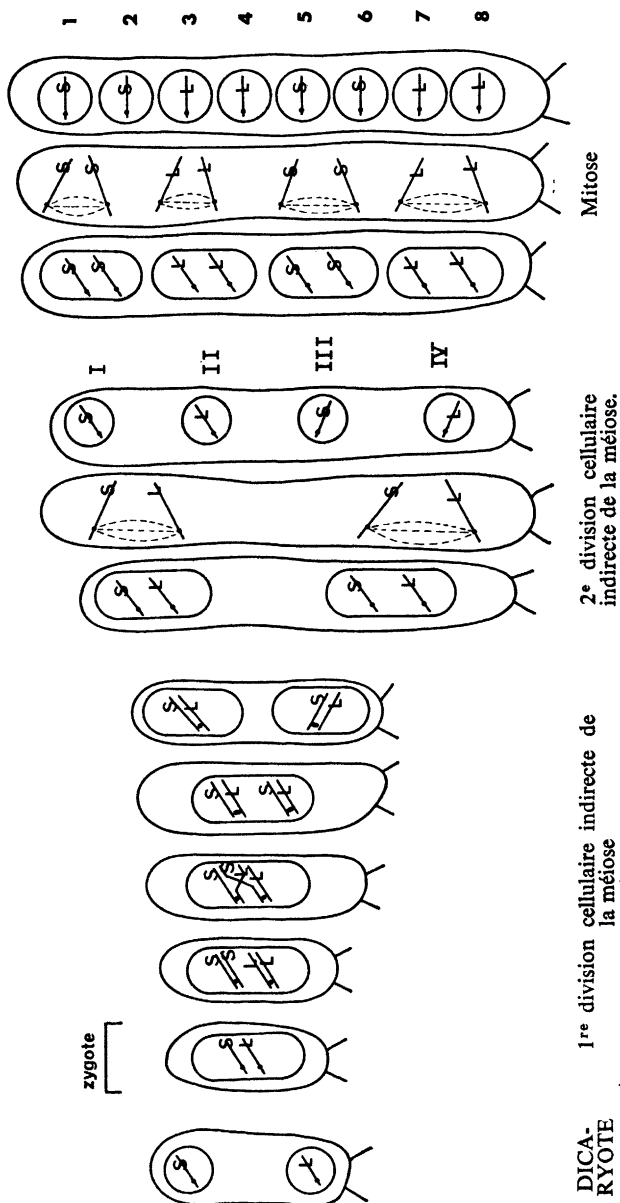
Les gènes 1, 2, 5, 6 dérivent par duplication du gène S, dont ils sont en l'absence de mutation la copie biochimique. Ces gènes sont par définition *identiques* entre eux. Les gènes 3, 4, 7, 8 sont pour les mêmes raisons identiques entre eux parce que copie d'un gène L. Il en est de même des gènes I, III, et des gènes II, IV, appartenant à l'asque issu de la méiose. Parce que les gènes I, II, III, IV s'excluent mutuellement dans les quatre noyaux appartenant à l'asque produit par la méiose, les gènes ancêtres S et L sont dits par définition *homologues*. Les gènes I, II, III, IV s'excluent mutuellement

TABLEAU I

Division cellulaire indirecte équationnelle par rapport au nombre de centromères et de chromatides.

2^e division cellulaire indirecte équationnelle par rapport au nombre de centromères, réductionnelle par rapport au nombre de chromatides (donc de gènes).

1^{re} division cellulaire indirecte réductionnelle par rapport au nombre de centromères, équationnelle par rapport au nombre de chromatides (donc de gènes).



Choix d'une issue parmi 4 possibles: gène S en position apicale.

Choix d'une issue parmi 2 possibles: nombre impair de « crossing-overs » réalisés entre le locus et le centromère.

Fig. 2.

dans les quatre noyaux produits par cette méiose, aussi les gènes ancêtres S et L qui ont des fonctions distinctes sont dits *allèles* par définition.

Si nous avons observé un asque issu d'un croisement entre deux souches de signes sexuels complémentaires de *Bombardia lunata* sauvage, les gènes S₊ et S₋ du dicaryote, puis du zygote, seraient dits *isoactifs* parce que leurs descendants s'excluent dans les quatre produits haploïdes de la méiose et qu'ils ont même action sur un même caractère.

III. — LES OBJETS DE LA THÉORIE GÉNÉTIQUE DE L'HÉRÉDITÉ ET LEURS RELATIONS

Nous voulons atteindre, au moyen de démonstrations, un certain nombre de résultats mathématiques applicables à la Génétique. Or, trop souvent, les démonstrations à l'usage des physiciens et des biologistes ne sont constituées que par des calculs qui apportent la conclusion d'une manière inopinée. Aussi, pour dégager les notions fondamentales dont les propriétés permettent ces calculs, tenterons-nous un exposé axiomatique de la génétique, c'est-à-dire que nous décrirons les concepts ou objets de la théorie génétique et chromosomique de l'hérédité et que nous énoncerons les relations qui existent entre ces objets.

Les objets sont décrits par le nombre minimum d'attributs nécessaires et suffisants au développement non contradictoire de la théorie et à l'obtention de résultats mathématiques utilisables dans l'interprétation d'expériences ou d'observations.

Ces attributs sont choisis indépendants entre eux, non contradictoires et tels que l'expérience biologique, dans certaines limites de précision, les assure comme certains.

Puisque les relations qui existent entre ces objets nous sont données par l'expérience, nous les étudierons systématiquement afin de les identifier aux relations définies par les mathématiciens.

1. Les notions de gènes, de gamètes, de zygotes.

Nous donnons du gène la définition suivante : « quantité d'information biologique quasi stable ⁽¹⁾ dont l'hypothèse d'existence est nécessaire (et

(1) Dans l'hypothèse de l'absence de mutations, cette quantité sera dite stable.

suffisante) pour expliquer la transmission des caractères qualitatifs repérables qu'elle conditionne au cours de la méiose et de la fécondation ».

Il est, de plus, nécessaire (et suffisant) de supposer que les gamètes contiennent chacun un gène — ils sont dits haploïdes — et les zygotes chacun deux gènes — ils sont dits diploïdes ⁽²⁾.

2. Les notions de population, de génération.

Nous appelons population la succession ordonnée de générations disjointes ⁽³⁾. Nous désignons par génération le cycle ordonné d'une génération zygotique, d'une génération zygotique utile, d'une génération gamétique, d'une génération gamétique utile. La génération zygotique est un ensemble de zygotes contemporains. La génération zygotique utile est un sous-ensemble de la génération zygotique, c'est l'ensemble des zygotes qui fournissent, à la suite des méioses, les gamètes de la même génération. La génération gamétique est un ensemble de gamètes contemporains. La génération gamétique utile est l'ensemble des gamètes qui, unis deux à deux, participent effectivement à la constitution de la génération zygotique suivante, dont les zygotes sont par définition contemporains. Nous appelons population zygotique l'ensemble des générations zygotiques d'une population. Nous appelons groupe zygotique un ensemble arbitraire de zygotes. Toute population et toute génération zygotique sont donc des groupes ⁽⁴⁾. Il en est de même pour les générations et les populations gamétiques.

3. L'homologie des gènes.

L'ensemble \mathcal{G} de tous les gènes présents dans une population zygotique est le *génotype général*. L'ensemble G de tous les gènes présents dans un zygote est le *génotype particulier* ou *individuel* de ce zygote. Il est immédiat que :

$$\left\{ \begin{array}{l} G_i \cap G_j = \emptyset \\ \bigcup_i G_i = \mathcal{G} \end{array} \right.$$

i et j étant des indices repérant les zygotes.

⁽²⁾ En excluant le cas des gènes liés au sexe, que nous pourrions d'ailleurs introduire.

⁽³⁾ Dans le cas d'une population quelconque, il est toujours possible de se réduire à ce schéma en considérant l'intervalle moyen entre générations.

⁽⁴⁾ Qualitatif sans rapport avec le « groupe » des mathématiciens.

Supposons que G soit un ensemble fini. Au cours d'une méiose nous dirons que *deux gènes sont liés par la relation d'homologie, de symbole \mathcal{A} , si, et seulement si* ⁽⁵⁾, *ils s'excluent dans les quatre produits haploïdes de cette méiose.*

La relation d'homologie est donc une relation binaire qui possède les propriétés suivantes :

Elle est réflexive :

$$\forall x \in G : x\mathcal{A}x.$$

Elle est transitive :

$$(\forall x, \forall y, \forall z) \in G : x\mathcal{A}y, y\mathcal{A}z \Rightarrow x\mathcal{A}z.$$

Elle est symétrique :

$$(\forall x, \forall y) \in G : x\mathcal{A}y \Leftrightarrow y\mathcal{A}x.$$

La relation d'homologie est donc une relation d'équivalence. En conséquence, cette relation réalise une partition de l'ensemble des gènes d'un génotype général en sous-ensembles disjoints chacun appelé *classe d'homologues*.

Si l'on considère ces sous-ensembles comme les éléments d'un nouvel ensemble, ce dernier est l'ensemble des classes d'homologues, ensemble quotient du génotype général par la relation d'équivalence qu'est la relation d'homologie. Le symbole d'une classe d'homologues quelconque sera C . Une classe d'homologues est un ensemble de gènes homologues. Deux gènes appartenant à une même classe d'homologues sont *homologues*.

Autrement dit, tout gamète possède pour chaque classe d'homologues un représentant qui est un gène appartenant à cette classe d'homologues. Chaque zygote possède pour chaque classe d'homologues deux représentants qui sont deux gènes appartenant à cette classe d'homologues.

4. L'isoactivité des gènes.

Deux gènes sont dits isoactifs ssi ils appartiennent à la même classe d'homologues et ont une action identique sur un même caractère. Deux gènes non isoactifs sont hétéroactifs, notons qu'ils peuvent être homologues ou non.

(5) Nous remplacerons l'expression « si et seulement si » par ssi.

Il s'agit encore d'une relation binaire dite relation d'isoaction, de symbole \mathcal{I} qui a les propriétés suivantes :

Elle est réflexive :

$$\forall x \in C_{ij} : x\mathcal{I}x.$$

Elle est transitive :

$$\forall (x, y, z) \in C_{ij} : \mathcal{I}xy, y\mathcal{I}z \Rightarrow x\mathcal{I}z.$$

Elle est symétrique :

$$\forall (x, y) \in C_{ij} : x\mathcal{I}y \Leftrightarrow y\mathcal{I}x.$$

La relation d'isoaction génique est une relation d'équivalence. En conséquence, elle assure une partition de l'ensemble des gènes d'une même classe d'homologues en sous-ensembles disjoints encore dénommés classes d'équivalence, classes de gènes *isoactifs*. C'est à ces classes de gènes isoactifs que les généticiens attribuent une nomenclature sous forme de lettres et d'abréviations et dont ils se servent pour décrire les génotypes individuels. Nous écrirons de façon symbolique :

$$C_{ij} = \{ A, a_1, a_2 \dots a_k \}.$$

$A, a_1, a_2 \dots a_k$ étant les classes de gènes isoactifs. Nous appelons *série allélique* une suite de classes de gènes isoactifs. Les sous-classes de gènes isoactifs d'une même classe d'homologues sont allèles. Par extension deux gènes homologues appartenant à des classes de gènes isoactifs distinctes sont allèles.

5. L'allélisme des gènes.

Deux gènes sont allèles ssi d'une part ils sont homologues et d'autre part ils ont une action différente sur un même caractère.

Cette relation n'est pas une relation d'équivalence.

Elle n'est pas réflexive :

$$\forall x \in C_{ij} : \nexists x\mathcal{A}x.$$

Elle n'est pas transitive :

$$\forall (x, y, z) \in C_{ij} : x\mathcal{A}y, y\mathcal{A}z \not\Rightarrow x\mathcal{A}z.$$

Elle est symétrique :

$$\forall (x, y) \in C_{ij} : x\mathcal{A}y \Leftrightarrow y\mathcal{A}x.$$

6. L'identité des gènes.

Nous introduisons l'identité entre deux gènes par la définition suivante : *Deux gènes sont identiques ssi ils dérivent par descendance mendélienne et sans mutation d'un même gène ancêtre.* Il s'agit d'une relation binaire du symbole I qui possède les propriétés suivantes :

Elle est réflexive :

$$\forall x \in \mathcal{G} : xIx.$$

Elle est transitive :

$$\forall (x, y, z) \in \mathcal{G} : xIy, yIz \Rightarrow xIz.$$

Elle est symétrique :

$$\forall (x, y) \in \mathcal{G} : xIy \Leftrightarrow yIx.$$

La relation d'identité est donc une relation d'équivalence. En conséquence, elle réalise une partition de l'ensemble des gènes du génotype général \mathcal{G} en classes disjointes de gènes identiques.

Si nous supposons que deux gènes identiques ont des effets identiques, il est immédiat que deux gènes identiques sont isoactifs, mais la réciproque n'est pas vraie, deux gènes isoactifs ne sont pas nécessairement identiques. Deux gènes identiques sont homologues, mais deux gènes homologues ne sont pas nécessairement identiques.

7. Le groupe de linkage ou groupe de dépendance.

Soit \mathcal{C} l'ensemble des classes d'homologues du génotype général \mathcal{G} . \mathcal{G} étant un ensemble fini, \mathcal{C} est un ensemble fini. *Deux classes d'homologues appartiennent au même groupe de linkage ssi au cours de la méiose elles sont dépendantes ou bien si elles sont chacune dépendantes d'au moins une classe d'homologues d'un ensemble de classes d'homologues appartenant au même groupe de linkage.* Il s'agit d'une relation binaire, de symbole \mathcal{D} , dont les propriétés sont les suivantes :

Elle est réflexive :

$$\forall x \in \mathcal{C} : x\mathcal{D}x.$$

Elle est transitive :

$$\forall (x, y, z) \in \mathcal{C} : x\mathcal{D}y, y\mathcal{D}z \Rightarrow x\mathcal{D}z.$$

Elle est symétrique :

$$\forall (x, y) \in \mathcal{C} : x\mathcal{D}y \Leftrightarrow y\mathcal{D}x.$$

La relation de dépendance est donc une relation d'équivalence. En conséquence, elle réalise une partition de l'ensemble C en sous-ensembles disjoints D_i appelés *groupe de linkage* ou *groupe de dépendance*, dont l'ensemble a pour symbole \mathcal{D} . Le nombre de groupes de dépendance D_i est petit, et l'expérience montre qu'il est fixe pour chaque espèce. D_i est un *groupe de dépendance*, ensemble de classes d'homologues liées que nous symbolisons par :

$$D_i = \{ C_{i1}, C_{i2}, C_{i3}, C_{ij} \dots C_{im} \} \quad \text{avec} \quad 1 \leq j \leq m$$

m étant le nombre maximum de classes d'homologues que comprend ce groupe de dépendance.

Le symbolisme d'une classe d'homologues est :

$$C_{ij} = \{ x_{ij1}, x_{ij2}, \dots, x_{ijk}, \dots, x_{ijn} \} \quad \text{avec} \quad 1 \leq k \leq n$$

$x \in \mathcal{G}$ étant un gène, n étant le nombre maximum de gènes de cette classe d'homologues, si la population comporte N zygotes, ce nombre est de $2N$.

On appelle *groupe de dépendance singulier* de symbole D_i^* , tout groupe de dépendance réalisé dans un gamète. Donc la connaissance du génotype individuel d'un gamète définit D_i^* pour tout i , ce qui assure une description correcte du génotype individuel de ce gamète. Par contre, la connaissance du génotype individuel d'un zygote permet de décrire $2^m D_i^*$, m limite supérieure de i , parmi lesquels nous ne retiendrons comme description du génotype individuel du zygote que les deux D_i^* présents dans les gamètes dont la fusion a créé ce zygote. Nous appelons groupes de dépendance *homologues* ⁽⁶⁾ tous les groupes singuliers D_i^* pour i fixe. Ainsi tout zygote possède $2h$ groupes de dépendance singuliers qui se répartissent en h paires de groupes de dépendance singuliers homologues.

Ce sont dans chaque individu ces groupes de dépendance singuliers D_i^* que les biologistes, suivant Darlington, associent aux chromosomes présents dans cet individu. Inversement, un chromosome d'une espèce doit être associé à un groupe de dépendance D_i . Cette différence essentielle entre les concepts analytiques et synthétiques qu'utilisent les généticiens est rarement soulignée.

Nous illustrerons cette remarque par un exemple. Considérons quatre individus haploïdes w, x, y, z porteurs chacun de deux gènes. Ces quatre individus forment un groupe. Le génotype général de ce groupe est la connais-

⁽⁶⁾ Il s'agit d'un emploi nouveau de l'adjectif « *homologue* » dont l'usage est accepté des généticiens.

sance des huit gènes que nous distinguons les uns des autres en les numérotant de 1 à 8. Les renseignements biologiques que nous possédons sont les suivants. Ils sont codifiés dans le tableau ci-dessous :

	1	2	3	4	5	6	7	8
1	\mathcal{H} $\mathcal{I}s$ $\mathcal{I}d$		\mathcal{H} \mathcal{A}		\mathcal{H} $\mathcal{I}s$		\mathcal{H} \mathcal{A}	
2		\mathcal{H} $\mathcal{I}s$ $\mathcal{I}d$		\mathcal{H} $\mathcal{I}s$ $\mathcal{I}d$		\mathcal{H} $\mathcal{I}s$		\mathcal{H} \mathcal{A}
3			\mathcal{H} $\mathcal{I}s$ $\mathcal{I}d$		\mathcal{H} \mathcal{A}		\mathcal{H} \mathcal{A}	
4				\mathcal{H} $\mathcal{I}s$ $\mathcal{I}d$		\mathcal{H} $\mathcal{I}s$		\mathcal{H} \mathcal{A}
5					\mathcal{H} $\mathcal{I}s$ $\mathcal{I}d$		\mathcal{H} \mathcal{A}	
6						\mathcal{H} $\mathcal{I}s$ $\mathcal{I}d$		\mathcal{H} \mathcal{A}
7							\mathcal{H} $\mathcal{I}s$ $\mathcal{I}d$	
8								\mathcal{H} $\mathcal{I}s$ $\mathcal{I}d$

\mathcal{H} : relation d'Homologie.

$\mathcal{I}s$: relation d'Isoaction.

\mathcal{A} : relation d'Allélisme.

$\mathcal{I}d$: relation d'Identité.

Nous aboutissons à la partition suivante :

1^{re} classe d'Homologie, { 1, 3, 5, 7 } dans laquelle nous avons les trois sous-classes d'isoaction { (1,5) (3) (7) }.

2^e classe d'Homologie { (2, 4, 6, 8) } dans laquelle nous avons les deux sous-classes d'isoaction { (2, 4, 6) (8) }.

Dans l'ensemble d'isoaction (2, 4, 6) nous avons une partition en 2 sous-ensembles d'identité ([2, 4] [6]).

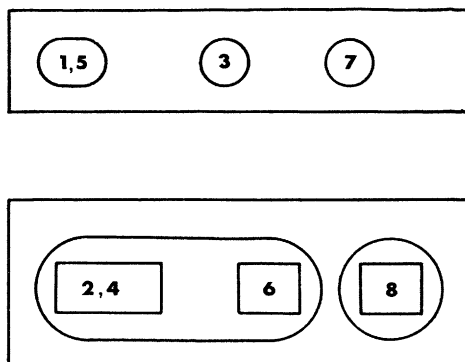


FIG. 3.

Prenons comme représentant de la première classe d'homologie les lettres A telles que :

$$\begin{aligned} 1 &\rightarrow \bar{A}_1 \\ 3 &\rightarrow A_2 \\ 5 &\rightarrow A_1 \\ 7 &\rightarrow A_3 \end{aligned}$$

$[A_1 A_2 A_3]$ est une série allélique.

Prenons comme représentant de la seconde classe d'homologie les lettres B telles que :

$$\begin{aligned} 2 &\rightarrow \bar{B}_1 \\ 4 &\rightarrow \bar{B}_1 \\ 6 &\rightarrow B_1 \\ 8 &\rightarrow B_2 \end{aligned}$$

$[B_1 B_2]$ est une série allélique d'où le schéma suivant des génotypes singuliers, sachant que les classes d'homologues sont liées :

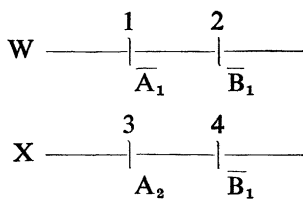
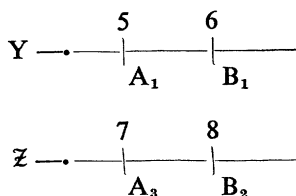
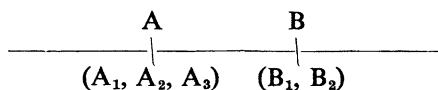


FIG. 4.



Si nous désirons résumer les renseignements génétiques recueillis pour le groupe d'individus W, X, Y, Z nous construirons la carte génétique suivante :



8. La carte factorielle.

Soient C_{ij} et $C_{ij'} \in D_i$, nous postulons qu'il existe une *distance* (?) entre C_{ij} et $C_{ij'}$, fonction numérique $d(C_{ij} \text{ à } C_{ij'})$, définie sur D_i^2 telle que :

- 1 * $d(C_{ij} \text{ à } C_{ij'}) \neq 0 \rightarrow C_{ij} \neq C_{ij'}$
- 2 * $d(C_{ij} \text{ à } C_{ij'}) = 0 \rightarrow C_{ij} = C_{ij'}$
- 3 * $d(C_{ij} \text{ à } C_{ij'}) = -d(C_{ij'} \text{ à } C_{ij})$
- 4 * $d(C_{ij} \text{ à } C_{ij'}) + d(C_{ij'} \text{ à } C_{ij'}) = d(C_{ij} \text{ à } C_{ij'})$.

Ces relations sont nécessaires et suffisantes pour représenter tout groupe de dépendance sur un segment de droite orientée.

Or $d(C_{ij} \text{ à } C_{ij'})$ peut être expérimentalement déterminée par la relation :

$$|d(C_{ij} \text{ à } C_{ij'})| = p$$

p étant la probabilité de recombinaison entre C_{ij} et $C_{ij'}$ si $p \leq \epsilon$; ϵ étant un nombre réel positif et petit. Le signe est choisi arbitrairement pour un premier couple ordonné. Par approche successive, il est possible d'ordonner tous les $C_{ij} \in D_i$.

Il est utile de définir dans D_i un élément $C_{i0} \in D$ appelé origine tel que nous pourrions nommer *carte factorielle* ou *droite chromosomique* du groupe de dépendance D_i l'ensemble des distances $d(C_{i0} \text{ à } C_{ij})$ $C_{i0} \in D_i$, $\forall C_{ij} \in D_i$. Le choix de C_{i0} est arbitraire, aussi choisissons-nous C_{i0} en fonction du critère suivant : « l'origine d'une droite chromosomique est l'élément du

(?) Le substantif *distance* est employé dans son sens commun et non dans son sens mathématique, synonyme de métrique.

groupe de dépendance qui est toujours *prééduit* au cours d'une méiose, cet élément est dénommé centromère, son symbole est C_{i0} ». L'expérience montre que cet élément est unique. Ainsi à toute classe d'homologues $C_{ij} \in D_i$ correspond un seul nombre réel.

9. Les attributs du gène.

Dans le cadre d'une théorie de génétique mathématique, le gène possède trois attributs :

1. Son action, simple ou multiple. Pour un seul caractère, cette action se décrit par une grandeur mesurable.
2. Sa quasi-stabilité de structure ou de nature à laquelle est attachée une probabilité de mutation.
3. Sa quasi-stabilité de position ⁽⁸⁾. La position d'un gène est déterminée par sa distance au centromère, c'est un nombre réel, celui qui est attaché à la classe d'homologues à laquelle il appartient.

10. La notion de locus.

Nous devons formellement considérer le *locus comme le représentant d'une classe d'homologues*, mais l'évolution historique de la pensée génétique en fait implicitement une forme amoindrie du gène, ce que cache la fonction de réceptacle du gène que l'on attribue au locus.

Considérons un élément à qui serait attaché un nombre réel. Un tel élément a les attributs d'un point de la droite orientée. Les généticiens appellent un tel élément un locus (pluriel loci) de symbole L . Il est immédiat qu'à tout locus L_{ij} nous pouvons associer un C_{ij} , c'est-à-dire qu'à un locus ne correspond qu'une seule classe d'homologues d'un groupe de dépendance et réciproquement, d'où l'écriture :

$$L_{ij} \Leftrightarrow C_{ij}.$$

Nous ajouterons comme attribut au locus *que c'est un élément qui accompagne tout gène, qui est stable et se duplique en même temps que le gène au cours de la méiose*. Ce nouveau locus de définition plus riche est appelé

⁽⁸⁾ Nous supposons par la suite que la position du gène est stable, éliminant de notre schéma des exemples comme celui du gène bar de la drosophile qui présente pourtant un gros intérêt dans le cas d'épistasie.

locus singulier ⁽⁹⁾. En un locus singulier ne peut être présent qu'un gène d'une classe d'homologues.

Deux loci singuliers, auxquels sont attachés les mêmes nombres réels, sont dits homologues ssi ils appartiennent au même groupe de dépendance.

Échange Annales

11. L'identité des loci.

Soit \mathcal{L} l'ensemble des loci du génotype général \mathcal{G} . En 1948, G. Malécot a défini l'identité de deux loci : « *Deux loci sont identiques ssi ils dérivent d'un même locus ancêtre par descendance mendélienne* ». Ce même auteur a donné le moyen de calculer la probabilité d'identité de deux loci dans le cas des populations à pedigrees et des populations panmictiques d'effectif limité. Cette relation d'identité de symbole \mathfrak{I} est binaire et possède les propriétés suivantes :

Elle est réflexive :

$$\forall x \in \mathcal{L} : x \mathfrak{I} x.$$

Elle est transitive :

$$\forall (x, y, z) \in \mathcal{L} : x \mathfrak{I} y, y \mathfrak{I} z \Rightarrow x \mathfrak{I} z.$$

Elle est symétrique :

$$\forall (x, y) \in \mathcal{L} : x \mathfrak{I} y \Leftrightarrow y \mathfrak{I} x.$$

Remarquons que ces propriétés ont été obtenues en supposant que le locus est stable, c'est-à-dire que ses attributs ne sont pas susceptibles de variation, en particulier le réel attaché à tout L_{ij} est fixe, indépendant des méioses ou — ce qui revient au même — du numéro de génération. Autrement dit, l'ordre défini sur la droite chromosomique ne peut pas être modifié par la méiose. Nous excluons donc de notre schéma les inversions, les duplications, les échanges réciproques entre groupes de dépendance.

12. Généralisation de la notion de locus.

La notion de locus, telle que nous l'avons introduite, peut parfois être insuffisante. G. Malécot, en 1948, a été le premier à généraliser cette notion lorsqu'il écrit dans les Mathématiques de l'hérédité (p. 9) : « Supposons

⁽⁹⁾ Dans la suite de notre exposé, et pour ne pas rompre avec l'usage courant fautif, nous dirons locus au lieu de locus singulier.

maintenant que les loci considérés puissent être affectés par des mutations de fréquence moyenne u par génération. La probabilité pour qu'un locus d'un individu reproduise sans modification le locus parental dont il dérive est alors $1 - u$ ».

Aussi nous attachons au locus deux attributs : l'un la position définie par rapport au centromère par un nombre réel, l'autre la stabilité définie par l'absence ou non de mutation. Aussi dans tous les cas l'identité des loci entraîne-t-elle celle des gènes.

Nous définissons ainsi la mutation d'un locus : « *Un locus mute ssi le gène qui l'accompagne mute.* »

La probabilité de cet événement est égale à la probabilité moyenne de mutation d'un gène quelconque de la classe d'homologues correspondant à ce locus. C'est cette définition plus riche du locus que nous utilisons dans l'étude des populations à loi de croisement.

IV. — LES DIFFÉRENTS ASPECTS DE LA RELATION D'IDENTITÉ

A la suite d'un cheminement qui va de 1941 à 1948, G. Malécot définit le *coefficient de consanguinité* comme la probabilité que deux loci homologues d'un zygote I soient identiques par descendance mendélienne. Soit f cette probabilité. Cet auteur étend la notion de coefficient de consanguinité par celle de *coefficient de parenté* qui est la probabilité que deux loci homologues pris au hasard, l'un parmi les deux loci homologues d'un zygote I, l'autre parmi les deux loci homologues d'un zygote J, soient identiques par descendance mendélienne.

Nous nous proposons de généraliser ces notions par celle de *coefficient d'identité* et d'utiliser ensuite ces nouveaux concepts introduits pour mieux décrire une population tant pour des caractères mendéliens qualitatifs que quantitatifs.

— La première étape considère simultanément n gènes homologues et non plus deux.

— La seconde étape est l'analyse de plusieurs classes d'homologues distinctes telles que dans chacune d'elles le nombre des gènes homologues est quelconque. Une telle étude permettra d'étudier la dépendance susceptible d'exister entre ces classes de gènes homologues.

