Ann. Génét. Sél. anim., 1971, **3** (3), 255-280.

GÉNÉTIQUE DES POPULATIONS DIPLOÏDES NATURELLES DANS LE CAS D'UN SEUL LOCUS

I. — Évolution de la fréquence d'un gène. Étude des variances et des covariances

G. MALÉCOT

Mathématiques Appliquées, (Université de Lyon)
B.P. n° 37, 69-Villeurbanne-Charpennes

RÉSUMÉ

Les populations naturelles sont souvent réparties en groupes, quelquefois isolés ou plus généralement communiquant par migration.

L'effectif de chaque groupe est limité : nous le désignons par N, en supposant, pour simplifier, qu'il varie peu d'un groupe à un autre.

Cet effectif limité est à l'origine de la « dérive génétique » : si l'on suppose, pour simplifier, que les générations sont séparées, les N individus diploïdes de la génération F_{n+1} résultent de la fusion de 2 N *gamètes utiles* tirés au sort parmi un nombre *beaucoup plus grand* de gamètes *produits* par la génération précédente F_n, ce dernier ensemble constituant ce qu'on appelle le *réservoir gamétique*.

Deux problèmes se posent alors pour l'étude du passage d'une génération à la suivante :

1) Constitution du réservoir gamétique.

2) Mode de tirage des gamètes utiles (en nombre 2N).

1) Si la fréquence du gène a parmi les N adultes reproducteurs de la génération F_n est connue et égale à q_n, la fréquence de a dans l'ensemble des gamètes qu'ils produisent (c'est-à-dire dans le réservoir gamétique pour F_{n+1}) sera en général différente, soit $q_n = q_n + \delta(q_n)$, en raison de la pression de mutation et de la pression de sélection, dont les effets (pour la production de gamètes infiniment nombreux, sont rappelés au chapitre II : il en résulte que, pour les petites fluctuations au voisinage d'une fréquence d'équilibre $\bar{q} = C$, $\delta(q_n)$ peut être linéarisée sous la forme :

$$\delta(q_n) = -k(q_n - C)$$

Ces formules sont à modifier s'il y a des migrations entre groupes ; le chapitre I, introductif de l'ensemble des problèmes qui sont traités par la suite, précise les notations comme suit (voir figure 1).

q_{nx} est la fréquence du gène a sur l'ensemble des N diploïdes constituant, dans la génération F_n, le groupe d'emplacement x.

La migration entraîne que le réservoir gamétique qui se constitue dans l'emplacement x pour la génération suivante est constituée par l'action de la mutation et de la sélection sur une fréquence q^*_{nx}, qui est une moyenne pondérée des fréquences sur le groupe x et sur les groupes voisins, désignés par z, soit

$$q^*_{nx} = \sum_z l_{xz}\, q_{nz}$$

l_{xz} étant le taux de migration du groupe z dans le groupe x $\left(\sum_z l_{xz} = 1\right)$.

Lorsque la migration est suivie de mutation et de sélection, la composition du réservoir gamétique est définie par :

$$q'_{nx} = q^*_{nx} + \delta(q^*_{nx})$$

ce qui donne, lorsque la linéarisation est possible :

$$q'_{nx} = C + (1 - k)(q^*_{nx} - C)$$

C'est à partir de cette formulation que se développe le chapitre III pour un groupe isolé (c'est-à-dire sans migration) : l'indice x peut alors être supprimé.

Le chapitre IV reprend le calcul en introduisant des migrations entre groupes.

2) La manière dont sont effectués, dans le réservoir gamétique, les 2N tirages au sort de gamètes utiles, dépend à la fois du mode de croisement (défini par la relation statistique entre les gamètes mâles et femelles qui s'unissent pour former chaque œuf) et de la variabilité de la fécondité individuelle (mesurée en gamètes utiles).

Lorsque la fécondité obéit à une loi de Poisson et lorsque la fusion de gamètes se fait au hasard, on se trouve dans le cas de *panmixie;* le cas d'un groupe panmictique isolé est étudié au chapitre III. La fréquence q_{n+1} de a dans la génération F_{n+n} se déduit de la probabilité q'_n (relative à chacun des 2N gamètes tirés au sort indépendamment dans le réservoir gamétique) par la formule :

$$q_{n+1} = q'_n + \varepsilon_n$$

ε_n étant une variable aléatoire dont la loi de probabilité conditionnée (quand q_n et q'_n sont connus) a, en vertu du théorème de Bernoulli une espérance nulle et une variance égale à

$$\frac{q'_n(lq - {}'_n)}{2N}$$

Si l'on raisonne « *a priori* » (c'est-à-dire en connaissant seulement la génération initiale F_0, avec sa fréquence q_0), les fréquences ultérieures $q_n (n \geqslant 1)$ sont des variables aléatoires liées en chaîne de Markoff, et dont les espérances *a priori* et moments *a priori* sont étudiés au chapitre III. L'espérance *a priori* q_n tend vers la limite $q = C;$ la variance *a priori* tend vers la limite

$$\sigma^2 = \frac{C - C^2}{1 + 4Nk}$$

qui définit la fluctuation dans l'état asymptotique (stationnaire) d'équilibre statistique (fluctuation estimable sur de nombreux groupes isolés de même effectif N) (§ C).

Cette variance est liée d'ailleurs au deuxième moment *a priori* de la fréquence q_n ou q'_n, donc, aux espérances des fréquences des trois génotypes *aa*, A*a*, AA (§ D).

Les troisièmes et quatrièmes moments *a priori* fournissent les espérances des fréquences d'accouplement de toutes les paires de génotypes (Morton et Yasuda, 1962) (§ E).

Le chapitre IV étudie la covariance entre deux groupes panmictiques en fonction de leur distance et donne une expression « quasi exponentielle » de la décroissance du coefficient de corrélation.

INTRODUCTION

Pour la clarification des notions de base de la Génétique de population (en ce qui concerne les populations limitées, en particulier), il y a un gros intérêt à utiliser des probabilités *a priori*. Dans tout ce qui suit, un calcul de probabilités *a priori* sera un calcul effectué lorsqu'on connaît seulement la « génération initiale » F_o (dont les loci sont occupés pour une proportion q_o par le gène a et pour la proportion $p_o = 1 - q_o$ par l'ensemble des gènes allèles, ensemble noté A). Un locus L_j tiré parmi les 2 loci homologues d'un individu I, lui-même pris au hasard dans une génération ultérieure F_n (nous dirons en bref, un « locus aléatoire dans F_n) a une probabilité *a priori* de porter « a » qui est calculable en fonction de n et des coefficients de mutation, sélection, et migration; cette probabilité est égale à q_o lorsque le locus aléatoire a égale probabilité de dériver de chacun des loci initiaux, ce qui est le cas lorsqu'il n'y a, ni mutation, ni sélection, ni migration. Dans le cas général, nous désignerons cette probabilité par $\overline{q_{xn}}$ et la probabilité contraire par $\overline{p_{xn}} = 1 - \overline{q_{nx}}$; elle peut dépendre de n (si l'état stationnaire

n'est pas atteint) et de l'emplacement x (s'il y a migration excluant la panmixie); nous nous placerons souvent — en le signalant chaque fois — dans le cas particulier où $\overline{q_{nx}}$ est constant, soit $\overline{q}$, souvent désigné aussi par C.

Que cette probabilité *a priori* q_{nx} soit constante ou variable, elle admet deux autres interprétations :

a) Elle est l'espérance mathématique *a priori* de la variable aléatoire X_j égale à 1 lorsque le locus singulier L_j (pris au hasard dans F_n et en x) porte le gène a, et a zéro lorsqu'il porte A.

b) Elle est aussi l'espérance mathématique d'une moyenne arithmétique de telles aléatoires : par exemple l'espérance mathématique de la fréquence q_{nx} du gène a sur un ensemble de 2N loci L_1, ...L_j, ...L_{2N} ayant même probabilité $\overline{q_{nx}}$ de porter a(*) (puisque $q_{nx} = \frac{r}{2N}$, r étant le nombre total $X_1 + \ldots X_j + \ldots X_{2N}$ de gènes a).

Nous écrirons donc $E(q_{nx}) = \overline{q_{nx}}$.

c) Il en résulte que cette fréquence q_{nx} peut être prévue comme sensiblement égale à $\overline{q_{nx}}$ lorsque 2N est assez grand er lorsque les 2N loci sont indépendants en probabilité (il suffit d'appliquer le théorème de BERNOULLI ou loi des grands nombres). Cela est le cas en particulier pour un groupe panmictique suffisamment nombreux, où la fréquence q_{nx} coïncide donc sensiblement avec la probabilité *a priori* $\overline{q_{nx}}$.

Dans ce cas, il n'y a pas d'incertitude, le modèle est « déterministe ». Mais si nous considérons un groupe panmictique d'effectif limité constant N, d'emplacement x, comprenant à chaque génération F_n, 2N loci singuliers homologues (apportés par 2N gamètes utiles tirés au sort parmi les gamètes beaucoup plus nombreux produits) et, si nous désignons par $L_j (1 \leqslant j \leqslant 2N)$ ces loci, et par X_j les aléatoires correspondantes, la fréquence de a est : $q_{nx} = \frac{X_1 + \ldots + X_{2N}}{2N}$ dont l'espérance *a priori* est $\overline{q_{nx}}$, et l'on a :

$$q_{nx} - \overline{q_{nx}} = \frac{X_1 - \overline{q_{nx}} + \ldots + X_{2N} - \overline{q_{nx}}}{2N}$$

En raison du nombre limité de tirages au sort, la fréquence q_{nx} fluctue maintenant autour de son espérance mathématique, avec une variance $\sigma_x^2(n) = E[(q_{nx} - \overline{q_{nx}})^2]$ que nous calculerons d'abord dans le cas d'isolement complet puis dans le cas de migration entre groupes.

Dans ce dernier cas si l'on considère deux groupes panmictiques d'emplacements x et y, il y a lieu d'étudier aussi, outre leurs variances $\sigma_x^2(n)$ et $\sigma_z^2(n)$, leur co-variance $E[(q_{nx} - \overline{q_{nx}})(q_{nz} - \overline{q_{nz}})]$ qui permettent l'étude du coefficient de corrélation et de sa variation avec la distance des deux groupes (chapitre IV.)

(*) Ceci suppose que les 2N loci considérés correspondent à la même génération F_n et au même emplacement x.

