

GÉNÉTIQUE DES POPULATIONS DE *SPHAEROMA SERRATUM* (F.)

II. Calcul des fréquences géniques

par

Georges Teissier

Faculté des Sciences de Paris

Résumé

Des formules dérivées de la relation de Hardy-Weinberg permettent de calculer les fréquences géniques dans les quatre couples d'allèles qui conditionnent les phénotypes structuraux de *Sphaeroma serratum*. Les estimations qu'elles fournissent des fréquences des gènes *S signatum*, *O ornatum*, *L lunulatum* et des divers gènes de couleur, ne peuvent pas être altérées sensiblement pas des écarts assez grands à la pannmixie, ou par une assez forte inégalité dans les valeurs sélectives des génotypes en présence. L'estimation de la fréquence du gène *D discretum* est beaucoup plus incertaine.

INTRODUCTION

L'étude comparative des fréquences des phénotypes en présence dans deux échantillons de *Sphaeroma serratum* permet de reconnaître si les populations dans lesquelles ils ont été prélevés diffèrent de façon significative ou si, au contraire, elles doivent être considérées comme appartenant à un ensemble plus vaste et génétiquement homogène (Bocquet et Teissier, 1960). Mais cette analyse phénotypique ne constitue que la première étape de la détermination de la structure génétique des populations étudiées, le but à atteindre restant toujours, en principe, la connaissance des fréquences géniques et de la distribution de tous les phénotypes. Il s'agit là d'un problème difficile, dont la solution rigoureuse ne pourrait être donnée que par l'analyse génétique complète d'un échantillon représentatif de la population étudiée, opération beaucoup trop laborieuse pour pouvoir être pratiquée de façon régulière. Mais il reste heureusement possible, en l'absence d'une telle analyse, d'estimer la fréquence des allèles en présence, au prix d'un certain nombre d'hypothèses. Celles qu'il est le plus naturel de faire pour les populations de Sphéromes sont classiques, et les formules que nous avons proposées en 1951 dérivent

de la façon la plus simple de celles qui sont à la base de la génétique des populations.

Il nous a semblé qu'il ne serait pas inutile de reprendre, en les explicitant, les raisonnements que nous avons suivis, et de préciser par là la signification des hypothèses que nous avons dû admettre. Nous essaierons ensuite d'estimer l'importance des conséquences d'un écart éventuel entre les situations génétiques idéales qui sont à la base de nos constructions théoriques, et les structures génétiques réelles des populations étudiées.

On verra que le résultat de cette confrontation est satisfaisant et que l'on peut sans grand risque, continuer à faire usage des formules que nous avons proposées.

I. GÉNOTYPES ET PHÉNOTYPES

Dans tout ce qui va suivre, il ne sera question, pour la simplicité de l'exposé, que des cinq types structuraux majeurs qui sont, rappelons-le, conditionnés par quatre couples d'allèles indépendants, constituant une série épistatique *Dd*, *Ll*, *Oo* et *Ss*. Ces cinq types structuraux correspondent théoriquement à 81 génotypes distincts dont la liste est donnée dans le tableau ci-dessous :

<i>albicans</i> (1 génotype) :	<i>dd</i>	<i>ll</i>	<i>oo</i>	<i>ss</i>
<i>discretum</i> (2 génotypes) :	<i>Dd</i> <i>DD</i>	<i>ll</i>	<i>oo</i>	<i>ss</i>
<i>lunulatum</i> (6 génotypes) :	<i>dd</i> <i>Dd</i> <i>DD</i>	<i>Ll</i> <i>LL</i>	<i>oo</i>	<i>ss</i>
<i>ornatum</i> (18 génotypes) :	<i>dd</i> <i>Dd</i> <i>DD</i>	<i>ll</i> <i>Ll</i> <i>LL</i>	<i>oo</i> <i>OO</i>	<i>ss</i>
<i>signatum</i> (54 génotypes) :	<i>dd</i> <i>Dd</i> <i>DD</i>	<i>ll</i> <i>Ll</i> <i>LL</i>	<i>oo</i> <i>oo</i> <i>OO</i>	<i>ss</i> <i>SS</i>

Un des allèles étant, dans chaque couple, dominant sur l'autre, le nombre maximum de phénotypes reconnaissables est de 16. Si l'on représente, pour simplifier l'écriture, les dominants et les récessifs par des majuscules et des minuscules en romain, la liste de ces 16 phénotypes peut s'écrire comme suit :