— La troisième étape passe du domaine chromosomique discontinu au domaine chromosomique continu. Pour cela nous considérons sur le chro-

mosome des segments sur lesquels existe un nombre quelconque, mais fixe, de classes d'homologues. Le nombre des segments chromosomiques homologues considérés peut être quelconque. La dernière étape de notre généralisation a pour but de lever l'hypothèse restrictive de segments fixés *a priori*. Nous rejoignons en ce point les travaux que R. A. Fisher développe dans *The theory of inbreeding* de 1949.

Les différents niveaux de généralisation.

La transitivité de l'identité de deux gènes entraîne que n loci homologues peuvent être identiques parce qu'ils dérivent d'un même gène ancêtre. Nous prendrons l'énoncé suivant comme définition de l'identité : « *Des gènes homologues, en nombre quelconque, sont identiques ssi ils dérivent tous d'un même gène ancêtre* ».

La relation binaire d'identité I réalise une partition d'un ensemble \mathcal{G} de n gènes ⁽¹⁰⁾ en sous-ensembles disjoints de gènes singuliers. Deux gènes pris au hasard dans deux sous-ensembles distincts ne dérivent pas d'un ancêtre commun, donc il n'y a aucune relation matérielle entre ces gènes, nous dirons qu'ils sont indépendants ⁽¹¹⁾. La connaissance de la nature de l'un ne modifie pas la connaissance de la nature de l'autre. Deux gènes pris au hasard dans un même sous-ensemble sont identiques car ils sont tous les deux copies d'un même modèle : le gène ancêtre. La connaissance de la nature de l'un nous donne la nature de l'autre.

Tout sous-ensemble de \mathcal{G} , quel qu'il soit, est une partie de \mathcal{G} , nous le désignons par \mathcal{F}_i . Il y a 2^n parties de \mathcal{G} qui constituent un ensemble \mathcal{P} . Nous pouvons décrire toute partition de l'ensemble \mathcal{G} comme l'ensemble des \mathcal{F}_i qui répondent aux critères suivants :

- 1) $\mathcal{F}_i \cap \mathcal{F}_j = \emptyset$
- 2) $\bigcup_i \mathcal{F}_i = \mathcal{G}.$

Toute partition fixe de façon totale les relations d'identité ou d'indépendance qui sont susceptibles d'exister entre les gènes de l'ensemble.

Toute partition d'un ensemble \mathcal{G} de gènes homologues, suivant la relation binaire d'équivalence qu'est la relation d'identité entre deux gènes, est par définition une *situation d'identité* restreinte. Toute probabilité attachée à une situation d'identité est un *coefficient d'identité restreinte*. Inversement,

⁽¹⁰⁾ Ces n gènes sont singuliers.

⁽¹¹⁾ C'est une restriction importante que de supposer l'existence d'une double implication entre la non-identité et l'indépendance.

une situation d'identité restreinte définie sur un ensemble \mathcal{G} de n gènes homologues est une partition de cet ensemble en sous-ensembles disjoints, tels que deux gènes quelconques pris chacun dans l'un de deux sous-ensembles quelconques sont toujours indépendants, tels que tout sous-ensemble ne compte qu'un gène ou plusieurs gènes tous identiques entre eux.

Si $\{x_i, x_j, x_k, x_l, x_m\} \in \mathcal{G} (1 \leq i, j, k, l, m \leq n \text{ quand l'ensemble } \mathcal{G} \text{ compte } n \text{ gènes homologues})$, toute situation d'identité restreinte est de la forme :

$$\mathfrak{S} \{ (x_i = x_j = \dots) \circ (x_k = x_l = \dots) \circ (x_m = \dots) \}.$$

Deux cas limites sont à concevoir : celui pour lequel tous les gènes homologues sont identiques :

$$\mathfrak{S} \{ (x_i = x_j = x_k = x_l = x_m = \dots) \}$$

et celui pour lequel tous les gènes homologues sont indépendants :

$$\mathfrak{S} \{ (x_i) \circ (x_j) \circ (x_k) \circ (x_l) \circ (x_m) \circ (\dots) \}.$$

Pour un ensemble \mathcal{G} de n gènes homologues, il existe un nombre x de situations d'identité restreinte tel que pour $n = 2$, $x = 2$; $n = 3$, $x = 5$; $n = 4$, $x = 15$.

Soit i un nombre entier tel que $1 \leq i \leq x$, soit \mathfrak{S} le symbole d'une situation d'identité. Nous pouvons désigner chaque situation d'identité restreinte par \mathfrak{S}_i . $\text{Pr} \{ \mathfrak{S}_i \}$ est la probabilité attachée à la réalisation de la situation \mathfrak{S}_i , d'où :

$$\sum_{i=1}^{i=x} \text{Pr} \{ \mathfrak{S}_i \} = 1.$$

Nous généralisons les situations d'identité restreinte de la manière suivante. Considérons deux ensembles non homologues de gènes homologues tels que chacun possède n gènes. Ces deux ensembles peuvent avoir un nombre de gènes différents.

Le premier ensemble permet de définir une série de x situations d'identité restreinte de la forme \mathfrak{S}_i^1 , et le second ensemble permet de définir une autre série de x situations d'identité de la forme \mathfrak{S}_j^2 .

Par définition nous appelons *situation d'identité généralisée* la situation $\mathfrak{S}_g = \mathfrak{S}_i^1 \cap \mathfrak{S}_j^2$.

Toute probabilité attachée à une telle situation d'identité généralisée est à un *coefficient d'identité généralisée*. Il est immédiat que :

$$\begin{aligned}\Pr \{ \mathfrak{S}_i^1 \cap \mathfrak{S}_j^2 \} &= \Pr \{ \mathfrak{S}_i^1 \} * \Pr \{ \mathfrak{S}_j^2 \} / \text{cond. } \mathfrak{S}_i^1 \\ &= \Pr \{ \mathfrak{S}_i^2 \} * \Pr \{ \mathfrak{S}_i^1 \} / \text{cond. } \mathfrak{S}_j^2\end{aligned}$$

l'écriture $\Pr \{ \mathfrak{S}_j^2 \} / \text{cond. } \mathfrak{S}_i^1$ ayant le sens suivant : « probabilité de réalisation de la situation d'identité restreinte \mathfrak{S}_j^2 sachant que la situation d'identité restreinte \mathfrak{S}_i^1 a été réalisée, il s'agit d'une probabilité conditionnée ».

Dans le cas où les deux ensembles de gènes seraient indépendants :

$$\Pr \{ \mathfrak{S}_i^1 \cap \mathfrak{S}_j^2 \} = \Pr \{ \mathfrak{S}_i^1 \} * \Pr \{ \mathfrak{S}_j^2 \}.$$

La généralisation à un nombre quelconque d'ensembles distincts de gènes singuliers ne présente pas de difficulté, toute situation d'identité généralisée étant de la forme :

$$\mathfrak{S}_i^1 \cap \mathfrak{S}_j^2 \cap \mathfrak{S}_k^3 \dots \cap \mathfrak{S}_p^m = \bigcap_{i=1}^m \mathfrak{S}_i^m.$$

La probabilité $\Pr \{ \bigcap_{i=1}^m \mathfrak{S}_i^m \}$ est égale à :

$$\begin{aligned}&= \Pr \{ \mathfrak{S}_i^1 \} * \Pr \{ \mathfrak{S}_j^2 \} / \text{cond. } \mathfrak{S}_i^1 * \dots * \Pr \{ \mathfrak{S}_p^m \} / \text{cond. } \bigcap_{i=1}^{m-1} \mathfrak{S}_i \\ &= \prod_{m=1}^{m=k} \Pr \{ \mathfrak{S}_i^m \} / \text{cond.}\end{aligned}$$

Le terme de rang m étant la probabilité conditionnée de la situation d'identité restreinte de l'ensemble des gènes homologues de rang m , conditionnée par la réalisation des situations d'identité restreinte de rang inférieur à m . Si tous les ensembles de gènes homologues sont indépendants deux à deux, cette expression de $\Pr \{ \bigcap_{i=1}^m \mathfrak{S}_i \}$ devient un produit de coefficient d'identité restreinte non conditionnée :

$$\Pr \{ \bigcap_{i=1}^m \mathfrak{S}_i \} = \prod_{m=1}^{m=k} \Pr \{ \mathfrak{S}_i^m \}.$$

Considérons le cas où tous ces ensembles de gènes singuliers appartiennent au même groupe de dépendance. Parmi tous les ordres possibles qui rangent les ensembles de gènes homologues il en existe un dont les pro-

priétés sont remarquables : c'est celui de la droite chromosomique complète, car les gènes se répartissent à droite et à gauche d'une origine le centromère, leur ordre étant le même que celui de leur distance au centromère. Toute carte factorielle complète est de la forme :

$$\begin{array}{cccccccccc}
 -k & -3 & -2 & -1 & 0 & 1 & 2 & 3 & k \\
 G & G & G & G & C & G & G & G & G \\
 \hline
 \circ & \circ & \circ & \circ & \circ & \circ & \circ & \circ & \circ
 \end{array}$$

$$\mathfrak{S} = \left\{ \bigcap_{-k \text{ à } -1} \mathfrak{S}_i \right\} \bigcap \mathfrak{S}_i^0 \bigcap \left\{ \bigcap_{1 \text{ à } k} \mathfrak{S}_i \right\} = \left\{ \bigcap_{-k \text{ à } -1} \mathfrak{S}_i \right\} \bigcap \left\{ \bigcap_{0 \text{ à } k} \mathfrak{S}_i \right\}$$

avec :

$$\Pr \{ \mathfrak{S} \} = \Pr \left\{ \bigcap_{0 \text{ à } k} \mathfrak{S}_i \right\} * \Pr \left\{ \bigcap_{-k \text{ à } -1} \mathfrak{S}_i \right\} / \text{cond.} \left\{ \bigcap_{0 \text{ à } k} \mathfrak{S}_i \right\}.$$

Il apparaît utile d'attacher de façon préférentielle au centromère les situations d'identité restreinte que nous avons précédemment définies, bien que cette règle ne soit pas toujours la plus efficace.

Considérons comme objets entre lesquels nous définissons une relation d'identité non pas des gènes, c'est-à-dire des points de la droite chromosomique, mais des segments de cette droite.

Deux segments sont dits homologues s'ils appartiennent au même groupe de linkage et s'ils ont mêmes extrémités.

Deux segments sont identiques si tous leurs gènes sont identiques.

Un segment chromosomique n'est donc finalement qu'une suite de gènes.

Un segment est défini par les coordonnées de ses extrémités qui peuvent être fixées *a priori* ou en fonction de divers paramètres. Dans le premier schéma, la longueur du segment est une constante; dans le second schéma, c'est une variable définie par sa fonction. Il est immédiat que les notions de situations d'identité et de coefficients d'identité s'appliquent aux segments chromosomiques, ainsi que les symboles utilisés. Nous appelons *situation d'identité chromosomique fixe* toute situation d'identité décrite entre n segments chromosomiques fixes, et *situation d'identité chromosomique variable* toute situation d'identité décrite entre n segments chromosomiques dont les extrémités sont variables.

V. — LES SITUATIONS ET LES COEFFICIENTS D'IDENTITÉ RESTREINTE

Nous avons, jusqu'à présent, considéré des gènes singuliers homologues ou non, sans jamais définir leur appartenance à une série de gamètes ou de zygotes. Dans ce chapitre nous nous intéressons à des situations d'identité

restreinte susceptibles d'exister entre des gènes homologues répartis dans un certain nombre de zygotes. Nous appelons *ordre* d'une situation d'identité restreinte ou d'un coefficient d'identité restreinte le nombre de gènes homologues considérés, et *rang* le nombre de zygotes entre lesquels se distribuent ces gènes. Si l'ordre est n , et le rang m , nous avons la double inégalité $\frac{n}{2} \leq m \leq n$. Remarquons que la limite inférieure de n est 2 puisque, pour qu'il y ait identité, il faut au moins deux gènes homologues, et que celle de m est 1.

Soient $n = 2$ et $m = 1$. Nous considérons les deux gènes homologues d'un même zygote, et définissons les situations et les coefficients d'identité restreinte d'ordre 2, de rang 1.

Soient $n = 2$ et $m = 2$. Nous considérons deux gènes homologues dont l'un est tiré au hasard parmi les deux gènes homologues d'un premier zygote et dont l'autre est tiré au hasard parmi les deux gènes homologues du second zygote. Entre ces deux gènes homologues nous décrivons et nous définissons des situations et des coefficients d'identité restreinte d'ordre 2, de rang 2.

Soient $n = 3$ et $m = 2$. Nous considérons trois gènes homologues tirés au hasard parmi les quatre gènes homologues de deux zygotes. Autrement dit, nous nous intéressons aux deux gènes homologues d'un zygote et à un gène pris au hasard parmi les deux gènes homologues d'un second zygote. Entre ces trois gènes homologues nous décrivons et nous définissons des situations et des coefficients d'identité restreinte d'ordre 3, de rang 2.

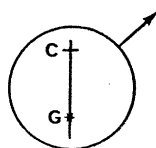
Soient $n = 3$ et $m = 3$. Nous considérons trois gènes homologues tirés au hasard chacun parmi les deux gènes homologues de l'un des trois zygotes considérés. Nous décrivons et définissons entre ces trois gènes homologues des situations et des coefficients d'identité restreinte d'ordre 3, de rang 3.

Soient $n = 4$ et $m = 2$. Nous considérons les quatre gènes homologues appartenant à deux zygotes. Nous décrivons et nous définissons entre ces quatre gènes homologues des situations et des coefficients d'identité restreinte d'ordre 4, et de rang 2. Remarquons que ce sont ces situations et ces coefficients d'identité restreinte qui permettront l'expression de la covariance entre apparentés quelconques.

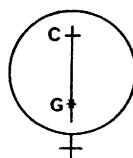
Soient $n = 4$ et $m = 3$. Nous considérons quatre gènes homologues dont deux appartiennent à un premier zygote et dont les deux autres gènes sont tirés au hasard chacun parmi les deux gènes homologues de l'un des deux zygotes restant. Nous décrivons et nous définissons entre ces quatre gènes homologues des situations et des coefficients d'identité restreinte d'ordre 4, de rang 3.

Soient $n = 4$ et $m = 4$. Nous considérons quatre gènes homologues chacun tiré au hasard parmi les deux gènes homologues de l'un des quatre zygotes. Nous décrivons et définissons entre ces quatre gènes homologues des situations et des coefficients d'identité restreinte d'ordre 4, de rang 4.

Nous représentons un gamète, un spermatozoïde ou un ovule, par les schémas ci-dessous, dans lesquels les groupes de dépendance singuliers D^* sont sous la forme d'un segment de droite, en y repérant le centromère C et un gène singulier G.



Gamète mâle



Gamète femelle

FIG. 5.

C'est d'une manière similaire que nous schématisons un zygote I en ajoutant, comme convention, que tout élément non indicé par une étoile est d'origine paternelle et que tout élément indicé par une étoile est d'origine maternelle.

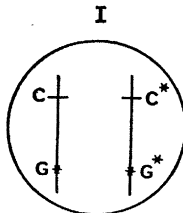


FIG. 6.

Nous symbolisons, comme précédemment, par \mathfrak{S} (s gothique majuscule) les situations d'identité; par le signe \equiv une relation d'identité; par le signe \circ une relation d'indépendance, c'est-à-dire l'exclusion de la relation d'identité; et par le signe $?$ la non-définition de l'une de ces deux relations.

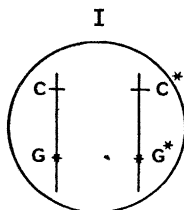


FIG. 7.

Les situations et les coefficients d'identité restreinte d'ordre 2, de rang 1.

Nous considérons les deux gènes homologues G_I et G_I^* d'un zygote I. Ces deux gènes peuvent être identiques ou indépendants.

Les deux situations d'identité restreinte possibles sont :

$$\mathfrak{S} \{ (G_I \equiv G_I^*) \} \quad \text{et} \quad \mathfrak{S} \{ (G_I) \circ (G_I^*) \}.$$

L'expression $\mathfrak{S} \{ (G_I \equiv G_I^*) \}$ se lit « situation pour laquelle les gènes G_I et G_I^* sont identiques »; l'expression $\mathfrak{S} \{ (G_I) \circ (G_I^*) \}$ se lit « situation pour laquelle les gènes G_I et G_I^* sont indépendants »; l'expression $\mathfrak{S} \{ (G_I ? G_I^*) \}$ se lit « situation pour laquelle les gènes G_I et G_I^* sont ou identiques ou indépendants », nous avons les deux relations suivantes :

$$\begin{aligned} \mathfrak{S} \{ (G_I \equiv G_I^*) \} \cap \mathfrak{S} \{ (G_I) \circ (G_I^*) \} &= \emptyset \\ \mathfrak{S} \{ (G_I \equiv G_I^*) \} \cup \mathfrak{S} \{ (G_I) \circ (G_I^*) \} &= \mathfrak{S} \{ (G_I ? G_I^*) \}. \end{aligned}$$

Toute probabilité attachée à une situation d'identité restreinte est un coefficient d'identité restreinte. On reconnaît aisément le coefficient de consanguinité f dans l'expression du coefficient d'identité restreinte suivante :

$$\Pr \{ (G_I \equiv G_I^*) \} = f_I$$

Nous avons aussi :

$$\Pr \{ (G_I) \circ (G_I^*) \} = 1 - f_I.$$

$$\Pr \{ (G_I ? G_I^*) \} = 1.$$

Pour des raisons d'homogénéité par rapport à une écriture symbolique que nous introduisons par la suite, nous posons :

$$\begin{aligned} \mathfrak{S} \{ (G_I \equiv G_I^*) \} &= \mathfrak{S}_{(21)}^{(2)} \quad \text{et} \quad \Pr \{ (G_I \equiv G_I^*) \} = \mathcal{P}_{(21)}^{(2)} \\ \mathfrak{S} \{ (G_I) \circ (G_I^*) \} &= \mathfrak{S}_{(21)}^{(0)} \quad \text{et} \quad \Pr \{ (G_I) \circ (G_I^*) \} = \mathcal{P}_{(21)}^{(0)}. \end{aligned}$$

Les situations et les coefficients d'identité restreinte d'ordre 2, de rang 2.

Soit I l'un des zygotes, soit J l'autre. Nous sommes en présence de quatre gènes homologues G_I , G_I^* , G_J , G_J^* . Nous désignons par $T(\overset{v}{G_I} \overset{v}{G_J})$ le tirage de deux gènes homologues, l'un $\overset{v}{G_I}$ parmi les deux gènes homologues G_I et G_I^* de I, l'autre $\overset{v}{G_J}$ parmi les deux gènes homologues G_J et G_J^* de J. Nous avons quatre tirages possibles dont chacun a pour probabilité de réalisation $\frac{1}{4}$, ce sont :

$$T(G_I G_J); \quad T(G_I G_J^*); \quad T(G_I^* G_J); \quad T(G_I^* G_J^*).$$

Supposons le tirage $T(\overset{v}{G}_I \overset{v}{G}_J)$ réalisé, alors nous avons les situations d'identité restreinte suivantes :

$$\mathfrak{S} \{ (\overset{v}{G}_I \equiv \overset{v}{G}_J) \} ; \quad \mathfrak{S} \{ (\overset{v}{G}_I) \circ (\overset{v}{G}_J) \} ; \quad \mathfrak{S} \{ (\overset{v}{G}_I ? \overset{v}{G}_J) \}$$

avec les relations :

$$\begin{aligned} \mathfrak{S} \{ (\overset{v}{G}_I \equiv \overset{v}{G}_J) \} \cap \mathfrak{S} \{ (\overset{v}{G}_I) \circ (\overset{v}{G}_J) \} &= \emptyset \\ \mathfrak{S} \{ (\overset{v}{G}_I \equiv \overset{v}{G}_J) \} \cap \mathfrak{S} \{ (\overset{v}{G}_I) \circ (\overset{v}{G}_J) \} &= \mathfrak{S} \{ (\overset{v}{G}_I ? \overset{v}{G}_J) \}. \end{aligned}$$

La généralisation aux autres tirages ne présente aucune difficulté :

$$T(\overset{v}{G}_I \overset{v}{G}_J) = T(G_I G_J) \cup T(G_I G_J^*) \cup T(G_I^* G_J) \cup T(G_I^* G_J^*)$$

d'où :

$$\begin{aligned} \mathfrak{S} \{ (\overset{v}{G}_I \equiv \overset{v}{G}_J) \} &= \mathfrak{S} \{ (G_I \equiv G_J) \} \cup \mathfrak{S} \{ (G_I \equiv G_J^*) \} \cup \mathfrak{S} \{ (G_I^* \equiv G_J) \} \\ &\quad \cup \mathfrak{S} \{ (G_I^* \equiv G_J^*) \}. \end{aligned}$$

Nous avons encore :

$$\begin{aligned} \Pr \{ (\overset{v}{G}_I \equiv \overset{v}{G}_J) \} &= \frac{1}{4} \Pr \{ (G_I \equiv G_J) \} + \frac{1}{4} \Pr \{ (G_I \equiv G_J^*) \} + \frac{1}{4} \Pr \{ (G_I^* \equiv G_J) \} \\ &\quad + \frac{1}{4} \Pr \{ (G_I^* \equiv G_J^*) \} \end{aligned}$$

$\Pr \{ (\overset{v}{G}_I \equiv \overset{v}{G}_J) \}$ est le coefficient de parenté φ_{II} de G. Malécot.

Nous avons d'autre part $\Pr \{ (\overset{v}{G}_I) \circ (\overset{v}{G}_J) \} = 1 - \varphi_{II}$.

Pour des raisons d'homogénéité, par rapport à une écriture symbolique que nous introduirons par la suite, nous posons :

$$\begin{aligned} \mathfrak{S} \{ (\overset{v}{G}_I \equiv \overset{v}{G}_J) \} &= \mathfrak{S} \binom{2}{22} \quad \text{et} \quad \Pr \{ (\overset{v}{G}_I \equiv \overset{v}{G}_J) \} = \mathcal{P} \binom{2}{22} \\ \mathfrak{S} \{ (\overset{v}{G}_I) \circ (\overset{v}{G}_J) \} &= \mathfrak{S} \binom{0}{22} \quad \text{et} \quad \Pr \{ (\overset{v}{G}_I) \circ (\overset{v}{G}_J) \} = \mathcal{P} \binom{0}{22}. \end{aligned}$$

Les situations et les coefficients d'identité restreinte d'ordre 3, de rang 2.

Il s'agit de situations d'identité restreinte susceptibles d'exister entre trois gènes homologues appartenant à deux zygotes I et J qui en possèdent quatre. Il est immédiat que l'un des zygotes a fourni deux gènes homologues tandis que l'autre n'a fourni qu'un seul de ses deux gènes homologues.

Désignons par I le zygote représenté par ses deux gènes homologues et par J celui représenté par un seul de ses deux gènes homologues.

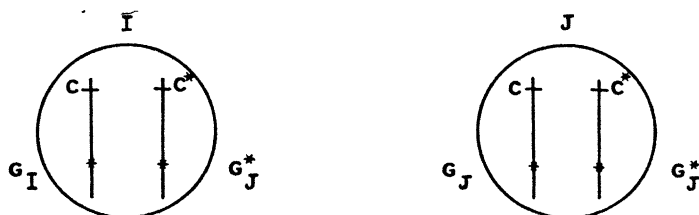


FIG. 8.

Nous pouvons réaliser deux tirages distincts, chacun de probabilité égale à $\frac{1}{2}$:

$$T(G_I, G_I^*, G_J) \quad \text{et} \quad T(G_I, G_I^*, G_J^*)$$

avec :

$$T(G_I, G_I^*, G_J) \cup T(G_I, G_I^*, G_J^*) = T(G_I, G_I^*, G_J^v).$$

Dans le cas de ces deux tirages, les différentes situations d'identité restreinte sont les suivantes :

Tirages	Situations d'identité restreinte	Probabilités
$T(G_I, G_I^*, G_J)$	$\{ (G_I \equiv G_I^* \equiv G_J) \}$	non désignée
	$\{ (G_I \equiv G_I^*) \circ (G_J) \}$	π_1
	$\{ (G_I \equiv G_J) \circ (G_I^*) \}$	π_2
	$\{ (G_I^* \equiv G_J) \circ (G_I) \}$	π_3
	$\{ (G_I) \circ (G_I^*) \circ (G_J) \}$	non désignée
$T(G_I, G_I, G_J)$	$\{ (G_I \equiv G_I^* \equiv G_J^*) \}$	non désignée
	$\{ (G_I \equiv G_I^*) \circ (G_J^*) \}$	π_4
	$\{ (G_I \equiv G_J^*) \circ (G_I^*) \}$	π_5
	$\{ (G_I^* \equiv G_J^*) \circ (G_I) \}$	π_6
	$\{ (G_I) \circ (G_I^*) \circ (G_J^*) \}$	non désignée

Considérons les situations d'identité restreinte :

$$\mathfrak{S} \{ (G_I = G_I^* = G_J) \} \text{ et } \mathfrak{S} \{ (G_I = G_I^* = G_J^*) \},$$

nous avons :

$$\mathfrak{S} \{ (G_I = G_I^* = G_J) \} \cup \mathfrak{S} \{ (G_I = G_I^* = G_J^*) \} = \mathfrak{S} \{ G_I = G_I^* = \overset{v}{G}_J \}.$$

Cette nouvelle situation est celle pour laquelle trois gènes homologues pris au hasard dans deux zygotes sont identiques. Nous écrivons plus brièvement :

$$\mathfrak{S} \{ (G_I = G_I^* = \overset{v}{G}_J) \} = \mathfrak{S}_{(32)}^3.$$

Nous attachons à une telle situation d'identité restreinte une probabilité ou coefficient d'identité restreinte de symbole $\mathcal{P}_{(32)}^3$.

Nous pouvons créer de même $\mathfrak{S}_{(32)}^2$ situation d'identité restreinte pour laquelle, parmi trois gènes homologues tirés au hasard dans deux zygotes, deux gènes sont identiques et le troisième gène indépendant des deux premiers. Il est tout aussi aisé de créer $\mathfrak{S}_{(32)}^0$ situation d'identité restreinte pour laquelle trois gènes homologues tirés au hasard dans deux zygotes sont indépendants entre eux. Toute probabilité attachée à ces situations d'identité restreinte est un coefficient d'identité restreinte. Nous avons les relations suivantes :

$$\mathcal{P}_{(32)}^3 + \mathcal{P}_{(32)}^2 + \mathcal{P}_{(32)}^0 = 1$$

$$\frac{1}{2}(\pi_1 + \pi_2 + \pi_3 + \pi_4 + \pi_5 + \pi_6) = \mathcal{P}_{(32)}^3.$$

Nous introduisons aussi deux autres coefficients d'identité restreinte :

$$\omega_1 \quad \text{et} \quad \omega_2.$$

ω_1 est la probabilité que le gène pris au hasard dans un zygote J soit indépendant de deux gènes identiques appartenant à un zygote I.

ω_2 est la probabilité que l'un des gènes homologues d'un zygote I soit indépendant de deux gènes identiques dont l'un est son homologue dans I et l'autre tiré au hasard dans le zygote J. D'où $\omega_1 + \omega_2 = \mathcal{P}_{(32)}^2$.

Les situations et les coefficients d'identité restreinte d'ordre 3, de rang 3.

Il s'agit de situations d'identité restreinte susceptibles d'exister entre trois gènes homologues chacun tiré au hasard parmi les deux gènes homologues de l'un des trois zygotes K, L, M. Chacun des zygotes est représenté par un gène.

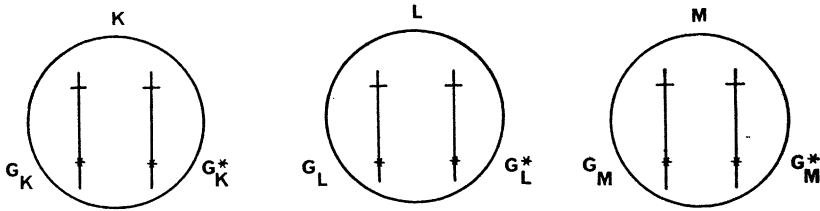


FIG. 9.

Il n'est pas nécessaire de considérer séparément les différents tirages possibles au nombre de $2^3 = 8$. Aussi nous introduisons l'union de tous les tirages sous la forme $T(\overset{v}{G}_K, \overset{v}{G}_L, \overset{v}{G}_M)$. Pour un tel tirage il existe trois situations d'identité restreinte distinctes qui sont :

$$1. \quad \mathfrak{S}(\overset{v}{3}_{33}) = \mathfrak{S} \{ (\overset{v}{G}_K \equiv \overset{v}{G}_L \equiv \overset{v}{G}_M) \},$$

situation d'identité restreinte pour laquelle les trois gènes homologues, chacun tiré au hasard parmi les deux gènes homologues de l'un des trois zygotes K, L, M, sont identiques.

$$2. \quad \mathfrak{S}(\overset{v}{2}_{33}) = \mathfrak{S} \{ (\overset{v}{G}_K \equiv \overset{v}{G}_L) \circ (\overset{v}{G}_M) \} \cup \mathfrak{S} \{ (\overset{v}{G}_K \equiv \overset{v}{G}_M) \circ (\overset{v}{G}_L) \} \\ \cup \mathfrak{S} \{ (\overset{v}{G}_M \equiv \overset{v}{G}_L) \circ (\overset{v}{G}_K) \}$$

situation d'identité restreinte pour laquelle parmi trois gènes homologues, chacun tiré au hasard parmi les deux gènes homologues de l'un des trois zygotes K, L, M, deux gènes sont identiques entre eux et le troisième indépendant des deux premiers.

$$3. \quad \mathfrak{S}(\overset{v}{0}_{33}) = \mathfrak{S} \{ (\overset{v}{G}_K) \circ (\overset{v}{G}_L) \circ (\overset{v}{G}_M) \},$$

situation d'identité restreinte pour laquelle les trois gènes homologues, chacun tiré au hasard parmi les deux gènes homologues de l'un des trois zygotes K, L, M, sont indépendants.