CHAPITRE I. — CONSTITUTION DU RÉSERVOIR GAMÉTIQUE

Les individus appartenant à une même génération F_n produisent des gamètes beaucoup plus nombreux qu'eux mêmes : ces gamètes constituent le « réservoir gamétique pour F_{n+1} » dans lequel seront puisés les « gamètes utiles » qui, par fusion deux à deux, constitueront les individus de la génération suivante F_{n+1}; si l'effectif est supposé constant de génération en génération, le nombre de gamètes utiles (et le nombre de loci singuliers qu'ils portent) sera constant et égal au double du nombre d'individus. Nous admettrons que c'est dans la formation du « réservoir gamétique pour F_{n+1} » que se produisent les phénomènes de migration, mutation, fécondité variable, fécondité différentielle, sélection gamétique, sélection zygotique (1). Il reste à préciser comment sont tirés, dans le réservoir gamétique, les gamètes utiles. Définissons un certain nombre de cas (que nous ne considérons pas tous forcément d'ailleurs).

A. — *Cas d'un groupe panmictique isolé*

Ce sera le cas où les gamètes produits se mélangent de telle sorte que les deux gamètes utiles donnant naissance à un individu résultent de deux tirages au sort indépendants dans le réservoir gamétique (2); les quatre gamètes utiles donnant naissance à deux individus qui s'accoupleront résultent de quatre tirages au sort indépendants, etc.

B. — *Cas d'un groupe isolé avec homogamie ou croisements suivant la parenté*

Ce sera le cas où les gamètes utiles donnant naissance à un même individu (ou à deux individus qui s'accoupleront) ne sont pas indépendants en probabilité, soit parce qu'ils sont systématiquement issus de géniteurs apparentés (consanguinité), soit parce qu'il y a tendance préférentielle à l'union de gamètes porteurs de gènes identiques ou différents (homogamie ou hétérogamie gamétiques), soit parce qu'il y a tendance préférentielle à l'accouplement d'individus porteurs de zygotes identiques ou différents (homogamie ou hétérogamie zygotiques (3).

(1) La sélection zygotique intervient pourtant dans une phase ultérieure du cycle vital; mais on sait qu'elle équivaut mathématiquement à une sélection gamétique en $sq(1-q)$, s dépendant de q (voir plus loin en II).

(2) En toute rigueur, les deux gamètes qui s'unissent étant l'un mâle et l'autre femelle, il faudrait distinguer le « réservoir gamétique mâle » et le « réservoir gamétique femelle ».

(3) En génétique l'homogamie phénotypique ne nous intéresse que pour l'homogamie zygotique qui en résulte.

C. — *Cas de groupes communiquant par migration*

Cas que nous définissons ainsi :

Un groupe panmictique d'effectif N sera dit « issu du réservoir gamétique d'emplacement x », ou en bref « d'emplacement x » s'il est constitué par la fusion de 2N gamètes résultant de 2N tirages au sort indépendants dans un réservoir gamétique constitué — avant que la mutation et la sélection n'aient agi — par le mélange de gamètes « autochtones » (produits par les individus occupant l'emplacement x ») auxquels se sont joints dans des proportions l_{y-x}, des gamètes produits par les individus occupant les emplacements y; si l'on désigne par l_o la proportion de gamètes autochtones, on a évidemment $\sum_y l_{y-x} = 1$;

Parmi les cas particuliers que nous traiterons, indiquons :

a) Le cas ou les groupes, de même effectif N, sont répartis de façon équidistants sur la droite où ils sont numérotés pour les entiers consécutifs, qui peuvent être regardés comme leurs « abscisses », soit :

$$\ldots x - 1,\ x,\ x + 1 \ldots$$

Nous supposerons que seuls les deux groupes « adjacents » à l'emplacement x peuvent lui fournir des gamètes, la probabilité étant m pour chacun, d'où :

$$l_1 = m,\ l_{-1} = m,\ l_o = 1 - 2m$$

Si les gènes A et a sont en proportion $p_{n,x}$ et $q_{n,x}$ (dans la génération F_n des individus d'emplacement x ou ce qui revient au même, chez les gamètes utiles (tirés dans le réservoir pour F_n) qui leur ont donné naissance) (1) (voir fig. 1). Le réservoir pour F_{n+1} sera constitué d'abord par une proportion de gènes a égale à :

$$q^*_{n,x} = mq_{n,x-1} + (1 - 2m)q_{n,x} + mq_{n,x+1} = q_{n,x} + m\,\Delta^2 q_{n,x} \qquad (2)$$

proportion sur laquelle joueront ensuite la mutation et la sélection dont l'action donnera finalement la proportion :

$$q'_{nx} = q^*_{n,x} + \delta(q^*_{n,x})$$

soit, dans le cas particulier où l'on linéarise (3) la sélection :

$$q'_{n,x} = q^*_{n,x} - k(q^*_{n,x} - C) = C + (1 - k)(q^*_{n,x} - C) = C + (1 - k)(q_{n,x} - C) + \\ + (1 - k)m\,\Delta^2 q_{n,x}$$

Dans ce cas, la composition du réservoir gamétique pour F_{n+1} s'obtiendra en appliquant un opérateur linéaire (opérateur de différence) aux proportions réalisées dans les groupes constituant la génération F_n.

b) Dans le cas de la « migration rectangulaire » à deux dimensions, les groupes seront numérotables par deux indices entiers x_1 et x_2, que nous regarderons comme les composantes d'un « vecteur » $\vec{x}$; si nous supposons que seuls les groupes adjacents à l'emplacement $\vec{x}$ peuvent lui fournir des gamètes avec probabilité m pour

(1) L'emplacement de chaque individu sera, pour la commodité, défini par son lieu de naissance (chez les animaux, le lieu de fécondation est différent du lieu de naissance).

(2) Le symbole Δ^2 désigne la différence seconde centrale par rapport à la variable entière x.

(3) Pour la signification de δ et sa linéarisation voir plus loin en II.

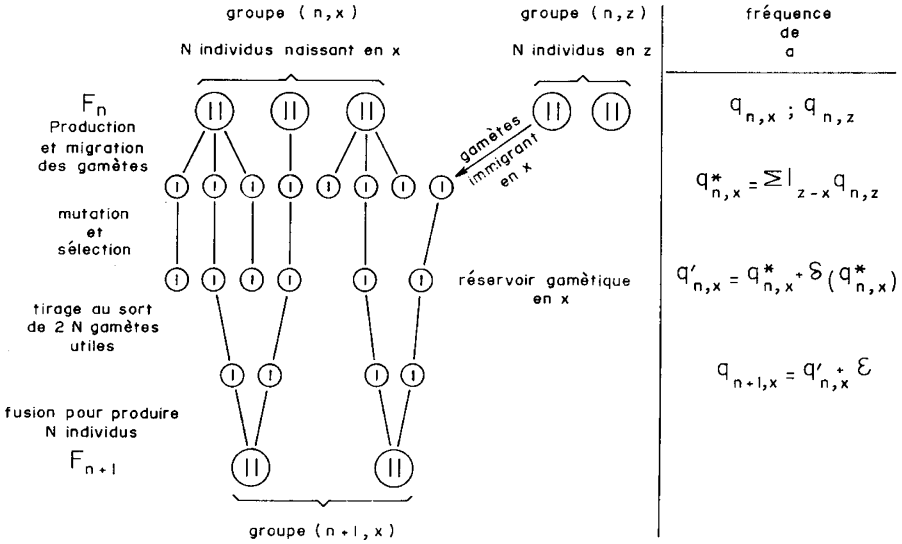

FIG. 1. — *Passage des diploïdes de* F_n *aux diploïdes de* F_{n+1}, *supposés en nombre constant* N *dans l'emplacement noté* x. *Un autre emplacement,* z, *est indiqué aussi, pour symboliser l'immigration en* x *de gamètes extérieurs. La colonne de droite indique la notation à chaque niveau, de la fréquence du gène* a. *Le réservoir gamétique est supposé très nombreux.* ε *est une varialble aléatoire « de Bernoulli » traduisant la dérive génétique due aux tirages au sort indépendants dans ce réservoir, de* 2N *gamètes « utiles ».*

chacun des 4 groupes de coordonnées $\begin{pmatrix} x_1 - 1 \\ x_2 \end{pmatrix}$; $\begin{pmatrix} x_1 + 1 \\ x_2 \end{pmatrix}$; $\begin{pmatrix} x_1 \\ x_2 - 1 \end{pmatrix}$; $\begin{pmatrix} x_1 \\ x_2 + 1 \end{pmatrix}$ et probabilité $1 - 4m$ pour le groupe $\vec{x}$, la probabilité q^*_{nx} avant sélection et mutation sera :

$$q^*_{nx} = q_{xn} + m\Delta^2_{n1} \quad q_{n,x} + m\Delta^2_{x2} \quad q_{nx} \tag{1}$$

et, après sélection, mutation et linéarisation éventuelle :

$$q'_{nx} = q^*_{n,x} + \delta(q^*_{n,x}) = C + (1 - k)(q^*_{n,x} - C) \tag{2}$$

c) Dans le cas *général*, nous écrivons :

$$q^*_{nx} = \sum_z l_{z-x} \quad q_{nz},$$

(la sommation étant étendue sur tous les emplacements possibles)

$$q'_{n,x} = \sum_z l_{z-x}\, q_{n,z} + \delta\left(\sum_z l_{z-x}\, q_{n,z}\right)$$

soit, dans le cas particulier où l'on *linéarise* la sélection et compte tenu de ce que $\sum_z l_z = 1$:

$$q'_{n,x} = C + (1 - k) \sum_z l_{z-y}(q_{n,z} - C) = C + (1-k)\ T(q_{nx} - C)$$

(1) Δ^2_{xi} désignant la différence seconde centrale par rapport à la coordonnée x_i (Δ^2_{xj} par rapport à x_j).

(2) C désignant la constante définie plus loin en IIA.

$T(q_{n,x} - C)$ désignant un opérateur linéaire agissant sur les $q_{n,x} - C$. Cet opérateur pourra être, comme dans les exemples précédents (*a*) et (*b*) un opérateur de différence finie; on pourra le remplacer par un opérateur d'intégration dans le cas où les coordonnées possibles des immigrants forment une suite continue; et, éventuellement si l'immigration ne provient que d'un très court rayon, on pourra le remplacer par un opérateur de dérivation.

D. — *Fusion des gamètes*

Considérons dans la génération suivante F_{n+1}, un zygote occupant l'emplacement x : il résulte de la fusion de deux gamètes utiles pris dans le réservoir d'emplacement x; chacun de ces gamètes a une probabilité conditionnée $q'_{n,x}$ de porter le gène *a*, et le coefficient de corrélation *conditionné* de ces deux gamètes peut être déterminée pour certains modèles d'homogamie ou de consanguinité; il est nul dans le cas où le groupe occupant l'emplacement x est un groupe panmictique. Nous verrons plus loin qu'il n'en est pas de même pour le coefficient de corrélation gamétique *a priori*, qui est différent de zéro, même dans un groupe panmictique lorsque son effectif n'est pas assez grand pour rendre pratiquement impossibles les accouplements consanguins. Pour commencer nous allons rappeler le cas classique d'un groupe panmictique infiniment grand où les tirages au sort des gamètes utiles fourniront des fréquences $q_{n+1,x}$ égales aux probabilités conditionnées $q'_{n,x}$.

