

MIGRATION CONTINUE DANS LE TEMPS ET STRUCTURE GÉNIQUE

par Daniel Courgeau *

Les modèles de génétique de population essaient de décrire l'évolution au cours du temps de populations soumises à un certain nombre d'hypothèses. Dans la plupart de ces modèles on pose, en particulier, la condition qu'il n'y a pas de croisements entre individus appartenant à des "générations" différentes, les "générations" correspondant à des ensembles, successifs d'individus.

Or, pour les populations humaines, au contraire continues dans le temps, cette hypothèse n'est pas vérifiée, et il importe de voir si son abandon entraîne des conséquences importantes.

On sait déjà que, lorsque toutes les conditions de la "panmixie parfaite" sont réalisées, hormis celle de générations séparées, la structure de la population tend asymptotiquement vers la structure panmictique correspondant aux proportions de la loi de Hardy-Weinberg (1). D.F. Roberts et R.W. Hiorns (2) ont essayé de prolonger cette loi, dans le cas de deux populations échangeant entre elles des migrants. Mais ils ont dû pour développer leur modèle, faire l'hypothèse, non explicitée, qu'aussitôt nés, les individus entraînent dans la population féconde. Le modèle présenté ici supprime cette hypothèse, peu réaliste, et s'applique à un nombre quelconque de populations échangeant entre elles des individus par migration. Il importe, en premier lieu, avant d'en développer les conséquences, de bien préciser toutes les hypothèses, afin de pouvoir ultérieurement modifier les moins réalistes.

(*) Institut national d'études démographiques, Paris.

(1) JACQUARD A. — Structure génétique des populations, 1970, 142-145, *Masson & C^o*.

(2) The dynamics of racial intermixture. *The American Journal of Human genetics*, septembre 1962, 14, n° 3.

Hypothèses du modèle Considérons une population infiniment nombreuse, décomposée en m groupes, chacun d'effectif infini. A un instant initial ($t = 0$) ces populations vont échanger entre elles des migrants, après avoir connu une période d'isolement complet. Elles sont donc considérées comme ayant atteint leurs structures panmictiques, pendant cette période d'isolement. Nous supposerons ensuite que les migrations s'effectuent entre la naissance des individus et le début de leur phase reproductrice : pendant cette phase, les divers groupes sont alors supposés panmictiques. Cela revient à dire que les croisements peuvent à tout moment se produire au hasard, parmi l'ensemble des individus en âge de procréer, à l'intérieur de chaque groupe.

Il est nécessaire, pour caractériser ces populations, de définir un taux d'entrée dans la population féconde, n , identique pour tous les groupes, cette entrée ayant lieu à l'âge de la puberté, p . Le taux de sortie de la population féconde sera supposé égal à n , mais cette sortie sera considérée comme uniformément répartie sur toute la population féconde. La migration intervenant avant la phase reproductrice, au lieu de faire entrer n individus du groupe j dans la population féconde de j , n'en fera entrer que $n \times m_j^j$, les autres provenant des divers groupes ($n \times m_i^j$ du groupe i , par exemple). On voit, que ces coefficients peuvent se représenter sous forme matricielle :

$$M = (m_i^j) \quad \text{avec} \quad \sum_j m_i^j = 1$$

Cette matrice M est une matrice carrée stochastique d'ordre m , les taux de migration étant supposés indépendants du temps.

De même la structure génique des divers groupes féconds à l'instant t pourra se représenter sous forme matricielle

$$\Omega(t) = (\omega_i^k(t)) \quad \text{avec} \quad \sum_i \omega_i^k(t) = 1$$

condensation en une matrice rectangulaire de type $(m \ n)$, des m structures géniques vectorielles correspondant à un locus pouvant contenir n allèles.

Evolution de la structure génique Entre les instants t et $t + dt$, la structure génique des diverses populations se modifie :

— par l'entrée de $n dt$ individus, dont la structure génique est, en l'absence de sélection et de mutation, mais en présence de migration avant la phase reproductrice :

$$M \Omega(t - p)$$

— par la sortie de $n dt$ individus, dont la structure génique est celle de la population féconde à la date t , soit $\Omega(t)$.

