

CONSANGUINITÉ PANMICTIQUE ET CONSANGUINITÉ SYSTÉMATIQUE

(COEFFICIENTS DE WRIGHT ET DE MALÉCOT)

G. MALÉCOT

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

SOMMAIRE

Dans des populations diploïdes qui sont soumises à un certain isolement géographique matérialisé par l'existence de « subdivisions » ou « sites », il est possible de définir 3 coefficients de consanguinité : F et α de S. WRIGHT et φ de MALÉCOT. F s'appelle coefficient de consanguinité totale, α coefficient de consanguinité systématique et φ coefficient de consanguinité panmictique. En utilisant la relation entre corrélations conditionnées (sur des sites où la composition du réservoir gamétique est connu) et corrélations *a priori* (pour un ensemble nombreux de sites dont les réservoirs gamétiques individuels sont inconnus) on démontre que $F = \alpha + \varphi - \alpha\varphi$. Cette relation avait déjà été énoncée par S. WRIGHT comme théorème de « multiplication des index panmictiques » dans une classification « hiérarchique » où la population totale est décomposée en « subdivisions ».

S. WRIGHT (1965) a montré que le « coefficient de consanguinité » d'un individu diploïde I , c'est-à-dire la corrélation entre les gamètes qui, dans un site géographique déterminé (désigné par x) s'unissent pour former un diploïde I , a des valeurs numériques différentes (α ou F) ⁽¹⁾ suivant que l'on prend pour espérance mathématique de chaque aléatoire gamétique la moyenne locale (fréquence génique locale q) ou la moyenne générale (moyenne C , calculée sur l'étendue la plus grande possible, c'est-à-dire sur l'ensemble de la population, des fréquences géniques).

Nous avons introduit la notion de « réservoir gamétique local » (dans le site géographique x) pour désigner l'ensemble de tous les gamètes disponibles dans le site x (MALÉCOT, 1960). Seule une faible portion de ces gamètes (les « gamètes utiles ») est utilisée pour produire les diploïdes qui naissent dans le site x . Il y a « panmixie locale » lorsque les deux gamètes qui s'unissent

⁽¹⁾ En notation de MORTON. Dans la notation de WRIGHT, $\alpha = F_{IS}$ et $F = F_{IT}$; S (« subdivision » ou « site ») désigne le réservoir gamétique local et T (« totalité ») l'ensemble de la population (gamétique ou zygotique, car les fréquences sont les mêmes dans les deux cas, en vertu de la loi des grands nombres).

pour former chaque diploïde sont tirés au sort *indépendamment* : alors leur corrélation α conditionnée par la connaissance du réservoir gamétique local (défini par la fréquence génique locale q) est nulle, et leur corrélation « *a priori* » F (calculée dans l'ignorance de q , donc dans l'ignorance de la composition du réservoir panmictique local) se réduit à la probabilité *a priori* $\varphi_{xx}(n)$ (1) pour que deux gamètes pris indépendamment dans le même site x lorsque n générations seront écoulées soient identiques : cette probabilité est le cas particulier (pour $y = x$) de ce que nous avons appelé le « coefficient de parenté » $\varphi_{xy}(n)$ (MALÉCOT, 1960), c'est-à-dire de la probabilité *a priori* pour que deux gamètes, pris respectivement et indépendamment dans les sites x et y lorsque n générations seront écoulées, soient identiques.

Nous avons montré ailleurs (MALÉCOT, 1967) comment $\varphi_{xy}(n)$ est calculable (2) lorsque l'on connaît le nombre $2N$ de gamètes utiles tirés au sort dans chaque site et les taux d'immigration qui, à chaque génération, transfèrent dans le réservoir gamétique local des gènes extérieurs. En particulier ces calculs fournissent $\varphi_{xx}(n)$ qui est > 0 et qui peut, pour la raison indiquée plus haut, être qualifié de « consanguinité panmictique » ; sa valeur stationnaire est, en bref, notée φ .

Lorsqu'il n'y a pas panmixie locale, mais « union préférentielle » des gamètes et lorsque cette union préférentielle peut être exprimée par une valeur déterminée α du coefficient de corrélation conditionné (connaissant le réservoir panmictique local), α sera appelé la « consanguinité systématique ».