Plus loin nous étudierons un groupe panmictique d'effectif limité (chapitre III) puis plusieurs de ces groupes (chapitre IV).

CHAPITRE II. — ÉVOLUTION DE LA FRÉQUENCE D'UN GÈNE DANS UN GROUPE PANMICTIQUE INFINIMENT NOMBREUX : MODÈLE DÉTERMINISTE

A. — *Composition du réservoir gamétique après mutation et sélection*

Désignons comme tout à l'heure par $q_{n,x}$ la fréquence de *a* dans F_n, dans le groupe d'emplacement x, par $q^*_{n,x}$ la fréquence après migration éventuelle et étudions la modification de $q^*_{n,x}$ (que pour le moment nous noterons q) par les mutations et la sélection.

a) Rappelons que les *mutations* peuvent créer une modification génétique de certains gamètes : si les mutations transforment une proportion u de gènes *a* en A et une proportion v de gènes A en *a*, la variation de la fréquence q de gènes *a* sous l'action des mutations est :

$$\delta_1(q) = -uq + v(1 - q) \tag{1}$$

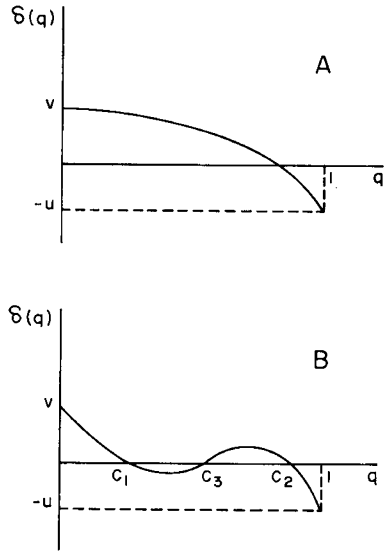

FIG. 2. — *Représentation graphique de la variation totale* δ(q) *due à la sélection et aux mutations.* A) *cas de 1 racine;* B) *cas de 3 racines.*

b) Si la *sélection gamétique* entraîne ensuite que les gamètes *a* et A ne sont conservés que dans les proportions de $1 + s$ à 1, cela entraîne la variation :

$$\delta_2(q) = \frac{(1+s)\,q}{(1+s)\,q + p} - q = \frac{q + sq}{1 + sq} - q = \frac{sq\,(1-q)}{1+sq}$$

qui est peu différent de $sq\,(1-q)$ si s est petit.

On sait que la *sélection zygotique* et la fécondité différentielle se traduit, en ce qui concerne la variation de fréquence q du gène *a*, par une formule analogue, mais où la constante s, si les taux de survie ou « valeurs sélectives » de *aa*, A*a*,AA, sont proportionnels à $1 + s$, $1 + hs$, 1, est remplacée par un polynôme du 1er degré $s(q) = t + wq = [h + q(1 - 2h)]$.

(La possibilité pour ce polynôme de s'annuler pour une valeur $= \dfrac{h}{2h-1}$ comprise entre O et 1 traduit la possibilité d'existence d'un « équilibre stable » entre les allèles; en fait, lorsque les hétérozygotes sont avantagés : $h > 1$).

Nous réunirons toutes ces variations dans la formule :

$$\delta(q) = \delta_1(q) + \delta_2(q) = -uq + v(1-q) + s(q)\;q(1-q) \tag{2}$$

D'où l'on déduit *la fréquence gamétique après mutation et sélection :*

$$q'_n = q^*_n + \delta(q^*_n) \tag{3}$$

c) Mais cette formule n'est exploitable conjointement avec la formule (1) que si elle est linéaire. Or, d'après (2), la sélection, même gamétique, exclut la linéarité. Mais *on peut linéariser la formule* (2) toutes les fois que l'on n'a à étudier

que des petites variations au voisinage d'une position d'équilibre stable C solution de l'équation $\delta(C) = 0$ avec la condition $\delta'(C) < 0$ (*). Si l'on pose alors $\delta'(C) = -k$ (**) $(k > 0)$.
les petites variations sont régies par la formule approchée (***) :

$$\delta(q) = -k(q - C) \tag{2'}$$

qui donne alors :

$$q'_n - C = q^*_n - C + \delta(q^*_n) = (1 - k)(q^*_n - C) \tag{3'}$$

B. — *Évolution de la fréquence génique*

Dans un groupe panmictique isolé dont l'effectif N est *très grand*, les 2N tirages indépendants dans le réservoir gamétique fournissent une fréquence q_{n+1}, sensiblement égale à q'_{nx}. Le groupe étant isolé, $q^*_n = q_n$, on a donc :

$$q_{n+1} = q_n + \delta(q_n)$$

$$q_{n+1} - C = (1 - k)(q_n - C) = (1 - k)^{n+1}(q_o - C)$$

ce qui montre que la fréquence q_n tend vers la limite C quand n augmente indéfiniment.

Mais dans un groupe d'effectif *limité*, le *tirage au sort des gamètes* entraîne des fluctuations de q_n autour de C, fluctuations qui seront étudiées dans le chapitre suivant.

CHAPITRE III. — ÉVOLUTION ALÉATOIRE DE LA FRÉQUENCE D'UN GÈNE DANS UN GROUPE PANMICTIQUE ISOLÉ : MODÈLE STOCHASTIQUE

A. — *Loi de probabilité* a priori, *son évolution*

Nous allons étudier maintenant les lois de probabilités *a priori* des fréquences aléatoires successives q_1, q_2... q_n et en particulier leurs espérances mathématiques $E(q_n) = q_n$, leurs variances $E[(q_n - \bar{q}_n)^2] = \sigma^2(n)$ leurs moments a priori $E(q_n{}^p)$. Les valeurs que nous obtiendrons pour les moments a priori permettront la prévision des moments expérimentaux mesurés à partir d'un grand nombre de groupes isolés.

Nous nous ramenons à un problème de « chaîne de Markoff » si nous introduisons la loi de probabilité conditionnée de q_{n+1} quand les fréquences précéden-

(*) Lorsque $s(q)$ est un polynôme de 1er degré au plus, donc $\delta(q)$ un polynôme du 3e degré au plus, il peut exister une racine, ou 3 racines, la racine intermédiaire C_3 ne correspond alors pas à un équilibre stable (fig. 2, A et B).

(**) $k = u + v - s(C)(1 - 2C) - s'(C)C(1 - C)$, se réduisant à $u + v$ s'il n'y a pas de sélection.

(***) S'il n'y a pas de sélection, $C = \dfrac{v}{u + v}$, $k = u + v$, et la formule (2') est rigoureuse.

tes $q_1 \ldots q_n$ sont connues (loi qui ne dépend que de q_n); cette loi sera désignée par $\mathfrak{L}_n$, toutes les espérances mathématiques conditionnées calculées à partir de cette loi seront désignées par le symbole E_n; nous aurons à faire fréquent usage du « théorème de l'espérance conditionnée » d'après lequel, pour calculer l'espérance à priori d'une fonction $\psi(q_1 \ldots q_n, q_{n+1})$, il suffit de calculer l'espérance de son espérance conditionnée, ce que nous écrivons en bref :

$$E[(\psi)] = E[E_n(\psi)]$$

Dans le calcul de $E_n(\psi)$, seule q_{n+1} est aléatoire (obéissant à la loi $\mathfrak{L}_n$ et $E_n(\psi)$ dépend donc des fréquences précédentes, $q_1, \ldots, q_n$; on regarde ensuite ces fréquences comme des aléatoires obéissant à la loi de probabilité *a priori* : $E_n(\psi)$ devient alors une aléatoire dont l'espérance est $E(\psi)$. Or, quand q_n est connu, nous savons que la fréquence aléatoire q_{n+1} est la moyenne arithmétique de 2N variables aléatoires X_j attachées aux 2N gamètes utiles qui, par leur fusion, constitueront F_{n+1}; ces 2N gamètes sont désignés — en vertu de la panmixie — par 2N tirages au sort indépendants, chaque tirage ayant une probabilité de fournir a qui est :

$$q'_n = q_n + \delta(q_n) = q_n - uq_n + v(1 - q_n) + q_n(1 - q_n)s(q_n)$$

et, dans le cas particulier où la sélection est linéarisée au voisinage d'une valeur d'équilibre stable C, avec un coefficient de rappel k :

$$q'_n = q_n + \delta(q_n) = q_n - k(q_n - C)$$

La loi conditionnée $\mathfrak{L}_n$ de q_{n+1} est donc une loi binomiale de moyenne q'_n, de variance $\dfrac{q'_n - q'^2_n}{2N}$; on peut d'ailleurs obtenir tous ses moments centrés $E_n[(q_{n+1} - q'_n)^p]$ par développement en série de la fonction caractéristique de la loi binomiale :

$$e^{-itq'_n}\left(q'_n\, e^{\frac{it}{2N}} + 1 - q_n\right)^{2N} = \left(q'_n\, e^{\frac{it}{2N}} p'_u + p'_n\, e^{-\frac{it}{2N}} q'_n\right)^{2N}$$

On constate aisément que les moments centrés d'ordre supérieur à deux sont de l'ordre de $\dfrac{1}{N^2}$, alors que la variance est de l'ordre de $\dfrac{1}{N}$. On a donc :

$$q_{n+1} = q'_n + \varepsilon_n \text{ avec } \begin{cases} E_n(\varepsilon_n) = 0 \\ E_n(\varepsilon_n^2 = \dfrac{q'_n - q'^2_n}{2N} \\ E_n(\varepsilon_n^p) = O\left(\dfrac{1}{N^2}\right) \text{ si } p > 2 \end{cases}$$

Une simplification d'écriture s'introduit si l'on calcule d'abord les moments à priori de $q_n - C$, soient $M^{(p)} = E[(q_n - C)^p]$. Les moments de q_n s'en déduiront aisément par le développement

$$E(q_n^p) = E[(q_n - C + C)^p]$$

On a, en effet, dans l'approximation linéaire (sélection linéarisée) que nous adopterons dorénavant

$$q_{n+1} - C = q'_n - C + \varepsilon_n = (1 - k)(q_n - C) + \varepsilon_n$$

D'où

$$(4)\quad E[(q_{n+1} - C)^p] = E[E_n[(q_{n+1} - C)^p]] = E[E_n[[(1 - k)(q_n - C) + \varepsilon_n]^p)]$$

$$(4')\quad M_{n+1}^{(p)} = (1 - k)^p M_n^{(p)} + C_p^2 E[(q'_n - C)^{p-2} E_n(\varepsilon_n^2)] + O\left(\frac{1}{N^2}\right)$$

B. — *Évolution de l'espérance mathématique* a priori

C'est le cas particulier $p = 1$

$$E(q_{n+1} - C) = (1 - k)[E(q_n - C)]$$

d'où :

$$q_{n+1} - C = (1 - k)(q_n - C) = (1 - k)^{n+1}(q_o - C)$$

q_n tend donc vers C (lentement si le coefficient de rappel k est petit). Il y a un cas d'exception : si $k = 0$, $E(q_n)$ reste constant et est donc égal à q_o.