Les probabilités attachées à ces situations sont des coefficients d'identité restreinte, tels que :

$$\Pr \{ \mathfrak{S}(\overset{v}{3}_{33}) \} = \mathcal{P}(\overset{v}{3}_{33})$$

$$\Pr \{ \mathfrak{S}(\overset{v}{2}_{33}) \} = \mathcal{P}(\overset{v}{2}_{33})$$

$$\Pr \{ \mathfrak{S}(\overset{v}{0}_{33}) \} = \mathcal{P}(\overset{v}{0}_{33})$$

avec la relation :

$$\mathcal{P}(\overset{v}{3}_{33}) + \mathcal{P}(\overset{v}{2}_{33}) + \mathcal{P}(\overset{v}{0}_{33}) = 1.$$

Les situations et les coefficients d'identité restreinte d'ordre 4, de rang 2.

Nous nous proposons de décrire les situations d'identité restreinte susceptibles d'exister entre les quatre gènes homologues de deux zygotes I et J.

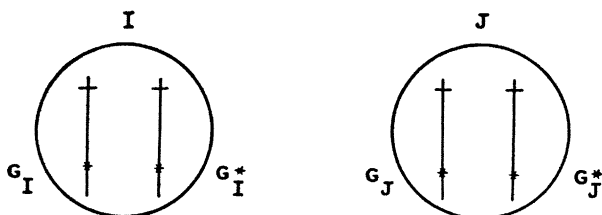


FIG. 10.

Nous avons réuni, dans le tableau ci-après, les situations et les probabilités d'identité restreinte possibles.

\mathfrak{S}_1 est une situation pour laquelle les quatre gènes homologues sont identiques, ce qui est une conséquence du fait que les deux zygotes sont consanguins et qu'ils sont apparentés.

\mathfrak{S}_2 est une situation pour laquelle les gènes homologues G_I , G_I^* , G_J sont identiques et indépendants du gène G_J^* , ce qui est la conséquence du fait que le zygote I est consanguin, que le zygote J est ou n'est pas consanguin, que les deux zygotes sont apparentés.

\mathfrak{S}_6 est une situation pour laquelle les gènes homologues G_I et G_I^* sont identiques, les gènes G_J et G_J^* sont identiques, et indépendants des deux précédents, ce qui est la conséquence du fait que les deux zygotes sont consanguins et sont ou non apparentés.

\mathfrak{S}_{12} est une situation pour laquelle les gènes homologues G_I et G_I^* sont identiques, les gènes homologues G_J et G_J^* sont identiques, et les gènes homologues appartenant à un même zygote indépendants. Les zygotes I et J sont apparentés, mais ils peuvent être ou non consanguins.

Nous pensons avoir montré ainsi toute la richesse de ce nouveau concept de situation d'identité restreinte. Soulignons qu'il ne faut pas confondre la consanguinité et la parenté avec la notion d'identité. Deux zygotes sont parents si et seulement si ils possèdent des ancêtres communs. Un zygote est consanguin si et seulement si son père et sa mère sont parents. Mais ce zygote consanguin peut avoir ces deux gènes homologues identiques ou indépendants ⁽¹²⁾. De façon générale, l'absence de consanguinité ou de

⁽¹²⁾ Nous supposons toujours que la non-identité entraîne l'indépendance, ce qui est une importante restriction que nous lèverons dans un autre article.

parenté entraîne l'indépendance de tous les gènes homologues en cause tandis que l'existence de la consanguinité et de la parenté entraîne l'une quelconque des situations d'identité possibles avec une certaine probabilité.

Situations d'identité restreinte		Probabilités
1	$\mathfrak{S}_1 = \mathfrak{S} \{ (G_I \equiv G_I^* \equiv G_J \equiv G_J^*) \}$	δ_1
2	$\mathfrak{S}_2 = \mathfrak{S} \{ (G_I \equiv G_I^* \equiv G_J) \circ (G_J^*) \}$	δ_2
3	$\mathfrak{S}_3 = \mathfrak{S} \{ (G_I \equiv G_I^* \equiv G_J^*) \circ (G_J) \}$	δ_3
4	$\mathfrak{S}_4 = \mathfrak{S} \{ (G_I \equiv G_J \equiv G_J^*) \circ (G_I^*) \}$	δ_4
5	$\mathfrak{S}_5 = \mathfrak{S} \{ (G_I^* \equiv G_J \equiv G_J^*) \circ (G_I) \}$	δ_5
6	$\mathfrak{S}_6 = \mathfrak{S} \{ (G_I \equiv G_I^*) \circ (G_J \equiv G_J^*) \}$	δ_6
7	$\mathfrak{S}_7 = \mathfrak{S} \{ (G_I \equiv G_I^*) \circ (G_J) \circ (G_J^*) \}$	δ_7
8	$\mathfrak{S}_8 = \mathfrak{S} \{ (G_I) \circ (G_I^*) \circ (G_J \equiv G_J^*) \}$	δ_8
9	$\mathfrak{S}_9 = \mathfrak{S} \{ (G_I \equiv G_J) \circ (G_I^* \equiv G_J^*) \}$	δ_9
10	$\mathfrak{S}_{10} = \mathfrak{S} \{ (G_I \equiv G_J) \circ (G_I^*) \circ (G_J^*) \}$	δ_{10}
11	$\mathfrak{S}_{11} = \mathfrak{S} \{ (G_I) \circ (G_J) \circ (G_I^* \equiv G_J^*) \}$	δ_{11}
12	$\mathfrak{S}_{12} = \mathfrak{S} \{ (G_I \equiv G_J^*) \circ (G_I^* \equiv G_J) \}$	δ_{12}
13	$\mathfrak{S}_{13} = \mathfrak{S} \{ (G_I \equiv G_J^*) \circ (G_I^*) \circ (G_J) \}$	δ_{13}
14	$\mathfrak{S}_{14} = \mathfrak{S} \{ (G_I) \circ (G_J^*) \circ (G_I^* \equiv G_J) \}$	δ_{14}
15	$\mathfrak{S}_{15} = \mathfrak{S} \{ (G_I) \circ (G_I^*) \circ (G_J) \circ (G_J^*) \}$	δ_{15}

Nous tirons aussi les relations suivantes :

$$f_I = \Pr \{ (G_I \equiv G_I^*) \} = \delta_1 + \delta_2 + \delta_3 + \delta_6 + \delta_7$$

$$f_J = \Pr \{ (G_J \equiv G_J^*) \} = \delta_1 + \delta_4 + \delta_5 + \delta_6 + \delta_8$$

$$\varphi_{IJ} = \frac{1}{4} \Pr \{ (G_I \equiv G_J) \} + \frac{1}{4} \Pr \{ (G_I \equiv G_J^*) \} + \frac{1}{4} \Pr \{ (G_I^* \equiv G_J) \} + \frac{1}{4} \Pr \{ (G_I^* \equiv G_J^*) \}$$

$$\varphi_{IJ} = \frac{1}{4} (\varphi_{NP} + \varphi_{NQ} + \varphi_{OP} + \varphi_{OQ})$$

$\varphi_{NP}, \varphi_{NQ}, \varphi_{OP}, \varphi_{OQ}$, étant les coefficients de parenté des parents N, O, P, Q de I et de J :

$$\Pr \{ (G_I \equiv G_J) \} = \delta_1 + \delta_2 + \delta_4 + \delta_9 + \delta_{10}$$

$$\Pr \{ (G_I \equiv G_J^*) \} = \delta_1 + \delta_3 + \delta_4 + \delta_{12} + \delta_{13}$$

$$\Pr \{ (G_I^* \equiv G_J) \} = \delta_1 + \delta_2 + \delta_5 + \delta_{12} + \delta_{14}$$

$$\Pr \{ (G_I^* \equiv G_J^*) \} = \delta_1 + \delta_3 + \delta_5 + \delta_9 + \delta_{11}$$

$$4\varphi_{IJ} = 4\delta_1 + 2\delta_2 + 2\delta_3 + 2\delta_4 + 2\delta_5 + 2\delta_9 + 2\delta_{12} + \delta_{10} + \delta_{11} + \delta_{13} + \delta_{14}$$

$$\varphi_{IJ} = \delta_1 + \frac{1}{2} (\delta_2 + \delta_3 + \delta_4 + \delta_5 + \delta_9 + \delta_{12}) + \frac{1}{4} (\delta_{10} + \delta_{11} + \delta_{13} + \delta_{14})$$

f_I étant le coefficient de consanguinité de I, f_J celui de J, et φ_{IJ} le coefficient de parenté de I et J.

Considérons les deux situations d'identité restreinte suivantes :

$$\mathfrak{S}_6 = \mathfrak{S} \{ (G_I \equiv G_I^*) \circ (G_J \equiv G_J^*) \}$$

et :

$$\mathfrak{S}_7 = \mathfrak{S} \{ (G_I \equiv G_I^*) \circ (G_J) \circ (G_J^*) \}.$$

La situation d'identité restreinte $\mathfrak{S}_6 \cup \mathfrak{S}_7$ est celle pour laquelle les relations entre G_I et G_J^* ne sont pas définies, ce que symbolise :

$$\mathfrak{T} = \mathfrak{S} \{ (G_I \equiv G_I^*) \circ (G_J ? G_J^*) \}.$$

Toute probabilité attachée à une telle situation est encore un coefficient d'identité restreinte. Nous donnons de ces situations et probabilités d'identité restreinte le tableau ci-contre, dont la lecture est immédiate.

Il est possible de définir d'autres situations et coefficients d'identité restreinte pour lesquels, dans un sous-ensemble donné, nous n'attachons aucune importance à la provenance des gènes homologues qui y sont réunis.

Ainsi $\mathfrak{S}_{(42)}^{(4)}$ est la situation d'identité restreinte pour laquelle les quatre loci homologues de deux zygotes sont identiques :

$$\mathfrak{S}_{(42)}^{(4)} = \mathfrak{S} \{ (G_I \equiv G_I^* \equiv G_J \equiv G_J^*) \}.$$

	Situations d'identité restreinte	Probabilités
1	$\mathfrak{S} \{ (G_i \equiv G_i^*) \circ (G_j ? G_j) \} = \mathfrak{T}_1$	$\Pr \{ \mathfrak{T}_1 \} = \delta_6 + \delta_7$
2	$\mathfrak{S} \{ (G_i \equiv G_j) \circ (G_j^* ? G_j^*) \} = \mathfrak{T}_2$	$\Pr \{ \mathfrak{T}_2 \} = \delta_9 + \delta_{10}$
3	$\mathfrak{S} \{ (G_i \equiv G_j^*) \circ (G_j^* ? G_j) \} = \mathfrak{T}_3$	$\Pr \{ \mathfrak{T}_3 \} = \delta_{12} + \delta_{13}$
4	$\mathfrak{S} \{ (G_i^* \equiv G_j) \circ (G_j ? G_j^*) \} = \mathfrak{T}_4$	$\Pr \{ \mathfrak{T}_4 \} = \delta_{12} + \delta_{14}$
5	$\mathfrak{S} \{ (G_i^* \equiv G_j^*) \circ (G_i ? G_j) \} = \mathfrak{T}_5$	$\Pr \{ \mathfrak{T}_5 \} = \delta_9 + \delta_{11}$
6	$\mathfrak{S} \{ (G_i \equiv G_j^*) \circ (G_i ? G_i^*) \} = \mathfrak{T}_6$	$\Pr \{ \mathfrak{T}_6 \} = \delta_6 + \delta_8$

Nous avons aussi :

$$\Pr \{ \mathfrak{S}(\binom{4}{42}) \} = \delta_1.$$

$\mathfrak{S}(\binom{3}{42})$ est la situation d'identité restreinte pour laquelle, parmi les quatre gènes homologues de deux zygotes, trois sont identiques et le quatrième indépendant des trois premiers. Nous avons immédiatement :

$$\Pr \{ \mathfrak{S}(\binom{3}{42}) \} = \delta_2 + \delta_3 + \delta_4 + \delta_5.$$

$\mathfrak{S}(\binom{22}{42})$ est la situation d'identité restreinte pour laquelle, parmi les quatre gènes homologues de deux zygotes, deux sont identiques et indépendants des deux autres, identiques entre eux. Nous avons immédiatement :

$$\Pr \{ \mathfrak{S}(\binom{22}{42}) \} = \delta_6 + \delta_9 + \delta_{12}.$$

$\mathfrak{S}(\binom{211}{42})$ est la situation d'identité restreinte pour laquelle, parmi les quatre gènes homologues de deux zygotes, deux sont identiques et indépendants des deux autres qui sont eux-mêmes indépendants entre eux. Nous avons immédiatement :

$$\Pr \{ \mathfrak{S}(\binom{211}{42}) \} = \delta_7 + \delta_8 + \delta_{10} + \delta_{11} + \delta_{13} + \delta_{14}$$

$$\Pr \{ \mathfrak{S}(\binom{22}{42}) \} + \Pr \{ \mathfrak{S}(\binom{211}{42}) \} = \Pr \{ \mathfrak{S}(\binom{2}{42}) \}$$

$\mathfrak{S}(\binom{0}{42})$ est la situation d'identité restreinte pour laquelle les quatre gènes homologues de deux zygotes sont tous indépendants deux à deux. Nous avons :

$$\Pr \{ \mathfrak{S}(\binom{0}{42}) \} = \delta_{15}.$$

Les situations d'identité et les coefficients d'identité d'ordre 4, de rang 3.

Soient K, L, M, les trois zygotes, désignons par K le zygote dans lequel sont tirés les deux loci homologues.

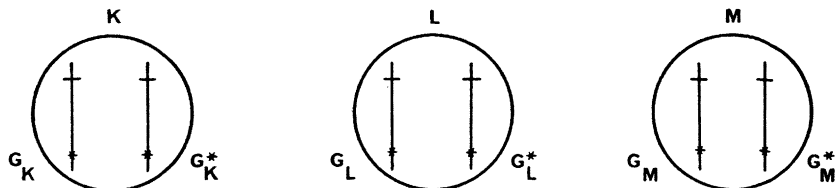


FIG. 11.

Nous utiliserons les coefficients d'identité suivants :

$$\mathfrak{I}_{(43)}^0; \mathfrak{I}_{(43)}^2; \mathfrak{I}_{(43)}^3; \mathfrak{I}_{(43)}^4; \mathfrak{I}_{(43)}^{22}; \mathfrak{I}_{(43)}^{211}$$

dont la signification est immédiate.

Les situations d'identité et les coefficients d'identité d'ordre 4, de rang 4.

Soient N, O, P, Q les quatre zygotes.

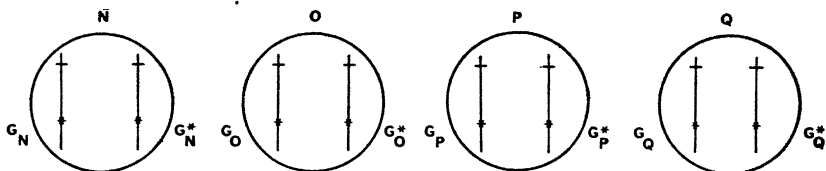


FIG. 12.

Nous utiliserons les coefficients d'identité suivants :

$$\mathfrak{I}_{(44)}^0; \mathfrak{I}_{(44)}^2; \mathfrak{I}_{(44)}^3; \mathfrak{I}_{(44)}^4; \mathfrak{I}_{(44)}^{22}; \mathfrak{I}_{(44)}^{211}$$

dont la signification est immédiate.

VI. — LES SITUATIONS ET LES COEFFICIENTS D'IDENTITÉ GÉNÉRALISÉE 2.

Soit un zygote H, dans lequel nous considérons deux classes d'homologues liées :

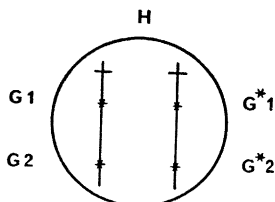


FIG. 13.

Les situations d'identité restreinte sont les suivantes :

$$\mathfrak{S} \{ (G1 \equiv G^*1) \} = \mathfrak{S}_{11}$$

$$\mathfrak{S} \{ (G1) \circ (G^*1) \} = \mathfrak{S}_{12}$$

$$\mathfrak{S} \{ (G2 \equiv G^*2) \} = \mathfrak{S}_{21}$$

$$\mathfrak{S} \{ (G2) \circ (G^*2) \} = \mathfrak{S}_{22}$$

Les situations d'identité généralisée 2 sont les suivantes :

$$\mathfrak{S}_{11} \cap \mathfrak{S}_{21} = \mathfrak{S}_1$$

$$\mathfrak{S}_{11} \cap \mathfrak{S}_{22} = \mathfrak{S}_2$$

$$\mathfrak{S}_{12} \cap \mathfrak{S}_{21} = \mathfrak{S}_3$$

$$\mathfrak{S}_{12} \cap \mathfrak{S}_{22} = \mathfrak{S}_4.$$

Les quatre situations d'identité ont pour probabilité respective Φ_1 , Φ_2 , Φ_3 , Φ_4 . Elles sont utilisées dans l'étude du linkage.

VII. — LES TIRAGES GAMÉTIQUES MULTIPLES DANS UN MÊME ZYGOTE

Un zygote quelconque possède deux états exclusifs : ou bien il possède ses deux gènes homologues identiques ou bien il les possède indépendants. Le raisonnement d'un tirage gamétique multiple est le suivant : supposer d'abord que le zygote a ses deux gènes dans l'état d'identité et effectuer alors les tirages gamétiques, puis supposer que les deux loci homologues du zygote sont dans l'état d'indépendance et alors effectuer les mêmes tirages gamétiques que précédemment. *Nous ne pouvons en aucune façon concevoir qu'un zygote puisse changer d'état entre deux tirages gamétiques successifs.*

Tirons au hasard n loci homologues parmi les deux gènes homologues d'un même zygote. Ce tirage équivaut à constituer des gamètes, nous l'appelons donc un tirage gamétique multiple.

Ces n gènes forment un ensemble qui ne peut être subdivisé au plus qu'en deux sous-ensembles et deux seuls ⁽¹³⁾ de gènes identiques.

Deux gènes, l'un pris dans un sous-ensemble, l'autre dans l'autre sous-ensemble, sont indépendants. Soit $\mathfrak{S}_{(n1)}^m$ le symbole d'une situation d'identité restreinte définie entre ces n gènes et telle que m ($m \leq n$) soit le nombre

⁽¹³⁾ Nous ne tenons pas compte dans ce paragraphe des mutations.

de gènes appartenant au premier sous-ensemble, $n - m$ étant celui de l'autre sous-ensemble. Le chiffre 1 rappelle qu'il s'agit de tirages gamétiques dans un seul zygote.

Calculons la probabilité de la situation d'identité restreinte $\mathfrak{S}(\binom{m}{n_1})$ pour $m = n$, autrement dit nous désirons que tous les n gènes tirés soient identiques.

Supposons que le zygote I dans lequel nous effectuons ces tirages soit en état d'identité, événement de probabilité f_1 , alors tous les tirages de n gènes fournissent n gènes identiques.

Supposons maintenant que le zygote I dans lequel nous effectuons ces n tirages soit en état d'indépendance, il n'y a que deux types de tirages qui assurent l'identité des n gènes, l'un étant le tirage de $n = m$ fois successives du même gène, l'autre étant le tirage de $n = m$ fois l'autre gène. Chacune de ces éventualités a la probabilité $(1/2)^n$.

D'où pour $m = n$:

$$\Pr \{ \mathfrak{S}(\binom{n}{n_1}) \} = f_1 + (1 - f_1)2(1/2)^n$$

$$\Pr \{ \mathfrak{S}(\binom{n}{n_1}) \} = (1/2)^n [2 + (2^n - 2)f_1].$$

Les applications immédiates de cette formule sont les suivantes :

$$\Pr \{ \mathfrak{S}(\binom{2}{2_1}) \} = (1/2)^2 [2 + (4 - 2)f_1] = 1/2 [1 + f_1]$$

$$\Pr \{ \mathfrak{S}(\binom{3}{3_1}) \} = (1/2)^3 [2 + (8 - 2)f_1] = (1/2)^2 [1 + 3f_1]$$

$$\Pr \{ \mathfrak{S}(\binom{4}{4_1}) \} = (1/2)^4 [2 + (16 - 2)f_1] = (1/2)^3 [1 + 7f_1].$$

Calculons la probabilité de la situation d'identité restreinte $\mathfrak{S}(\binom{m}{n_1})$ pour $m \neq n$. Supposons que le zygote I dans lequel nous effectuons ces n tirages soit en état d'identité, événement de probabilité f_1 , aucune suite de tirages n'assure $\mathfrak{S}(\binom{m}{n_1})$. Supposons que le zygote I dans lequel nous effectuons ces n tirages soit dans l'état d'indépendance, événement de probabilité $1 - f_1$.

Toute suite de n gènes issue d'un tirage multiple de n gènes homologues a une probabilité de $(1/2)^n$. Il y a $2C_n^m = C_n^m + C_n^{n-m}$ suites de gènes qui conduisent à la situation $\mathfrak{S}(\binom{m}{n_1})$ quand $m \neq n - m$ et C_n^m quand $m = n - m$.

Nous avons donc pour $m \neq n - m$:

$$\Pr \{ \mathfrak{S}(\binom{m}{n_1}) \} = (1 - f_1)(1/2)^n 2C_n^m$$

et pour $m = n - m$:

$$\Pr \{ \mathfrak{S}(\binom{m}{n_1}) \} = (1 - f_1)(1/2)^n C_n^m.$$

Les applications immédiates de ces formules sont les suivantes :

$$\Pr \{ \mathfrak{S}_{(21)}^{(1)} \} = (1 - f_1)(1/2)^2(1 + 1) = 1/2(1 - f_1)$$

$$\Pr \{ \mathfrak{S}_{(31)}^{(2)} \} = (1 - f_1)(1/2)^3(3 + 3) = (1/2)^3 6(1 - f_1)$$

$$\Pr \{ \mathfrak{S}_{(41)}^{(2)} \} = (1 - f_1)(1/2)^4 6 = (1/2)^4 6(1 - f_1).$$

$$\Pr \{ \mathfrak{S}_{(41)}^{(3)} \} = (1 - f_1)(1/2)^4(4 + 4) = (1/2)^4 8(1 - f_1) = 1/2(1 - f_1).$$

Remarquons que :

$$\Pr \{ \mathfrak{S}_{(21)}^{(2)} \} + \Pr \{ \mathfrak{S}_{(21)}^{(1)} \} = 1$$

$$\Pr \{ \mathfrak{S}_{(31)}^{(3)} \} + \Pr \{ \mathfrak{S}_{(31)}^{(2)} \} = 1$$

$$\Pr \{ \mathfrak{S}_{(41)}^{(4)} \} + \Pr \{ \mathfrak{S}_{(41)}^{(3)} \} + \Pr \{ \mathfrak{S}_{(41)}^{(2)} \} = 1.$$

VIII. — CONCLUSION

Echange Annales

L'interprétation génique et chromosomique de l'hérédité possède une économie des moyens logiques remarquable : un objet le gène, trois postulats et quatre relations binaires. La codification des situations d'identité s'impose car ces écritures symboliques seront utilisées dans d'autres articles pour établir les corrélations entre apparentés quelconques, pour prédire l'évolution de la structure génique des populations soumises à des lois régulières d'appariement et pour analyser les ascendances des populations pedigree.

Il est difficile de réaliser une axiomatique complète de la génétique. En fait nous avons choisi un ensemble d'expériences biologiques dont l'analyse permet de décrire un objet à l'aide de postulats et de définir les relations entre ces objets à l'aide d'axiomes. Ainsi, il n'existe d'axiomatique en biologie que relative à un ensemble d'expériences. Pour être efficace il est nécessaire, que dans le cadre de la logique des mathématiques, qui est une logique binaire, ces postulats et axiomes soient indépendants et non contradictoires. On peut alors se demander quelles propriétés doivent posséder les ensembles expérimentaux ainsi choisis pour que les postulats et les axiomes qui en découlent, soient indépendants et non contradictoires dans une telle logique ?

Si l'axiomatique apparaît en biologie, comme une méthode permettant d'ordonner les conclusions de certains ensembles d'expériences afin d'exploiter toutes les conséquences qu'ouvre le raisonnement déductif, elle est aussi

la source d'une exigence plus grande dans l'étude des logiques expérimentales et de celles des phénomènes biologiques.

C'est une restriction importante que de supposer comme nous l'avons fait qu'il existe en génétique des populations une double implication entre la non-identité et l'indépendance d'une part, l'identité et la dépendance absolue d'autre part. Il semble qu'il n'y ait plus beaucoup de difficultés à s'affranchir de cette restriction, de telle sorte que si l'identité entraîne la dépendance absolue, la dépendance absolue n'implique pas obligatoirement l'identité.

IX. — RÉSUMÉ

Nous présentons une étude logique des objets de l'interprétation génique et chromosomique de l'hérédité et des relations susceptibles d'exister entre ces objets. Nous donnons du gène la définition suivante : « C'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 \mathcal{G} l'ensemble, supposé fini, des gènes de tous les zygotes d'un groupe. Soit \mathcal{G} l'ensemble, supposé fini, de tous les gènes d'un zygote. Le premier ensemble est le génotype général, le second le génotype particulier. Nous définissons entre deux gènes appartenant à l'un de ces ensembles quatre relations binaires.

1. *La relation d'homologie* : « Deux gènes sont homologues ssi 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 ssi ils sont homologues et ssi ils 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'homologue, 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 ssi ils sont homologues et ssi ils 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 ssi ils dérivent par descendance mendélienne, en l'absence de mutation, d'un même gène ancêtre ». C'est une relation binaire d'équivalence qui induit sur tout ensemble de gènes une partition en sous-ensembles disjoints appelés : *classes d'identité*. Des 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 d'un même modèle, le gène ancêtre. Deux gènes appartenant à deux classes d'identité distinctes sont indépendants, si la connaissance de la nature de l'un ne permet pas de déterminer avec certitude la nature de l'autre. Ces deux gènes ne sont pas les copies non modifiées d'un même modèle.

Toute partition d'un ensemble de gènes par la relation d'identité est une *situation d'identité*, dont nous donnons un mode d'écriture symbolique. La probabilité de réalisation d'une *situation d'identité* est un *coefficient d'identité*. Si l'ensemble des gènes considérés est une classe d'homologues, la partition de cet ensemble par la relation d'identité est une *situation d'identité restreinte*, et la probabilité de sa réalisation un *coefficient d'identité restreinte*. Nous donnons des situations d'identité restreinte une description formelle pour les cas des ensembles de gènes dont nous aurons besoin dans l'étude de la covariance entre apparentés quelconques. Nous généralisons la notion de situation d'identité restreinte par celles de *situation d'identité généralisée* et de *situation d'identité chromosomique*. Cette dernière notion rejoint celle des « junctions » de R. A. Fisher.

X. — BIBLIOGRAPHIE

- N. BOURBAKI, *Théorie des ensembles*. Ch. 4 : Structures. Hermann, Paris, 1957, 124 p.
 N. BOURBAKI, *Théorie des ensembles*. Fascicule de résultats. Hermann, Paris, 1958, 62 p.
 N. BOURBAKI, *Théorie des ensembles*. Ch. 1 : Description de la mathématique formelle. Ch. 2 : Théorie des ensembles. Hermann, Paris, 1960, 134 p.
 N. BOURBAKI, *Théorie des ensembles*. Ch. 3 : Ensembles ordonnés, cardinaux, nombres entiers. Hermann, Paris, 1963, 146 p.
 R. A. FISHER, *The theory of Inbreeding*. Oliver and Boyd, Edinburgh, London, 1949, VIII + 120 p.
 M. GILLOIS, La relation d'identité en génétique. *Thèse Fac. Sciences*, Paris, 1964.
 M. GILLOIS, 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.

