

## MUTATIONS, MIGRATIONS ET STRUCTURES GÉNIQUES

*En progressant constamment, la génétique de population est amenée à abandonner quelques hypothèses adoptées pour simplifier des questions qui paraissaient à l'abord fort délicates à démêler en raison du grand nombre des éventualités. C'est ainsi que les généticiens ont généralement admis que deux allèles distincts seulement sont présents en un même locus. En fait il n'en est pas toujours ainsi; comme le remarque M. A. JACQUARD un locus peut comporter plus d'une dizaine d'allèles <sup>(1)</sup>. Cette multiplicité conduit à se placer dans le cas le plus général d'un nombre d'allèles  $n$ . M. Daniel COURGEAU, chargé de recherches à l'I.N.E.D., étudie dans cette hypothèse une population soumise à des mutations et à des migrations.*

On envisagera dans cet article la généralisation des modèles déterministes, développés par G. Malécot <sup>(2)</sup>, dans le cas où la population subit des mutations et des migrations, toutes les autres conditions de la panmixie parfaite étant supposées réalisées.

1. Mutations. Envisageons une population dont la structure génique est :

$$s = (\omega_1, \dots, \omega_i, \dots, \omega_n) \text{ avec } \sum_i \omega_i = 1$$

Chaque gène  $i$  aura, à chaque génération, une probabilité de muter en l'un des autres gènes  $j$ . Nous noterons cette probabilité, ou taux de mutation, sous la forme  $v_j^i$ . De même, la probabilité qu'aura un gène  $i$  de rester identique à lui-même sera :

$$v_i^i = 1 - \sum_{j \neq i} v_j^i$$

La fréquence de l'allèle  $i$  deviendra alors après mutation :

---

(1) « Panmixie et consanguinité », *Population*, 1968, novembre-décembre 1968.

(2) « Probabilités et hérédité » I.N.E.D., 1966, 163-189.

$$\begin{aligned}
 \omega_i(g) &= \omega_i(g-1) - \sum_{j \neq i} v_j^i \omega_i(g-1) + \sum_{j \neq i} v_i^j \omega_j(g-1) \\
 &= [1 - \sum_{j \neq i} v_j^i] \omega_i(g-1) + \sum_{j \neq i} v_i^j \omega_j(g-1) \\
 &= \sum_j v_i^j \omega_j(g-1)
 \end{aligned}$$

Appelant matrice de mutation  $V$ , la matrice stochastique de terme général  $v_j^i$  et l'ordre  $n$ , cette relation s'écrit :

$$s(g) = s(g-1)V = s(o)V^g$$

Ainsi, la structure génique de la population à la génération  $g$  est entièrement déterminée par sa structure génique initiale et la matrice de mutation.

L'étude de la limite vers laquelle  $s(g)$  pourrait tendre, lorsque  $g \rightarrow \infty$ , est liée à l'étude de la limite de la matrice  $V^g$ . Or, cette matrice peut se mettre sous la forme suivante :

$$V^g = US^gU^{-1}$$

où  $U$  est une matrice non singulière et  $S$  est une matrice partitionnée. Elle est de la forme :

$$S = \begin{pmatrix} S_{n_1}(\rho_1) & 0 & \dots & \dots & 0 \\ 0 & S_{n_2}(\rho_2) & \dots & \dots & 0 \\ \vdots & \vdots & & & \vdots \\ 0 & 0 & & & S_{n_p}(\rho_p) \end{pmatrix}$$

où  $S_{n_r}(\rho_r)$  est une matrice d'ordre  $n_r$ ,  $\rho_r$  étant une valeur propre de l'ordre de multiplicité  $n_r$  :

$$S_{n_r}(\rho_r) = \begin{pmatrix} \rho_r & 1 & 0 & \dots & 0 & 0 \\ 0 & \rho_r & 1 & \dots & 0 & 0 \\ \vdots & \vdots & \vdots & & \vdots & \vdots \\ 0 & 0 & 0 & \dots & \rho_r & 1 \\ 0 & 0 & 0 & \dots & 0 & \rho_r \end{pmatrix}$$

qui donne, élevée à la puissance  $g$  :

$$S_{n_r}^g(\rho_r) = \begin{pmatrix} \rho_r^g & \binom{g}{1} \rho_r^{g-1} & \dots & \dots & \binom{g}{n_r-1} \rho_r & \rho_r^{g-n_r+1} \\ 0 & \rho_r^g & \dots & \dots & \binom{g}{n_r-2} \rho_r & \rho_r^{g-n_r+2} \\ \vdots & \vdots & & & \vdots & \vdots \\ 0 & 0 & & & \rho_r^g & \rho_r^g \end{pmatrix}$$

La condition nécessaire et suffisante, pour que ces matrices aient une limite lorsque  $g \rightarrow \infty$ , est donc :

— que les valeurs propres  $\rho_r$  soient de module  $\leq 1$ ;

— que pour les valeurs propres  $\rho_r$  de module 1, les matrices  $S_{nr}^g$  soient d'ordre 1;

— qu'il n'y ait pas de valeurs propres de module 1 autres que 1.