C. — *Évolution du 2e moment et de la variance* a priori

Pour $p = 2$, (4') donne l'équation de récurrence du 2e moment.

$$E[(q_{n+1} - C)^2] = (1 - k)^2 E[(q_n - C)^2] + E[E_n(\varepsilon_n^2)]$$

$$(5)\quad E_n(\varepsilon_n^2) = \frac{q'_n - C - (q'_n - C)^2 + C - C^2 - 2C(q' - C)}{2N}$$

dont la moyenne à priori est :

$$E[E_n(\varepsilon^2)] = \frac{(1 - 2C)(1 - k)(q_n - C) + C - C^2 - E[(q'_n - C)^2]}{2N}$$

$$= \frac{C - C^2 - (1 - k)^2 E[(q_n - C)^2] + (1 - 2C)(1 - k)(\overline{q_n} - C)}{2N}$$

ce qui donne la récurrence :

$$M_{n+1}^{(2)} = (1 - k)^2 M_n^{(2)} \left((1 - \frac{1}{2N}\right) + \frac{C - C^2}{2N} + \frac{1 - 2C)(1 - k)}{2N}(q_n - C)$$

$M_n^{(2)}$ est donc une fonction de n qui est la somme de la solution générale de l'équation homogène, solution qui est $K\left[(1 - k)^2\left(1 - \frac{1}{2N}\right)\right]^n$ et de la solution particulière $M^{(2)} + H(1 - k)^n$ avec les conditions

$$M^{(2)} = (1 - k)^2\left((1 - \frac{1}{2N}\right) M^{(2)} + \frac{C - C^2}{2N}$$

$$H(1-k)^{n+1} = H(1-k)^2\left(1-\frac{1}{2N}\right)(1-k)^n + \frac{(1-2C)(1-k)}{2N}(1-k)^n \qquad (q_o - C)$$

D'où (en négligeant k^2 vis-à-vis de k) :

$$M_n^{(2)} = \frac{C-C^2}{1+4Nk} + K\left(1-2k-\frac{1}{2N}\right)^n + \frac{1-2C}{1+2Nk}(1-k)^n(q_o - C)$$

ce qui montre que $M_n^{(2)}$ tend vers $\frac{C-C^2}{1+4Nk}$ (lentement si k est petit); il en est de même de $\sigma^2_{(n)} = M_n^{(2)} - (q_n - C)^2$ dont la limite pour n très grand est aussi $\frac{C-C^2}{1+4Nk}$. Cette variance a, évidemment, l'intérêt d'indiquer l'importance de la fluctuation persistante de q_n autour de son espérance mathématique limite C c'est là la fluctuation qui pourra être observée au bout d'un nombre suffisant de générations du même groupe suivi au cours du temps.

D. — *Fréquences des zygotes*

Mais le 2e moment *a priori* $E[(q_n - C)^2]$ et le moment $E[q_n'^2]$ qui s'y ramène ont une autre application importante : en raison de la panmixie, la probabilité conditionnée, pour un zygote pris au hasard dans F_{n+1}, d'être *aa* est $q_n'^2$.

La fréquence R_{n+1} du zygote *aa* que l'on observe dans F_{n+1} résulte donc de N tirages au sort indépendants ayant chacun la probabilité conditionnée $q_n'^2$ d'être *aa*. La loi conditionnée de cette fréquence est donc une loi binomiale de moyen $q_n'^2$ de variance $\frac{q_n'^2(1-q_n'^2)}{N}$; l'espérance mathématique *a priori* de cette fréquence est donc :

$$E[E_n(R_{n+1})] = E(q'^2_n) = E[(q_n' - C)^2] + C^2 + 2C\,E(q_n' - C)$$

et elle tend vers $C^2 + (1-k)^2 \times \lim E[(q_n - C)^2]$ c'est-à-dire très sensiblement vers $C^2 + \frac{C-C^2}{1+4Nk}$ [1].

Comme C est la limite de l'espérance *a priori* $\overline{q_n}$ de la fréquence du gène *a* dans F_n, limite que nous noterons ici q, on voit que l'on a sensiblement, *lorsque* n *est suffisamment* grand, les formules :

$$E(R_n) = q^2 + \frac{q-q^2}{1+4Nk} \qquad \text{(Zygotes } aa\text{)}$$

et de même

$$E(P_n) = p^2 + \frac{q-q^2}{1+4Nk} \qquad \text{(Zygotes AA)}$$

Ainsi que, pour la fréquence $2Q_n$ du zygote A*a* :

$$E(2Q_n) = 1 - E(P_n) - E(R_n) = 2\overline{pq} - 2\frac{\overline{q}-\overline{q^2}}{1+4Nk} = 2(\overline{q}-\overline{q^2})\frac{4Nk}{1+4Nk}$$

(1) Cette formule assimile donc l'espérance de R_n au 2e moment de q_n' et à celui, très voisin, de q_n.

Les différences avec les valeurs fournies par la loi de HARDY-WEINBERG, traduisent la « consanguinité » due à l'effectif limité; il suffit d'introduire le « coefficient de consanguinité » $\varphi = \frac{1}{1 + 4N\dot{k}}$ (qui sera obtenu plus loin, en l'absence cette fois de sélection, comme valeur limite du coefficient de parenté de deux gamètes pris au hasard ou de deux gamètes s'unissant pour former un zygote) pour retrouver les relations classiques entre les proportions des trois zygotes et le coefficient de consanguinité. Remarquons que la variance de R_n peut se calculer à partir de son 2e moment qui n'est autre que le 4e moment de q'_n et, sensiblement, de q_n.

E. — *Généralisation de MORTON et YASUDA* (1962)[1].

Les moments d'ordres 3 et 4 conduisent aussi à une interprétation en termes de croisements : quand l'on connaît q_n la probabilité conditionnée d'un accouplement dans F_{n+1} entre, par exemple deux hétérozygotes A*a*, est :

$$\pi_{hh} = (2p'_n q'_n)^2 = 4(q'_n - q'^2_n)^2 = 4q'_n - 8q'^3_n + 4q'^4_n$$

La fréquence dans F_{n+1} des accouplements entre deux hétérozygotes A*a* (il y a N accouplements en tout, si l'on fait correspondre un accouplement à chaque enfant de la génération suivante) résulte de N tirages au sort indépendants dont chacun a une probabilité conditionnée π_{hh} d'être de ce type. La loi conditionnée de cette fréquence est donc une loi binomiale d'espérance π_{hh}, de variance $\frac{\pi_{hh}(1 - \pi_{hh})}{N}$.

L'espérance mathématique *a priori* de cette fréquence est donc :

(6) $\quad E(\pi_{hh}) = 4E(q'^2_n) - 8E(q'^3_n) + 4E(q'^4_n)$

Cette espérance mathématique s'exprime aisément (2) à partir des moments (d'ordre $p \leqslant 4$) (3) :

$$M^{(p)}_{1n} = E[(q'_n - C)^p] = (1 - k)^p \, E[(q_n - C)^p] = (1 - k)^p \, M^{p}_{n}$$

dont nous allons calculer, pour $p = 3$ et $p = 4$, les limites en utilisant l'approximation (4') sous la forme :

$$M^{(p)}_{1,n+1} = (1 - k)^p \, M^{(p)}_{1,n} + C^2_p (1 - k)^p \, E[(q'_n - C)^{p-2} \, E_n(\varepsilon^2_n)]$$

$$M^{(3)}_{1,+1} = (1 - k)^3 M^{(3)}_{1,n} + 3(1 - k)^3 E[(q'_n - C) E_n(\varepsilon^2_n)]$$

$$M^{(4)}_{1,n+1} = (1 - k)^4 M^{(4)}_{1,n} + 6(1 - k)^4 E[(q'_n - C)^2 E_n(\varepsilon^2_n)].$$

Ce qui donne, compte tenu de (3) :

$$M^{(3)}_{1,n+1} = (1 - k)^3 \, M^{(3)}_{1,n} + \frac{3(1 - k)}{2N} [(1 - 2C) M^{(2)}_{1n} - M^{(3)}_{1n}]$$

$$M^{(4)}_{1,n+1} = (1 - k)^4 M^{(4)}_{1,n} + \frac{6(1 - k)^2}{2N} (C - C^2) M^{(2)}_{1n} + (1 - 2C) [M^{(3)}_{1n} - M^{(4)}_{1n}]$$

(1) Cf. aussi MORTON (1969).

(2) Par la formule :

$$E(\pi_{hh}) = 4\left[M^{(1)}_{1,n} + 2CM^{(2)}_{1,n} + C^2\right] - 8\left[M^{(3)}_{1,n} + 3CM^{(2)}_{1,n} + 3C^2M^{(1)}_{1,n} + C^3\right] + 4\left[M^{(4)}_{1,n} + 4CM^{(3)}_{1,n} + 6C^2M^{(2)}_{1,n} + 4C^3M^{(1)}_{1,n} + C^4\right]$$

(3) p est ici un entier sans rapport avec une fréquence.