Il en résulte la relation suivante :

$$\Omega(t + dt) = \Omega(t) + n M \Omega(t - p) dt - n \Omega(t) dt$$

soit encore :

$$(1) \quad \frac{d[\Omega(t)]}{dt} = n M \Omega(t - p) - n \Omega(t)$$

Pendant la période initiale ($0 \leq t \leq p$), $\Omega(t - p)$ reste constant et égal à $\Omega(0)$.

On voit alors facilement que :

$$\Omega(t) = M \Omega(0) + e^{-nt} (I - M) \Omega(0)$$

où I est la matrice unité.

Montrons que pendant la période, $p(r - 1) \leq t \leq pr$, on aura une solution du type :

$$\Omega_r(t) = \{[\alpha_r(t) I + \beta_r(t) M + \dots + \rho_r(t) M^{r-2} + \sigma_r(t) M^{r-1}] (I - M) e^{-nt} + M^r\} \Omega(0)$$

où les fonctions $\alpha_r(t), \dots, \sigma_r(t)$ sont des polynômes en t , à déterminer.

On vérifie facilement que l'équation (1) conduit aux relations suivantes :

$$\frac{d[\sigma_r(t)]}{dt} = en \rho_{r-1}(t - p), \dots, \frac{d[\beta_r(t)]}{dt} = en \alpha_{r-1}(t - p), \frac{d[\alpha_r(t)]}{dt} = 0$$

De plus ces fonctions, du fait que $\Omega_r[(r - 1)p] = \Omega_{r-1}[(r - 1)p]$, doivent vérifier les conditions :

$$\sigma_r[(r - 1)p] = e^{r-1}, \rho_r[(r - 1)p] = \rho_{r-1}[(r - 1)p], \dots$$

$$\alpha_r[(r - 1)p] = \alpha_{r-1}[(r - 1)p]$$

Or, on a vu plus haut que $\alpha_1(t) = 1$, on en déduit facilement les deux relations :

$$\alpha_r(t) = 1 \quad \text{et} \quad \frac{d[\beta_2(t)]}{dt} = en$$

La deuxième équation entraîne :

$$\beta_2(t) = e[n(t - p) + 1] = \beta_r(t)$$

On voit ainsi de proche en proche que :

$$\sigma_r(t) = e^{r-1} \left\{ -\frac{n^{r-1}}{(r-1)!} [t - (r-1)p]^{r-1} + \frac{n^{r-2}}{(r-2)!} [t - (r-1)p]^{r-2} + \dots + n[t - (r-1)p] + 1 \right\}$$

et que $\sigma_s(t) = \sigma_r(t)$ pour tout s supérieur à r .

On peut donc calculer la valeur de $\Omega_r(t)$, en particulier lorsque $t = rp$.

$$\begin{aligned} \Omega_r(rp) = & \left\{ \left[1 + e \left(1 + \frac{r-1}{1!} \right) M \right. \right. \\ & + e^2 \left(1 + \frac{r-2}{r!} + \frac{(r-2)^2}{2!} \right) M^2 + \dots \\ (2) \quad & + e^{r-2} \left(1 + \frac{2}{1!} + \dots + \frac{2^{r-2}}{(r-2)!} \right) M^{r-2} \\ & \left. + e^{r-1} \left(1 + \frac{1}{1!} + \dots + \frac{1}{(r-1)!} \right) M^{r-1} \right] [1 - M] e^{-r} + M^r \left. \right\} \Omega(o). \end{aligned}$$