Le présent travail est consacré à établir la relation entre le coefficient F (corrélation *a priori* entre les gamètes qui s'unissent ou « coefficient de consanguinité *a priori* » ou « consanguinité totale ») et les coefficients φ et α (« consanguinité panmictique » et « consanguinité systématique »).

Utilisée pour la définition du coefficient de parenté $\varphi_{xy}(n)$, la notion d'identité — ou de non identité — de 2 gamètes, ou plutôt de 2 loci homologues portés par deux gamètes, demande à être précisée soigneusement. Nous dirons d'abord que deux loci sont identiques s'ils dérivent, par descendance mendélienne et sans aucune mutation, d'un même locus d'un ancêtre commun : dans ce cas ils portent forcément le même gène : soit tous deux a , avec probabilité *a priori* C , soit tous deux A , avec probabilité $1 - C$; le cas contraire (de « non-identité ») est celui où les deux loci, ou bien ont subi au moins une mutation depuis un ancêtre commun, ou bien n'ont aucun ancêtre commun aussi loin que l'on remonte en arrière, c'est-à-dire jusqu'à la génération F_0 incluse; dans ce dernier cas, les deux loci proviennent d'individus différents de F_0 , donc (s'il n'y a pas de sélection) de deux individus pris au hasard indépendamment et que nous supposons sans parenté : les deux loci sont alors indépendants.

Par contre, dans le cas où il y a eu au moins une mutation, l'indépendance ne peut être admise; il y a au contraire entre les deux loci une dépendance qui ne donne le même résultat que l'identité que lorsque, dans le cas de diallélisme, il s'est effectué un nombre pair de mutations (GILLOIS, 1966). Une voie de calcul plus simple permet de se ramener *quand même* à l'indépendance en introduisant la notion de « *bétamutation* » : il s'agit là d'un mécanisme mathématique équivalent à la mutation de a vers A ou de A vers a avec une probabilité u ou v . Supposons que — toujours dans la transmission d'un locus d'une génération à la suivante — il y ait une probabilité k de transformation du gène présent dans le locus en un gène qui soit, non plus forcément différent, mais *tiré au hasard* avec probabilités C et $1 - C$ d'être a ou A .

Dès lors la probabilité, pour chaque gène A de se transformer par ce mécanisme en un gène a sera kC , que nous égalons à v ; la probabilité, pour chaque gène a , de se transformer en un gène A , sera $k(1 - C)$, que nous égalons à u : cela est possible à condition que $k = u + v$ et $\frac{v}{k} = \frac{v}{u + v} = C$, ce qui est réalisé lorsque n est assez grand pour que la probabilité *a priori* C ait atteint sa valeur d'équilibre $\frac{v}{u + v}$.

(1) Dans la notation de WRIGHT, $\varphi_{xx}(n) = F_{ST}$, S symbolisant ici le « choix au hasard » dans le même site.

(2) Si n est assez grand $\varphi_{xx}(n)$ et $\varphi_{xy}(n)$ sont indépendants de n et leurs produits par $C(1 - C)$ sont la variance et la covariance des fréquences géniques dans l'état « stationnaire » d'équilibre statistique.

On peut donc alors regarder $k = u + v$, comme la probabilité à chaque changement de génération d'une « bétamutation », mutation fictive transformant le gène en un gène *indépendant* tiré dans un réservoir de composition constante (définie par C).

Nous redéfinirons donc le coefficient de parenté $\varphi_{xy}(n)$ de deux loci en x et y , comme la probabilité *a priori* pour qu'ils dérivent, *sans avoir subi de bétamutation*, d'un même locus d'un ancêtre commun, c'est-à-dire soient identiques : $\varphi_{xy}(n)$ est une probabilité d'identité.

La probabilité complémentaire $1 - \varphi_{xy}(n)$ est alors la probabilité qu'ils dérivent d'ancêtres indépendants (dans F_0) ou bien qu'ils aient subi des bétamutations les rendant indépendants; C'est une probabilité d'indépendance si tous les tirages indépendants ont lieu dans le même réservoir et si l'on admet que n est assez grand pour que la probabilité *a priori* (pour un tirage) ait atteint sa limite C.