- M. GILLOIS, Corrélation génotypique théorique entre apparentés. *Journées d'Études de la Commission de Génétique de la Fédération européenne de Zootechnie*, Lisbonne, octobre 1964.
- G. MALÉCOT, Le calcul des probabilités et les problèmes de l'hérédité. *Ann. Univ. Lyon, Sciences*, section A, t. 2, 1939, p. 25-37.
- G. MALÉCOT, a) Étude mathématique des populations « mendéliennes ». *Ann. Univ. Lyon, Sciences*, section A, t. 4, 1941, p. 45-60. b) Mendélisme et consanguinité. *C. R. Acad. Sci.*, Paris, t. 215, 1941, p. 313-314.
- G. MALÉCOT, *Les mathématiques de l'hérédité*. Masson et Cie, Paris, 1 vol., 1948, 60 p.
-

II. — Corrélation génétique dans le cas de dominance

SOMMAIRE

I. Introduction	39
II. Rappel des situations et des coefficients d'identité.	41
III. Les variables aléatoires génétiques	45
IV. Les corrélations génétiques entre apparentés quelconques dans le cas de dominance	52
V. Les corrélations génétiques entre apparentés quelconques dans le cas d'effet maternel génétique	68
VI. Généralisation au cas de multiallélisme.	83
VII. Estimation des paramètres génétiques dans une population . . .	86
VIII. Conclusion.	89
IX. Résumé.	90
X. Bibliographie	93

I. — INTRODUCTION

Le mécanisme de la transmission des caractères quantitatifs n'est pas connu, car les supports héréditaires qui les conditionnent n'ont pas encore été isolés, décrits, analysés. Pourtant dès le début de ce siècle, F. Galton avait réuni une remarquable information statistique concernant les caractères quantitatifs et leur transmission héréditaire. Ces résultats semblaient en contradiction avec la conception de l'hérédité discontinue ou génique des caractères qualitatifs née des expériences de G. Mendel.

Néanmoins, dès 1904, K. Pearson, puis Udny Yule en 1906, Weinberg en 1908 et 1910 et surtout R. A. Fisher en 1918-1921 postulent que l'hérédité continue pourrait avoir un mécanisme héréditaire génique. En particulier R. A. Fisher, utilisant cette hypothèse, donne l'expression du coefficient de corrélation théorique pour des cas particuliers tels que des couples de plein-frères, de demi-frères, d'oncle-neveu ou de cousins, lorsqu'il s'agit d'un

caractère avec dominance et que ces individus appartiennent à une population d'effectif infini et panmictique.

En 1921-1923 S. Wright introduit le *coefficient of inbreeding* d'un zygote, c'est-à-dire la corrélation linéaire entre les deux gamètes chacun porteur d'un gène, dont l'union a donné ce zygote consanguin. Le calcul de ce coefficient de corrélation se fait par la seule connaissance des ascendances. Pour cela S. Wright emploie la notion des « path-coefficients » qui est bien confuse. L'usage de ce coefficient « of inbreeding » lui permet de généraliser les expressions de la corrélation théorique entre zygotes apparentés et consanguins mais tout en excluant l'hypothèse de la dominance.

G. Malécot, à la suite d'un cheminement allant de 1941 à 1948 présente dans son livre *Les mathématiques de l'hérédité* l'aspect fondamental et aujourd'hui classique de la consanguinité. Cet auteur considère la probabilité que deux loci homologues, pris au hasard parmi tous les couples de loci homologues d'un zygote I, soient identiques parce que dérivant d'un même locus appartenant à un zygote ancêtre commun au père et à la mère de I. Soit f_I cette probabilité attachée au zygote I et que G. Malécot nomme *coefficient de consanguinité*. Il ne faut pas confondre le coefficient de consanguinité qui est la probabilité d'un événement réalisable dans un zygote avec le coefficient « of inbreeding » qui est un coefficient de corrélation. L'analogue du coefficient « of inbreeding », dans la théorie de la consanguinité de G. Malécot, est le *coefficient de corrélation gamétique*.

G. Malécot étend la notion de coefficient de consanguinité par celle de *coefficient de parenté* de deux zygotes. Le coefficient de parenté est la probabilité que deux loci homologues tirés au hasard l'un dans un zygote I, l'autre dans un zygote J soient identiques parce que dérivant d'un même locus ancêtre. Ces notions probabilistes ont permis à O. Kempthorne de donner en 1957 un calcul très élégant du coefficient de corrélation théorique entre deux zygotes apparentés non consanguins dans le cas de dominance. Cette expression avait été établie par G. Malécot dès 1939 par une méthode très proche dans son esprit de celles qu'emploie S. Wright.

Ce travail est la démonstration du coefficient de corrélation théorique entre deux zygotes consanguins et apparentés dans le cas de dominance. Pour obtenir ce résultat nous utilisons une notion nouvelle : *les coefficients d'identité*. Ces coefficients d'identité sont issus de la généralisation des coefficients de consanguinité et de parenté. Nous présentons ensuite la démonstration du coefficient de corrélation théorique entre deux zygotes consanguins et apparentés dans le cas de dominance et d'effet maternel génotypique.

II. — RAPPEL DES SITUATIONS ET DES COEFFICIENTS D'IDENTITÉ

Nous avons présenté (Gillois M., 1964) une étude logique des objets de l'interprétation génétique et chromosomique de l'hérédité et des relations susceptibles d'exister entre ces objets. Nous donnons du gène la définition suivante : « C'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 \mathcal{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. Le premier ensemble est le génotype général, le second le génotype particulier. Nous définissons entre deux gènes appartenant à l'un de ces ensembles quatre relations binaires.

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 si ils 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 si ils 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. » C'est une relation binaire d'équivalence qui induit sur tout ensemble de gènes une partition en sous-ensembles disjoints appelés *classes d'identité*. Des 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 d'un même modèle : le gène ancêtre. Deux gènes appartenant à deux classes distinctes sont indépendants ⁽¹⁴⁾, car la connaissance de la nature de l'un ne permet pas de déterminer avec certitude la nature de l'autre. Ces deux gènes ne sont pas les copies non modifiées d'un même modèle.

Toute partition d'un ensemble de gènes par la relation d'identité est une *situation d'identité*, dont nous donnons ci-dessous un mode d'écriture symbolique. La probabilité de réalisation d'une *situation d'identité* est un *coefficient d'identité*. Si l'ensemble des gènes considérés est une classe d'homologues, la partition de cet ensemble par la relation d'identité est une *situation d'identité restreinte*, et la probabilité de sa réalisation un *coefficient d'identité restreinte*. Nous donnons (Gillois, 1964) des situations d'identité restreinte une description formelle pour les cas des ensembles de gènes dont nous aurons besoin dans l'étude de la covariance entre apparentés quelconques. Nous généralisons (Gillois, 1964) la notion de situation d'identité restreinte par celles de *situation d'identité généralisée* et de *situation d'identité chromosomique*. Cette dernière notion rejoint celle des « junctions » de R. A. Fisher.

Appliquons ces définitions générales au cas de deux zygotes I et H, ce qui nous permettra de calculer ultérieurement la covariance théorique entre I et J quels que soient leurs liens de parenté. Le schéma ci-contre, symbolise les deux zygotes I et J dans lesquels nous identifions quatre gènes homologues G_I et G_J , gènes transmis par le père ; G_I^* et G_J^* , gènes transmis par la mère. Décrivons les situations d'identité restreinte susceptibles d'exister entre les quatre gènes homologues des zygotes I et J.

Nous avons réuni, dans le tableau ci-contre, les situations et les probabilités d'identité restreinte possibles.

(14) Cet énoncé suppose, ce qui est une restriction importante, l'existence d'une double implication entre la non-identité et l'indépendance.

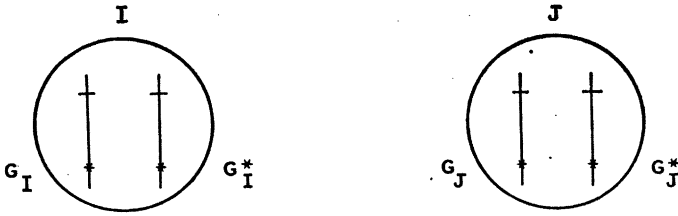


FIG. 14.

Situations d'identité restreinte (de symbole \mathfrak{S}_i)		Probabilités
1	$\mathfrak{S}_1 = \mathfrak{S} \{ (G_I \equiv G_I^* \equiv G_J \equiv G_J^*) \}$	δ_1
2	$\mathfrak{S}_2 = \mathfrak{S} \{ (G_I \equiv G_I^* \equiv G_J) \circ (G_J^*) \}$	δ_2
3	$\mathfrak{S}_3 = \mathfrak{S} \{ (G_I \equiv G_I^* \equiv G_J^*) \circ (G_J) \}$	δ_3
4	$\mathfrak{S}_4 = \mathfrak{S} \{ (G_I \equiv G_J \equiv G_J^*) \circ (G_I^*) \}$	δ_4
5	$\mathfrak{S}_5 = \mathfrak{S} \{ (G_I^* \equiv G_J \equiv G_J^*) \circ (G_I) \}$	δ_5
6	$\mathfrak{S}_6 = \mathfrak{S} \{ (G_I \equiv G_I^*) \circ (G_J \equiv G_J^*) \}$	δ_6
7	$\mathfrak{S}_7 = \mathfrak{S} \{ (G_I \equiv G_I^*) \circ (G_J) \circ (G_J^*) \}$	δ_7
8	$\mathfrak{S}_8 = \mathfrak{S} \{ (G_I) \circ (G_I^*) \circ (G_J \equiv G_J^*) \}$	δ_8
9	$\mathfrak{S}_9 = \mathfrak{S} \{ (G_I \equiv G_J) \circ (G_I^* \equiv G_J^*) \}$	δ_9
10	$\mathfrak{S}_{10} = \mathfrak{S} \{ (G_I \equiv G_J) \circ (G_I^*) \circ (G_J^*) \}$	δ_{10}
11	$\mathfrak{S}_{11} = \mathfrak{S} \{ (G_I) \circ (G_J) \circ (G_I^* \equiv G_J^*) \}$	δ_{11}
12	$\mathfrak{S}_{12} = \mathfrak{S} \{ (G_I \equiv G_J^*) \circ (G_I^* \equiv G_J) \}$	δ_{12}
13	$\mathfrak{S}_{13} = \mathfrak{S} \{ (G_I \equiv G_J^*) \circ (G_I^*) \circ (G_J) \}$	δ_{13}
14	$\mathfrak{S}_{14} = \mathfrak{S} \{ (G_I) \circ (G_J^*) \circ (G_I^* \equiv G_J) \}$	δ_{14}
15	$\mathfrak{S}_{15} = \mathfrak{S} \{ (G_I) \circ (G_I^*) \circ (G_J) \circ (G_J^*) \}$	δ_{15}

\mathfrak{S}_1 est une situation pour laquelle les quatre gènes homologues sont identiques, ce qui est une conséquence du fait que les deux zygotes sont consanguins et qu'ils sont apparentés.

\mathfrak{S}_2 est une situation pour laquelle les gènes homologues G_I , G_I^* , G_J sont identiques et indépendants du gène G_J^* , ce qui est la conséquence du fait que le zygote I est consanguin, que le zygote J est ou n'est pas consanguin, que les deux zygotes sont apparentés.

\mathfrak{S}_6 est une situation pour laquelle les gènes homologues G_I et G_I^* sont identiques, les gènes G_J et G_J^* sont identiques et indépendants des deux précédents, ce qui est la conséquence du fait que les deux zygotes sont consanguins et sont ou non apparentés.

\mathfrak{S}_{12} est une situation pour laquelle les gènes homologues G_I et G_I^* sont identiques, les gènes homologues G_J et G_J^* sont identiques, et les gènes homologues appartenant à un même zygote indépendants. Les zygotes I et J sont apparentés, mais ils peuvent être ou non consanguins.

Nous pensons avoir montré ainsi toute la richesse de ce nouveau concept de situation d'identité restreinte. Soulignons qu'il ne faut pas confondre la consanguinité et la parenté avec la notion d'identité. Deux zygotes sont parents si et seulement si ils possèdent des ancêtres communs. Un zygote est consanguin si et seulement si son père et sa mère sont parents. Mais ce zygote consanguin peut avoir ses deux gènes homologues identiques ou indépendants. De façon générale, l'absence de consanguinité ou de parenté entraîne l'indépendance de tous les gènes homologues en cause, tandis que l'existence de la consanguinité et de la parenté entraîne l'une quelconque des situations d'identité possibles avec une certaine probabilité.

Le coefficient de parenté de Malécot ne décrit que les états d'identité et d'indépendance de deux gènes pris au hasard, l'un parmi les deux gènes homologues d'un zygote, l'autre parmi les deux gènes homologues de l'autre zygote. Le concept de situation d'identité que nous avons introduit décrit tous les états d'identité et d'indépendance existant entre les quatre gènes homologues de deux zygotes. Cette propriété de la situation d'identité rend possible l'étude génétique des populations consanguines, car les coefficients d'identité apportent toute l'information génétique nécessaire au calcul des covariances théoriques entre zygotes apparentés. Ce résultat, obtenu dans le cas de la dominance entre gènes homologues et de multiallélisme, est tout à fait général, si l'on suppose l'existence d'une double implication entre la non-identité et l'indépendance.

III. — LES VARIABLES ALÉATOIRES GÉNÉTIQUES

Nous allons supposer, ce qui est le cas des caractères quantitatifs, que la classe d'isoaction du gène présent à un locus singulier donné L est inconnue. Nous pouvons seulement dire que le gène appartient à l'une des classes d'isoaction de sa classe d'homologues. A chacune de ces possibilités nous attachons une probabilité p_i telle que $\sum_i p_i = 1$. Nous ajoutons que l'effet d'un gène se manifeste par une contribution numérique à l'expression d'un caractère.

1. Les variables aléatoires gamétiques.

Soit une série biallélique $S = \{A, a\}$ avec pour A la probabilité p et pour a la probabilité q d'être situés à un locus singulier. Chiffrons à s la contribution numérique de A et à t celle de a . Nous définissons ainsi la variable aléatoire gamétique simple, d'indicateur X , qui peut prendre l'une ou l'autre des deux valeurs particulières possibles s et t quand un événement aléatoire E_i de probabilité p_i est réalisé. Cet événement aléatoire est l'état génique d'un locus singulier, ces états pouvant être : A avec la probabilité p et a avec la probabilité q .

Variable aléatoire ou indicateur	Réalisations possibles et exclusives de l'événement aléatoire, présence d'un gène A ou a au locus singulier	Probabilité <i>a priori</i> de la réalisation d'un événement aléatoire	Valeurs numériques attachées à la réalisation de l'événement aléatoire
X	A	p	s
	a	q	t

Calcul de quelques paramètres de X :

$$E(X) = ps + qt,$$

par changement d'origine nous poserons :

$$\begin{aligned} E(X) &= 0 \\ E(X^2) &= ps^2 + qt^2 = \text{VAR}(X). \end{aligned}$$

Cette quantité est la variance gamétique ou additive.

La variable aléatoire gamétique est par définition une somme de variables aléatoires gamétiques simples.

2. Les variables aléatoires zygotiques.

Un organisme diploïde est réalisé par la fusion de deux gamètes, chacun apportant un des deux gènes homologues, qui constitueront un couple ou « zygote ». On peut supposer que la valeur numérique d'un caractère réalisé chez l'organisme diploïde est la somme des contributions numériques à ce caractère de chacun des deux gènes situés sur le couple de loci homologues. Or, les contributions numériques de chacun de ces deux gènes ont été précédemment définies comme celles des variables aléatoires attachées à ces loci. Ceci se traduit par l'écriture symbolique suivante :

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

Y possède trois valeurs possibles $2s$, $s + t$, $2t$ correspondant chacune à l'un des trois états zygotiques possibles : AA, Aa, aa, ayant chacun une certaine probabilité d'apparition, qu'il s'agit de déterminer. Déterminer ces probabilités, c'est définir complètement une nouvelle variable aléatoire, dite « variable aléatoire zygote simple ».

Envisageons deux loci singuliers homologues d'un individu I. Considérés ensemble, ces deux loci singuliers définissent trois états zygotiques AA, Aa, aa, ce qui entraîne les trois valeurs numériques correspondantes $2s$, $s + t$, $2t$ pour le caractère envisagé.

La variable aléatoire attachée à I et prenant ces trois valeurs numériques possibles sera, par définition, la variable aléatoire zygote simple attachée à I, dont le symbole est Y_I :

$$Y_I = X_I + X_I^*.$$

La variable aléatoire gamétique X_I prend la valeur numérique s quand le locus singulier auquel elle est attachée est occupé par un gène de la classe d'isoaction A, événement qui a une probabilité p . La variable aléatoire gamétique X_I prend la valeur numérique t quand ce locus singulier auquel elle est attachée est occupé par un gène de la classe d'isoaction a, événement

qui a une probabilité q . Ces deux événements sont exclusifs et tels que :
 $p + q = 1$.

Il en est de même pour la variable aléatoire X_I^* .

Les rapports entre les variables aléatoires gamétiques X_I et X_I^* sont illustrés par le schéma suivant :

Variable aléatoire X_I			Variable aléatoire X_I^*		
Événements ou états géniques possibles au locus	Probabilités des événements ou loi de probabilité marginale de la variable aléatoire X_I	Valeurs numériques de la variable aléatoire X_I	Événements ou états géniques possibles au locus	Probabilités des événements ou loi de probabilité marginale de la variable aléatoire X_I^*	Valeurs numériques de la variable aléatoire X_I^*
A	p	s	A	p	s
a	q	t	a	q	t

Connaissant les lois des probabilités marginales des variables aléatoires gamétiques X_I et X_I^* , rechercher leur loi de probabilité conjointe fournit la loi de probabilité de la variable aléatoire zygotique, puisque par hypothèse :

$$Y_I = X_I + X_I^*.$$

Pour établir la loi de probabilité conjointe des variables aléatoires X_I et X_I^* , il est nécessaire d'introduire leur coefficient de corrélation qui est aussi le coefficient de consanguinité f_I de l'individu I.

L'identité des états géniques de deux loci homologues de I ($X_I = X_I^*$) se réalise avec la probabilité f_I .

L'indépendance ⁽¹⁵⁾ des états géniques des deux loci homologues de I se réalise avec la probabilité $(1 - f_I)$.

⁽¹⁵⁾ C'est ici qu'apparaît avec le plus de netteté l'hypothèse de l'existence de la double implication entre la non-identité et l'indépendance. Cette hypothèse n'est vérifiée que pour des populations panmictiques par exemple.

Il est alors possible d'introduire la variable aléatoire zygotique :

$$Y_I = X_I + X_I^*.$$

$X_I \backslash X_I^*$		A	a
		s	t
A	s	Génotype : AA Valeur numérique : $2s$ Probabilité : $p^2 + f_1pq$	Génotype : Aa Valeur numérique : $s + t$ Probabilité : $pq - f_1pq$
a	t	Génotype : Aa Valeur numérique : $s + t$ Probabilité : $pq - f_1pq$	Génotype : aa Valeur numérique : $2t$ Probabilité : $q^2 + f_1pq$

La variable aléatoire zygotique Y_I attachée à l'individu I prend les valeurs numériques $2s$, $s + t$, $2t$, quand le génotype de l'individu est AA, Aa, aa et ceci avec les probabilités correspondantes :

$$p^2 + f_1pq, 2(pq - f_1pq), q^2 - f_1pq.$$

Supposons une population dont tous les gènes sont indépendants. Cette condition sera désignée par p . Elle revient à dire que cette population est sans consanguinité, sans parenté.

$$E_p(Y) = E(X + X^*) = E(X) + E(X^*) = 0$$

$$E_p(Y^2) = E_p(X + X^*)^2 = E(X^2) + 2E(XX^*) + E(X^{*2})$$

$$E_p(Y^2) = E(X^2) + E(X^{*2}) = 2E(X^2).$$

Supposons une population dont tous les gènes sont identiques. Cette condition sera désignée par c .

$$E_c(Y) = p2s + q2t = 0$$

$$E_c(Y) = E_c[(X + X^*)] = E(X) + E(X^*) = 0$$

$$E_c(Y^2) = E_c(X + X^*)^2 = E_c(X^2) + E_c2(XX^*) + E_c(X^{*2})$$

$$E_c(Y^2) = 4E(X^2).$$

La variable aléatoire zygotique est par définition une somme de variables aléatoires zygotiques simples.

3. Les variables aléatoires génotypiques et de dominance.

Nous supposons que les deux gènes homologues réunis dans un organisme diploïde n'additionnent pas leurs effets potentiels dans la réalisation quantitative d'un caractère. Nous utiliserons les notations de G. Malécot, pour décrire une telle action, en désignant par i, j et k les valeurs numériques ou contributions numériques des génotypes AA, Aa, aa à la réalisation d'un caractère.

De façon générale, i, j, k ne peuvent pas être mis sous la forme de $i = 2s$; $j = s + t$; $k = 2t$ (en effet, supposons j de valeur quelconque, et non $i = \frac{i+k}{2}$). i, j et k étant donnés, nous utilisons la méthode des moindres carrés pour exprimer s et t , contributions numériques des gènes A et a en fonction de i, j, k .

Désignons comme « résidus de dominance » les trois différences suivantes :

$$i - 2s, \quad j - (s + t), \quad k - 2t.$$

Nous imposons à la somme des carrés des résidus de dominance d'être minimum.

Autrement dit, nous décomposons la contribution numérique d'un génotype à la constitution d'un caractère en deux parts additives dont l'une est le résultat de l'addition des effets de chacun des deux gènes, et dont l'autre est le résultat de l'interaction entre ces deux gènes due à leur juxtaposition dans le zygote.

Cet énoncé se traduit symboliquement ainsi :

$i = 2s$	+	$(i - 2s)$
$j = s + t$	+	$(j - s - t)$
$k = 2t$	+	$(k - 2t)$
contribution		résidu
numérique		de
additive		dominance

Nous voulons que soit minimum en moyenne le carré du résidu de dominance pour les individus d'une population ou d'un groupe. Pour cela, nous allons désigner au hasard un individu I dans la population ou dans le groupe et rendre minimum la moyenne du carré de son résidu de dominance.

L'individu I sera tiré au hasard dans la population ou dans le groupe. La probabilité *a priori* est donc de :

$$\begin{aligned} p^2 + f_1 pq & \text{ pour I, d'appartenir au génotype AA;} \\ 2pq - 2f_1 pq & \text{ pour I, d'appartenir au génotype Aa;} \\ q^2 + f_1 pq & \text{ pour I, d'appartenir au génotype aa;} \end{aligned}$$

f_1 est le coefficient de consanguinité dans la population ou dans le groupe d'un individu I pris au hasard à la génération n . Dans le cas où $f_1 = 0$, nous retrouvons la formule de Hardy-Weinberg. *Dans la suite du calcul des valeurs de s et t , nous supposons que $f_1 = 0$, ce qui assure l'indépendance de tous les gènes homologues.*

Nous voulons rendre minimum l'expression suivante :

$$p^2(i - 2s)^2 + 2pq(j - s - t)^2 + q^2(k - 2t)^2.$$

Ce qui est réalisé si les deux dérivées g'_s et g'_t sont simultanément nulles :

$$g'_s = p(i - 2s) + q(j - s - t) = 0$$

$$g'_t = p(j - s - t) + q(k - 2t) = 0$$

d'où les relations classiques :

$$ps + qt = 0$$

$$pi + qj = s$$

$$pj + qk = t.$$

Considérons un individu I pris au hasard dans une population dont tous les gènes sont indépendants.

Le génotype de cet individu possède trois états possibles AA, Aa, aa et ceci avec une probabilité p^2 , $2pq$, q^2 . Les contributions numériques de chacun de ces génotypes à la réalisation d'un caractère quantitatif sont i , j , k .

Par définition nous appellerons « variable aléatoire génotypique » de l'individu I, la variable aléatoire Z_1 qui prend les valeurs numériques i , j , k pour les états AA, Aa, aa (réalisés avec les probabilités p^2 , $2pq$, q^2).

D'autre part, nous avons décomposé i , j , k chacun en deux éléments additifs, dont le premier estime l'action indépendante des gènes aux loci homologues ($2s$, $s + t$, $2t$) et dont le second mesure l'action d'interférence ou résidu de dominance entre les deux mêmes gènes ($i - 2s$), ($j - s - t$), ($k - 2t$),

d'où :

$$i = 2s + (i - 2s)$$

$$j = s + t + (j - s - t)$$

$$k = 2t + (k - 2t).$$

Pour cela, nous avons imposé à s et à t une condition des moindres carrés.

Interprétons la variable aléatoire dont les valeurs numériques sont :

$$2s, s + t, 2t$$

quand le génotype est AA, Aa, aa avec les probabilités p^2 , $2pq$, q^2 . Il s'agit de la « variable aléatoire zygotique », de symbole Y_I , de l'individu I. Elle a été précédemment définie comme somme de deux variables génétiques :

$$Y_I = X_I + X_I^*$$

Interprétons la variable aléatoire dont les valeurs numériques sont :

$$(i - 2s), (j - s - t), (k - 2t)$$

quand le génotype de l'individu est AA, Aa, aa avec les probabilités p^2 , $2pq$, q^2 , comme une nouvelle variable aléatoire que nous appellerons « variable aléatoire du résidu de dominance » dont le symbole est D_I .

De ce qui précède, nous tirons l'écriture symbolique :

$$Z_I = Y_I + D_I$$

Nous prendrons p comme symbole indiciel de la condition d'indépendance totale entre les gènes, et c comme symbole indiciel de la condition d'identité totale entre les gènes.

Ne confondons pas $E(Z)$ et $E_p(Z)$ car :

$$E(Z) = (1 - f)E_p(Z) + fE_c(Z)$$

cette remarque s'étendant sans difficulté aux autres espérances mathématiques de variables aléatoires génétiques.

Supposons une génération (ou une population ou un groupe) dont les gènes sont tous indépendants :

$$E_p(Z) = p^2i + 2pqj + q^2k = 0$$

$$E_p(Z) = E_p(Y) + E_p(D) = 0$$

$$E_p(D) = 0 \quad \text{car} \quad E_p(Y) = 0$$

$$E_p(Z^2) = p^2i^2 + 2pqj^2 + q^2k^2 \neq 0$$

$$E_p(D^2) = p^2(i - 2s)^2 + 2pq(j - s - t)^2 + q^2(k - 2t)^2$$

or :

$$s + t = pi + qk + j$$

donc :

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

$$E_p(YD) = 0$$

$$E_p(ZD) = 0.$$

Supposons une population dont les gènes sont tous identiques :

$$\begin{aligned}
 E_c(Z) &= pi + qk \\
 E_c(Z) &= p^2i + pqi + pqk + q^2k = pq(i - 2j + k) \\
 E_c(D) &= p(i - 2s) + q(k - 2t) \\
 E_c(D) &= E_c(Z) - E_c(Y) \\
 E_c(Z^2) &= pi^2 + qk^2 \\
 E_c(D^2) &= p(i - 2s)^2 + q(k - 2t)^2 \\
 E_c(X^2) &= psi + qkt \\
 E_c(XD) &= ps(i - 2s) + qt(k - 2t) = E_c(XZ) - 2E(X^*).
 \end{aligned}$$

La variable aléatoire génotypique est une somme de variables aléatoires génotypiques simples et la variable aléatoire de dominance est une somme de variables aléatoires de dominance simples.

IV. — LES CORRÉLATIONS GÉNÉTIQUES ENTRE APPARENTES QUELCONQUES

1. Expression de la covariance génotypique entre I et J.

Soient N, O, P, Q les pères et mères de I et de J. Soit φ le coefficient de parenté de l'un des 6 couples possibles obtenus avec N, O, P, Q. Le croquis ci-contre schématise les relations de dépendance entre I et J.

Nous attachons à I les variables aléatoires suivantes :

$$\begin{aligned}
 X_I &\text{ à } L_I \\
 X_I^* &\text{ à } L_I^* \\
 Y_I &= X_I + X_I^* \text{ à } I \\
 Z_I &= Y_I + D_I \text{ à } I.
 \end{aligned}$$

Nous attachons de même à J :

$$\begin{aligned}
 X_J &\text{ à } L_J \\
 X_J^* &\text{ à } L_J^* \\
 Y_J &= X_J + X_J^* \text{ à } J \\
 Z_J &= Y_J + D_J \text{ à } J.
 \end{aligned}$$

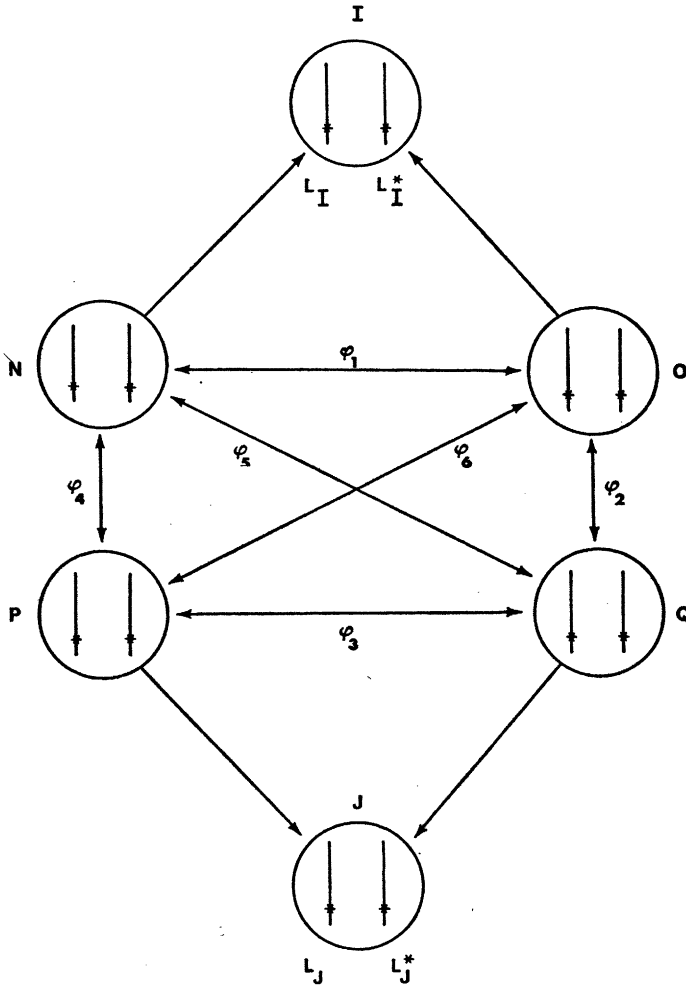


FIG. 15.