Si cette dernière condition n'est pas réalisée les matrices  $S^g$  seront cycliques.

Or, la matrice  $V$  étant stochastique, la première condition est réalisée. De plus, la relation :  $\sum_j v_j^i = 1$  montre que l'on a au moins une valeur propre égale à 1, à laquelle correspondra un vecteur propre gauche  $u_1$ , dont toutes les composantes seront égales. Enfin, un allèle ayant toujours une probabilité non nulle de rester identique à lui-même, aucun des  $v_j^i$  ne sera nul : cette condition a pour conséquence que les autres valeurs propres sont de module inférieur à 1. Ainsi la matrice  $S^g$  aura, à la limite, tous ses éléments nuls, à l'exception de la valeur propre 1. Le produit :  $\lim_{g \rightarrow \infty} US^g$  tendra donc vers, une matrice, dont seule la colonne correspondant au vecteur propre  $u_1$ , ne sera pas nulle. Enfin, le produit :  $\lim_{g \rightarrow \infty} US^g U^{-1}$  tendra vers une matrice, dont toutes les lignes seront identiques. En effectuant le produit  $s(o)V^g$ , on obtiendra un vecteur ligne indépendant de  $s(o)$  et dépendant uniquement d'une quelconque des lignes identiques de la matrice :  $\lim_{g \rightarrow \infty} V^g$ .

*Ainsi la limite de  $S(g)$  n'est fonction que des taux de mutation et est indépendante de la structure génique initiale de la population.*

Le calcul de cette limite ne nécessite que la détermination du vecteur propre droit, correspondant à la valeur propre 1. Ce vecteur ainsi déterminé, à un facteur multiplicatif près, sera totalement défini par la condition :  $\sum_i \omega_i = 1$ .

Dans le cas particulier où les taux de mutation directs et reverses sont égaux, on voit tout de suite que la solution sera :

$$\lim_{g \rightarrow \infty} s(g) = \left( \frac{1}{n}, \dots, \frac{1}{n}, \dots, \frac{1}{n} \right)$$

L'étude de la vitesse avec laquelle  $s(g)$  tend vers sa limite est fonction de la plus grande valeur propre  $\rho_M$ , inférieure à 1, de la matrice  $V$ , ainsi que de son degré de multiplicité. On sait, que les  $v_j^i$  sont en général très faibles pour  $i \neq j$  (de l'ordre de  $10^{-5}$  environ). Pour cerner l'ordre de grandeur de  $\rho_M$ , on peut étudier la somme des valeurs propres qui est égale à :  $\sum_i v_i^i$ . Une des valeurs propres étant égale à 1, il vient l'inégalité suivante :

$$\rho_M > \frac{\sum_i v_i^i - 1}{n - 1} \approx 1 - n 10^{-5}$$

Si, de plus,  $\rho_M^g$  est une racine simple de l'équation caractéristique,  $\rho_M$  sera sensiblement nul, lorsque  $g$  dépassera l'ordre de grandeur de  $\frac{10^5}{n}$  génération. Le temps nécessaire pour atteindre l'équilibre sera donc, en général, considérable.

2. Migrations. On supposera toujours la population infiniment nombreuse, décomposée en  $m$  groupes, chacun de population infinie. Nous supposons qu'à chaque génération,  $g$ , des migrations ont lieu entre ces groupes, les autres conditions de la panmixie parfaite étant réalisées : Notons par  $l_r^k(g)$  la proportion de migrants originaires du groupe  $r$  et présents dans le groupe  $k$  à la génération  $g$ , parmi les individus se reproduisant dans le groupe  $k$ . La proportion d'autochtones du groupe  $k$  est égale à  $l_k^k(g)$ . Il est commode d'utiliser la notation matricielle :

$$L(g) = \left( l_r^k(g) \right) \text{ avec } \sum_r l_r^k(g) = 1$$

où  $L(g)$  est une matrice carrée stochastique d'ordre  $m$ , l'indice supérieur étant l'indice ligne, l'indice inférieur l'indice colonne. Par la suite, on supposera, pour simplifier, la migration indépendante de la génération : on la notera  $L$ .