En particulier $\varphi_{xx}(n)$ est la probabilité *a priori* d'identité de deux gamètes pris au hasard dans le même emplacement x dans F_n ; donc, dans l'hypothèse de panmixie locale ⁽¹⁾, la probabilité d'identité des deux gamètes utiles se réunissant pour fournir un individu, probabilité que nous avons appelée le coefficient de consanguinité (MALÉCOT, 1948). Dans le cas d'identité, ces deux gamètes ont donc la probabilité *a priori* C d'être tous deux porteurs de a (génotype aa) et la probabilité $(1 - C)$ d'être tous deux porteurs de A (génotype AA). Mais dans le cas d'indépendance (cas qui se présente avec la probabilité $1 - \varphi_{xx}(n)$), les probabilités *a priori* des trois génotypes sont données par la loi de HARDY-WEINBERG :

$$\frac{aa}{C^2} \qquad \frac{Aa}{2C(1-C)} \qquad \frac{AA}{(1-C)^2}$$

Les probabilités *a priori* des trois génotypes que peut présenter un individu de F_n (dans l'emplacement x) sont donc les suivantes (en allégeant les notations : φ pour φ_{xx}) :

$$\begin{aligned} aa &: \varphi C + (1-\varphi)C^2 = C^2 + \varphi C(1-C) \\ Aa &: 2(1-\varphi) C(1-C) \\ AA &: \varphi(1-C) + (1-\varphi)(1-C)^2 = (1-C)^2 + \varphi C(1-C). \end{aligned}$$

Ce qui est bien conforme au tableau 1 d'association des gamètes.

TABLEAU I

Loi de probabilité conjointe des deux variables indicatrices gamétiques X et X' ⁽²⁾

| | | (a) I | (A) O | |
|----|-------|---------------------------------|---|---------|
| X' | (a) I | $\varphi C + (1 - \varphi) C^2$ | $(1 - \varphi) C (1 - C)$ | (C) |
| | (A) O | $(1 - \varphi) C (1 - C)$ | $\varphi(1 - C) + (1 - \varphi)(1 - C)^2$ | (1 - C) |
| | | (C) | (1 - C) | |

Ce sont là les espérances mathématiques *a priori* des fréquences P, 2Q, R des trois génotypes dans chaque groupe, alors que C et (1-C) sont, rappelons-le, les espérances *a priori* des fréquences q et p de deux gènes a et A dans chaque groupe : si l'on peut estimer toutes ces espérances par les fréquences moyennes sur un grand nombre de groupes, leur comparaison donne une estimation du coefficient de consanguinité $\varphi_{xx}(n)$ dans le cas de panmixie locale.

⁽¹⁾ C'est-à-dire panmixie à l'intérieur de chaque site géographique.

⁽²⁾ La variable indicatrice gamétique prend la valeur I ou O suivant que le gamète porte le gène a ou A .

Ces formules ont été généralisées par YASUDA (1966), en cherchant les probabilités *a priori* des six couples de génotypes (1) qui peuvent se présenter chez deux conjoints (tabl. 2).

TABLEAU 2

Les six couples possibles de génotypes qui peuvent se présenter chez deux conjoints

| | AA | Aa | aa |
|----|----|----|----|
| AA | X | X | X |
| Aa | | X | X |
| aa | | | X |

En les regardant comme les probabilités relatives à un « quadruplet » de gamètes de même site géographique, on obtient des polynômes du 4^e degré en C et $1-C$ dont les coefficients sont les probabilités *a priori* des diverses relations d'identité ou d'indépendance entre quatre loci de même site géographique. Ces « coefficients de parenté quaternaires » ont été calculés par GILLOIS (1964) et HARRIS (1964) dans le cas où il n'y a pas de migration (1).

Ce n'est que par approximation, si φ est petit, que les probabilités *a priori* des six couples de conjoints (et leurs fréquences moyennes) peuvent être exprimées en fonction de φ , les autres coefficients de parenté étant de l'ordre de φ^2 et φ^3 .