Il reste maintenant à montrer que lorsque $t \rightarrow \infty$, $\Omega_r(t)$ tend vers une limite, que nous allons calculer. On sait que, dans le cas général, la matrice M aura une valeur propre égale à 1 unique, toutes les autres valeurs propres étant de module inférieur à un. Elle pourra s'écrire sous la forme

$$M = U S U^{-1}$$

où S est une matrice diagonale, dont les termes sont les valeurs propres de M et U une matrice inversible. Il suffit alors de montrer que pour les valeurs propres α de module inférieur à 1, la suite indiquée ci-dessous, va tendre vers 0 lorsque $r \rightarrow \infty$:

$$\begin{aligned} u_r = & \left(1 + \alpha e \left[1 + \frac{(r-1)}{1!} \right] + \alpha^2 e^2 \left[1 + \frac{(r-2)}{1!} + \frac{(r-2)^2}{2!} \right] + \dots + \right. \\ & \left. + \alpha^{r-1} e^{r-1} \left[1 + \frac{1}{1!} + \dots + \frac{1}{(r-1)!} \right] \right) (1 - \alpha) e^{-r} \end{aligned}$$

Considérons le terme général

$$\varphi_r(n) = \frac{\alpha^n}{e^{r-n}} \left[1 + \frac{(r-n)}{1!} + \frac{(r-n)^2}{2!} + \dots + \frac{(r-n)^n}{n!} \right]$$

Si n est compris entre r et $r/2$, on voit facilement que

$$|\varphi_r(n)| < |\alpha|^n \leq |\alpha|^{r/2}.$$

Or, on sait qu'il existe une valeur r_0 de r telle que pour $r > r_0$:

$$|\alpha|^{r/2} \leq \frac{1}{r^2}$$

Supposons maintenant n compris entre 0 et $r/2$ on peut alors écrire :

$$|\varphi_r(n)| \leq \left[1 + |\alpha| \frac{(r-n)}{1!} + \dots + |\alpha|^n \frac{(r-n)^n}{n!} \right] \frac{1}{e^{r-n}} < e^{(r-n)(|\alpha|-1)}$$

$$\leq e^{(|\alpha|-1)r/2}$$

On sait également, puisque $|\alpha|$ est inférieur à 1, qu'il existe une valeur r_1 de r , telle que pour $r > r_1$

$$\frac{(|\alpha|-1)r}{e^2} \leq \frac{1}{r^2}$$

En prenant la plus grande des valeurs r_0 et r_1 , r_m , on voit facilement que pour $r > r_m$: $|u_r| < \frac{|1-\alpha|}{r}$. Ainsi lorsque $r \rightarrow \infty$, $u_r \rightarrow 0$ et :

$$\Omega_r(rp) \xrightarrow[r \rightarrow \infty]{} M^r \Omega(o)$$

On voit donc que lorsque t tend vers l'infini, $\Omega(t)$ devient équivalent à $M^r \Omega(o)$ et tendra vers la même limite que dans le cas des générations séparées (3), seule la façon de tendre vers cette limite sera modifiée. En particulier il est nécessaire de prendre comme durée d'une génération, p années, pour que les deux types de modèles soient comparables. Il est intéressant de constater que cette durée est nettement inférieure à la durée de vie moyenne d'un individu, et n'est pas fonction du taux d'entrée des individus dans la période féconde. Le résultat est important car il permet de *passer de données sur une population continue, habituellement connues, à un traitement discontinu de cette population, la période à faire intervenir étant la durée séparant la naissance d'un individu et son arrivée à la phase féconde.*

L'écart entre le modèle continu et le modèle discontinu équivalent, peut cependant être important pour les faibles valeurs de t . En particulier si les taux de migration sont assez élevés, la matrice (I-M) sera très différente de 0 et entraînera une importance accrue des termes fonction de t . Mais, en général, ces taux de migration seront faibles et permettront de négliger rapidement ces termes. Seule l'étude de chaque cas particulier permettrait de préciser l'état entre les deux types de modèles.