Nous avons pour expression de la corrélation génotypique r_D entre I et J :

$$r_D = \frac{\text{COVAR} (Z_I Z_J)}{\sqrt{\text{VAR} (Z_I) \cdot \text{VAR} (Z_J)}}$$

avec :

$$\text{COV} (Z_I Z_J) = E(Z_I Z_J) - [E(Z_I) \cdot E(Z_J)]$$

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

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

$E(Z_i Z_j)$ ne peut pas être calculé directement, mais si l'on connaît la probabilité de chacune des situations qui fixent sans ambiguïté les relations d'indépendance et d'identité entre les gènes G_i, G_i^*, G_j, G_j^* , il est possible de calculer l'espérance mathématique, conditionnée par la réalisation de chacune des situations d'identité restreinte, du produit des variables aléatoires génotypiques de I et de J.

Ces situations d'identité restreinte, et leur probabilité, sont réunies dans le tableau de la page 43.

Remarquons bien, avant d'entreprendre ce calcul, que nous cherchons à nous réduire strictement soit à un état d'identité, soit à un état d'indépendance et cela quel que soit l'état du groupe considéré. Ceci nous permettra d'exprimer la covariance de $Z_i Z_j$, ou les variances de Z_i et de Z_j , comme des fonctions linéaires des moments du second ordre des variables aléatoires génétiques X et D, moments calculés sous la condition d'indépendance totale ou de dépendance totale.

En prenant $Ci (1 \leq i \leq 15)$ pour symbole d'une condition d'une situation d'identité restreinte réalisée entre les quatre gènes homologues G_i, G_i^*, G_j, G_j^* , $E(Z_i Z_j)$ devient par application du théorème des moyennes conditionnées :

$$\begin{aligned} E(Z_i Z_j) = & \delta_1 E_{C1}(Z_i Z_j) + \delta_2 E_{C2}(Z_i Z_j) + \delta_3 E_{C3}(Z_i Z_j) + \delta_4 E_{C4}(Z_i Z_j) \\ & + \delta_5 E_{C5}(Z_i Z_j) + \delta_6 E_{C6}(Z_i Z_j) + \delta_7 E_{C7}(Z_i Z_j) + \delta_8 E_{C8}(Z_i Z_j) \\ & + \delta_9 E_{C9}(Z_i Z_j) + \delta_{10} E_{C10}(Z_i Z_j) + \delta_{11} E_{C11}(Z_i Z_j) + \delta_{12} E_{C12}(Z_i Z_j) \\ & + \delta_{13} E_{C13}(Z_i Z_j) + \delta_{14} E_{C14}(Z_i Z_j) + \delta_{15} E_{C15}(Z_i Z_j). \end{aligned}$$

Nous remarquons d'autre part que :

$$\begin{aligned} E(Z_i Z_j) &= E[(X_i + X_i^* + D_i) \cdot (X_j + X_j^* + D_j)] \\ E(Z_i Z_j) &= E(X_i X_j) + E(X_i X_j^*) + E(X_i^* X_j) + E(X_i^* X_j^*) + E(X_i D_j) + E(X_i^* D_j) \\ &\quad + E(X_j D_i) + E(X_j^* D_i) + E(D_i D_j). \end{aligned}$$

Supposons la condition C1 remplie :

$$\mathfrak{S} \{ (G_i \equiv G_i^* = G_j \equiv G_j^*) \} = \mathfrak{S}_1.$$

La probabilité de cette condition est δ_1 . La situation d'identité restreinte \mathfrak{S}_1 entraîne entre les variables aléatoires gamétiques simples X_i, X_i^*, X_j, X_j^* les relations d'identité suivantes ⁽¹⁶⁾ :

$$\{ (X_i \equiv X_i^* = X_j \equiv X_j^*) \}.$$

⁽¹⁶⁾ Nous conservons le même symbolisme, soit \equiv et $=$, que pour les relations entre gènes homologues.

Nous avons :

$$E_{C1}(Z_i Z_j) = E_{C1}(X_i X_j) + E_{C1}(X_i X_j^*) + E_{C1}(X_i^* X_j) + E_{C1}(X_i^* X_j^*) \\ + E_{C1}(X_i D_i) + E_{C1}(X_i^* D_i) + E_{C1}(X_j D_i) + E_{C1}(X_j^* D_i) + E_{C1}(D_i D_i).$$

Les espérances mathématiques, conditionnées par C1, des produits de deux variables aléatoires gamétiques simples ne sont pas nulles; car les variables aléatoires gamétiques étant sous la condition C1 toutes identiques, nous avons :

$$E_{C1}(X_i X_i) + E_{C1}(X_i X_i^*) + E_{C1}(X_i^* X_i) + E_{C1}(X_i^* X_i^*) = 4E(X^2).$$

Les espérances mathématiques, conditionnées par C1, des produits d'une variable aléatoire gamétique simple et d'une variable aléatoire de dominance simple ne sont pas nulles.

Calcul de $E_{C1}(X_i D_i)$.

La condition C1 entraîne $D_i \equiv D_j$, d'où :

$$E_{C1}(X_i D_i) = E_{C1}(X_i D_j) \\ E_{C1}(X_i D_i) = E_c(XD)$$

Nous démontrerions de même que :

$$E_{C1}(X_i^* D_i) = E_c(XD) \\ E_{C1}(X_j D_i) = E_c(XD) \\ E_{C1}(X_j^* D_i) = E_c(XD).$$

En conclusion :

$$E_{C1}(X_i D_i) + E_{C1}(X_i^* D_i) + E_{C1}(X_j D_i) + E_{C1}(X_j^* D_i) = 4E_c(XD).$$

L'espérance mathématique, conditionnée par C1, du produit de deux variables aléatoires de dominance simples n'est pas nulle.

La condition C1 entraîne $D_i \equiv D_j$, d'où :

$$E_c(D_i D_i) = E_c(D^2).$$

En conclusion, l'expression de $E_{C1}(Z_i Z_j)$ est :

$$E_{C1}(Z_i Z_j) = 4E(X^2) + 4E_c(XD) + E_c(D^2).$$

Supposons la condition C2 remplie :

$$\subseteq \{ (G_i \equiv G_i^* \equiv G_j) \circ (G_j^*) \} = \subseteq_2.$$

La probabilité de cette condition est δ_2 . La situation d'identité \mathfrak{S}_2 entraîne entre les variables aléatoires gamétiques simples X_i, X_i^*, X_j, X_j^* les relations d'identité et d'indépendance suivantes :

$$\{ (X_i = X_i^* = X_j) \circ (X_j^*) \}.$$

Les espérances mathématiques $E_{C_2}(X_i X_j^*)$ et $E_{C_2}(X_i^* X_j^*)$, conditionnées par C2, sont nulles car les variables aléatoires gamétiques de chaque produit sont indépendantes et l'espérance mathématique de chacune nulle, sous la condition C2.

Les espérances mathématiques des produits $E_{C_2}(X_i X_j)$ et $E_{C_2}(X_i^* X_j)$ ne sont pas nulles, car les deux variables aléatoires de chaque produit sont identiques entre elles, d'où :

$$E_{C_2}(X_i X_j) = E_{C_2}(X_i^* X_j) = E(X^2).$$

Nous allons démontrer que l'espérance mathématique, conditionnée par C2, du produit de la variable aléatoire gamétique simple X_i et de la variable aléatoire de dominance simple D_i est nulle.

Quand les variables aléatoires gamétiques X_i, X_i^*, X_j sont fixées, la variable aléatoire D_i ne dépend plus que de la variable aléatoire X_j^* , laquelle est indépendante des trois autres. En conséquence X_i et D_i sont des variables aléatoires indépendantes, d'où :

$$E_{C_2}(X_i D_i) = E_{C_2}(X_i) \cdot E_{C_2}(D_i) = 0 \quad \text{car} \quad E_{C_2}(X_i) = 0$$

Il en est de même pour :

$$E_{C_2}(X_i^* D_i) = 0 \quad \text{et pour} \quad E_{C_2}(X_j^* D_i) = 0.$$

Nous allons démontrer que l'espérance mathématique $E_{C_2}(X_i D_i)$ n'est pas nulle. La condition C2 entraîne :

$$X_i = X_i^* = X_j \quad \text{d'où} \quad E_{C_2}(X_j D_i) = E_{C_2}(X_i D_i).$$

Donc les résultats obtenus pour $E_{C_1}(XD)$ nous permettent d'écrire :

$$E_{C_2}(X_i D_i) = E_c(XD).$$

En conséquence :

$$E_{C_2}(X_j D_i) = E_c(XD).$$

Nous allons démontrer que l'espérance mathématique, conditionnée par C2, du produit de deux variables aléatoires de dominance simples est nulle. Quand les variables aléatoires gamétiques X_i, X_i^*, X_j sont fixées, la

variable aléatoire D_i ne dépend plus que de la variable aléatoire X_i^* , laquelle est indépendante des trois autres. En conséquence D_i et D_j sont des variables aléatoires indépendantes, d'où :

$$E_{C2}(D_i D_j) = E_{C2}(D_i) * E_{C2}(D_j)$$

or :

$$E_{C2}(D_i) = p(i - 2s) + q(j - s - t) = 0$$

donc :

$$E_{C2}(D_i D_j) = 0$$

l'expression de E_{C2} est donc :

$$E_{C2}(Z_i Z_j) = 2E(X^2) + E_c(XD).$$

Remarquons que les conditions C3, C4, C5 sont équivalentes à la condition C2, à une permutation près des indices i et j , aussi pouvons-nous écrire :

$$E_{C3}(Z_i Z_j) = 2E(X^2) + E_c(XD)$$

$$E_{C4}(Z_i Z_j) = 2E(X^2) + E_c(XD)$$

$$E_{C5}(Z_i Z_j) = 2E(X^2) + E_c(XD).$$

Supposons la condition C6 remplie :

$$\mathfrak{S} \{ (G_i \equiv G_i^*) \circ (G_j \equiv G_j^*) \} = \mathfrak{S}_6.$$

La probabilité de cette condition est δ_6 . La situation d'identité restreinte \mathfrak{S}_6 entraîne entre les variables aléatoires gamétiques simples X_i , X_i^* , X_j , X_j^* les relations d'identité ou d'indépendance suivantes :

$$\{ (X_i \equiv X_i^*) \circ (X_j \equiv X_j^*) \}.$$

Les espérances mathématiques, conditionnées par C6, des produits de deux variables aléatoires gamétiques simples sont toutes nulles puisque dans chaque produit les deux variables aléatoires gamétiques simples sont indépendantes.

Les espérances mathématiques, conditionnées par C6, des produits d'une variable aléatoire gamétique simple et d'une variable aléatoire de dominance simple sont aussi nulles; car les deux variables aléatoires de chaque produit sont indépendantes, et dans tous les cas l'une d'entre elles est une variable aléatoire gamétique simple d'espérance mathématique nulle. Les termes de la forme $E_{C6}(XD)$ sont tous nuls.

L'espérance mathématique, conditionnée par C6, du produit de deux

variables aléatoires de dominance simple n'est pas nulle. Ces deux variables aléatoires sont indépendantes, donc :

$$E_{c6}(D_1 D_2) = E_{c6}(D_1) \cdot E_{c6}(D_2)$$

or :

$$E_{c6}(D_1) = E_{c6}(D_2) = p(i - 2s) + q(k - 2t)$$

il vient alors :

$$\begin{aligned} E_{c6}(D_1 D_2) &= [E_c(D)]^2 \\ E_{c6}(D_1 D_2) &= p^2(i - 2s)^2 + 2pq(i - 2s)(k - 2t) + q^2(k - 2t)^2 \end{aligned}$$

or :

$$2pq(i - 2s)(k - 2t) = 2pq(pi + qk)^2 = 2pq(j - s - t)^2$$

car :

$$s + t = pi + qk + j$$

donc :

$$E_{c6}(D_1 D_2) = E_p(D^2).$$

En conclusion :

$$E_{c6}(Z_1 Z_2) = E_p(D^2).$$

Ajoutons que la relation $E_p(D^2) = [E_c(D)]^2$ sera démontrée de façon générale p. 86.

Supposons la condition C7 remplie :

$$\subseteq \{ (G_1 \equiv G_1^*) \circ (G_2) \circ (G_2^*) \} = \mathfrak{S}_7.$$

La probabilité de cette condition est δ_7 . La situation d'identité \subseteq entraîne entre les variables aléatoires gamétiques simples X_i , X_i^* , X_j , X_j^* les relations d'identité ou d'indépendance suivantes :

$$\{ (X_i \equiv X_i^*) \circ (X_j) \circ (X_j^*) \}.$$

Les espérances mathématiques, conditionnées par la condition C7, des produits de deux variables aléatoires gamétiques simples sont toutes nulles, car dans chaque produit les termes sont indépendants et l'espérance mathématique de chacun d'entre eux nulle.

Les espérances mathématiques, conditionnées par C7, des produits d'une variable aléatoire gamétique simple et d'une variable aléatoire de dominance sont toutes nulles, car les deux variables aléatoires de chaque produit sont indépendantes, et dans tous les cas l'une d'entre elles est une variable aléatoire gamétique d'espérance mathématique nulle. Les termes de la forme $E_{c7}(XD)$ sont donc tous nuls.

L'espérance mathématique conditionnée, par C7, du produit de deux variables aléatoires de dominance est nulle.

Ces deux variables aléatoires sont indépendantes, donc :

$$E_{C7}(D_I D_J) = E_{C7}(D_I) \cdot E_{C7}(D_J)$$

or :

$$E_{C7}(D_J) = 0$$

il vient alors :

$$E_{C7}(D_I D_J) = 0.$$

En conclusion, nous avons :

$$E_{C7}(Z_I Z_J) = 0$$

Remarquons qu'à la permutation près des indices I et J la condition C8 est équivalente à la condition C7, aussi avons-nous :

$$E_{C8}(Z_I Z_J) = 0.$$

Supposons la condition C9 remplie :

$$\mathfrak{S} \{ (G_I = G_J) \circ (G_I^* = G_J^*) \} = \mathfrak{S}_9.$$

La probabilité de cette condition est δ_9 . La situation d'identité restreinte \mathfrak{S}_9 entraîne entre les variables aléatoires gamétiques simples X_I , X_I^* , X_J , X_J^* les relations d'identité et d'indépendance suivantes :

$$\{ (X_I = X_J) \circ (X_I^* = X_J^*) \}.$$

Les espérances mathématiques $E_{C9}(X_I X_J^*)$ et $E_{C9}(X_I^* X_J)$, conditionnées par C9, sont nulles car les variables aléatoires gamétiques de chaque produit sont indépendantes et l'espérance mathématique de chacune nulle sous la condition C9.

Les espérances mathématiques des produits $E_{C9}(X_I X_J)$ et $E_{C9}(X_I^* X_J^*)$ ne sont pas nulles, car les deux variables aléatoires de chaque produit sont identiques entre elles, d'où :

$$E_{C9}(X_I X_J) = E_{C9}(X_I^* X_J^*) = E(X)^2.$$

Les espérances mathématiques, conditionnées par C9, des produits d'une variable aléatoire gamétique et d'une variable aléatoire de dominance sont toutes nulles car dans chaque zygote les variables aléatoires géniques sont indépendantes, ce qui assure l'indépendance entre variables aléatoires gamétique et de dominance :

$$E_{C9}(X_I D_J) = E_{C9}(X_I^* D_J) = E_{C9}(X_J D_I) = E_{C9}(X_J^* D_I) = 0.$$

L'espérance mathématique, conditionnée par C9, du produit de deux variables aléatoires de dominance n'est pas nulle. Il est immédiat que $D_i = D_j$, d'où :

$$E_{C9}(D_i D_j) = E_p(D^2).$$

En conclusion, il vient :

$$E_{C9}(Z_i Z_j) = 2E(X^2) + E_p(D^2).$$

A la permutation près de L_i et de L_j^* , la condition C12 est équivalente à la condition C9, d'où il vient :

$$E_{C12}(Z_i Z_j) = 2E(X^2) + E_p(D^2).$$

Supposons la condition C10 remplie :

$$\mathfrak{S} \{ (G_i \equiv G_j) \circ (G_i^*) \circ (G_j^*) \} = \mathfrak{S}_{10}.$$

La probabilité de cette condition est δ_{10} . La situation d'identité restreinte \mathfrak{S}_{10} entraîne entre les variables aléatoires gamétiques simples X_i, X_i^*, X_j, X_j^* les relations d'identité et d'indépendance suivantes :

$$\{ (X_i \equiv X_j) \circ (X_i^*) \circ (X_j^*) \}.$$

La condition C10 entraîne $X_i \equiv X_j$, d'où :

$$E_{C10}(X_i X_j) = E_{C10}(X_i^2) = E(X^2).$$

Les espérances mathématiques $E_{C10}(X_i X_j^*), E_{C10}(X_i^* X_j), E_{C10}(X_i^* X_j^*)$, conditionnées par C10, sont nulles car les variables aléatoires gamétiques de chaque produit sont indépendantes et l'espérance mathématique de chacune nulle.

Les espérances mathématiques, conditionnées par C10, des produits d'une variable aléatoire gamétique et d'une variable aléatoire de dominance sont nulles parce que ces deux variables aléatoires sont indépendantes et l'une d'entre elles est d'espérance mathématique nulle :

$$E_{C10}(X_i D_j) = E_{C10}(X_i^* D_j) = E_{C10}(X_j D_i) = E_{C10}(X_j^* D_i) = 0.$$

L'espérance mathématique du produit de deux variables aléatoires de dominance, conditionnée par C10, est nulle, car si l'on fixe l'état génique de G_i (et par conséquent de G_j) la variable aléatoire D_i ne dépend plus que de G_i^* et la variable aléatoire D_j ne dépend plus que de G_j^* . Or G_i^* et G_j^* sont des gènes indépendants, donc les variables aléatoires D_i et D_j sont indépendantes. Dans ces conditions l'espérance mathématique de D_i et celle

de D_i sont nulles puisque à la fois G_i est indépendant de G_i^* et G_i indépendant de G_i^* . D'où :

$$E_{c10}(D_i D_j) = 0.$$

En conclusion, il vient :

$$E_{c10}(Z_i Z_j) = E(X^2).$$

Remarquons que les conditions C11, C13, C14 sont équivalentes à la condition C10, ce qui permet d'écrire :

$$E_{c11}(Z_i Z_j) = E(X^2)$$

$$E_{c13}(Z_i Z_j) = E(X^2)$$

$$E_{c14}(Z_i Z_j) = E(X^2).$$

Supposons la condition C15 remplie :

$$\mathfrak{S} \{ (G_i) \circ (G_i^*) \circ (G_j) \circ (G_j^*) \} = \mathfrak{S}_{15}.$$

La probabilité de cette condition est δ_{15} . Elle entraîne l'indépendance entre Z et Z_j , si bien que :

$$E_{c15}(Z_i Z_j) = E_{c15}(Z_i) \cdot E_{c15}(Z_j) = 0.$$

Les résultats partiels obtenus sont réunis dans le tableau ci-dessous :

Probabilité de la condition C_i	Espérance mathématique conditionnée de $E_{ci}(Z_i Z_j)$
δ_1	$E_{c1}(Z_i Z_j) = 4E(X^2) + 4E_c(XD) + E_c(D^2)$
δ_2	$E_{c2}(Z_i Z_j) = 2E(X^2) + E_c(XD)$
δ_3	$E_{c3}(Z_i Z_j) = 2E(X^2) + E_c(XD)$
δ_4	$E_{c4}(Z_i Z_j) = 2E(X^2) + E_c(XD)$
δ_5	$E_{c5}(Z_i Z_j) = 2E(X^2) + E_c(XD)$
δ_6	$E_{c6}(Z_i Z_j) = 0 + E_p(D^2)$
δ_7	$E_{c7}(Z_i Z_j) = 0$
δ_8	$E_{c8}(Z_i Z_j) = 0$
δ_9	$E_{c9}(Z_i Z_j) = 2E(X^2) + E_p(D^2)$
δ_{10}	$E_{c10}(Z_i Z_j) = E(X^2)$
δ_{11}	$E_{c11}(Z_i Z_j) = E(X^2)$
δ_{12}	$E_{c12}(Z_i Z_j) = 2E(X^2) + E_p(D^2)$
δ_{13}	$E_{c13}(Z_i Z_j) = E(X^2)$
δ_{14}	$E_{c14}(Z_i Z_j) = E(X^2)$
δ_{15}	$E_{c15}(Z_i Z_j) = 0$

Il vient alors après regroupement :

$$\begin{aligned} E(Z_I Z_J) = & \{ 4\delta_1 + 2\delta_2 + 2\delta_3 + 2\delta_4 + 2\delta_5 + 2\delta_9 + \delta_{10} \\ & + \delta_{11} + 2\delta_{12} + \delta_{13} + \delta_{14} \} E(X^2) \\ & + \{ 4\delta_1 + \delta_2 + \delta_3 + \delta_4 + \delta_5 \} E_c(XD) \\ & + \{ \delta_6 + \delta_9 + \delta_{12} \} E_p(D^2) + \delta_1 E_c(D^2). \end{aligned}$$

Expression de l'espérance mathématique des variables aléatoires génotypiques de I et de J, et d'un zygote pris au hasard dans une population :

$$E(Z_I), E(Z_J), E(Z).$$

Soit f_I le coefficient de consanguinité du zygote I, nous avons :

$$E(Z_I) = (1 - f_I)E_p(Z) + f_I E_c(Z_I).$$

Or nous avons posé :

$$E_p(Z_I) = 0$$

donc :

$$E(Z_I) = f_I E_c(Z_I)$$

ce qui donne aussi :

$$E(Z_J) = f_J E_c(Z_J).$$

Nous sommes maintenant en mesure de calculer les termes de correction suivants :

$$\begin{aligned} E(Z_I) \cdot E(Z_J) &= f_I f_J [E_c(Z_I) \cdot E_c(Z_J)] \\ [E(Z_I)]^2 &= f_I^2 [E_c(Z_I)]^2 \\ [E(Z_J)]^2 &= f_J^2 [E_c(Z_J)]^2. \end{aligned}$$

Nous remarquerons que :

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

car :

$$E_c(Y) = 0 \quad \text{d'où} \quad E(Z_I) \cdot E(Z_J) = f_I f_J E_c(D_I) \cdot E_c(D_J)$$

or :

$$E_c(D_I) = E_c(D_J) \quad \text{d'où} \quad E(Z_I) \cdot E(Z_J) = f_I f_J [E_c(D_I)]^2 = f_I f_J [E_c(D)]^2$$

or, par définition :

$$\begin{aligned} E_c(D) &= pi + qk \\ [E_c(D)]^2 &= (pi + qk)^2 \end{aligned}$$

et nous avons de plus :

$$E_p(D^2) = p^2(i - 2s)^2 + 2pq(j - s - t)^2 + q^2(k - 2t)^2$$

avec :

$$\begin{aligned} p^2(i - 2s)^2 &= q^2(pi + qk)^2 \\ 2pq(j - s - t)^2 &= 2pq(pi + qk)^2 \\ q^2(k - 2t)^2 &= p^2(pi + qk)^2 \end{aligned}$$

d'où il vient :

$$\begin{aligned} E_p(D^2) &= (pi + qk)^2 = [E_c(D)]^2 \\ E_p(D^2) &= [E_c(D)]^2. \end{aligned}$$

Nous pouvons donc écrire que :

$$E(Z_i) \cdot E(Z_j) = f_i f_j E_p(D^2).$$

Nous démontrons de la même façon que :

$$\begin{aligned} E(Z_i)^2 &= f_i^2 E_p(D^2) \\ E(Z_j)^2 &= f_j^2 E_p(D^2). \end{aligned}$$

Nous avons la relation $\text{COV}(Z_i Z_j) = E(Z_i Z_j) - [E(Z_i) \cdot E(Z_j)]$ qui peut être explicitée comme suit à l'aide des résultats précédents :

$$\begin{aligned} \text{COV}(Z_i Z_j) &= \{ 4\delta_1 + 2\delta_2 + 2\delta_3 + 2\delta_4 + 2\delta_5 + 2\delta_9 + \delta_{10} \\ &\quad + \delta_{11} + 2\delta_{12} + \delta_{13} + \delta_{14} \} E(X^2) \\ &\quad + \{ 4\delta_1 + \delta_2 + \delta_3 + \delta_4 + \delta_5 \} E_c(XD) \\ &\quad + \{ \delta_6 + \delta_9 + \delta_{12} - f_i f_j \} E_p(D^2) + \delta_1 E_c(D^2) \end{aligned}$$

mais :

$$4\delta_1 + 2\delta_2 + 2\delta_3 + 2\delta_4 + 2\delta_5 + 2\delta_9 + \delta_{10} + \delta_{11} + 2\delta_{12} + \delta_{13} + \delta_{14} = 4\varphi_{11}$$

donc :

$$\begin{aligned} \text{COV}(Z_i Z_j) &= 4\varphi_{11} E(X^2) + \{ 4\delta_1 + \delta_2 + \delta_3 + \delta_4 + \delta_5 \} E_c(XD) \\ &\quad + \{ \delta_6 + \delta_9 + \delta_{12} - f_i f_j \} E_p(D^2) + \delta_1 E_c(D^2). \end{aligned}$$

En posant :

$$\begin{aligned} 4\varphi_{11} &= A, \quad 4\delta_1 + \delta_2 + \delta_3 + \delta_4 + \delta_5 = B, \\ \delta_6 + \delta_9 + \delta_{12} - f_i f_j &= C, \quad \delta_6 + \delta_9 + \delta_{12} = C^*, \quad \delta_1 = D. \end{aligned}$$

Nous avons :

$$\text{COV}(Z_i Z_j) = A E(X^2) + B E_c(XD) + C E_p(D^2) + D E_c(D^2)$$

lorsque les gènes de la population (ou du groupe) sont totalement identiques, nous avons :

$$\delta_1 = 1, \quad f_i = f_j = 1.$$

Tous les autres coefficients d'identité restreinte étant nuls, l'expression COV ($Z_i Z_j$) se modifie ainsi :

$$\text{COV} (Z_i Z_j) = 4E(X^2) + 4E_c(XD) - E_p(D^2) + E_c(D^2).$$

Le développement de ces différents termes donne :

$$\begin{aligned} 4E_c(X^2) &= 4ps^2 + 4qt^2 \\ 4E_c(XD) &= -8ps^2 - 8qt^2 + 4psi + 4qtk \\ -E_p(D^2) &= -(pi + qk)^2 \\ E_c(D^2) &= 4ps^2 + 4qt^2 - 4psi - 4qtk + pi^2 + qk^2 \end{aligned}$$

d'où :

$$\begin{aligned} \text{COV} (Z_i Z_j) &= pi^2 + qk^2 - (pi + qk)^2 = E_c(Z^2) - [E_c(Z)]^2 \\ \text{COV} (Z_i Z_j) &= \text{VAR}_c(Z). \end{aligned}$$

Résultat qui ne saurait surprendre car sous la condition d'identité entre tous les loci $Z^i = Z_j$, d'où :

$$\text{COV} (Z, Z) = \text{VAR} (Z).$$

Dans le schéma de G. Malécot de 1939, les zygotes I et J sont apparentés et non consanguins, ce qui entraîne que I et J sont non consanguins, donc $f_i = f_j = 0$:

$$\begin{aligned} \delta_1 &= 0 \\ \delta_2 &= \delta_3 = \delta_4 = \delta_5 = 0 \\ \delta_6 &= \delta_7 = \delta_8 = 0 \end{aligned}$$

et que :

$$\delta_{12} = \delta_{13} = \delta_{14} = 0$$

si :

$$\delta_9 \neq 0, \quad \delta_{10} \neq 0, \quad \delta_{11} \neq 0.$$

Il est alors évident que :

$$\delta_9 = \varphi\varphi'; \quad \delta_{10} = \varphi(1 - \varphi'); \quad \delta_{11} = \varphi'(1 - \varphi)$$

en considérant φ et φ' les coefficients de parenté de G. Malécot, définis entre les parents de I et J, apparentés.