L'existence des limites s'établit comme précédemment; ces limites, que nous noterons $M_1^{(3)}$ et $M_1^{(4)}$, sont données par les équations suivantes (ou nous remplaçons la limite de $M_{1n}^{(2)} = (1-k)^2 M_n^{(2)}$ par $\frac{C-C^2}{1+4Nk}$, et négligeons k^2 et $\frac{k}{N}$:

$$(7)\quad 3\left(k+\frac{1}{2N}\right)M_1^{(3)} = \frac{3}{2N}\frac{(1-2C)(C-C^2)}{1+4Nk} \quad \text{d'où} \quad M_1^{(3)} = \frac{(1-2C)(C-C^2)}{(1+4Nk)(1+2Nk)}$$

$$4k + \frac{3}{N}M_1^{(4)} = \frac{3}{N}\left[\frac{(C-C^2)^2}{1+4Nk} + \frac{(1-2C)^2(C-C^2)}{(1+4Nk)(1+2Nk)}\right]$$

$$\text{d'où}\quad M_1^{(4)} = \frac{(C-C^2)^2}{(1+4Nk)\left(1+\frac{4Nk}{3}\right)} + \frac{(1-2C)^2(C-C)^2}{(1+4Nk)(1+2Nk)\left(1+\frac{4Nk}{3}\right)}$$

On en déduit à l'aide de (6) et (7) (dans l'état stationnaire) :

$$(7')\quad E(\pi_{hh}) = 4C^2 - 8C^3 + 4C^4 + (4 - 24C + 24C^2)\frac{C-C^2}{1+4Nk} - 8\frac{(1-2C)^2(C-C^2)}{(1+4Nk)(1+2Nk)} + 4\frac{(1-2C)^2(C-C^2)}{(1+4Nk)(1+2Nk)\left(\frac{1+4Nk}{3}\right)}$$

Si $\frac{1}{1+4Nk}$ (qui est aussi, rappelons-le, le « coefficient de consanguinité » φ) est assez petit pour que l'on puisse négliger les deux derniers termes de cette somme, on retrouve la *formule de Yasuda* (1966) :

$$E(\pi_{hh}) = 4C^2(1-C)^2 + 4C(1-C)[(1-6C(1-C)]\varphi$$

Les autres termes de (7') étant équivalents à

$$-16(1-2C)^2(C-C^2)\varphi^2 \quad \text{et} \quad 24(1-2C)^2(C-C^2)\varphi^3$$

Remarque.

Comme il a été indiqué au début du § D [1], le calcul que nous venions de faire peut aussi fournir les *variances a priori* des fréquences des génotypes : la variance de $2Q_n = 2p'_n q'_n$, par exemple, n'est autre que

$E\{(2p'_n q'_n)^2\} - \{E(2p'_n q'_n)\}^2 = E(\pi_{hh}) - \{E(2Q_n)\}^2$

Or, d'après l'expression obtenue pour $E(2Q_n)$ la limite de $[E(2Q_n)]^2$ est $4(C-C^2)^2\frac{16N^2k^2}{(1+4Nk)^2}$, qu'il suffit de retrancher de (7'). En négligeant encore ce qui est du 2e ordre en $\varphi = \frac{1}{1+4Nk}$, on trouve pour partie principale de la variance *a priori* de $2Q_n$

$$4C^2(1-C)^2\frac{8Nk+1}{(1+4Nk)^2} + 4C(1-C)[1-6C(1-C)]\varphi \sim 4C(1-C)[1-4C(1-C)]\varphi$$

[1] Rappelons que C et $\bar{q}$ sont deux notations équivalentes, la fréquence d'équilibre C étant la limite de l'espérance *a priori*

On constate ainsi que, quand φ est petit, l'écart-type de $2Q_n$ l'emporte de beaucoup sur la différence $2C(1-C)\varphi$ entre l'espérance de $2Q_n$ et la valeur $2C(1-C)$ qui correspondrait à $\varphi = 0$ (« loi de Hardy »). Les différences par rapport à la loi de Hardy et la valeur du coefficient de consanguinité φ ne peuvent être testées qu'en évaluant la moyenne arithmétique des fréquences $2Q_n$ observées sur un *grand nombre* de groupes panmictiques isolés de même effectif N.

Autrement dit, le rapport $\frac{2Q_n}{2C(1-C)}$, qui a une espérance égale à « l'index panmictique » $1-\varphi$ de WRIGHT (1965) fluctue beaucoup trop pour permettre d'estimer cet index autrement qu'en mesurant sa moyenne sur un grand nombre de groupes.

CHAPITRE IV. — ÉVOLUTION DES FRÉQUENCES D'UN GÈNE DANS UN ENSEMBLE DE GROUPES PANMICTIQUES COMMUNIQUANT PAR MIGRATION

Nous allons étudier la loi de probabilité *a priori* des fréquences aléatoires q_{nx} dans les divers emplacements x et dans les générations successives F_n. Nous supposerons que le groupe panmictique d'emplacement x comprend N individus, N étant indépendant de x et de n.

Quand les q_{nx} sont connus, les fréquences aléatoires $q_{n+1,x}$ sont (problème de Bernouilli) les moyennes arithmétiques de 2N variables aléatoires indépendantes ayant chacune les probabilités q'_{nx} et p'_{nx} d'être égales à 1 et à 0.

On a donc :

$$q_{n+1,x} = q'_{nx} + \varepsilon_{nx} \text{ avec } \begin{cases} E_n(\varepsilon_{nx}) = 0 \\ E_n(\varepsilon_{nx}^2) = \dfrac{q'_{nx} - q'^2_{nx}}{2N} \\ E_n(\varepsilon_{nx}\,\varepsilon_{nz}) = 0 \text{ si } x \neq z, \text{ en raison de l'indépendance des tirages} \\ E_n(\varepsilon_{nx}^p) = 0\left(\dfrac{1}{N^2}\right) \end{cases}$$

Il y a lieu encore de calculer les moments *a priori* de $q_{n+1,x} - C$ et d'utiliser le théorème de l'espérance conditionnée :

$$(1) \qquad \boxed{q_{n+1,x} - C = q'_{nx} - C + \varepsilon_{nx} = (1-k)\,Tq_{nx} + \varepsilon_{nx}}$$

T désignant l'opérateur linéaire introduit en I, C, *a*, *b* ou *c*.

$$E_n(q_{n+1,x} - C) = (1-k)\,Tq_{nx} = (1-k)\sum_z l_{z-x}(q_{nz} - C)$$

$$\overline{q_{n+1,x}} - C = E[E_n(q_{n+1,x} - C)] = (1-k)\,T(\overline{q_{nx}})$$

(puisque l'opérateur T est linéaire).

Si nous désignons par A_n le plus grand module des différences $\overline{q_{ny}} - C$, on obtient :

$$|\overline{q_{n+1,x}} - C| \leqslant (1 - k)\, A_n \sum_z l_{z-x} = (1 - k)\, A_n$$

Le plus grand module est donc réduit, d'une génération à la suivante, au moins dans le rapport $(1-k)$: il tend vers zéro quand $n \to \infty$, donc toutes les moyennes *a priori* $\overline{q_{n+1,x}}$ tendent vers la limite commune C (*).

Considérons maintenant :

$$E_n\,[(q_{n+1,x} - C)^2] = (1-k)^2\,(Tq_{nx})^2 + E_n(\varepsilon_{nx}^2)$$

qui donne :

$$E[(q_{n+1,x} - C)^2] = (1-k)^2 \sum_{zw} l_{z-x}\, l_{w-x}\, [E(q_{nz} - C)\,(q_{nw} - C)] + E\left(\frac{q'_{nx} - q'^2_{nx}}{2N}\right)$$

$$E_n\,[(q_{n+1,x} - C)\,(q_{n+1,y} - C)] = (1-k)^2\, Tq_{nx}\, Tq_{ny} \text{ si } y \neq x.$$

$$E\,[(q_{n+1,x} - C)\,(q_{n+1,y} - C) = (1-k)^2 \sum_{zw} l_{z-x}\, l_{w-y}\, E\,[(q_{nz} - C)\,(q_{nw} - C)].$$

Si l'on désigne par $V_{zw}(n)$ le 2e moment :

$E\,[(q_{nz} - C)\,(q_{nw} - C)]$ (dont la limite sera la même que celle de la covariance). on a :

$$(2) \qquad V_{xy}(n+1) = (1-k)^2 \sum_{zw} l_{z-x}\, l_{w-y}\, V_{zw}\,(n) + \delta\,(y-x)\,\frac{E(q'_{nx}) - E(q'^2_{nx})}{2N}$$

$\delta(y-x)$ étant nul quand $y \neq x$, et égal à 1 quand $y = x$.

$$E[(q'_{nx})^2] = C^2 + E[(q'_{nx} - C)^2] + 2CE(q'_{nx} - C)$$

et
$$E[(q'_{nx} - C)^2] = (1-k)^2\, E[(Tq_{nx})^2]$$

Donc, lorsque $E[q'_{nx} - C]$ est remplaçable par sa limite O, $E[(q'_{nx})^2]$ peut s'écrire :

$$C^2 + (1-k)^2 \sum_{zw} l_{z-x}\, l_{w-x}\, V_{zw}(n)$$

et (2) devient alors :

$$(2') \qquad V_{xy}(n+1) = (1-k)^2 \left[1 - \frac{\delta(y-x)}{2N}\right] \sum_{zw} l_{z-x}\, l_{w-y}\, V_{zw}(n) + \delta(y-x)\,\frac{C - C^2}{2N} \quad (**)$$

Des majorations faciles montrent que quand n tend vers l'infini la matrice $V(n)$ d'élément $V_{xy}(n)$ tend vers une matrice limite V dont les éléments V_{xy} (éléments qui sont les limites des covariances) ne dépendent que de la « distance » $y - x = d$ et sont la solution du système d'équations linéaires :

$$(3) \qquad V_{xy} = (1-k)^2 \left(1 - \frac{\delta(y-x)}{2N}\right) \sum_{zw} l_{z-x}\, l_{w-y}\, V_{zw} + \delta(y-x)\,\frac{C - C^2}{2N}$$

(*) Le calcul peut être étendu au cas où les taux de mutation et de sélection dépendent de l'emplacement x : les valeurs d'équilibre sont alors différentes, les limites aussi.

**) Puisque $E(q'_{nx}) \to C$ et $E[(q'_{nx})^2] - E[(q'_{nx} - C)^2] \to C^2$.

Cette solution est unique car la solution générale du système homogène associé à (2′) tend vers 0 quand n tend vers l'infini.

Ce système (3) est d'ailleurs, par rapport aux indices x, y, z, w une équation aux différences finies :

$$h_1^2 - \frac{1}{h_1^2}$$

1. — *Cas unidimensionnel à migration symétrique entre colonies adjacentes*

L'équation (I, C, 2) est seulement « du 2e ordre », dans le cas particulier unidimensionnel si l'opérateur T se ramène à un « opérateur de différence seconde ».

$$Tq_{nx} = \sum_z l_{z-x}\,(q_{nz}-C) = q_{nx}-C+m\,\Delta^2\,(q_{nx}-C)$$

MALÉCOT (1950, 1951, 1954).