<i>albicans</i> (1 phénotype) :	<i>d</i>	<i>l</i>	<i>o</i>	<i>s</i>
<i>discretum</i> (9 phénotypes) :	<i>D</i>	<i>l</i>	<i>o</i>	<i>s</i>
<i>lunulatum</i> (2 phénotypes) :	<i>d</i> <i>D</i>	<i>L</i>	<i>o</i>	<i>s</i>
<i>ornatum</i> (4 phénotypes) :	<i>d</i> <i>D</i>	<i>L</i>	<i>O</i>	<i>s</i>
<i>signatum</i> (8 phénotypes) :	<i>d</i> <i>D</i>	<i>L</i>	<i>o</i> <i>O</i>	<i>s</i>

Dans la pratique, et s'il n'intervient pas d'autres gènes que ceux conditionnant les cinq types structuraux moyens, le classement le plus minutieux d'une population ne peut comporter plus de 9 catégories. Il est toujours impossible, en effet, de reconnaître si les *signatum* et les *ornatum* sont *discretum* ou *albicans* et de classer sans erreur les *lunulatum* en *lunulatum-albicans* et *lunulatum-discretum*. Les phénotypes utilisables pour l'étude de la structure génétique de la population sont donc en définitive : *albicans* A, *discretum* D, *lunulatum* L, *ornatum* O, avec deux phénotypes O1 et O L suivant qu'ils ne sont pas ou sont, en même temps *lunulatum*; *signatum* S, avec quatre phénotypes, *signatum* proprement dit S o l, *signatum-lunulatum* S o L, *signatum-ornatum* S O l et *signatum-ornatum-lunulatum* S O L. Encore faut-il remarquer que les dominants doubles ou triples sont rarement faciles à identifier et que l'on doit le plus souvent se contenter d'un classement en cinq catégories correspondant aux cinq phénotypes définis par les relations d'épistasie.

Il n'est pas inutile de noter, à ce propos, que le nombre de génotypes correspondant à un même phénotype augmente des termes inférieurs aux termes supérieurs de la série épistatique. Pour un génotype *albicans*, il existe 2 génotypes *discretum*, 6 génotypes *lunulatum*, 18 génotypes *ornatum*, dont 12 *ornatum-lunulatum*, 54 génotypes *signatum*, dont 12 *signatum-lunulatum*, 12 *signatum-ornatum*, et 24 *signatum-ornatum-lunulatum*. Or, il se trouve que, dans les quelques dizaines de populations étudiées jusqu'à présent, les fréquences des allèles dominants vont, dans l'ensemble, en décroissant de D à S. Il en résulte qu'en général les phénotypes les plus rares sont ceux qui peuvent correspondre au plus grand nombre de génotypes : les *signatum-lunulatum-ornatum* qui représentent à eux seuls plus du quart des génotypes possibles, sont presque introuvables.

Quelqu'important que soit un échantillon, il ne sera jamais assez grand pour que l'on puisse être assuré qu'il donne une image complète de la structure génotypique de la population dans laquelle il a été prélevé.

II. CALCUL DES FRÉQUENCES GÉNIQUES

Avant d'indiquer les procédés de calcul que nous avons appliqués aux populations de Sphéromes, il nous faut rappeler les principes de la détermination des fréquences géniques *p* et *q* dans un couple d'allèles quelconques *A*, *a*. Nolons, pour commencer, que cette détermination est immédiate, lorsque tous les génotypes peuvent être identifiés et dénombrés. Si nous classons les *N* individus qui constituent un échantillon en homozygotes *AA*, hétérozygotes *Aa* et homozygotes *aa*, en faisant abstraction de toutes les autres différences génétiques qui peuvent exister entre les individus étudiés, et si les nombres respectifs d'*AA*, *Aa* et *aa* sont *x*, *y* et *z*, le nombre des gènes *A* est évidemment égal à $2x + y$ et celui des gènes *a* à $y + 2z$, le nombre total des allèles *A* et *a* étant lui-même égal à $2(x + y + z)$. Les fréquences respectives *p* et *q* de *A* et de *a* sont alors égales à $(2x + y)/2(x + y + z)$

et $(y + 2z)/(x + y + z)$. Ce procédé de calcul très simple, et qui n'implique aucune hypothèse d'aucune sorte, ne peut malheureusement pas être appliqué au cas des Sphéromes, en raison des faits de dominance qui, pour chacun des couples d'allèles étudiés ne permettent de reconnaître que deux catégories d'individus, les dominants *A*, et les récessifs *aa*. Nous devons alors construire un modèle génétique qui nous permette d'estimer les fréquences géniques à partir d'un certain nombre d'hypothèses préalables, dont il faudra, naturellement, s'il est possible, éprouver la validité.