Le fait que I et J ne soient apparentés que par deux de leurs gènes distincts entraîne (en supposant que $\delta_9 \neq 0$ et que $\delta_{12} = 0$) que la situation :

$$\mathfrak{S} \{ (G_i \equiv G_j) \circ (G_i^* \equiv G_j^*) \},$$

réalisation simultanée de deux événements indépendants, l'un $\mathfrak{S}(G_i \equiv G_j)$, l'autre $\mathfrak{S}(G_i^* \equiv G_j^*)$, a une probabilité de réalisation δ_9 non nulle.

En conséquence :

$$\begin{aligned} \Pr \{ \mathfrak{S}_9 \} &= \Pr \{ \mathfrak{S}(G_i \equiv G_j) \} * \Pr \{ \mathfrak{S}(G_i^* \equiv G_j^*) \} \\ \Pr \{ \mathfrak{S}_9 \} &= \varphi \varphi' = \delta_9. \end{aligned}$$

Nous démontrerions par un même raisonnement que :

$$\begin{aligned} \Pr \{ \mathfrak{S}_{10} \} &= \varphi(1 - \varphi') = \delta_{10} \\ \Pr \{ \mathfrak{S}_{11} \} &= \varphi'(1 - \varphi) = \delta_{11}. \end{aligned}$$

En portant ces valeurs dans l'expression de la COV ($Z_i Z_j$), il vient :

$$\text{COV} (Z_i Z_j) = E(Z_i Z_j) = (2\delta_9 + \delta_{10} + \delta_{11})E(X^2) + \delta_9 E_p(D^2)$$

or :

$$2\delta_9 + \delta_{10} + \delta_{11} = 2\varphi\varphi' + \varphi(1 - \varphi') + \varphi'(1 - \varphi) = \varphi + \varphi'$$

et :

$$\begin{aligned} \delta_9 &= \varphi\varphi' \\ \text{COV} (Z_i Z_j) &= (\varphi + \varphi')E(X^2) + \varphi\varphi'E_p(D^2) \end{aligned}$$

qui est la formule de G. Malécot de 1939.

Dans le cas où les zygotes I et J ne sont plus apparentés que par un seul de leurs gènes homologues, ce qui entraîne, comme condition supplémentaire à celles du schéma de G. Malécot, si $\delta_{11} = 0$, que $\delta_9 = 0$ et $\delta_{10} \neq 0$. Dans un tel schéma, seule peut être réalisée une situation de la forme :

$$\mathfrak{S}_{10} = \mathfrak{S} \{ (G_i \equiv G_j) \circ (G_i^*) \circ (G_j^*) \}.$$

\mathfrak{S}_{10} est la réalisation simultanée de deux événements indépendants, l'un $\mathfrak{S}(G_i \equiv G_j)$, l'autre $\mathfrak{S}(G_i^* \circ G_j^*)$. Or, dans ce schéma :

$$\begin{aligned} \Pr \{ \mathfrak{S}(G_i \equiv G_j) \} &= \varphi \\ \Pr \{ \mathfrak{S}(G_i^* \circ G_j^*) \} &= 1 \end{aligned}$$

d'où :

$$\Pr \{ \mathfrak{S}_{10} \} = \varphi.$$

Autrement dit, tout revient à faire $\varphi' = 0$ dans le schéma de G. Malécot de 1939.

$$\text{COV} (Z_i Z_j) = E(Z_i Z_j) = \varphi E(X^2).$$

2. Expression de la variance génotypique de I et de J.

Nous avons :

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

L'espérance mathématique $E(Z_i^2)$ de la variable aléatoire zygotique Z_i se calcule ainsi :

Soient f_i et f_j les coefficients de consanguinité de I et J et f le coefficient de consanguinité moyen d'une population :

$$E(Z^2) = (1 - f)E_p(Z^2) + fE_c(Z^2)$$

or $Z = Y + D$:

$$\begin{aligned} E(Z^2) &= (1 - f)E_p[(Y + D)^2] + fE_c[(Y + D)^2] \\ E_p[(Y + D)^2] &= E_p(Y^2) + 2E_p(YD) + E_p(D^2) \\ E_p(Y^2) &= 2E(X^2) \\ E_p(YD) &= 0 \\ E_c[(Y + D)^2] &= E_c(Y^2) + 2E_c(YD) + E_c(D^2) \\ E_c(Y^2) &= 4E(X^2) \\ 2E_c(YD) &= 2E_c[(X + X') \cdot D] = 4E_c(XD) \end{aligned}$$

d'où :

$$E(Z^2) = (1 + f)2E(X^2) + 4fE_c(XD) + fE_c(D^2) + (1 - f)E_p(D^2).$$

Ce qui permet d'écrire :

$$E(Z_i^2) = 2(1 + f_i)E(X^2) + 4f_iE_c(XD) + (1 - f_i)E_p(D^2) + f_iE_c(D^2)$$

et :

$$E(Z_j^2) = 2(1 + f_j)E(X^2) + 4f_jE_c(XD) + (1 - f_j)E_p(D^2) + f_jE_c(D^2).$$

Le carré de l'espérance mathématique de la variable aléatoire génotypique a été calculé précédemment :

$$[E(Z_i)]^2 = f_i^2 E_p(D^2)$$

et :

$$[E(Z_j)]^2 = f_j^2 E_p(D^2).$$

De ces résultats nous tirons les expressions des variances de I et de J :

$$\begin{aligned} \text{VAR}(Z_i) &= 2(1 + f_i)E(X^2) + 4f_iE_c(XD) \\ &\quad + (1 - f_i)E_p(D^2) + f_iE_c(D^2) - f_i^2 E_p(D^2) \end{aligned}$$

$$\text{VAR}(Z_i) = 2(1 + f_i)E(X^2) + 4f_i E_c(XD) + (1 - f_i - f_i^2)E_p(D^2) + f_i E_c(D^2)$$

de même il vient pour la variance de Z_j :

$$\text{VAR}(Z_j) = 2(1 + f_j)E(X^2) + 4f_j E_c(XD) + (1 - f_j - f_j^2)E_p(D^2) + f_j E_c(D^2).$$

On remarque que si $f = 0$:

$$E(Z^2) = 2E(X^2) + E_p(D^2)$$

qui est l'expression de l'analyse de la variance donnée par R. A. Fisher en 1918.

Par ailleurs, si $f_i = f_j = 1$, il vient :

$$\begin{aligned}\text{VAR}(Z_i) &= 4E(X^2) + 4E_c(XD) - E_p(D^2) + E_c(D^2) = \text{VAR}_c(Z) \\ \text{VAR}(Z_j) &= \text{VAR}(Z_i).\end{aligned}$$

Il faut souligner que d'avoir par convention centré toutes nos variables aléatoires génétiques sur leurs moyennes dans une population dont tous les gènes sont indépendants conduit à une variance a priori non nulle pour une population dont tous les gènes sont identiques.

Car il s'agit de la *variance a priori* pour le calcul de laquelle nous supposons que les gènes de la population, *tous identiques, peuvent appartenir à une classe d'isoaction, avec pour chacune une probabilité qui était celle de la classe d'isoaction dans la population d'origine*, d'où l'expression en rappelant que $p + q = 1$:

$$\begin{aligned}\text{VAR}_c(Z) &= pi^2 + qk^2 - (pi + qk)^2 \\ &= pq(i + k)^2.\end{aligned}$$

3. Expression des corrélations génotypique et zygotique entre I et J.

La corrélation génotypique peut être obtenue aisément à partir des résultats précédents :

$$r_u = \frac{\text{COV}(Z_i Z_j)}{\sqrt{\text{VAR}(Z_i) \cdot \text{VAR}(Z_j)}}.$$

Remarquons que pour une population dont tous les gènes sont identiques, nous avons :

$$r_u = 1.$$

Si nous supposons que le déterminisme du caractère ne comporte pas d'effet de dominance, les termes suivants s'annulent :

$$E_c(XD), \quad E_p(D^2), \quad E_c(D^2)$$

ce qui conduit à une nouvelle expression de la corrélation entre I et J appelée corrélation zygotique :

$$r_{IJ} = \frac{4\varphi_{IJ}E(X^2)}{\sqrt{2(1+f_I)E(X^2) \cdot 2(1+f_J)E(X^2)}}$$

$$r_{IJ} = \frac{2\varphi_{IJ}}{\sqrt{(1+f_I)(1+f_J)}}.$$

Il est remarquable que cette corrélation zygotique ne dépende que des coefficients d'identité restreinte.

Les formules obtenues dans le cas d'une classe d'homologues peuvent être généralisées au cas d'un nombre quelconque de classes d'homologues en redéfinissant chacune des composantes de la variance ou de la covariance comme la somme de la composante de chaque classe d'homologues, pourvu qu'il n'y ait pas d'épistasie et que l'équilibre asymptotique ait été atteint dans le cas de linkage.

V. — LES CORRÉLATIONS GÉNÉTIQUES ENTRE APPARENTÉS QUELCONQUES DANS LE CAS D'EFFET MATERNEL GÉNÉTIQUE

La valeur d'un caractère quantitatif chez un individu est la résultante de son patrimoine héréditaire et du milieu où il a vécu. Or parmi les effets d'environnement il peut être souhaitable d'estimer l'effet maternel génétique. C'est le résultat qui est présenté.

Lorsque la dépendance du jeune vis-à-vis de sa mère est aussi étroite que dans le cas des mammifères, les qualités maternelles ne manquent pas d'avoir sur le phénotype du jeune une certaine influence.

Dans les problèmes d'élevage ou zootechniques il est souvent nécessaire d'estimer la potentialité génétique d'un caractère et l'influence du génotype de la mère sur ce caractère. C'est le cas, par exemple, de la vitesse de croissance qui dépend en particulier de l'aptitude laitière de la mère.

Le schéma le plus général consiste à supposer que le caractère mesuré chez le jeune est sous la dépendance d'un ensemble fini de classes d'homologues, que l'effet génétique maternel est sous la dépendance d'un autre

ensemble fini de classes d'homologues; les deux ensembles pouvant ne pas être disjoints.

Soient L_1 et L_2 le premier et le second ensemble de loci. Soient $I_1 \subset L_1$ et $I_2 \subset L_2$ les sous-ensembles de loci communs à L_1 et L_2 , avec $I_1 \equiv I_2$.

Si l'on considère l'ensemble produit $L_1 \times L_2$, c'est-à-dire les couples ordonnés formés de deux classes d'homologues, le premier pris dans L_1 , le second dans L_2 , ils se répartissent en deux sous-ensembles, celui pour lequel les deux classes d'homologues sont distinctes, celui pour lequel les deux classes d'homologues ne sont que la même classe d'homologues répétée deux fois.

Il est donc plus efficace d'étudier d'abord deux schémas simples :

— le premier consiste à supposer que le caractère en cause est déterminé par une seule série biallélique L et que l'effet maternel sur ce caractère est déterminé par une autre et seule série biallélique \mathcal{L} ;

— le second consiste à supposer que le caractère en cause est déterminé par une seule série biallélique L et que l'effet maternel sur ce caractère est déterminé par la même série biallélique L ;

— puis, nous nous proposons d'exposer une étude du schéma le plus général en supposant néanmoins que l'état stationnaire d'indépendance a pu être atteint malgré le linkage susceptible d'exister entre les classes d'homologues en cause.

I. Nous supposons que le caractère en cause est déterminé par une série biallélique L , et que l'effet maternel sur ce caractère est déterminé par une autre et seule série biallélique \mathcal{L} .

Soient deux individus I et J apparentés suivant le schéma général de l'étude du coefficient de corrélation génotypique dans une population quelconque. Ajoutons comme convention que N est la mère de I , et que Q est la mère de J . A propos des individus N et Q il est possible de définir l'ensemble des coefficients d'identité restreinte, comme nous l'avons défini pour I et J . En particulier, nous considérons les ensembles des coefficients d'identité restreinte définis entre les couples de zygotes suivants : IJ , NQ , IQ , IN , JN , JQ . Nous numérotions ces couples de 1 à 6. Nous avons précédemment posé les coefficients A , B , C , C^* , D égaux à des fonctions linéaires de δ_i . Nous indiquons ces coefficients de la manière suivante :

- $A_1, B_1, C_1, C^*_1, D_1$ pour le couple IJ ;
- $A_2, B_2, C_2, C^*_2, D_2$ pour le couple NQ ;
- $A_3, B_3, C_3, C^*_3, D_3$ pour le couple IQ ;
- $A_4, B_4, C_4, C^*_4, D_4$ pour le couple IN ;

A5, B5, C5, C*5, D5 pour le couple JN;
et A6, B6, C6, C*6, D6 pour le couple JQ.

Nous attachons à L la variable aléatoire gamétique X, et à \mathfrak{L} la variable aléatoire gamétique \mathfrak{X} . Nous définissons de plus deux variables aléatoires zygotiques :

$$Y = X + X^* \quad \text{et} \quad \mathfrak{Y} = \mathfrak{X} + \mathfrak{X}^*;$$

deux variables aléatoires génotypiques :

$$Z = X + X^* + D \quad \text{et} \quad \mathfrak{Z} = \mathfrak{X} + \mathfrak{X}^* + \mathfrak{D}$$

et deux variables aléatoires de dominance :

$$D \quad \text{et} \quad \mathfrak{D}.$$

Elles satisferont aux relations habituelles que nous rappelons brièvement :

$$\begin{aligned} E(X) &= ps + qt = 0 \\ E(X^2) &= ps^2 + qt^2 \\ E(Y) &= p^2s + 2pq(s + t) + q^2t = 0 \\ E_p(Y^2) &= 2E(X^2) & E_c(Y^2) &= 4E(X^2) \\ E_p(Z) &= p^2i + 2pqj + q^2k = 0 \\ E_c(Z) &= pi + qk \neq 0 \\ E_p(Z^2) &= p^2i^2 + 2pqj^2 + q^2k^2 \neq 0 \\ E_c(Z^2) &= pi^2 + qk^2 \neq 0 \\ E_c(XZ) &= psi + qkt \neq 0 \\ E_p(D) &= p^2(i - 2s) + 2pq(j - s - t) + q^2(k - 2t) = 0 \\ E_c(D) &= p(i - 2s) + q(k - 2t) = E_c(Z) \neq 0 \\ E_p(D^2) &= p^2(i - 2s)^2 + q^2(k - 2t)^2 + 2pq(j - s - t)^2 \neq 0 \\ E_c(D^2) &= p(i - 2s)^2 + q(k - 2t)^2 \neq 0. \end{aligned}$$

En supposant que les constantes p, q, s, t, i, j, k prennent d'autres valeurs quelconques, nous pourrions définir les mêmes relations pour les variables aléatoires $\mathfrak{X}, \mathfrak{Z}, \mathfrak{D}$ en indiquant p, q, s, t, i, j, k d'une apostrophe. Nous introduisons de plus les deux variables aléatoires P_I et P_J attachées aux individus I et J et définies par la relation suivante :

$$\begin{aligned} P_I &= Z_I + \mathfrak{Z}_N \\ P_J &= Z_J + \mathfrak{Z}_Q. \end{aligned}$$

Le coefficient de corrélation phénotypique entre I et J en supposant l'influence du milieu réduit à l'influence génétique maternelle est :

$$r_{II} = \frac{\text{COV}(P_I P_J)}{\sqrt{\text{VAR}(P_I) \cdot \text{VAR}(P_J)}}.$$

Dans l'hypothèse de l'absence d'épistasie entre les loci de l'individu et entre les loci maternels qui influent sur le même caractère, et dans l'hypothèse supplémentaire que l'équilibre asymptotique ait été atteint entre tous les couples de loci malgré le linkage susceptible d'exister, nous avons indépendance entre Z_i et \bar{Z}_N , Z_i et \bar{Z}_Q entre Z_j et \bar{Z}_Q , Z_j et \bar{Z}_N donc :

$$\text{COV}(P_i P_j) = \text{COV}(Z_i Z_j) + \text{COV}(\bar{Z}_N \bar{Z}_Q).$$

Or, nous avons :

$$\text{COV}(Z_i Z_j) = A1 E(X^2) + B1 E_c(XD) + C1 E_p(D^2) + D1 E_c(D^2)$$

$$\text{COV}(\bar{Z}_N \bar{Z}_Q) = A2 E(\mathfrak{X}^2) + B2 E_c(\mathfrak{X}\mathfrak{D}) + C2 E_p(\mathfrak{D}^2) + D2 E_c(\mathfrak{D}^2).$$

De même :

$$\text{VAR}(P_i) = \text{VAR}(Z_i) + \text{VAR}(\bar{Z}_N)$$

$$\text{VAR}(P_j) = \text{VAR}(Z_j) + \text{VAR}(\bar{Z}_Q).$$

Or, nous avons précédemment calculé ces variances, d'où :

$$r_D = \frac{\text{COV}(Z_i Z_j) + \text{COV}(\bar{Z}_N \bar{Z}_Q)}{\sqrt{[\text{VAR}(Z_i) + \text{VAR}(\bar{Z}_N)] * [\text{VAR}(Z_j) + \text{VAR}(\bar{Z}_Q)]}}.$$

Une telle expression serait longue à développer. Supposons que l'effet génétique du locus L et celui du locus ℓ soient strictement additifs, nous avons l'expression suivante de la corrélation phénotypique :

$$r_D = \frac{A1 E(X^2) + A2 E(\mathfrak{X}^2)}{\sqrt{[2(1-f_i)E(X^2) + 2(1-f_N)E(\mathfrak{X}^2)] \cdot [2(1-f_j)E(X^2) + 2(1-f_Q)E(\mathfrak{X}^2)]}}$$

avec :

$$A1 = 4\varphi_{ii} \quad \text{et} \quad A2 = 4\varphi_{NQ}.$$

II. Nous supposons maintenant que le caractère en cause est contrôlé génétiquement par une seule série biallélique L, et que l'effet maternel génétique sur ce caractère est dû à cette même série biallélique. Le schéma conduit à créer deux variables aléatoires gamétiques X et \mathfrak{X} attachées à un même gène, donc non indépendantes. Nous définissons de plus deux variables aléatoires zygotiques :

$$Y = X + X^* \quad \text{et} \quad \mathfrak{Y} = \mathfrak{X} + \mathfrak{X}^*$$

deux variables aléatoires génotypiques :

$$Z = Y + D \quad \text{et} \quad \mathfrak{Z} = \mathfrak{Y} + \mathfrak{D}$$

et deux variables aléatoires de dominance :

$$D \quad \text{et} \quad \mathfrak{D}$$

comme nous l'avons fait précédemment. Ces variables aléatoires satisfont aux relations habituelles que nous avons présentées page 70, et pour lesquelles $p = p'$, etc.

Nous introduisons les variables aléatoires P_i et P_j telles que :

$$P_i = Z_i + \bar{Z}_N$$

$$P_j = Z_j + \bar{Z}_Q.$$

Le coefficient de corrélation phénotypique entre I et J , en supposant l'influence du milieu réduit à l'effet génétique maternel, est égal à :

$$r_{ij} = \frac{\text{COV}(P_i P_j)}{\sqrt{\text{VAR}(P_i) * \text{VAR}(P_j)}}$$

$$\text{COV}(P_i P_j) = \text{COV}(Z_i Z_j) + \text{COV}(\bar{Z}_N \bar{Z}_Q) + \text{COV}(Z_i \bar{Z}_Q) + \text{COV}(Z_j \bar{Z}_N).$$

Le calcul de $\text{COV}(Z_i Z_j)$ et de $\text{COV}(\bar{Z}_N \bar{Z}_Q)$ a déjà été réalisé précédemment. *Le calcul de $\text{COV}(Z_i \bar{Z}_Q)$ et celui de $\text{COV}(Z_j \bar{Z}_N)$ peuvent être conduits en remarquant que les variables aléatoires génotypiques Z_i et \bar{Z}_Q dépendent des états de quatre gènes homologues appartenant à deux zygotes distincts I et Q , et qu'il en est de même pour Z_j et \bar{Z}_N . La connaissance des coefficients d'identité restreinte pour chacun des couples d'individus I , Q et J , N permet d'attribuer à chacune des situations d'identité ou d'indépendance des gènes G_i , G_i^* , G_Q , G_Q^* ou des loci G_i , G_j^* , G_N , G_N^* une probabilité, d'où le calcul suivant :*

$$\text{COV}(Z_i \bar{Z}_Q) = E(Z_i \bar{Z}_Q) - E(Z_i) \cdot E(\bar{Z}_Q).$$

Le calcul de $E(Z_i \bar{Z}_Q)$ est conduit en utilisant le théorème des moyennes conditionnées comme dans le cas de $E(Z_i Z_j)$.

Les résultats sont les suivants :

Probabilité de la condition C_i	Espérance mathématique conditionnée de $E_{ci}(Z_i \bar{Z}_Q)$
δ_1	$E_{c1}(Z_i \bar{Z}_Q) = 4E(X\mathfrak{X}) + 2E_c(X\mathfrak{D}) + 2E_c(\mathfrak{X}\mathfrak{D}) + E_c(\mathfrak{D}\mathfrak{D})$
δ_2	$E_{c2}(Z_i \bar{Z}_Q) = 2E(X\mathfrak{X}) + E_c(\mathfrak{X}\mathfrak{D})$
δ_3	$E_{c3}(Z_i \bar{Z}_Q) = 2E(X\mathfrak{X}) + E_c(\mathfrak{X}\mathfrak{D})$
δ_4	$E_{c4}(Z_i \bar{Z}_Q) = 2E(X\mathfrak{X}) + E_c(X\mathfrak{D})$
δ_5	$E_{c5}(Z_i \bar{Z}_Q) = 2E(X\mathfrak{X}) + E_c(X\mathfrak{D})$
δ_6	$E_{c6}(Z_i \bar{Z}_Q) = + E_p(\mathfrak{D}\mathfrak{D})$
δ_7	$E_{c7}(Z_i \bar{Z}_Q) = 0$
δ_8	$E_{c8}(Z_i \bar{Z}_Q) = 0$
δ_9	$E_{c9}(Z_i \bar{Z}_Q) = 2E(X\mathfrak{X}) + E_p(\mathfrak{D}\mathfrak{D})$
δ_{10}	$E_{c10}(Z_i \bar{Z}_Q) = E(X\mathfrak{X})$
δ_{11}	$E_{c11}(Z_i \bar{Z}_Q) = E(X\mathfrak{X})$
δ_{12}	$E_{c12}(Z_i \bar{Z}_Q) = 2E(X\mathfrak{X}) + E_p(\mathfrak{D}\mathfrak{D})$
δ_{13}	$E_{c13}(Z_i \bar{Z}_Q) = E(X\mathfrak{X})$
δ_{14}	$E_{c14}(Z_i \bar{Z}_Q) = E(X\mathfrak{X})$
δ_{15}	$E_{c15}(Z_i \bar{Z}_Q) = 0$

Il vient après regroupement et en posant :

$$2\delta_1 + \delta_2 + \delta_3 = \tilde{B}$$

$$2\delta_1 + \delta_4 + \delta_5 = \tilde{B}$$

$$E(Z_i Z_o) = A3 E(X\mathfrak{X}) + \tilde{B}3 E_c(X\mathcal{D}) + \tilde{B}3 E_c(\mathfrak{X}\mathcal{D}) + C*3 E_p(D\mathcal{D}) + D3 E_c(D\mathcal{D})$$

avec :

$$E(X\mathfrak{X}) = p ss' + q tt'$$

$$E_c(X\mathcal{D}) = ps(i' - 2s') + qt(k' - 2t')$$

$$E_c(\mathfrak{X}\mathcal{D}) = ps'(i - 2s) + qt'(k - 2t)$$

$$E_p(D\mathcal{D}) = p^2(i - 2s)(i' - 2s') + 2pq(j - s - t)(j' - s' - t') \\ + q^2(k - 2t)(k' - 2t')$$

$$E_c(D\mathcal{D}) = p(i - 2s)(i' - 2s') + q(k - 2t)(k' - 2t')$$

$E(X\mathfrak{X}) = p ss' + q tt'$ peut surprendre, il s'agit d'une expression analogue à $E(X^2) = ps^2 + qt^2$, c'est-à-dire que l'on considère le même gène, et les deux variables qui lui sont attachées.

De même nous obtenons pour $E(Z_j Z_N)$ l'expression :

$$E(Z_j Z_N) = A5 E(X\mathfrak{X}) + \tilde{B}5 E_c(X\mathcal{D}) + \tilde{B}5 E_c(\mathfrak{X}\mathcal{D}) + C*5 E_p(D\mathcal{D}) + D5 E_c(D\mathcal{D})$$

L'espérance mathématique du produit des deux variables aléatoires P_i et P_j est donc :

$$E(P_i P_j) = A1 E(X^2) + B1 E_c(X\mathcal{D}) + C*1 E_p(D^2) + D1 E_c(D^2) \\ + A2 E(\mathfrak{X}^2) + B2 E_c(\mathfrak{X}\mathcal{D}) + C*2 E_p(\mathcal{D}^2) + D2 E_c(\mathcal{D}^2) \\ + (A3 + A5)E(X\mathfrak{X}) + (\tilde{B}3 + \tilde{B}5)E_c(X\mathcal{D}) + (\tilde{B}3 + \tilde{B}5)E_c(\mathfrak{X}\mathcal{D}) \\ + (C*3 + C*5)E_p(D\mathcal{D}) + (D3 + D5)E_c(D\mathcal{D}).$$

Il nous reste maintenant à calculer le terme de correction $E(P_i)*E(P_j)$.

$$E(P_i) = E(Z_i + Z_N) = E(Z_i) + E(Z_N)$$

$$E(Z_i) = (1 - f_i)E_p(Z) + f_i E_c(Z) = f_i E_c(Z)$$

or :

$$f_i E_c(Z) = f_i E_c(D)$$

donc :

$$E(P_i) = f_i E_c(D) + f_N E_c(\mathcal{D})$$

de même :

$$E(P_j) = f_j E_c(D) + f_o E_c(\mathcal{D})$$

d'où :

$$E(P_I) \cdot E(P_I) = f_I f_I E_c(D) \cdot E_c(D) + f_I f_Q E_c(D) \cdot E_c(\mathcal{D}) \\ + f_I f_N E_c(D) \cdot E_c(\mathcal{D}) + f_N f_Q E_c(\mathcal{D}) \cdot E_c(\mathcal{D})$$

or, nous avons par ailleurs :

$$E_c(D) \cdot E_c(D) = E_s(D^2) = E_p(D^2) \\ E_c(\mathcal{D}) \cdot E_c(\mathcal{D}) = E_s(\mathcal{D}^2) = E_p(\mathcal{D}^2)$$

et il est aisé de montrer que :

$$E_c(D) \cdot E_c(\mathcal{D}) = E_s(D\mathcal{D}) = E_p(D\mathcal{D})$$

s étant le symbole de la condition suivante : les deux paires de loci sont simultanément en état d'identité. D'où :

$$E(P_I) \cdot E(P_I) = f_I f_I E_p(D^2) + (f_I f_Q + f_I f_N) E_p(D\mathcal{D}) + f_N f_Q E_p(\mathcal{D}^2).$$

La covariance entre les variables aléatoires P_I , P_I a donc l'expression suivante :

$$\text{COV}(P_I, P_I) = A1 E(X^2) + B1 E_c(XD) + C1 E_p(D^2) + D1 E_c(D^2) \\ + A2 E(\mathcal{X}^2) + B2 E_c(\mathcal{X}\mathcal{D}) + C2 E_p(\mathcal{D}^2) + D2 E_c(\mathcal{D}^2) \\ + (A3 + A5) E(X\mathcal{X}) + (\tilde{B}3 + \tilde{B}5) E_c(X\mathcal{D}) + (\tilde{B}3 + \tilde{B}5) E_c(\mathcal{X}\mathcal{D}) \\ + (C3 + C5) E_p(D\mathcal{D}) + (D3 + D5) E_c(D\mathcal{D}).$$

Il nous reste à calculer les variances des variables aléatoires phénotypiques P_I et P_I :

$$\text{VAR}(P_I) = \text{VAR}(Z_I + \mathcal{Z}_N) = \text{VAR}(Z_I) + \text{VAR}(\mathcal{Z}_N) + 2\text{COV}(Z_I \mathcal{Z}_N).$$

Or, nous connaissons déjà l'expression de la variance de Z_I et de celle de \mathcal{Z}_N , de plus l'expression de la covariance entre Z_I et \mathcal{Z}_N est la suivante puisque les variables aléatoires Z_I et \mathcal{Z}_N dépendent des états de quatre gènes homologues dont deux appartiennent au zygote I et deux au zygote N; or nous avons précédemment défini pour ces quatre gènes homologues des situations d'identité restreinte et leurs coefficients. Aussi, après application du théorème des moyennes conditionnées, nous pouvons écrire :

$$\text{COV}(Z_I \mathcal{Z}_N) = A4 E(X\mathcal{X}) + \tilde{B}4 E_c(X\mathcal{D}) + \tilde{B}4 E_c(\mathcal{X}\mathcal{D}) + C*4 E_p(D\mathcal{D}) \\ + D4 E_c(D\mathcal{D}) - f_I f_N E_p(D\mathcal{D})$$

d'où l'expression de la variance de P_I suivante :

$$\text{VAR}(P_I) = \text{VAR}(Z_I) + \text{VAR}(\mathcal{Z}_N) + 2\text{COV}(Z_I \mathcal{Z}_N)$$

$$\begin{aligned}\text{VAR}(P_i) = & 2(1 + f_i)E(X^2) + 4f_iE_c(XD) + (1 - f_i - f_i^2)E_p(D^2) + f_iE_c(D^2) \\ & + 2(1 - f_N)E(X^2) + 4f_NE_c(XD) + (1 - f_N - f_N^2)E_p(D^2) + f_NE_c(D^2) \\ & + 2(A4)E(XX) + 2(\overset{<}{B}4)E_c(XD) + 2(\overset{>}{B}4)E_c(XD) \\ & + 2(C^*4)E_p(DD) + 2(D4)E_c(DD) - 2f_i f_N E_p(DD).\end{aligned}$$

Un raisonnement simple permet de retrouver l'expression de la variance de la variable aléatoire P_i quand les variables aléatoires génotypiques Z_1 et Z_N sont indépendantes (quelle qu'en soit la cause biologique).