Le système (3) s'écrit alors, en posant $V_{xy} = v(y-x) = v(d)$, d étant un entier de signe quelconque [on a évidemment $v(-d) = v(d)$].

$$(4)\quad v(d) = (1-k)^2\left[1 - \frac{\delta(d)}{2N}\right]\{(1-2m)^2\,v(d) + 2m(1-2m)\,[v(d+1) + v(d-1)] + m^2\,[2v(d) + v(d-2) + v(d+2)]\} + \delta(d)\,\frac{C-C^2}{2N}.$$

C'est là une équation aux différences finies linéaires du 4e ordre à coefficients constants et homogènes lorsque $d \neq 0$; $v(d)$, devant resté borné quand $d \to +\infty$, sera, pour $d > 0$ (*) une combinaison linéaire de solutions de la forme h^d, h étant une racine de module < 1 de l'équation caractéristique :

$$1 = (1-k)^2\left\{(1-2m)^2 + 2m^2 + 2m\,(1-2m)\left(h + \frac{1}{h}\right) + m^2\left(h^2 + \frac{1}{h^2}\right)\right\}$$

en posant $h + \frac{1}{h} = u$, on est ramené à l'équation du 2e degré :

$$m^2u^2 + 2m\,(1-2m)\,u + (1-2m)^2 - \frac{1}{(1-k)^2} = 0$$

Le discriminant est positif et égal à $\frac{m^2}{(1-k)^2}$; d'où :

$$u = \frac{-m\,(1-2m) \pm m/(1-k)}{m^2}$$

ou, en posant $1/(1-k) = 1+k_1$ (**)

$$u = 2 - \frac{1}{m} \pm \frac{1+k_1}{m}$$

(*) Hypothèse permise puisque, si d est < 0, $w(d) = w(-d)$.
(**) k étant petit, k_1 est très peu différent de k.

à la première racine $u_1 = 2 + \frac{k_1}{m}$ correspond pour h l'équation du deuxième degré: $h^2 - \left(2 + \frac{k_1}{m}\right) h + 1 = 0$, de racines positives, la seule inférieure à 1 étant :

$$h_1 = 1 + \frac{k_1}{2m} - \sqrt{\frac{k_1}{m} + \frac{k_1^2}{4m^2}}$$

elle est voisine de 1 si $\frac{k_1}{m}$ est petit; à la 2e racine $u_2 = 2 - \frac{2}{m} - \frac{k_1}{m}$ correspond pour h l'équation du 2e degré :

$h^2 - \left(2 - \frac{2}{m} - \frac{k}{m}\right) h + 1 = 0$ de racines négatives $\left(\text{puisque } \frac{2}{m} > 2\right)$; la seule racine supérieure à -1 est :

$$h^2 = 1 - \frac{1}{m} - \frac{k}{2m} + \sqrt{\left(\frac{1+k/2}{m}\right)^2 - \frac{2+k}{m}}$$

Cette racine est petite en valeur absolue si m est grand, car la somme des racines est alors grande en valeur absolue et elles sont inverses l'une de l'autre.

On résout donc (4), lorsque $d > 0$, par :

$$v(d) = \lambda h_1^d + \mu h_2^d \tag{5}$$

les constantes λ et μ étant déterminées par la condition de parité, qui abaisse l'ordre de la récurrence lorsque $d = 1$ et 2, et par l'équation non homogène (4) écrite lorsque $d = 0$:

$$(4_0)\quad v(0) = (1-k)^2 \left(1 - \frac{1}{2N}\right) \{[(1-2m)^2 + 2m^2] v(0) + 2m(1-2m)\,[v(1) + v(-1)] + m^2\,[v(2) + v(-2)]\} + \frac{C - C^2}{2N}$$

$$(4_1)\quad v(1) = (1-k)^2 \{[(1-2m)^2 + 2m^2] v(1) + 2m(1-2m)\,[v(0) + v(2)] + m^2\,[v(-1) + v(3)]\}$$

$$(4_2)\quad v(2) = (1-k)^2 \{[(1-2m)^2 + 2m^2] v(2) + 2m(1-2m)\,[v(1) + v(3)] + m^2\,[v(0) + v(4)]\}$$

Or la solution (5) vérifie toujours l'équation homogène associée à (4), lorsqu'on traite cette équation homogène comme une récurrence *descendante* à partir de $d \geqslant 1$. Donc (4_2) entraîne que $w(0) = \lambda + \mu$, et (4_1) entraîne que $w(-1) = \frac{\lambda}{h_1} + \frac{\mu}{h_2}$; mais il n'en est pas de même pour $w(-2)$, puisque (4_0) n'est plus homogène; en fait, la récurrence homogène descendante sur (5) donne :

$$v(0) = (1-k)^2 \{[(1-2m)^2 + 2m^2] v(0) + 2m(1-2m)\,[v(1) + v(-1)] + m^2 \left[v(2) + \frac{\lambda}{h_1^2} + \frac{\mu}{h_2^2}\right]\}$$

en substituant dans (4_0), on obtient une relation entre $v(0) = \lambda+\mu$ et $v(2) = \lambda h_1^2+\mu h_2^2$:

$$\frac{2Nv(0)}{2N-1} = \frac{C-C^2}{2N-1} + (1-k)^2m^2\left[v(-2) - \frac{\lambda}{h_1^2} - \frac{\mu}{h_2^2}\right] + v(0)$$

$$(\lambda+\mu) = v(0) = (2N-1)(1-k)^2m^2\left[\lambda\left(h_1^2 - \frac{1}{h_1^2}\right) + \mu\left(h_2^2 - \frac{1}{h_2^2}\right)\right] + C-C^2$$

$$\text{or} \quad h_2^2 - \frac{1}{k_1^2} = \left(h_1 + \frac{1}{h_1}\right)\left(h_1 - \frac{1}{h_1}\right) = -2\left(2 + \frac{k_1}{m}\right)\sqrt{\frac{k_1}{m} + \frac{k_1^2}{4m^2}}$$

$$h_2^2 - \frac{1}{h_2^2} = \left(h_2 + \frac{1}{h_2}\right)\left(h_2 - \frac{1}{h_2}\right) = 2\left(2 - \frac{2}{m} - \frac{k}{m}\right)\sqrt{\left(\frac{1+k/2}{m}\right)^2 - \frac{2+k}{m}}$$

La résolution s'achève aisément. Dans le cas particulier où m^2 est négligeable ainsi que k^2, le résultat ainsi obtenu est le même que si l'on néglige directement m^2 et k^2 dans l'équation (3) qui s'écrit alors (pour $d \geq 1$) :

$$(3') \quad v(d) = (1-2k)v(d) + 2m[v(d-1)-2v(d)+v(d+1)]$$

Cette équation aux différences finies de 2e ordre a pour solution générale une combinaison linéaire de deux solutions « exponentielles » de la forme α^d, α étant l'une ou l'autre des racines de l'équation :

$$m\alpha^2 - (k + 2m)\alpha + m = 0$$

Mais entre les racines $\left.\begin{matrix}\alpha_1\\ \alpha_2\end{matrix}\right\} = 1 + \frac{k}{2N} \pm \sqrt{\frac{(k+2m)^2-4m^2}{4m^2}}$ seule la plus petite (soit α_1) est à retenir; car la plus grande α_2 est > 1 et fournit une exponentielle α_2 qui augmente indéfiniment.

On a donc $v(d) = \lambda\alpha_1^d$ $(\forall d \geqslant 0)$ (1), et le coefficient $\lambda = v(0)$ se détermine en écrivant l'équation (3) pour $d = y-x = 0$:

$$(3'') \quad v(0) = \{(1-2k)v(0)+2m[v(-1)-2v(0)+v(1)]\}\left(1 - \frac{1}{2N}\right) + \frac{C-C^2}{2N}$$

Puisque $v_{xy} = v_{yx}$, on a $v(-1) = v(1) = \alpha_1 v(0)$, d'où, (en retranchant de deux membres $\frac{v(0)}{2N}$, pour alléger le calcul) :

$$(2k + 4m - 4m\alpha_1)\left(1 - \frac{1}{2N}\right)v(0) = \frac{C-C^2-v(0)}{2N}$$

ce qui donne, en négligeant $\frac{1}{2N}$ vis-à-vis de 1 :

$$v(0) = \frac{C-C^2-v(0)}{4Nk + 8Nm(1-\alpha_1)} = \frac{C-C^2-v(0)}{4Nk + 4N[\sqrt{4mk + k^2}-k]} = \frac{C-C^2-v(0)}{4N\sqrt{4mk + k^2}}$$

(1) L'équation homogène (3') restant valable pour $d = 1$, et introduisant alors dans le second membre $v(0)$, assure le prolongement de la solution jusqu'à la valeur $d = 0$.

et $$v(0) = \frac{C-C^2}{1+4N\sqrt{4mk+k^2}}$$

Lorsque $m = 0$, on retrouve bien la variance limite $\frac{C-C^2}{1+4Nk}$ obtenue pour un groupe isolé; lorsque m est > 0, la variance est plus faible; ou, ce qui revient au même, elle est la même que dans un groupe isolé dont l'effectif serait supérieur à N et égal à $N\sqrt{1+\frac{4m}{k}}$.

Pour deux groupes dont les emplacements sont distants de $d > 0$, nous avons vu que leur covariance limite $v(d)$ est égale à $v(0)\alpha_1^d$. Il en résulte que leur coefficient de corrélation est :

$$\frac{v(d)}{v(0)} = \alpha_1^d = \alpha_2^{-d} = \left[1 + \frac{k}{2m} + \sqrt{\frac{k}{m} + \frac{k^2}{4m^2}}\right]^{-d}$$

Il décroît donc en fonction exponentielle de la distance d (qui est, rappelons-le, le nombre d'intervalles séparant les 2 groupes). Lorsque $\frac{k}{m}$ est petit, la parenthèse $1 + \sqrt{\frac{k}{m}} + o\left(\frac{k}{m}\right)$ est voisine de 1, donc, puisque $(1+\varepsilon)^{\frac{1}{\varepsilon}} \sim e$, on a : $(1+\varepsilon)^{-d} \sim e^{-d\varepsilon}$: la corrélation est réduite par un facteur égal à $\frac{1}{e}$ lorsque la distance d augmente de $\frac{1}{\varepsilon} \sim \sqrt{\frac{m}{k}}$; sa décroissance est donc lente lorsque le taux de migration est grand par rapport au coefficient de rappel k.