Il est également intéressant de voir sous quelles conditions, le modèle proposé permet de retrouver le modèle de Roberts et Hiorns. En suppo-

(3) On sait que lorsque 1 est la seule valeur de module 1 de M , M^r tend, lorsque $r \rightarrow \infty$, vers une matrice dont toutes les lignes sont identiques, et son produit par $\Omega(o)$ donnera une matrice du même type. Les structures géniques des divers sous-groupes tendent vers une structure unique, qui sera, en général, fonction de la structure initiale des populations et de la matrice de migration.

sant le délai entre naissance et phase féconde nul et un taux d'entrée dans la population égal à 1, le modèle s'écrit :

$$\frac{d[\Omega(t)]}{dt} = (M - I) \Omega(t).$$

Il est bien identique à la généralisation, pour un nombre quelconque de sous-populations, du modèle de Roberts et Hiorns, que nous avons effectuée (4). On voit ainsi que deux des conditions du présent modèle n'y sont plus vérifiées.

Validité des hypothèses du modèle Diverses hypothèses ont dû être faites pour permettre de développer ce modèle. Il importe maintenant d'en voir la validité.

La première considère les populations comme d'effectif infini : on devra envisager, dans une étape ultérieure, des populations d'effectif fini. Les résultats obtenus ici correspondraient alors à l'espérance mathématique de la structure génique, qui resterait inchangée. Seule la variance entre les diverses sous-populations serait à étudier.

Une autre condition consistant à supposer toutes les populations isolées avant le début de la migration, est particulièrement adaptée au traitement des populations noires, blanches et Amerindiennes des Etats-Unis, entre lesquelles aucun contact n'existait avant la migration des deux premières populations vers l'Amérique. Par contre, étudiant des populations échangeant des migrants depuis des temps très reculés, le modèle sera d'autant moins applicable que les taux de migrations entre les populations ont varié au cours du temps.

L'hypothèse de la stabilité résidentielle de la population féconde est également discutable. Cependant, le fait que les migrations matrimoniales constituent une grande part des migrations humaines, va à son avantage, et on peut penser que l'introduction des migrations multiples, ne changerait pas outre mesure les résultats.

Par contre, le fait de considérer les populations panmictiques pendant leur phase féconde sera, sans doute, peu vérifié par les populations humaines. Le choix du conjoint aura un rôle important dans cette phase, et la fécondité sera fonction de l'âge des individus. Mais on peut dire que ce cas extrême permet, avec le cas des générations discontinues, d'encadrer le cas des populations réelles. D'où l'intérêt de développer un tel modèle.

(4) in A. JACQUARD « Structures génétiques des populations », pp. 330-332. La matrice M, que nous y introduisons, correspond en effet à la matrice (M-I) de cet article.

Enfin, le fait que les sorties de la population féconde sont supposées uniformément réparties sur toute cette population, est également discutable. Il serait possible de remplacer cette condition par une sortie de la population féconde à un instant unique ou par une combinaison des deux.

Conclusion Pour essayer de cerner la réalité de plus près, il est nécessaire d'élaborer des modèles de plus en plus complexes. Celui qui est développé ici envisage l'effet d'une migration continue dans le temps entre diverses populations, de façon déterministe. L'étude de la limite vers laquelle tend la structure génique de ces populations, montre qu'elle est identique à celle du cas discontinu. L'intérêt du modèle se trouve en fait, dans la détermination, à partir du cas continu, le seul sur lequel on puisse mesurer les caractéristiques des populations, des paramètres à faire intervenir pour traiter le problème de façon discontinue. Une discussion des diverses hypothèses faites pour élaborer le modèle, permet d'envisager les étapes ultérieures de l'étude de l'effet des migrations sur la structure génétique des populations.

SUMMARY

The effect of a continuous-time migration between various populations is examined by means of a deterministic model. The study of the limit towards which the genic structure of these populations tends, shows it is identical to that of the discontinuous case. As the continuous case is the only one in which the characteristics of populations can be measured, it is advantageous to resort to this model which makes it possible to determine parameters that can be used to deal with the problem in a discontinuous way. A discussion of the various assumptions made to elaborate the model allows us to envision the further phases of this study of the effect of migrations on the genetic structure of populations.