Si Z_1 est indépendante de Z_N , nous avons :

$$A4 = \overset{<}{B}4 = \overset{>}{B}4 = D4 = 0$$

de plus :

$$C^*4 = \delta_8 + \delta_9 + \delta_{12} \quad \text{avec} \quad \delta_9 = \delta_{12} = 0$$

et $\delta_8 = f_i f_N$, ce qui est immédiat si l'on se reporte aux situations d'identité restreinte auxquelles sont attachés les coefficients $\delta_8, \delta_{12}, \delta_8$.

En conséquence $2(C^*4)E_p(DD) = 2f_i f_N E_p(DD)$ ce qui donne pour la variance de P_i l'expression du schéma précédent.

Il est aisé d'obtenir l'expression de la variance de la variable aléatoire phénotypique P_i .

Nous sommes en mesure d'exprimer le coefficient de corrélation phénotypique entre les individus I et J en remplaçant les termes de l'expression ci-dessous par leur valeur analytique que nous avons obtenus :

$$r_u = \frac{\text{COV}(P_i P_j)}{\text{VAR}(P_i) \cdot \text{VAR}(P_j)}.$$

Supposons que le déterminisme génétique du caractère soit strictement additif, il vient alors :

$$\begin{aligned}\text{COV}(P_i P_j) &= A1E(X^2) + A2E(X^2) + (A3 + A5)E(XX) \\ \text{VAR}(P_i) &= 2(1 + f_i)E(X^2) + 2(1 + f_N)E(X^2) + 2(A4)E(XX) \\ \text{VAR}(P_j) &= 2(1 + f_j)E(X^2) + 2(1 + f_N)E(X^2) + 2(A6)E(XX).\end{aligned}$$

III. *Il nous est maintenant facile d'aborder l'étude du schéma le plus général. Le caractère mesuré chez le jeune est sous la dépendance d'un ensemble fini de classes d'homologues, l'effet génétique maternel est sous la dépendance d'un autre ensemble fini de classes d'homologues, les deux ensembles pouvant être non disjoints.* Soient L_1 et L_2 le premier et le second ensemble de classes d'homologues. Soient $l_1 \subset L_1$ et $l_2 \subset L_2$ les sous-ensembles de classes d'homologues communes à L_1 et L_2 , avec $l_1 = l_2$, autrement dit :

$$L_1 \cap L_2 = l_1 = l_2.$$

Si nous considérons l'ensemble produit $L_1 \times L_2$, c'est-à-dire les couples ordonnés formés de deux classes d'homologues, le premier pris dans L_1 , le second dans L_2 , ils se répartissent en deux sous-ensembles, celui pour lequel les deux classes d'homologues sont distinctes, celui pour lequel les deux classes d'homologues ne sont que la même répétée deux fois. Pour identifier les classes d'homologues nous les indiquerons de 1 à k , en remarquant que si l'ensemble L_1 compte m éléments, l'ensemble L_2 en compte n , les ensembles l_j et l_2 en comptent g , nous avons la relation :

$$k = m + n - g.$$

k est le nombre de classes d'homologues de l'ensemble $L_1 \cup L_2$. Ainsi pour tout couple de classes d'homologues pris dans $L_1 \times L_2$, nous avons deux indices différents si les deux classes sont distinctes, deux indices égaux s'il s'agit de la même répétée deux fois.

Soit α l'indice des classes d'homologues appartenant à L_1 , donc présents chez le jeune. Soit β l'indice de celles appartenant à L_2 , donc présents chez la mère. α prend m valeurs distinctes et β , n .

Nous introduisons les deux nouvelles variables aléatoires phénotypiques suivantes : \hat{P}_I attachée à I et \hat{P}_J attachée à J. Elles satisfont aux relations suivantes :

$$\hat{P}_I = \sum_{\alpha}^{\alpha} \hat{Z}_I = \sum_{\beta}^{\beta} \hat{Z}_N$$

\sum_{α} étant le symbole de la sommation de tous les termes \hat{Z}_I^{α} pour toutes les valeurs que peut prendre l'indice α :

$$\hat{P}_J = \sum_{\alpha}^{\alpha} \hat{Z}_J + \sum_{\beta}^{\beta} \hat{Z}_Q.$$

Posons :

$$\hat{Z}_I = \sum_{\alpha}^{\alpha} \hat{Z}_I$$

$$\hat{Z}_J = \sum_{\alpha}^{\alpha} \hat{Z}_J$$

$$\hat{Z}_N = \sum_{\beta}^{\beta} \hat{Z}_N$$

$$\hat{Z}_Q = \sum_{\beta}^{\beta} \hat{Z}_Q.$$

La covariance phénotypique entre I et J s'obtient de la manière suivante :

$$\begin{aligned}\text{COV}(\widehat{P}_I \widehat{P}_J) &= E(\widehat{P}_I \widehat{P}_J) - E(P_I) \cdot E(P_J) \\ E(\widehat{P}_I \widehat{P}_J) &= E[(\widehat{Z}_I + \widehat{Z}_N) \cdot (\widehat{Z}_J + \widehat{Z}_Q)] \\ E(\widehat{P}_I \widehat{P}_J) &= E(\widehat{Z}_I \widehat{Z}_J) + E(\widehat{Z}_I \widehat{Z}_Q) + E(\widehat{Z}_N \widehat{Z}_J) + E(\widehat{Z}_N \widehat{Z}_Q) \\ E(\widehat{Z}_I \widehat{Z}_J) &= E\left[\left(\sum_{\alpha} \widehat{Z}_{I\alpha}\right) \cdot \left(\sum_{\alpha} \widehat{Z}_{J\alpha}\right)\right] = E\left(\sum_{\alpha\alpha} \widehat{Z}_{I\alpha} \widehat{Z}_{J\alpha}\right).\end{aligned}$$

Le symbole $\sum_{\alpha\alpha}$ est celui d'une somme dont les termes sont des produits du type $\widehat{Z}_{I\alpha} \widehat{Z}_{J\alpha}$, et tels que le premier α est attribuable au gène présent dans I et le second α au gène présent dans J, chacun des indices α prenant indépendamment toutes les valeurs qui leur sont possibles.

$$E(\widehat{Z}_{I\alpha} \widehat{Z}_{J\alpha}) = \sum_{\alpha\alpha} E(\widehat{Z}_{I\alpha} \widehat{Z}_{J\alpha})$$

si $\alpha \neq \alpha$, les variables aléatoires génotypiques \widehat{Z}_I et \widehat{Z}_J sont indépendantes si l'équilibre asymptotique a été atteint malgré le linkage susceptible d'exister entre les classes d'homologues en cause :

$$E(\widehat{Z}_{I\alpha} \widehat{Z}_{J\alpha}) = E(\widehat{Z}_{I\alpha}) \cdot E(\widehat{Z}_{J\alpha})$$

si $\alpha = \alpha$,

$$E(\widehat{Z}_{I\alpha} \widehat{Z}_{J\alpha}) \neq 0$$

c'est une expression que nous avons étudiée précédemment.

Il en est de même pour $E(\widehat{Z}_N \widehat{Z}_Q)$, d'où :

$$\begin{aligned}E(\widehat{Z}_I \widehat{Z}_J) &= \sum_{\substack{\alpha\alpha \\ \alpha=\alpha}} E(\widehat{Z}_{I\alpha} \widehat{Z}_{J\alpha}) + \sum_{\substack{\alpha\alpha \\ \alpha \neq \alpha}} [E(\widehat{Z}_{I\alpha}) \cdot E(\widehat{Z}_{J\alpha})] \\ E(\widehat{Z}_N \widehat{Z}_Q) &= \sum_{\substack{\beta\beta \\ \beta=\beta}} E(\widehat{Z}_{N\beta} \widehat{Z}_{Q\beta}) + \sum_{\substack{\beta\beta \\ \beta \neq \beta}} [E(\widehat{Z}_{N\beta}) \cdot E(\widehat{Z}_{Q\beta})].\end{aligned}$$

Il nous reste à étudier les termes $E(\widehat{Z}_I \widehat{Z}_Q)$ et $E(\widehat{Z}_J \widehat{Z}_N)$:

$$E(\widehat{Z}_I \widehat{Z}_Q) = E\left[\left(\sum_{\alpha} \widehat{Z}_{I\alpha}\right) \cdot \left(\sum_{\beta} \widehat{Z}_{Q\beta}\right)\right] = E\left[\sum_{\alpha\beta} \widehat{Z}_{I\alpha} \widehat{Z}_{Q\beta}\right]$$

$$E(\widehat{Z}_t \widehat{Z}_0) = E\left(\sum_{\substack{\alpha\beta \\ \alpha=\beta}}^{\alpha} Z_t \bar{Z}_0\right) + E\left(\sum_{\substack{\alpha\beta \\ \alpha \neq \beta}}^{\alpha} Z_t \bar{Z}_0\right)$$

$$E(\widehat{Z}_t \widehat{Z}_0) = \sum_{\substack{\alpha\beta \\ \alpha=\beta}} E(Z_t \bar{Z}_0) + \sum_{\substack{\alpha\beta \\ \alpha \neq \beta}} E(Z_t \bar{Z}_0)$$

de même :

$$E(\widehat{Z}_j \widehat{Z}_N) = \sum_{\substack{\alpha\beta \\ \alpha=\beta}} [E(Z_j \bar{Z}_N)] + \sum_{\substack{\alpha\beta \\ \alpha \neq \beta}} [E(Z_j \bar{Z}_N)].$$

Remarquons que dans ces deux dernières expressions le premier terme du second membre est une somme de n termes (si $m > n$) et que le second terme est une somme de $m(m-1)$ termes.

En conséquence :

$$E(\widehat{P}_i \widehat{P}_j) = \sum_{\substack{\alpha\alpha \\ \alpha=\alpha}} [E(Z_i \bar{Z}_j)] + \sum_{\substack{\alpha\alpha \\ \alpha \neq \alpha}} [E(Z_i) \cdot E(\bar{Z}_j)]$$

$$+ \sum_{\substack{\beta\beta \\ \beta=\beta}} [E(\bar{Z}_N \bar{Z}_0)] + \sum_{\substack{\beta\beta \\ \beta \neq \beta}} [E(\bar{Z}_N) \cdot E(\bar{Z}_0)]$$

$$+ \sum_{\substack{\alpha\beta \\ \alpha=\beta}} [E(Z_t \bar{Z}_0)] + \sum_{\substack{\alpha\beta \\ \alpha \neq \beta}} [E(Z_t \bar{Z}_0)]$$

$$+ \sum_{\substack{\alpha\beta \\ \alpha=\beta}} [E(Z_j \bar{Z}_N)] + \sum_{\substack{\alpha\beta \\ \alpha \neq \beta}} [E(Z_j \bar{Z}_N)].$$

Calculons dès à présent le terme correcteur $E(\widehat{P}_i) \cdot E(\widehat{P}_j)$.

L'espérance mathématique de la variable aléatoire phénotypique \widehat{P}_i est :

$$E(P_i) = E\left[\left(\sum_{\alpha}^{\alpha} Z_i\right) + \left(\sum_{\beta}^{\beta} \bar{Z}_N\right)\right]$$

$$= \sum_{\alpha} E(Z_i) + \sum_{\beta} E(\bar{Z}_N).$$

de même :

$$E(\widehat{P}_j) = \sum_{\alpha} E(Z_j) + \sum_{\beta} E(\bar{Z}_0)$$

d'où :

$$\begin{aligned}
 E(\widehat{P}_I) \cdot E(\widehat{P}_j) &= \left[\sum_{\alpha} E(\overset{\alpha}{Z}_I) + \sum_{\beta} E(\overset{\beta}{Z}_N) \right] \cdot \left[\sum_{\alpha} E(\overset{\alpha}{Z}_I) + \sum_{\beta} E(\overset{\beta}{Z}_O) \right] \\
 &= \left[\sum_{\alpha} E(\overset{\alpha}{Z}_I) \right] \cdot \left[\sum_{\alpha} E(\overset{\alpha}{Z}_I) \right] + \left[\sum_{\alpha} E(\overset{\alpha}{Z}_I) \right] \cdot \left[\sum_{\beta} E(\overset{\beta}{Z}_O) \right] \\
 &= \left[\sum_{\beta} E(\overset{\beta}{Z}_N) \right] \cdot \left[\sum_{\alpha} E(\overset{\alpha}{Z}_I) \right] + \left[\sum_{\beta} E(\overset{\beta}{Z}_N) \right] \cdot \left[\sum_{\beta} E(\overset{\beta}{Z}_O) \right] \\
 E(\widehat{P}_I) \cdot E(\widehat{P}_j) &= \sum_{\substack{\alpha\alpha \\ \alpha=\alpha}} [E(\overset{\alpha}{Z}_I) \cdot E(\overset{\alpha}{Z}_I)] + \sum_{\substack{\alpha\alpha \\ \alpha \neq \alpha}} [E(\overset{\alpha}{Z}_I) \cdot E(\overset{\alpha}{Z}_I)] \\
 &+ \sum_{\substack{\alpha\beta \\ \alpha=\beta}} [E(\overset{\alpha}{Z}_I) \cdot E(\overset{\beta}{Z}_O)] + \sum_{\substack{\alpha\beta \\ \alpha \neq \beta}} [E(\overset{\alpha}{Z}_I) \cdot E(\overset{\beta}{Z}_O)] \\
 &+ \sum_{\substack{\alpha\beta \\ \alpha=\beta}} [E(\overset{\beta}{Z}_N) \cdot E(\overset{\alpha}{Z}_I)] + \sum_{\substack{\alpha\beta \\ \alpha \neq \beta}} [E(\overset{\beta}{Z}_N) \cdot E(\overset{\alpha}{Z}_I)] \\
 &+ \sum_{\substack{\beta\beta \\ \beta=\beta}} [E(\overset{\beta}{Z}_N) \cdot E(\overset{\beta}{Z}_O)] + \sum_{\substack{\beta\beta \\ \beta \neq \beta}} [E(\overset{\beta}{Z}_N) \cdot E(\overset{\beta}{Z}_O)].
 \end{aligned}$$

Chaque fois que nous avons $\alpha\alpha$ ($\alpha \neq \alpha$), $\alpha\beta$ ($\alpha \neq \beta$) et $\beta\beta$ ($\beta \neq \beta$) les loci en cause sont indépendants si l'équilibre asymptotique a été atteint malgré le linkage susceptible d'exister entre les classes d'homologues, aussi de façon générale :

$$E(Z_I) \cdot E(Z_I) = E(Z_I Z_I), \quad \alpha \neq \alpha, \text{ etc.}$$

donc :

$$\begin{aligned}
 \text{COV}(\widehat{P}_I, \widehat{P}_j) &= E(\widehat{P}_I \widehat{P}_j) - [E(\widehat{P}_I) \cdot E(\widehat{P}_j)] \\
 &= \sum_{\substack{\alpha\alpha \\ \alpha=\alpha}} [E(\overset{\alpha}{Z}_I \overset{\alpha}{Z}_I) - \{ E(\overset{\alpha}{Z}_I) \cdot E(\overset{\alpha}{Z}_I) \}] \\
 &+ \sum_{\substack{\beta\beta \\ \beta=\beta}} [E(\overset{\beta}{Z}_N \overset{\beta}{Z}_O) - \{ E(\overset{\beta}{Z}_N) E(\overset{\beta}{Z}_O) \}] \\
 &+ \sum_{\substack{\alpha\beta \\ \alpha=\beta}} [E(\overset{\alpha}{Z}_I \overset{\beta}{Z}_O) - \{ E(\overset{\alpha}{Z}_I) \cdot E(\overset{\beta}{Z}_O) \}] \\
 &+ \sum_{\substack{\alpha\beta \\ \alpha=\beta}} [E(\overset{\alpha}{Z}_I \overset{\beta}{Z}_N) - \{ E(\overset{\alpha}{Z}_I) E(\overset{\beta}{Z}_N) \}]
 \end{aligned}$$

$$\begin{aligned} \text{COV}(\widehat{P}_I \widehat{P}_I) &= \sum_{\substack{\alpha\alpha \\ \alpha=\alpha}} \text{COV}(\widehat{Z}_I^\alpha \widehat{Z}_I^\alpha) + \sum_{\substack{\beta\beta \\ \beta=\beta}} \text{COV}(\widehat{Z}_N^\beta \widehat{Z}_N^\beta) \\ &+ \sum_{\substack{\alpha\beta \\ \alpha=\beta}} \text{COV}(\widehat{Z}_I^\alpha \widehat{Z}_N^\beta) + \sum_{\substack{\alpha\beta \\ \alpha=\beta}} \text{COV}(\widehat{Z}_I^\beta \widehat{Z}_N^\alpha). \end{aligned}$$

Nous avons, dans le schéma I et II, donné les expressions de ces différentes covariances génotypiques, aussi après mise en facteur commun des coefficients d'identité restreinte nous obtenons :

$$\begin{aligned} &\text{COV}(\widehat{P}_I \widehat{P}_I) \\ &= A1 \left[\sum_{\alpha} E(X_{\alpha}^2) \right] + B2 \left[\sum_{\alpha} E_c(X_{\alpha} D_{\alpha}) \right] + C1 \left[\sum_{\alpha} E_p(D_{\alpha}^2) \right] + D1 \left[\sum_{\alpha} E_c(D_{\alpha}^2) \right] \\ &+ A2 \left[\sum_{\beta} E(X_{\beta}^2) \right] + B2 \left[\sum_{\beta} E_c(X_{\beta} D_{\beta}) \right] + C2 \left[\sum_{\beta} E_p(D_{\beta}^2) \right] + D2 \left[\sum_{\beta} E_c(D_{\beta}^2) \right] \\ &+ (A3 + A5) \left[\sum_{\substack{\alpha\beta \\ \alpha=\beta}} E(X_{\alpha} X_{\beta}) \right] + (\check{B}3 + \check{B}5) \left[\sum_{\substack{\alpha\beta \\ \alpha=\beta}} E_c(X_{\alpha} D_{\beta}) \right] \\ &+ (\check{B}3 + \check{B}5) \left[\sum_{\substack{\alpha\beta \\ \alpha=\beta}} E_c(X_{\beta} D_{\alpha}) \right] + (C3 + C5) \left[\sum_{\substack{\alpha\beta \\ \alpha=\beta}} E_p(D_{\alpha} D_{\beta}) \right] \\ &+ (D3 + D5) \left[\sum_{\substack{\alpha\beta \\ \alpha=\beta}} E_c(D_{\alpha} D_{\beta}) \right]. \end{aligned}$$

Il nous reste à généraliser, selon le même schéma, la variance des variables aléatoires phénotypiques P_I et P_I .

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

Le calcul de $E(\widehat{P}_I^2)$ présente quelques difficultés. Nous avons :

$$\begin{aligned} \widehat{P}_I &= \sum_{\alpha} \widehat{Z}_I^{\alpha} + \sum_{\beta} \widehat{Z}_N^{\beta} \\ \widehat{P}_I^2 &= \left(\sum_{\alpha} \widehat{Z}_I^{\alpha} \right)^2 + 2 \left(\sum_{\alpha} \widehat{Z}_I^{\alpha} \right) \cdot \left(\sum_{\beta} \widehat{Z}_N^{\beta} \right) + \left(\sum_{\beta} \widehat{Z}_N^{\beta} \right)^2 \end{aligned}$$

$$\begin{aligned}
\widehat{P}_I^2 &= \sum_{\substack{\alpha\alpha \\ \alpha=\alpha}} (Z_I^\alpha Z_I^\alpha) + \sum_{\substack{\alpha\alpha \\ \alpha \neq \beta}} (Z_I^\alpha Z_I^\alpha) \\
&+ 2 \sum_{\substack{\alpha\beta \\ \alpha=\beta}} (Z_I^\alpha Z_N^\beta) + 2 \sum_{\substack{\alpha\beta \\ \alpha \neq \beta}} (Z_I^\alpha Z_N^\beta) \\
&+ \sum_{\substack{\beta\beta \\ \beta=\beta}} (Z_N^\beta Z_N^\beta) + \sum_{\substack{\beta\beta \\ \beta \neq \beta}} (Z_N^\beta Z_N^\beta) \\
E(\widehat{P}_I^2) &= \sum_{\substack{\alpha\alpha \\ \alpha=\alpha}} E(Z_I^\alpha Z_I^\alpha) + \sum_{\substack{\alpha\alpha \\ \alpha \neq \alpha}} E(Z_I^\alpha Z_I^\alpha) \\
&+ 2 \sum_{\substack{\alpha\beta \\ \alpha=\beta}} E(Z_I^\alpha Z_N^\beta) + 2 \sum_{\substack{\alpha\beta \\ \alpha \neq \beta}} E(Z_I^\alpha Z_N^\beta) \\
&+ \sum_{\substack{\beta\beta \\ \beta=\beta}} E(Z_N^\beta Z_N^\beta) + \sum_{\substack{\beta\beta \\ \beta \neq \beta}} E(Z_N^\beta Z_N^\beta).
\end{aligned}$$

Le terme correctif s'obtient de même :

$$\begin{aligned}
E(\widehat{P}_I) \cdot E(\widehat{P}_I) &= \left[E\left(\sum_{\alpha} Z_I^\alpha + \sum_{\beta} Z_N^\beta \right) \right]^2 \\
&= \left[\sum_{\alpha} E(Z_I^\alpha) + \sum_{\beta} E(Z_N^\beta) \right]^2 \\
&= \sum_{\substack{\alpha\alpha \\ \alpha=\beta}} E(Z_I^\alpha) E(Z_I^\alpha) + \sum_{\substack{\alpha\alpha \\ \alpha \neq \beta}} E(Z_I^\alpha) E(Z_I^\alpha) \\
&+ 2 \sum_{\substack{\alpha\beta \\ \alpha=\beta}} E(Z_I^\alpha) E(Z_N^\beta) + 2 \sum_{\substack{\alpha\beta \\ \alpha \neq \beta}} E(Z_I^\alpha) E(Z_N^\beta) \\
&+ \sum_{\substack{\beta\beta \\ \beta=\beta}} E(Z_N^\beta) E(Z_N^\beta) + \sum_{\substack{\beta\beta \\ \beta \neq \beta}} E(Z_N^\beta) E(Z_N^\beta)
\end{aligned}$$

or si $\alpha \neq \alpha$, il vient :

$$E(Z_I^\alpha Z_I^\alpha) = E(Z_I^\alpha) E(Z_I^\alpha), \text{ etc.}$$

d'où l'expression de $\text{VAR}(\hat{P}_1)$:

$$\begin{aligned}\text{VAR}(\hat{P}_1) &= \sum_{\substack{\alpha\alpha \\ \alpha=\alpha}} [E(\bar{Z}_1^\alpha \bar{Z}_1^\alpha) - E(\bar{Z}_1^\alpha) E(\bar{Z}_1^\alpha)] \\ &\quad + 2 \sum_{\substack{\alpha\beta \\ \alpha=\beta}} [E(\bar{Z}_1^\alpha \bar{Z}_N^\beta) - E(\bar{Z}_1^\alpha) E(\bar{Z}_N^\beta)] \\ &\quad + \sum_{\substack{\beta\beta \\ \beta=\beta}} [E(\bar{Z}_N^\beta \bar{Z}_N^\beta) - E(\bar{Z}_N^\beta) E(\bar{Z}_N^\beta)] \\ \text{VAR}(\hat{P}_1) &= \sum_{\substack{\alpha\alpha \\ \alpha=\alpha}} \text{COV}(\bar{Z}_1^\alpha, \bar{Z}_1^\alpha) + 2 \sum_{\substack{\alpha\beta \\ \alpha=\beta}} \text{COV}(\bar{Z}_1^\alpha, \bar{Z}_N^\beta) + \sum_{\substack{\beta\beta \\ \beta=\beta}} \text{COV}(\bar{Z}_N^\beta, \bar{Z}_N^\beta) \\ \text{VAR}(\hat{P}_1) &= \sum_{\substack{\alpha\alpha \\ \alpha=\alpha}} \text{VAR}(\bar{Z}_1^\alpha) + 2 \sum_{\substack{\alpha\beta \\ \alpha=\beta}} \text{COV}(\bar{Z}_1^\alpha, \bar{Z}_N^\beta) + \sum_{\substack{\beta\beta \\ \beta=\beta}} \text{VAR}(\bar{Z}_N^\beta).\end{aligned}$$

Tous les termes du second membre de cette expression de la variance de P_1 sont connus, d'où :

$$\begin{aligned}\text{VAR}(\bar{Z}_1^\alpha) &= 2(1+f_I) E(X_\alpha^2) + 4f_I E_c(X_\alpha D_\alpha) + (1-f_I-f_I^2) E_p(D_\alpha^2) + f_I E_c(D_\alpha^2) \\ \text{VAR}(\bar{Z}_N^\beta) &= 2(1+f_N) E(\mathcal{X}_\beta^2) + 4f_N E_c(\mathcal{X}_\beta \mathcal{D}_\beta) + (1-f_N-f_N^2) E_p(\mathcal{D}_\beta^2) + f_N E_c(\mathcal{D}_\beta^2) \\ \text{COV}(\bar{Z}_1^\alpha, \bar{Z}_N^\beta) &= (A4) E(X_\alpha \mathcal{X}_\beta) + (\check{B}4) E_c(X_\alpha \mathcal{D}_\beta) + (\hat{B}4) E_c(\mathcal{X}_\beta D_\alpha) + (C4) E_p(D_\alpha \mathcal{D}_\beta).\end{aligned}$$

Rappelons que $(C4) = (C^*4) - f_I f_N$. En conséquence :

$$\begin{aligned}\text{VAR}(\hat{P}_1) &= 2(1+f_I) \left[\sum_{\alpha} E(X_\alpha^2) \right] + 4f_I \left[\sum_{\alpha} E_c(X_\alpha D_\alpha) \right] \\ &\quad + (1-f_I-f_I^2) \left[\sum_{\alpha} E_p(D_\alpha^2) \right] + f_I \left[\sum_{\alpha} E_c(D_\alpha^2) \right] \\ &\quad + 2(1+f_N) \left[\sum_{\beta} E(\mathcal{X}_\beta^2) \right] + 4f_N \left[\sum_{\beta} E_c(\mathcal{X}_\beta \mathcal{D}_\beta) \right] \\ &\quad + (1-f_N-f_N^2) \left[\sum_{\beta} E_p(\mathcal{D}_\beta^2) \right] + f_N \left[\sum_{\beta} E_c(\mathcal{D}_\beta^2) \right] \\ &\quad + 2(A4) \left[\sum_{\substack{\alpha\beta \\ \alpha=\beta}} E(X_\alpha \mathcal{X}_\beta) \right] + 2(\check{B}4) \left[\sum_{\substack{\alpha\beta \\ \alpha=\beta}} E_c(X_\alpha \mathcal{D}_\beta) \right]\end{aligned}$$

$$\begin{aligned}
& + 2(\text{B}4) \left[\sum_{\substack{\alpha\beta \\ \alpha=\beta}} E_c(\mathfrak{X}_\beta D_\alpha) \right] + 2(\text{C}4) \left[\sum_{\substack{\alpha\beta \\ \alpha=\beta}} E_p(D_\alpha \mathfrak{D}_\beta) \right] \\
& + 2(\text{D}4) \left[\sum_{\substack{\alpha\beta \\ \alpha=\beta}} E_c(D_\alpha \mathfrak{D}_\beta) \right].
\end{aligned}$$

Une expression semblable peut être donnée pour la variance de \widehat{P} , en fonction de (A6), (B6), (C6), (D6) et de f_i et f_o .