Revenant au coefficient de parenté binaire $\varphi_{xx}(n)$, il exprime la « corrélation gamétique *a priori* » entre deux gamètes de même emplacement x , donc en particulier, dans le cas de panmixie locale, entre les gamètes qui s'unissent. La démonstration est immédiate : les variables indicatrices gamétiques X et X' attachées à deux gamètes pris au hasard dans le même emplacement x ont chacune, rappelons-le, pour loi de probabilité *a priori* $1/(C)$; $0/(1-C)$ (cf. tabl. 1) et pour écart-type $\sqrt{C(1-C)}$; leur covariance, calculée à partir du tableau 1, est $\varphi_{xx}C(1-C)$, leur coefficient de corrélation est donc φ_{xx} , c. q. f. d. (théorème de WRIGHT-MALÉCOT).

Il est important de remarquer que c'est une corrélation *a priori* et non une corrélation conditionnée. En effet, lorsqu'on tire au hasard et indépendamment deux gamètes dans ce réservoir gamétique local (panmixie locale) leur corrélation conditionnée est nulle (la loi de HARDY-WEINBERG est localement vérifiée). Il n'y a pas là de contradiction avec le fait que leur corrélation *a priori* $\varphi_{xx}(n)$ est positive, car l'interprétation expérimentale de cette dernière est différente. C'est la corrélation entre les écarts de deux variables gamétiques X et X' , tirées au hasard dans le même site géographique, par rapport à leur espérance *a priori* qui peut souvent être estimée par la moyenne générale sur l'ensemble des sites géographiques (alors que la corrélation conditionnée était calculée par rapport à la moyenne locale, définie par la composition du réservoir gamétique local). $\varphi_{xx}(n)$ caractérise ce que l'on peut appeler « consanguinité panmictique » ou, avec LAMOTTE (1951), « consanguinité de position ».

Ce phénomène présente un aspect plus général, dont la formulation rigoureuse n'est pas encore explicitée :

1° Si l'on évalue le coefficient de corrélation par rapport à la moyenne sur une vaste région, entre deux gamètes tirés dans une même sous-région, ce coefficient est d'autant plus grand que la sous-région est restreinte par rapport à la région. A la limite, il est nul quand la sous-région coïncide avec la région et il est maximum au contraire quand la sous-région est assez petite pour correspondre à un réservoir gamétique unique (« groupe panmictique » ou « isolat »).

(1) Dans le cas de panmixie locale.

2° En fait (voir les données brésiliennes non publiées de MORTON (1)) on ne peut pas observer de sous-région assez petite pour correspondre à un groupe panmictique unique, parce qu'il y a toujours, semble-t-il, une tendance préférentielle en faveur du choix de cousins (proches ou éloignés) : la valeur maximum de φ_{xy} , soit φ_{xx} (consanguinité due au hasard ou « coefficient de consanguinité panmictique ») est de l'ordre de 5 p. 1000 alors que le coefficient de parenté mesuré entre conjoints (coefficient de consanguinité total) est de l'ordre de 10 p. 1000; la différence correspond à un choix préférentiel de gamètes (homogamie) à l'intérieur de chaque groupe local, c'est-à-dire à une corrélation conditionnée (par la connaissance de ce groupe) qui reste positive (environ 5 p. 1000) et se traduit localement par un écart à la loi de HARDY-WEINBERG (déficit d'hétérozygotes).

3° C'est WRIGHT (1965) qui a donné la meilleure formulation de l'incidence de la non-panmixie par homogamie locale; la voici simplifiée : si l'on appelle α le coefficient de corrélation conditionné (c'est-à-dire évalué par rapport à la moyenne connue d'un groupe local) des gamètes qui s'unissent deux à deux dans ce groupe, et si les fréquences qui y sont réalisées sont p et q , la probabilité conditionnée pour que le couple de gamètes qui s'unissent donne un hétérozygote est, non pas $2pq$, mais $2pq(1-\alpha)$: c'est aussi, sensiblement, la fréquence des hétérozygotes dans ce groupe; donc leur fréquence moyenne sur l'ensemble de tous les groupes d'une vaste région sera sensiblement $E[2pq(1-\alpha)]$. Or, dans l'état stationnaire, on a :