2. — *Deuxième Méthode*

Il existe pour retrouver ces résultats une autre méthode que nous appliquerons d'abord au cas particulier précédent, puis au cas général de migration unidimensionnelle. Nous allons associer à la fonction $v(d)$, regardée maintenant comme une fonction paire de l'entier d positif ou négatif, une « fonction génératrice » $F(\alpha)$ définie par la « série de Laurent » :

$$F(\alpha) = \sum_d \alpha^d v(d)$$

que nous considérons seulement sur la circonférence C où $|\alpha| = 1$, où elle définit [1] la « série de Fourier » :

$$F(e^{i\Theta}) = \sum_d v(d)\, e^{i\Theta d}$$

[1] Sa convergence absolue et uniforme sur ce cercle (pour $k > 0$) peut être démontrée directement, voir plus loin (*c*), mais elle est assurée si l'on se reporte à la valeur de $v(d)$ déduite par la première méthode.

avec $v(d) = v(-d) = \frac{1}{2\pi}\int_0^{2\pi} e^{i\theta d}\,F(e^{i\theta})d\theta = \frac{1}{2\pi i}\int_C \alpha^{d-1}F(\alpha)d\alpha$

a) En multipliant les deux membres de (3′) par α^d pour toutes les valeurs de d et ajoutant à (3″) on obtient :

$$F(\alpha) = (1-2k)F(\alpha) + 2m\left[\alpha - 2 + \frac{1}{\alpha}\right]F(\alpha) + \frac{C-C^2-(1-2k)v(0)-4m[v(1)-v(0)]}{2N}$$

dans le dernier terme, le coefficient de $v(0)$ est :

$-[1-2k+4m(h_1-1)]$ que nous pouvons réduire à -1 si $\sqrt{4mk+k^2}$ est petit. On obtient alors :

$$F(\alpha) = \frac{C-C^2-v(0)}{4N\left[k + 2m - m\left(\alpha + \frac{1}{\alpha}\right)\right]}$$

$F(\alpha)$ est une fonction rationnelle dont les pôles sont les quantités α_1 et α_2 rencontrées précédemment; elle est donc décomposable en deux éléments simples proportionnels à $\frac{1}{\alpha-\alpha_1}$ et $\frac{1}{\alpha-\alpha_2}$; ces deux fractions sont elles-mêmes développables, sur la circonférence $|\alpha| = 1$, en les séries $\frac{1}{\alpha} + \frac{\alpha_1}{\alpha^2} + \frac{(\alpha_1)^{d-1}}{\alpha^d} + \ldots$

et $$-\left[\frac{1}{\alpha_2} + \frac{\alpha}{\alpha^2} + \ldots + \frac{\alpha^d}{\alpha_2^{d+1}} + \ldots\right]$$

Une combinaison de ces deux séries fournit la série de Laurent cherchée. Il suffit d'écrire :

$$F(\alpha) = \frac{C-C^2-v(0)}{-4Nm}\,\frac{\alpha}{(\alpha-\alpha_1)(\alpha-\alpha_2)} = \frac{A_1}{\alpha-\alpha_1} + \frac{A_2}{\alpha-\alpha_2}$$

avec $$A_1 = \frac{C-C^2-v(0)}{4Nm} \times \frac{\alpha_1}{\alpha_2-\alpha_1} \text{ et } A_2 = \frac{C-C^2-v(0)}{4Nm}\,\frac{\alpha^2}{\alpha_1-\alpha_2}$$

pour obtenir $$v(d) = v(-d) = -\frac{A_2}{\alpha^{d-1}} = A_1\alpha_1^{d-1} = \frac{C-C^2-v(0)}{4Nm(\alpha_2-\alpha_1)}\,\alpha_1^d$$

ce qui redonne le résultat précédent obtenu.

b) On peut traiter de la même façon le cas de p groupes répartis de façon équidistante sur un *milieu unidimensionnel fermé* (cercle).

La distance d de deux groupes n'est alors définie que modulo p, et il faut définir $F(\alpha)$ en sommant $\alpha^d\,v(d)$ sur une suite de p valeurs entières, et prenant pour α les racines de l'unité.

c) On peut traiter de la même façon *le cas général unidimensionnel* défini par la formule (3) que l'on récrit sous la forme :

$$V_{xy}\alpha^{y-x} = (1-k)^2\sum_{zw}\alpha^{z-x}l_{z-x}\alpha^{w-z}V_{zw}\alpha^{y-w}l_{w-y} + \delta_1(y-x)\alpha^{y-x}$$

en désignant, pour simplifier, par $\delta_1(y-x)$ la fonction nulle pour $y-x \neq 0$ est égale, lorsque $y = x$, à :

$$\frac{C-C^2-(1-k)^2\sum_{zw} l_{z-x}l_{w-x}V_{zw}}{2N} = \delta_1(0)$$

En sommant sur toutes les valeurs positives et négatives de la distance algébrique $d = y-x$ [1], et intervertissant l'ordre des sommations, ce qui est licite puisque l'on a affaire à une série triple, absolument convergente lorsque $|\alpha| = 1$ et $k > 0$ on obtient, compte tenu du fait que $V_{xy} = v(y-x) = v(d)$:

$$F(\alpha) = \sum_x \alpha^d(d) = \sum_x \alpha^{y-x}V_{xy} = (1-k)^2 \sum_{zw} \left(\sum_x \alpha^{z-x}l_{z-x}\right)\alpha^{w-z}V_{zw}\alpha^{y-w}l_{w-y} + \delta_1(0).$$

La série $\sum_x \alpha^{z-x}l_{z-x} = \sum_y \alpha^y l_y$ est, pour $|\alpha| = 1$, absolument convergente (puisque $\sum_y l_y = 1$); sa somme, que nous appellerons $L(\alpha)$, est la fonction génératrice de la loi de migration parentale (ou loi de probabilité de provenance des gamètes).

$L(\alpha)$ ne dépendant pas de z, se met en facteur; la sommation suivante, effectuée par rapport à z, fait apparaître le facteur $\sum_z \alpha^{w-z}v_{zw} = F(\alpha)$.

Après mise en facteur de $L(\alpha)$ $F(\alpha)$, la dernière sommation, effectuée par rapport à w, ne porte que sur $\alpha^{y-w}l_{w-y} = \left(\frac{1}{\alpha}\right)^{w-y}l_{w-y}$; elle donne donc un troisième facteur $L\left(\frac{1}{\alpha}\right)$. D'où :

$$F(\alpha) = (1-k)^2\, L(\alpha)\, F(\alpha)\, L\left(\frac{1}{\alpha}\right) + \delta_1(0)$$

$$(5)\quad F(\alpha) = \frac{\delta_1(0)}{1-(1-k)^2 L(\alpha) L\left(\frac{1}{\alpha}\right)} \quad \text{et} \quad v(d) = v(-d) = \frac{1}{2\pi i}\int_C \alpha^{d-1}F(\alpha)d\alpha$$

(Formule de Fourier)

Dans le cas usuel où la fonction génératrice de migration $L(\alpha)$ ne comporte qu'un nombre fini de monomes, $F(\alpha)$ est une fonction rationnelle, dont les pôles sont les racines α de l'équation $L(\alpha)\, L\left(\frac{1}{\alpha}\right) = \frac{1}{(1-k)^2} = 1+k_1$ [2]

Posons $L(\alpha)\, L\left(\frac{1}{\alpha}\right) = H(\alpha)$; $H(\alpha)$ étant une fonction qui, égale à 1 pour $\alpha = 1$, y possède un « point selle » au voisinage duquel $H(\alpha)$ est > 1 si α est réel, et < 1 si $|\alpha| = 1$;

En effet : $\frac{dH}{d\alpha} = L'(\alpha)\, L\left(\frac{1}{\alpha}\right) - \frac{1}{\alpha^2} L(\alpha)\, L'\left(\frac{1}{\alpha}\right)$ pour $\alpha = 1$ [3]

[1] Ou bien sur p valeurs consécutives quelconques dans le cas de p groupes sur le cercle, α étant alors racine d'ordre p de l'unité.

[2] Si k est très petit, on peut prendre $k_1 = 2k$.

[3] L'apostrophe indique une dérivation par rapport à l'argument.

$$\frac{d^2H}{d\alpha^2} = L''(\alpha)\,L\left(\frac{1}{\alpha}\right) - \frac{2}{\alpha^2}L'(\alpha)\,L'\left(\frac{1}{\alpha}\right) + \frac{1}{\alpha^4}L(\alpha)\,L''\left(\frac{1}{\alpha}\right) + \frac{2}{\alpha^3}L(\alpha)\,L'\left(\frac{1}{\alpha}\right)$$

qui, pour $\alpha = 1$, est égal à $2L''(1)+2L'(1)-2[L'(1)]^2 =$
$= 2\sum_y y(y-1)l_y + 2\sum_y yly - 2\left(\sum_y yly\right)^2$ ce qui n'est autre que le double de la « variance de migration » que nous noterons σ^2.

On a donc, au voisinage de $\alpha = 1$:

$$H(\alpha) = 1+\sigma^2(\alpha-1)^2+0[(\alpha-1)^3]$$

(1 est donc racine double de l'équation $H(\alpha) = 1$)

Parmi les pôles, racines de l'équation $H(\alpha) = 1+k_1$, il y en a, si k_1 est très petit, (donc équivalent à $2k$), deux et deux seulement très voisins de 1, réels et donnés par l'équation de 2e degré $(\alpha-1)^2 = \frac{k_1}{\sigma^2}$; appelons les α_2 et $\alpha_1(0 < \alpha_1 < 1 < \alpha_2)$.

Appelons les autres α_j $(j > 2)$; les éléments simples de $F(\alpha)$ ont pour numérateurs :

$$A_j = \frac{\delta_1(0)}{(1-k)^2}\lim_{\alpha\to\alpha_j}\frac{\alpha-\alpha_j}{1+k_1-H(\alpha)} = -\frac{\delta_1(0)}{(1-k)^2}\frac{1}{\frac{dH}{d\alpha_j}}$$

pour $j = 1$ ou 2, $\frac{dH}{d\alpha_j} = 2\sigma^2(\alpha_j-1)+0\,[(\alpha_j-1)^2] = (-1)^j\,2\sigma\sqrt{k_1} + 0\left(\frac{4k}{\sigma^2}\right)$.

Il s'agit donc de dérivées très petites qui donnent des résidus grands; il n'en est pas de même en général pour les autres racines de l'équation $H(\alpha) = 1+k_1$ [1].

Plutôt que de développer en séries de Laurent les éléments simples, il est plus clair d'introduire les résidus de ceux des pôles, qui sont intérieurs au cercle $C(|\alpha|=1)$ en écrivant :

$$v(d) = \frac{1}{2\pi i}\sum_j\int_c \frac{\alpha_j^{d-1}A_j}{\alpha-\alpha_j}\,d\alpha = \sum_{(\text{int. de C})}\alpha_j^{d-1}A_j$$

Or en peut montrer que, si $l_0l_1 > 0$, il n'y a pas d'autre pôle que α_1, qui soit à la fois intérieur à C et très voisin de 1 en module (0 n'est pas pôle si $d > 0$); on a donc, lorsque d est assez grand [2] :

$$v(d) \sim A_1\,\alpha_1^{d-1} \sim \frac{\delta_1(0)}{2\sigma\sqrt{2k}}\left(1-\frac{\sqrt{2k}}{\sigma}\right)^d$$

ce qui met, encore, en évidence une décroissance exponentielle, l'augmentation de distance nécessaire pour que la corrélation soit réduite par un facteur égal à $\frac{1}{e}$ étant sensiblement égale à $\frac{\sigma}{\sqrt{2k}}$; cette formule englobe celle du cas particulier

[1] Ces racines ne donneraient de résidus grands que si elles étaient voisines de racines doubles; or il est exceptionnel que cela se produise pour d'autres racines que celles α_1 et α_2, qui sont voisines de 1.