Nous sommes en mesure d'exprimer le coefficient de corrélation phénotypique entre I et J, apparentés quelconques, en remplaçant les termes de l'expression ci-dessous par leur valeur analytique :

$$r_{IJ} = \frac{\text{COV}(P_I P_J)}{\sqrt{\text{VAR}(P_I) \cdot \text{VAR}(P_J)}}.$$

Si l'on suppose que le déterminisme génétique du caractère est strictement additif, l'expression de la covariance entre les variables aléatoires phénotypiques P_i et P_j et de leurs variances devient :

$$\begin{aligned}
\text{COV}(P_i P_j) &= A1 \left[\sum_{\alpha} E(X_{\alpha}^2) \right] + A2 \left[\sum_{\beta} E(\mathfrak{X}_{\beta}^2) \right] + (A3 + A5) \left[\sum_{\substack{\alpha\beta \\ \alpha=\beta}} E(X_{\alpha} \mathfrak{X}_{\beta}) \right] \\
\text{VAR}(P_i) &= 2(1 + f_i) \left[\sum_{\alpha} E(X_{\alpha}^2) \right] + 2(1 + f_n) \left[\sum_{\beta} E(\mathfrak{X}_{\beta}^2) \right] \\
&\quad + 2(A4) \left[\sum_{\substack{\alpha\beta \\ \alpha=\beta}} E(X_{\alpha} \mathfrak{X}_{\beta}) \right] \\
\text{VAR}(P_j) &= 2(1 + f_j) \left[\sum_{\alpha} E(X_{\alpha}^2) \right] + 2(1 + f_o) \left[\sum_{\beta} E(\mathfrak{X}_{\beta}^2) \right] \\
&\quad + 2(A6) \left[\sum_{\substack{\alpha\beta \\ \alpha=\beta}} E(X_{\alpha} \mathfrak{X}_{\beta}) \right].
\end{aligned}$$

VI. — GÉNÉRALISATION AU CAS DU MULTIALLÉLISME

Les formules que nous avons obtenues restent valables dans le cas du multiallélisme, pourvu que soient définies à nouveau les variables aléatoires génétiques. Soit n le nombre de classes d'allèles de la série allélique. Soient k et l les indices ($1 \leq k \leq n$, $1 \leq l \leq n$) attribués à ces classes.

Un homozygote sera donc de la forme $[a_k a_k]$ ou $[a_l a_l]$ tandis qu'un hétérozygote sera de la forme $[a_k a_l]$ ou $[a_l a_k]$, sans que nous puissions distinguer $[a_k a_l]$ de $[a_l a_k]$. Soit p_k la probabilité qu'un gène pris au hasard dans la population (la génération ou le groupe suivant les nécessités) appartienne à la classe d'allèles d'indice k . Nous avons :

$$\sum_{k=1}^{k=n} p_k = 1 \quad \text{ou bien} \quad \sum_{l=1}^{l=n} p_l = 1.$$

Il est alors immédiat que :

$$E_p(D^2) = \sum_{k=1}^{k=n} p_k^2 (i_k - 2s)^2 + \sum_{\substack{l=1 \\ k=1}}^{l=n \\ k=n} 2p_k p_l (j_{kl} - s_k - s_l)^2$$

expression de la variance de dominance d'une génération dont tous les loci sont indépendants, dont la moyenne $E_p(D) = 0$ et où i_k est la contribution numérique du couple de gènes $[a_k a_k]$ au caractère considéré, $j_{kl} = j_{lk}$ celle du couple $[a_k a_l]$ ou $[a_l a_k]$.

s_k est la contribution numérique additive du gène a_k , que nous allons fixer par la méthode des moindres carrés, en annulant les dérivées de $E_p(D^2)$ par rapport à s_k , d'où :

$$\begin{aligned} \varphi'_{s_k} &= p_k(i_k - 2s_k) + \sum_{\substack{l=1 \\ l \neq k}}^{l=n} p_l(j_{kl} - s_k - s_l) = 0 \\ 0 &= p_k i_k - 2p_k s_k + \sum_{\substack{l=1 \\ l \neq k}}^{l=n} p_l j_{kl} - \sum_{\substack{l=1 \\ l \neq k}}^{l=n} p_l s_k - \sum_{\substack{l=1 \\ l \neq k}}^{l=n} p_l s_l. \end{aligned}$$

Remarquons que :

$$\begin{aligned} \sum_{\substack{l=1 \\ l \neq k}}^{l=n} p_l s_k &= (1 - p_k) s_k \\ \sum_{\substack{l=1 \\ l \neq k}}^{l=n} p_l s_l &= -p_k s_k \end{aligned}$$

car :

$$\sum_{l=1}^{l=n} p_l s_l = 0 = E(X)$$

d'où :

$$2p_k s_k + (1 - p_k) s_k - p_k i_k = p_k i_k + \sum_{\substack{l=1 \\ l \neq k}}^{l=n} p_l j_{kl}$$

$$s_k = p_k i_k + \sum_{\substack{l=1 \\ l \neq k}}^{l=n} p_l j_{kl}.$$

Cette dernière expression étant la forme générale des n équations qui traduisent la condition des moindres carrés appliqués au résidu de dominance. Si $n = 2$:

$$s_1 = p_1 i_1 + p_2 j_{12}$$

avec :

$$p_1 + p_2 = 1$$

$$s = p i + q j$$

et :

$$s_2 = p_2 i_2 + p_1 j = t = q k + p j$$

expressions déjà rencontrées.

Nous définissons à nouveau les moments du premier ordre et du second ordre de la population, en conservant les mêmes conventions que dans le cas du bi-allélisme.

$$E(X) = \sum_{k=1}^{k=n} p_k s_k = 0$$

$$E_p(Y) = 2E(X) = 0$$

$$E_c(Y) = 2E(X) = 0$$

$$E(X^2) = \sum_k p_k s_k^2 \quad (17)$$

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

$$E_c(Y^2) = 4E(X^2)$$

$$E_p(Z) = \sum_k p_k^2 i_k + \sum_{\substack{kl \\ k \neq l}} 2p_k p_l j_{kl} = 0$$

(17) Nous remplaçons pour des raisons de commodité $\sum_{k=1}^{k=n}$ par \sum_k .

avec $j_{kl} = j_{lk}$:

$$\begin{aligned}
 E_c(Z) &= \sum_k p_k i_k \\
 E_p(Z^2) &= \sum_k p_k^2 i_k^2 + \sum_{\substack{kl \\ k \neq l}} 2p_k p_l j_{kl}^2 \\
 E_c(Z^2) &= \sum_k p_k i_k^2 \\
 E_p(D) &= \sum_k p_k^2 (i_k - 2s_k) + \sum_{\substack{kl \\ k \neq l}} 2p_k p_l (j_{kl} - s_k - s_l) = 0 \\
 E_c(D) &= \sum_{kk} p_k (i_k - 2s_k) = E_c(Z) \\
 E_p(D^2) &= \sum_k p_k^2 (i_k - 2s_k)^2 + \sum_{\substack{kl \\ k \neq l}} 2p_k p_l (j_{kl} - s_k - s_l)^2 \\
 E_c(D^2) &= \sum_k p_k (i_k - 2s_k)^2.
 \end{aligned}$$

Nous avons toujours la relation :

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

L'expression de la corrélation génotypique est donc toujours valable, car les relations entre les moments sont inchangées et les démonstrations des expressions de la covariance et de la variance génotypiques sont indépendantes de la fréquence des gènes.

VII. — ESTIMATION DES PARAMÈTRES GÉNÉTIQUES DANS UNE POPULATION

Nous supposons, dans tout ce qui suit, l'absence d'effet d'environnement. Les résultats que nous présentons pourront toujours être modifiés pour tenir compte de facteurs de variation non génétiques mais contrôlés, que ces facteurs aient une action additive ou non par rapport au génotype. *Nous considérons un groupe d'individus quelconques, mais tels que nous possédions pour tout couple d'individus l'ensemble des coefficients d'identité restreinte,*

et pour chaque individu du couple la mesure du caractère dont nous nous proposons d'étudier le déterminisme héréditaire. Si nous nous intéressons à un effet maternel génétique, il sera aussi nécessaire de connaître le père et la mère de tout individu, et de pouvoir exprimer tous les coefficients d'identité restreinte utiles de tous les couples d'individus pris dans l'ensemble commun des parents et des enfants. La connaissance des paramètres des variables aléatoires génétiques s'obtient en deux étapes, l'une est l'estimation des paramètres du premier ordre, l'autre l'estimation des paramètres du second ordre.

L'étude des moments du deuxième ordre nécessite l'estimation de $E_p(Z)$ et $E_c(Z)$. Nous avons :

$$\begin{aligned} E(Z) &= (1 - f)E_p(Z) + fE_c(Z) \\ E(Z_i) &= (1 - f_i)E_p(Z) + f_iE_c(Z). \end{aligned}$$

Nous voulons déduire de l'ensemble des mesures m_i effectuées sur tous les zygotes des estimations m_p et m_c de $E_p(Z)$ et de $E_c(Z)$.

Nous appliquons la méthode des moindres carrés en rendant minimum la fonction :

$$\varphi = \sum_i [m_i - (1 - f_i)m_p - f_i m_c]^2.$$

Pour cela annulons les dérivées φ' par rapport aux inconnues m_p et m_c pour obtenir le système suivant :

$$\begin{vmatrix} \sum_i (1 - f_i)^2 & \sum_i (1 - f_i)f_i \\ \sum_i (1 - f_i)f_i & \sum_i f_i^2 \end{vmatrix} * \begin{vmatrix} m_p \\ m_c \end{vmatrix} = \begin{vmatrix} \sum_i (1 - f_i)m_i \\ \sum_i f_i m_i \end{vmatrix}$$

La solution en est immédiate :

$$\begin{aligned} \frac{\left[\sum_i (1 - f_i)m_i \right] \cdot \left[\sum_i f_i^2 \right] - \left[\sum_i (1 - f_i)f_i \right] \cdot \left[\sum_i f_i m_i \right]}{\left[\sum_i (1 - f_i)^2 \right] \cdot \left[\sum_i f_i^2 \right] - \left[\sum_i (1 - f_i)f_i \right] \cdot \left[\sum_i (1 - f_i)f_i \right]} &= m_p \\ \frac{\left[\sum_i (1 - f_i)^2 \right] \cdot \left[\sum_i f_i m_i \right] - \left[\sum_i (1 - f_i)m_i \right] \cdot \left[\sum_i (1 - f_i)f_i \right]}{\left[\sum_i (1 - f_i)^2 \right] \cdot \left[\sum_i f_i^2 \right] - \left[\sum_i (1 - f_i)f_i \right] \cdot \left[\sum_i (1 - f_i)f_i \right]} &= m_c. \end{aligned}$$

Pour l'étude des moments de deuxième ordre, il suffit de considérer comme mesure du caractère d'un zygote l'expression $m^1 - m_p$. Dans ces conditions toutes les hypothèses simplificatrices que nous avons introduites sont satisfaites.

Dans le cas où nous considérons plusieurs facteurs de variation non génétiques, chacun à plusieurs niveaux, l'établissement de la matrice serait du même type et, une fois obtenues les estimations des inconnues, nous attribuons à chaque individu la valeur $m_1 - m_p$. C'est cette valeur notre variable aléatoire Z .

Nous avons établi précédemment l'expression suivante pour tous les couples de zygotes de la population :

$$\text{COV}(\hat{Z}_i \hat{Z}_j) = A \sum_{\alpha} E(X_{\alpha}^2) + B \sum_{\alpha} E_c(X_{\alpha} D_{\alpha}) + C \sum_{\alpha} E_p(D_{\alpha}^2) + D \sum_{\alpha} E_c(D_{\alpha}^2)$$

et ceci en supposant le caractère conditionné par n séries alléliques, en l'absence d'épistasie, et en supposant atteint l'équilibre asymptotique malgré le linkage susceptible d'exister entre les classes d'homologues.

Indiquons A_{ij} , B_{ij} , C_{ij} , D_{ij} , les fonctions d'identité restreinte définies par le couple ij .

L'application de la méthode des moindres carrés donne le système matriciel suivant : $A * B = C$, avec :

$$A = \begin{bmatrix} \sum_i A_{ij}^2 & \sum_{ij} A_{ij} B_{ij} & \sum_{ij} A_{ij} C_{ij} & \sum_{ij} A_{ij} D_{ij} \\ \sum_{ij} A_{ij} B_{ij} & \sum_{ij} B_{ij}^2 & \sum_{ij} B_{ij} C_{ij} & \sum_{ij} B_{ij} D_{ij} \\ \sum_{ij} A_{ij} C_{ij} & \sum_{ij} B_{ij} C_{ij} & \sum_{ij} C_{ij}^2 & \sum_{ij} C_{ij} D_{ij} \\ \sum_{ij} A_{ij} D_{ij} & \sum_{ij} B_{ij} D_{ij} & \sum_{ij} C_{ij} D_{ij} & \sum_{ij} D_{ij}^2 \end{bmatrix}$$

$$B = \begin{bmatrix} \sum_{\alpha} E(X_{\alpha}^2) \\ \sum_{\alpha} E_c(X_{\alpha} D_{\alpha}) \\ \sum_{\alpha} E_p(D_{\alpha}^2) \\ \sum_{\alpha} E_c(D_{\alpha}^2) \end{bmatrix}$$

$$C = \begin{bmatrix} \sum_{ij} A_{ij} \text{COV}(\hat{Z}_i \hat{Z}_j) \\ \sum_{ij} B_{ij} \text{COV}(\hat{Z}_i \hat{Z}_j) \\ \sum_{ij} C_{ij} \text{COV}(\hat{Z}_i \hat{Z}_j) \\ \sum_{ij} D_{ij} \text{COV}(\hat{Z}_i \hat{Z}_j) \end{bmatrix}$$

La solution de cette équation matricielle fournit les estimations des différents moments du deuxième ordre des variables aléatoires génétiques de la population.

La généralisation au cas d'effet maternel génétique ne présente pas de difficultés.

VIII. — CONCLUSION

Le coefficient de parenté de G. Malécot ne décrit que les états d'identité et d'indépendance de deux gènes pris au hasard, l'un parmi les deux gènes homologues d'un zygote, l'autre parmi les deux gènes homologues de l'autre zygote. Le concept de situation d'identité que nous avons introduit décrit tous les états d'identité et d'indépendance existant entre les quatre gènes homologues de deux zygotes. Cette propriété de la situation d'identité rend possible l'étude génétique des populations consanguines. Les coefficients d'identité apportent toute l'information génétique nécessaire au calcul des covariances théoriques entre zygotes apparentés. Ce résultat, obtenu dans le cas de multiallélisme et de dominance entre gènes homologues, est tout à fait général et les résultats de R. A. Fisher (1918) et de G. Malécot (1939) n'en représentent que des cas particuliers. Nous avons considéré par surcroît dans notre étude l'effet maternel car la plupart des caractères quantitatifs des animaux domestiques, du moins chez les jeunes, sont influencés par ce dernier.

Pour établir les expressions théoriques de la covariance entre apparentés quelconques et de la variance de zygotes consanguins, nous avons introduit une description de la population autre que celle de R. A. Fisher. Ce dernier considérait une population d'effectif illimité et panmictique, ce qui avait pour conséquence d'assurer l'indépendance entre gènes homologues dès la première génération et son maintien pour toutes les générations sui-

vantes, la génération d'origine étant numérotée zéro. Par un changement d'origine — ce qui annule la moyenne de la population — et en imposant une condition de moindres carrés à la dominance — ce qui définit l'effet additif des gènes — R. A. Fisher introduisait d'une part la variance additive, que nous désignons par $2E(X^2)$, et d'autre part la variance de dominance, que nous désignons par $E_p(D^2)$. Dans le cas d'une population dont tous les gènes homologues sont indépendants, nous utilisons la méthode de R. A. Fisher mais en soulignant qu'ils s'agit d'une description conditionnée par la réalisation de l'indépendance entre les gènes homologues, sans aucune référence au nombre d'individus et au mode de croisement de la population. Cette définition de l'effet additif des gènes est choisie, plutôt que toute autre, pour sa généralité et son efficacité, car elle est encore valable dans le cas de populations dont les gènes homologues ne sont pas tous indépendants. Nous écrivons :

$$E(Z) = (1 - f)E_p(Z) + fE_c(Z)$$

pour un zygote tiré au hasard dans une population faisant apparaître, pondérées chacune par leur probabilité $1 - f$ et f , les espérances mathématiques de la variable aléatoire génotypique calculée sous la condition de l'indépendance pour la première et sous celle de l'identité des gènes pour la seconde. C'est ainsi que pour décrire la variance génotypique d'une population, nous introduisons $E(X^2)$, $E_p(D^2)$, $E_c(XD)$, $E_c(D^2)$, soit :

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

dont les termes $E(X^2)$ et $E_p(D^2)$ étaient déjà connus. Le terme $E_c(D^2)$ est le moment d'ordre deux de la variable aléatoire génotypique, calculée sous la condition d'identité entre gènes; sous la même condition, apparaît l'espérance mathématique du produit des variables aléatoires gamétique et de dominance $E_c(XD)$.

Une telle description vaut pour toutes les populations. Mais l'identification entre l'identité et la dépendance absolue d'une part, entre la non-identité et l'indépendance d'autre part, n'est possible que pour certaines populations telles que les populations panmictiques.

IX. — RÉSUMÉ

Avant d'aborder le calcul de la corrélation génotypique entre deux zygotes apparentés de façon quelconque, nous définissons les variables aléatoires génétiques et nous présentons la description génétique d'un groupe quelconque.

Soit un zygote I désigné au hasard dans le groupe. Considérons une classe d'homologues tirée au hasard, le zygote I possède deux gènes appartenant chacun à l'une des classes d'isoaction. Soit n le nombre de classes d'isoaction de la classe d'homologues désignée, et soient les deux indices k et l tels que ($1 \leq k \leq n$, $1 \leq l \leq n$) attribuables à ces classes d'isoaction. Nous ne connaissons pas la nature du gène G_i , provenant du père de I, mais nous pouvons dire qu'il a une probabilité p_i d'appartenir à la classe d'isoaction a_i . Soit s_i la contribution de ce gène à un caractère quantitatif. Nous introduisons la variable aléatoire gamétique X , attachée à G_i , telle que :

$$E(X^t) = \sum_k p_k s_k^t, \quad t \text{ entier positif.}$$

De la même façon, pour le gène G_i^* provenant de la mère de I, nous introduisons la variable aléatoire X^* , telle que :

$$E(X^{*t}) = \sum_k p_k s_k^t.$$

Nous avons de plus :

$$\sum_k p_k = 1.$$

Pour les mêmes raisons que précédemment, nous ne connaissons pas le génotype de ce zygote I, mais nous savons que I peut être homozygote $[a_k a_k]$, tandis qu'hétérozygote il aura un génotype de la forme $[a_k a_l]$ ou $[a_l a_k]$, sans que nous puissions distinguer l'ordre des indices. La probabilité que I soit $[a_k a_k]$ est $f p_k + (1 - f) p_k^2$, et qu'il soit $[a_k a_l]$ est $(1 - f) 2 p_k p_l$, f étant le coefficient de consanguinité de I.

Soit i_k la contribution du couple $[a_k a_k]$ mesurable, en l'absence d'influence du milieu, sur le zygote I et soit $j_{kl} = j_{lk}$ celle du couple $[a_k a_l]$ ou $[a_l a_k]$, nous introduisons la variable aléatoire génotypique $Z = X + X^* + D$, telle que :

$$E(Z^t) = \sum_{k=1}^{k=n} [f p_k + (1 - f) p_k^2] i_k^t + \sum_{\substack{k=1 \\ l=1 \\ k \neq l}}^{k=n \\ l=n} (1 - f) 2 p_l p_k j_{kl}^t$$

ce qui permet de définir sans ambiguïté la variable aléatoire de dominance D . Si $Z = X + X^*$, alors $Z \neq Y$, variable aléatoire zygotique.

Le zygote I étant tiré au hasard dans le groupe, $E(Z')$ est attribuable à ce groupe, nous avons :

$$E(Z) = fE_c(Z) + (1-f)E_p(Z)$$

f étant le coefficient de consanguinité du groupe puisque I est tiré au hasard dans ce groupe, $E_c(Z)$ l'espérance mathématique de Z calculée sous la condition de l'identité réalisée entre les deux gènes homologues en cause, et $E_p(Z)$ l'espérance mathématique de Z calculée sous la condition de l'indépendance réalisée entre les mêmes gènes. Par changement d'origine, nous posons $E_p(Z) = 0$, ce qui entraîne :

$$E(Z) = fE_c(Z).$$

De même nous avons :

$$E(D^2) = fE_c(D^2) + (1-f)E_p(D^2).$$

Nous pouvons déterminer les valeurs de s_k en imposant une condition de moindres carrés à $E_p(D^2)$, c'est-à-dire en rendant $E_p(D^2)$ minimum. Si $\varphi = E_p(D^2)$, nous avons un système de n équations normales de la forme :

$$\begin{aligned} \varphi'_{s_k} &= p_k i_k - 2p_k s_k + \sum_{\substack{l=1 \\ l \neq k}}^{l=n} p_l j_{kl} - \sum_{\substack{l=1 \\ l \neq k}}^{l=n} p_l s_k - \sum_{\substack{l=1 \\ l \neq k}}^{l=n} p_l s_l = 0 \\ &= p_k^2 i_k + \sum_{\substack{l=1 \\ l \neq k}}^{l=n} p_k p_l j_{kl} - 2p_k^2 s_k - \sum_{\substack{l=1 \\ l \neq k}}^{l=n} p_k p_l s_k - \sum_{\substack{l=1 \\ l \neq k}}^{l=n} p_k p_l s_k = 0 \end{aligned}$$

après sommation sur k , et simplification, il vient :

$$E_p(Z) - 2E(X) = 0$$

d'où $E(X) = 0$, ce qui entraîne $E_p(D) = 0$.

Ayant introduit certaines variables aléatoires génétiques et présenté la description d'un groupe, sans faire aucune hypothèse sur ce groupe nous établissons l'expression de la corrélation génotypique entre deux zygotes I et J apparentés de façon quelconque.

$$r_{ij} = \frac{\text{COV}(Z_i Z_j)}{\sqrt{\text{VAR}(Z_i) \cdot \text{VAR}(Z_j)}}$$

avec :

$$\begin{aligned} \text{COV}(Z_i Z_j) &= E(Z_i Z_j) - E(Z_i) \cdot E(Z_j) \\ \text{VAR}(Z_i) &= E(Z_i^2) - [E(Z_i)]^2. \end{aligned}$$

L'application du théorème des moyennes conditionnées nous permet d'écrire :

$$E(Z_i Z_j) = \sum_{i=1}^{i=15} \delta_i E_{Ci}(Z_i Z_j)$$

expression où δ_i est un coefficient d'identité restreinte, probabilité de réalisation de la situation d'identité restreinte C_i réalisée entre les gènes G_i , G_i^* , G_j , G_j^* . Toute situation d'identité restreinte définit les relations d'identité ou d'indépendance entre les variables aléatoires gamétiques et de dominance. Le résultat du calcul est :

$$\text{COV}(Z_i Z_j) = 4\varphi_{ij} E(X^2) + (4\delta_1 + \delta_2 + \delta_3 + \delta_4 + \delta_5) E_c(XD) \\ + (\delta_6 + \delta_9 + \delta_{12} - f_1 f_j) E_p(D^2) + \delta_1 E_c(D^2)$$

$$\text{VAR}(Z_i) = 2(1 + f_i) E(X^2) + 4f_i E_c(XD) + (1 - f_i - f_i^2) E_p(D^2) + f_i E_c(D^2).$$

Ces résultats généralisent les expressions de R. A. Fisher, de S. Wright et de G. Malécot. Nous remarquons que dans cette expression la dominance apparaît non seulement avec le moment d'ordre deux centré $E_p(D^2)$ introduit par R. A. Fisher mais avec les moments $E_c(D^2)$ et $E_c(XD)$. Pour un caractère avec dominance, un fort niveau de l'identité des gènes augmente la corrélation génotypique qui atteint l'unité pour l'identité complète

$$(\delta_1 = 1).$$

Nous avons aussi calculé la corrélation génétique dans le cas d'un caractère avec effet maternel génotypique. Nous donnons finalement une méthode d'estimation des moments de second ordre, $E(X^2)$, $E_c(XD)$, $E_c(D^2)$, $E_p(D^2)$ pour une population réelle. Cette estimation suppose connus les coefficients d'identité restreinte. Ces résultats sont encore valables si plusieurs classes d'homologues sont en cause, si d'une part il n'existe pas d'épistasie et si, d'autre part, en cas de linkage, l'équilibre asymptotique a été atteint.

X. — BIBLIOGRAPHIE

- R. A. FISHER, The correlations between relatives on the supposition of mendelian inheritance. *Trans. Roy. Soc.*, Edinburgh, t. 52, 1918, p. 399-433.
 R. A. FISHER, The theory of inbreeding. *Oliver and Boyd*, Edinburgh, London, 1949, VIII + 120 p.
 M. GILLOIS, La relation d'identité en génétique. *Thèse Fac. Sciences*, Paris, 1964.
 M. GILLOIS, 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.
 M. GILLOIS, Corrélation génotypique théorique entre apparentés. *Journées d'Études de la Commission de Génétique de la Fédération européenne de Zootechnie*, Lisbonne, octobre 1964.

- O. KEMPTHORNE, The correlations between relatives in a random mating population. *Proc. Roy. Soc.*, t. 143, 8, 1954, p. 103-113.
- O. KEMPTHORNE, a) The theoretical values of correlations between relatives in random mating populations. *Genetics*, t. 40, 1955, p. 153-167; b) The correlations between relatives in random mating populations. *Cold Spring Harbor Symp. quant. Biol.*, t. 20, 1955, p. 60-75.
- O. KEMPTHORNE, *An Introduction to Genetic Statistics*. John Wiley and Sons, Inc., New York, Chapman and Hall, Ltd., London, 1957, xvii + 545 p.
- G. MALÉCOT, a) Sur l'analyse des aléatoires et le problème de l'hérédité. *C. R. Acad. Sci.*, Paris, t. 206, 1938, p. 153-155; b) Sur les aléatoires mendéliennes et les corrélations de l'hérédité. *C. R. Acad. Sci.*, Paris, t. 206, 1938, p. 404-406; c) Sur la biométrie et les lois de Mendel. *C. R. des séances et Conférences de la Soc. Math. de France*, 1938, p. 44-45.
- G. MALÉCOT, a) Loi de Mendel et homogamie. *C. R. Acad. Sci.*, Paris, t. 208, 1939, p. 407-409; b) Les corrélations entre individus apparentés dans l'hypothèse d'homogamie. *C. R. Acad. Sci.*, Paris, t. 208, 1939, p. 552-554; c) *Théorie mathématique de l'hérédité mendélienne généralisée*. Guilhot, 52, Boulevard Malesherbes, Paris, 1 vol., in-8°, 1939, 100 p.; d) Le calcul des probabilités et les problèmes de l'hérédité. *Ann. Univ. Lyon, Sciences*, section A, t. 2, 1939, p. 25-37.
- G. MALÉCOT, Étude mathématique des populations « mendéliennes ». *Ann. Univ. Lyon, Sciences*, section A, t. 4, 1941, p. 45-60.
- G. MALÉCOT, *Les mathématiques de l'hérédité*. Masson et Cie, Paris, 1 vol., 1948, 60 p.
- G. MALÉCOT, a) *Relation entre probabilités et fréquences d'un gène*. Cours professé en Sorbonne, édité par le Bureau de Génétique et de Biologie Quantitative 1 vol., 1961, 80 p.; b) *Les corrélations en génétique*. Cours professé en Sorbonne, édité par le Bureau de Génétique et de Biologie Quantitative, 1 vol., 1961, 80 p.; c) *Relationship and correlations*. Max-Planck Instituts für Tierzucht und Tierernährung. Scientific Problems of Recording Systems and breeding plans of Domestic Animals. Lectures of II. International Summer School., sept. 12-sept. 27. Special, vol. 129, 1961, p. 146.
- K. PEARSON, On a generalized theory of alternative inheritance with special reference to Mendel's laws. *Phil. Trans. Roy. Soc.*, A, t. 203, 1904, p. 53-86.
- W. WEINBERG, Ueber Vererbungsgesetze beim Menschen. *Zeit. Ind. Abst. Ver.*, t. 1, 1908, p. 377-292, 440-460; t. 2, p. 276-330.
- W. WEINBERG, Weitere Beiträge zur Theorie der Vererbung. *Archiv für Rassen und Gesellschafts-biologie*, t. 7, 1910, p. 35-49, 169-173.
- S. WRIGHT, On the nature of size factors. *Genetics*, t. 3, 1918, p. 367-374.
- S. WRIGHT, The relative importance of heredity and environment in determining the piebald pattern of guinea pigs. *Proc. Nat. Acad. Sci.*, t. 6, 1920, p. 320-332.
- S. WRIGHT, Correlation and causation. *J. Agric. Res.*, t. 20, 1921, p. 557-585.
- S. WRIGHT, Coefficients of Inbreeding and Relationship. *Amer. Nat.*, t. 56, 1922, p. 330-338.
- S. WRIGHT, The theory of path coefficients. A reply to Niles' criticism. *Genetics*, t. 8, 1923, p. 239-255.
- S. WRIGHT, The Method of path coefficients. *Ann. Math. Statist.*, t. 5, 1934, p. 161-215.
- G. YULE UDNY, On the theory of inheritance of quantitative compound characters on the basis of Mendel's law. A preliminary note. *Report of Third International Conference on Genetics*, 1906, p. 140-142.