$$E(pq) = E(q-q^2) = C - C^2 - \sigma_x^2 = C(1-C)(1-\varphi)$$

La fréquence moyenne des hétérozygotes sur l'ensemble de la vaste région sera donc : $2C(1-C)(1-\varphi)(1-\alpha)$ alors que la fréquence moyenne du gène a est C . La comparaison montre que le coefficient de corrélation entre les gamètes qui s'unissent, calculé par rapport à la moyenne générale (coefficient de corrélation *a priori* entre les gamètes qui s'unissent) est F tel que :

$$1 - F = (1-\varphi)(1-\alpha), \text{ soit } F = \varphi + \alpha - \varphi\alpha$$

C'est le théorème de WRIGHT énoncé par cet auteur (1965) comme théorème de « multiplication des index panmictiques » $(1-\alpha)$ et $(1-\varphi)$ dans une « classification hiérarchique » où la « population totale » est décomposée en « subdivisions » (ou « sites »).

La présente démonstration relie directement les corrélations α et F (qui peuvent varier de -1 à $+1$) à la probabilité φ (qui est comprise entre 0 et 1) et qui est aussi interprétable comme une corrélation (positive) entre deux gamètes pris indépendamment dans le même « site » (dans la même « subdivision »).

Signalons quelques cas particuliers : si $\alpha = 1$, $F = 1$; si $\alpha = 0$, $F = \varphi$; si $\alpha = -1$, $F = 2\varphi - 1$.

Il y a une autre interprétation : le complément à 1 d'un coefficient de consanguinité a été appelé par WRIGHT « index panmictique » : la formule de WRIGHT correspond au cas où les index panmictiques se multiplient : $1 - \alpha$ est la probabilité d'indépendance conditionnée par la connaissance de la fréquence locale; $1 - \varphi$ est la probabilité d'indépendance de deux gamètes pris au hasard dans un même groupe local par rapport à la moyenne générale. Le produit $1 - F$ est la probabilité d'indépendance des deux gamètes qui s'unissent par rapport à la moyenne générale.

Reçu pour publication en juillet 1969.

SUMMARY

RANDOM INBREEDING AND NONRANDOM INBREEDING

In diploid populations which undergo a certain geographic isolation manifested by the existence of subdivisions or sites, it is possible to define three coefficients of inbreeding: F and α of Sewall Wright and of Malecot. F is called the coefficient of total inbreeding, α the coefficient of nonrandom inbreeding, and the coefficient of random inbreeding. By using a relation between conditional correlation coefficients (on sites where the composition of the gene pool is known) and *a priori* correlations (for a large number of sites whose individual gametic reservoirs are unknown), it can be shown that $F = \alpha + -\alpha$. That relation had already been put forward by Sewall Wright as a theorem of « multiplication of panmictic indices » in a hierarchical classification where the total population is partitioned into subdivisions.

(1) Prof. E. N. MORTON, Institute of Genetics, University of Hawaii, Honolulu, Hawaii (U.S.A.).

RÉFÉRENCES BIBLIOGRAPHIQUES

- GILLOIS M., 1966. La notion de dépendance en génétique. *Ann. Inst. Henri Poincaré*, Sect. B, 2, 261-278.
- HARRIS D. L., 1964. Genotypic covariances between inbred relatives. *Genetics*, 50, 1319-1348.
- LAMOTTE M., 1951. Recherches sur la structure génétique des populations naturelles de *Cepaea Nemoralis* L. *Bull. biol. Fr. Belg.*, Suppl. n° 35.
- MALÉCOT G., 1948. *Les mathématiques de l'hérédité*. Masson, Paris.
- MALÉCOT G., 1967. Conséquences statistiques de la parenté. 36th Meet. int. Inst. Statistics (sous presse).
- YASUDA N., 1966. *The genetical structure of North eastern Brazil*. Thesis, University of Hawaii, Honolulu.
- WRIGHT S., 1965. The interpretation of population structure by F-statistics with special regard to systems of matings. *Evolution, Lawrence, Kans*, 9, 395-420.
-