Appelons  $s^k(g)$  la structure génique du  $k^{\text{ième}}$  groupe à la génération  $g$ , et condensons les structures géniques des  $m$  groupes sous forme matricielle :

$$\Omega(g) = \left( \omega_i^k(g) \right)$$

$\Omega(g)$  est alors une matrice rectangulaire de type  $(m, n)$ , telle que :

$$\sum_i \omega_i^k(g) = 1$$

En supposant que les fréquences des allèles sont les mêmes dans chaque lot de migrants que dans le groupe, dont il est issu, il vient :

$$\omega_i^k(g) = \sum_r l_r^k \omega_i^r(g-1)$$

cette relation s'écrit sous forme matricielle :

$$\boxed{\Omega(g) = L\Omega(g-1) = L^g\Omega(o)}$$

Ainsi, les structures géniques des populations considérées, à la génération  $g$ , sont entièrement déterminées par leurs structures géniques initiales et la matrice de migration. De plus, cette relation montre, que, contrairement au cas de la mutation, l'étude faite par G. Malécot est encore valable dans le cas où le nombre d'allèles est quelconque : en effet la connaissance du  $i^{\text{ème}}$  allèle est indépendante de celles des autres. Écrivant :

$$s'_i = \begin{pmatrix} \omega_i^1 \\ \omega_i^2 \\ \vdots \\ \omega_i^n \end{pmatrix}$$

la relation précédente donne :

$$s'_i(g) = L s'_i(g-1) = L^g s'_i(o)$$

L'étude de la limite de  $\Omega(g)$  quand  $g \rightarrow \infty$ , est liée à celle de la matrice  $L^g$ . Le problème est le même que le précédent, la matrice  $L$  étant stochastique. Dans le cas général où un groupe garde toujours un certain pourcentage d'autochtones, la matrice  $L^g$  tendra vers une matrice dont toutes les lignes sont identiques. Le produit  $L^g \Omega(o)$  donnera encore une matrice du même type mais fonction de  $\Omega(o)$ .

*Ainsi la limite de  $\Omega(g)$ , lorsqu'elle existe, est fonction de la matrice de migration et de la structure génique initiale des diverses populations. Mais cette limite sera la même pour toutes les sous-populations. Son calcul pourra être effectué comme celui de  $V^g$ , qu'il sera nécessaire ensuite de multiplier par  $\Omega(o)$ .*

L'étude de la vitesse avec laquelle  $\Omega(g)$  tend vers sa limite est encore fonction de la plus grande valeur propre, inférieure à 1, de la matrice  $L$ , ainsi que de son degré de multiplicité. Mais, dans ce cas, les  $l_j^i$  pourront être beaucoup plus importants que les  $v_j^i$ , et leur ordre de grandeur très variable. Il sera donc en général nécessaire de calculer les valeurs propres de la matrice  $L$ , pour connaître la vitesse avec laquelle  $L^g$  tendra vers sa limite. G. Malécot a traité le cas particulier où il existe au moins une colonie qui envoie des migrants dans toutes les autres : dans ce cas, la limite sera sensiblement atteinte, lorsque  $g$  dépassera l'ordre de grandeur de  $\frac{1}{\eta}$ ,  $\eta$  étant le pourcentage minimal de migrants envoyé, de cette colonie, dans les autres.

3. Mutations et migrations. On se placera dans le cas du paragraphe 2, en supposant en plus que des mutations se produisent dans chaque groupe. Les probabilités des diverses mutations seront

supposées indépendantes des groupes où elles ont lieu. On aura donc encore, dans ce cas, une matrice de mutation  $V$ , identique à celle du paragraphe 1.

La fréquence  $\omega'_i{}^k(g)$  de l'allèle  $i$  dans le  $k^{\text{ième}}$  groupe avant mutation, mais après migration, est toujours donnée par :

$$\omega'_i{}^k(g) = \sum_r l_r^k \omega'_i{}^r(g-1)$$

après mutation, elle devient égale à :

$$\omega_i^k(g) = \sum_j v_j^i \omega'_i{}^k(g)$$

ces relations peuvent s'écrire, plus simplement, sous forme matricielle :

$$\Omega(g) = L\Omega(g-1)V = L^g\Omega(o)V^g$$

Un raisonnement identique aux précédents montre que la limite de  $\Omega(g)$ , quand elle existe, ne sera fonction que de la matrice de mutation  $V$  et sera indépendante de la migration et de la structure initiale de la population. Dans le cas où les taux de migration sont beaucoup plus importants que les taux de mutation, les populations tendront, dans un premier temps, vers un état d'équilibre dont la structure génique sera fonction des taux de migration et de sa structure génique initiale, toutes les sous-populations devenant identiques entre elles. Dans un deuxième temps, la population tendra vers une structure génique d'équilibre qui ne dépendra plus que des taux de mutation.

Notons cependant que, si l'hypothèse d'indépendance des taux de mutation par rapport au temps et à l'espace est raisonnable, l'hypothèse d'indépendance des taux de migration par rapport au temps est beaucoup moins acceptable. On pourrait alors écrire :

$$\Omega(g) = L(g)\Omega(g-1)V = L(g)L(g-1)\dots L(1)\Omega(o)V^g$$

Ce modèle permettrait d'étudier l'évolution d'une population pendant une période où les variations des taux de migration seraient connues.

Daniel COURGEAU.