[2] Et même $\forall d \geqslant 0$ si, vis-à-vis de A_1, les autres résidus des pôles intérieurs sont négligeables.

précédemment traité, pour lequel σ^2 était égal à $2m$ supposé petit, ce qui donnait

$$\sigma/\sqrt{2k} = \sqrt{m/k}$$

Pour calculer $v(0)$, remarquons que $\delta_1(0)$ a été défini comme égal à :

$$\frac{C-C^2-(1-k)^2\sum_{zw} l_{z-x}l_{w-x}V_{zw}}{2N}$$

Or la somme $\sum_{zw}$ s'écrit aussi [1] :

$$\sum_{zw} \alpha^{z-x} l_{z-x}\alpha^{x-w} l_{w-x}\alpha^{w-z}V_{zw}$$

qui est le coefficient de α^0 dans la série :

$$\sum_{yzw} \alpha^{z-y} l_{z-y}\alpha^{x-w} l_{w-x}\alpha^{w-z}V_{zw} = L(\alpha)\,F(\alpha)\,L\left(\frac{1}{\alpha}\right)$$

et est donc égal à l'intégrale $\frac{1}{2\pi i}\int_c L(\alpha)\,F(\alpha)\,L\left(\frac{1}{\alpha}\right)\frac{d\alpha}{\alpha}$

alors que $v(0)$ est égal à l'intégrale $\frac{1}{2\pi i}\int_c F(\alpha)\,\frac{d\alpha}{\alpha}$

Nous avons vu que cette dernière intégrale est approchée par le résidu relatif au pôle α_1, résidu qui est $\frac{A_1}{\alpha_1}$, alors que la première est approchée par le résidu $L(\alpha_1)\,L\left(\frac{1}{\alpha_1}\right)\frac{A_1}{\alpha_1} = H(\alpha_1)\,\frac{A_1}{\alpha_1} = (1+k_1)\,\frac{A_1}{\alpha_1}$, en la multipliant par $(1-k)^2$ qui est aussi $\frac{1}{1+k_1}$, on trouve :

$$\delta_1(0) = \frac{C-C^2-v(0)}{2N}$$

$v(0)$ est donc donné par l'équation :

$$v(0) \sim \frac{\delta_1(0)}{2\sigma\sqrt{2k}} = \frac{C-C^2-v(0)}{4N\sigma\sqrt{2k}}$$

D'où :

$$v(0) \sim \frac{C-C^2}{1+4N\sigma\sqrt{2k}} \quad \text{et} \quad v(d) \sim v(0)\left(1-\frac{\sqrt{2k}}{\sigma}\right)^d$$

généralisation (en remplaçant $\sqrt{2m}$ par σ, et négligeant k^2) des formules obtenues dans le cas de migration entre colonies adjacentes seulement.

[1] L'intervention de cette somme tient à ce que la variance de ε_x se déduit de celle de q'_{nx}, *a priori* différente de la variance de $q_{n+1,x}$, que nous choisissons pour inconnue sous le nom de $v(0)$.

Nous avons établi que, dans le cas bidimensionnel, la décroissance asymptotique reste « quasi exponentielle », le 2e membre de $v(d)$ étant modifié par l'introduction d'un facteur égal à $\frac{1}{\sqrt{d}}$; la covariance $v(d)$ possède un « décrément logarithmique » $\frac{v(d+l)}{v(d)}$ qui, pour d grand, est équivalent à $\left(1 - \frac{\sqrt{2k}}{\sigma}\right)^l$; cette formule très simple permet l'estimation du « coefficient de rappel » k lorsque l'écart-type de migration σ peut être estimé.

Reçu pour publication en février 1971.

SUMMARY

GENETICS OF NATURAL DIPLOID POPULATIONS FOR ONE LOCUS :

I. — EVOLUTION OF GENE FREQUENCIES IN AN ISOLATED PANMICTIC GROUPS

Natural populations are often distributed in groups which are sometimes isolated, but generally communicate by migration.

The number in each group is limited. In order to simplify, we designate the number by N, supposing that it varies little from one group to another.

This limited number is the cause of " genetic drift ". If, to simplify, we suppose that generations are separated, the N diploid individuals of the F_{n+1} generation then result from the fusion of 2N *useful gametes* drawn at random from a *much greater* number of gametes *produced* by the preceeding F_n generation, this latter whole constituting what is called the *gametic pool.*

This poses two problems concerning the passage from a generation to the following one :

1) constitution of the gametic pool;

2) method of drawing 2N useful gametes;

1) If the frequency of gene a in N adult breeding animals of the F_n generation is known, and if it is equal to q_n, the frequency of a in all the gametes they produce: (that is, in the gametic pool for F_{n+1}) will generally be different ($q_n = q_n + \delta(q_n)$) because of the pressures of mutation and selection. The effects of these latter (for the production of an infinite number of gametes, are discussed in chapter II. Thus, for small variations having a balanced frequency of about $\bar{q} = C$, $\delta(q_n)$ may be expressed as:

$$\delta(q_n) = -k(q_n - C)$$

These formules must be changed if migration between groups occurs. The following notations (see fig. 1) are described in detail in chapter I which in an introduction to the problems treated in following chapters.

q_{nx} is the frequency of gene a on the whole of N diploids which constitute the group at site x in the F_n generation.

The result of migration if that the gametic pool, built up at site x for the following generation, is constituted by the action of mutation and selection on a q^*_{nx} frequency which is a weighted average of frequencies on group x and on neighboring groups, designated by z, or

$$q^*_{nx} = \sum_z l_{xz}\, q_{nz}$$

l_{xz} being the migration rate of group z in the group x ($\sum_z l_{xz} = 1$).

When migration if followed by mutation and selection, the composition of the gametic pool is defined by:

$$q'_{nx} = q^*_{nx} + \delta(q^*_{nx})$$

when linearization is possible, this gives:

$$q'_{nx} = C + (1 - k)(q^*_{nx} - C)$$

Chapter III develops this formulation for an isolated group; that is, without migration. Thus, index x may be omitted.

Chapter IV (next to appear) repeats the computation, introducing migration between groups.

2) The way in which 2N useful gametes are drawn at random in the gametic pool depends both on the method of crossing (defined by the statistical relationship between male and female gametes which unite to form each egg) and on individual variability in fertility (measured among useful gametes).

When fertility obeys Poisson's law and when gamete fusion occurs at random, the result is *panmixy.* An isolated panmictic group is studied in chapter III.

The q_{n+1} frequency of a in the F_{n+1} generation is deduced from the probability q'_n (relative to each 2N gamete drawn at random independent ly in the gametic pool) by the formula:

$$q_{n+1} = q'_n + \varepsilon_n$$

ε_n being the random variable having a conditional law of probability (when q_n and q'_n are known) which, according to Bernoulli's theorem has a zero mean and a variance equal to:

$$\frac{q'_n(1 - q'_n)}{2N}$$

Using *a priori* reasoning (that is, knowing only the initial F_0 generation with its q_0 frequency) future $q_n (n \geqslant 1)$ frequencies are random variables linked in Markoff chain. The *a priori* means and moments of these variables are studied in chapter III. The *a priori* $\bar{q}_n$ mean goes toward the limit $\bar{q} = C$; *a priori* variance goes toward the limit $\sigma^2 = \frac{C - C^2}{1 + 4Nk}$ which defines the fluctuation of statistical balance (fluctuation estimable on numerous isolated groups of the same number N) in the asymptotic state (stationary) (§ C).

Moreover, this variance is related to the second *a priori* moment of the q_n or q_n frequency, and thus to frequency means of three genotypes — *aa*, A*a*, and AA (§ D).

The third and fourth *a priori* moments give frequency means of pairing of all the genotypic pairs (MORTON and YASUDA) (§ E).

Chapter IV studies the covariance between two panmictic groups in relation with their distance giving a " quasi exponential " expression of the observed decreasing of correlation coefficient.

RÉFÉRENCES BIBLIOGRAPHIQUES

Outre les références indiquées dans le texte, l'auteur s'est basé sur des travaux de FISHER, WRIGHT et HALDANE cités dans la bibliographie.

FISHER, R.A., 1918. The correlation between relatives on the supposition of Mendelian inheritance. *Trans. Roy. Soc. Edinburgh*, **52**, 399-433.

FISHER, R.A., 1930. *The Genetical Theory of Natural Selection.* Clarendon Press. London.

HALDANE, J.B.S., 1939. The spread of harmful autosomal recessive genes in human population. *Annals Eugenics*, **9**, 232-237.

MALÉCOT G., 1948. *Les mathématiques de l'hérédité.* Masson, Paris.

MALÉCOT G., 1950. Quelques schémas probabilistes sur la vsriabilité des populations naturelles. *Ann. Univ. Lyon*, Section A, **13**, 37-60.

MALÉCOT G., 1951. Un traitement stochastique des problèmes linéaires (mutation, linkage, migration) en génétique de populations. *Ann. Univ. Lyon.* Section A, **14**, 79-117.

MALÉCOT G., 1954. Sur les modèles stochastiques, linéaires, asymptotiquement stationnaires. *Ann. Univ. Lyon*, Section A, **17**, 19-35.

MORTON, N.E., 1969. Human population structure. *Ann. Rev. Genet.* **3**, 53-74.

MORTON N.E., YASUDA, 1962. Structure of human populations. in J. SUTTER. Les déplacements humains, Entretiens de Monaco.

WRIGHT S., 1931. « Evolution in Mendelian population ». *Genetics*, **16**, 97-159.

WRIGHT S., 1939. *Statistical Genetics in relation to Evolution.* Actualités Scientifiques et Industrielles, 802, Hermann Paris.

WRIGHT S., 1946. Isolation by distance under diverse systems of mating. *Genetics*, **31**, 39-59.

WRIGHT S. 1965. The interpretation of population structure by F Statistics with special regards to systems of matings. Evolution, *Lawrence, Kans*, **9**, 395-420.

YASUDA, 1966. *The Genetical Structure of North-Eastern Brazil.* Thes. Univ. of Honolulu. Hawaï.