Nous supposerons que le gène considéré n'est pas lié au sexe, qu'il ne comporte que deux allèles *A* et *a* et qu'il y a panmixie, c'est-à-dire que chaque gamète de l'un des sexes peut, en principe, se conjuguer indifféremment avec l'un quelconque des gamètes de l'autre sexe, ou encore, ce qui revient exactement au même, que les individus se croisent absolument au hasard et que leurs capacités reproductrices ne dépendent pas de leur génotype. Dans ces conditions, et si l'effectif de la population est assez grand pour que l'on puisse confondre sans erreur les probabilités et les fréquences, si *p* et *q* sont les fréquences respectives de *A* et *a* dans la population des gamètes supposés dénombrés au moment de la fécondation, les probabilités de conjugaison des *p* gamètes *A* et des *q* gamètes *a* du mâle, aux *p* gamètes *A* et aux *q* gamètes *a* de la femelle, s'expriment par un produit symbolique $(pA + qa)^2$ dont le développement $p^2 AA + 2pq Aa + q^2 aa$, donne les fréquences des œufs fécondés appartenant aux trois génotypes. Ces fréquences sont dites mendéliennes, parce qu'elles sont une conséquence immédiate de l'application de la première loi de Mendel à la population hétérallélique considérée.

La relation qui précède, connue sous le nom de loi de Hardy-Weinberg, joue un rôle fondamental en génétique des populations et sera constamment utilisée dans ce qui va suivre.

Si les trois génotypes *AA*, *Aa* et *aa* sont également viables, s'ils manifestent une égale aptitude à la concurrence vitale dans le milieu où ils se trouvent, et si, par conséquent, le taux de mortalité est indépendant de la structure génétique, cette structure sera la même dans la population des adultes et dans celle des œufs, et les fréquences génétiques *p* et *q* ne changeront pas d'une génération à la suivante. C'est ce que l'on vérifie sans peine par application des formules données au début de ce chapitre. Mais, dans le cas qui nous occupe, l'intérêt majeur de ces formules est de permettre le calcul des fréquences de *A* et de *a*, puisque, s'il y a panmixie et si les trois génotypes ont même valeur sélective, la fréquence des récessifs *aa* est toujours q^2 , quelque soit le degré de dominance de *A* sur *a*. Si l'on représente par *(a)* le nombre des récessifs et par *(A + a)* l'effectif total de l'échantillon, la meilleure estimation de *q* est $q^2 = (a)/(A + a)$. L'erreur commise sur l'estimation de *q* étant, par ailleurs, comme il est facile de l'établir, la fraction $1/2 q$ de celle qui est commise sur l'estimation de *q*, qui est donnée elle-même par une formule bien connue, on peut écrire en définitive :

$$q = \sqrt{(a)/(A+a)} = \sqrt{(A)} / 2(A+a)$$

relation qui résoud le problème posé.

Deux remarques doivent être faites à propos de cette formule, classique et fréquemment utilisée. Lorsqu'un des allèles est beau-

coup plus fréquent que l'autre, ce qui est précisément le cas pour la plupart des gènes intervenant dans le déterminisme des types de coloration des Sphéromes, les erreurs d'échantillonnage ne sont pas réparties symétriquement autour de la valeur estimée et il faut, pour être rigoureux, faire intervenir la notion d'intervalle de confiance. Cet inconvénient serait sensible si les écarts-types devaient être utilisés pour comparer deux échantillons et pour déterminer si ceux-ci diffèrent de façon significative ou si, au contraire, ils peuvent être considérés comme ayant été prélevés dans la même population, le calcul devant être répété pour chacun des mutants en présence. En fait, de tels calculs sont rarement utiles et il est assez exceptionnel d'avoir à déterminer avec précision les limites d'un intervalle de confiance.

Une deuxième remarque est due à Haldane, qui a attiré l'attention sur le fait que la meilleure estimation de q n'est pas la racine carrée de la meilleure estimation de q^2 , mais une valeur très légèrement supérieure. Pour corriger cette erreur systématique, on doit ajouter 0,25 au nombre de récessifs figurant au numérateur, et au nombre total d'individus figurant au dénominateur de la fraction servant à estimer q^2 , la meilleure estimation de q devenant alors $\sqrt{(a + 0,25)/(A + a + 0,25)}$. Cette correction est d'ailleurs presque toujours insignifiante.

Pour pouvoir être appliqués aux Sphéromes, les résultats précédents doivent être étendus au cas de plusieurs gènes. Là encore, il s'agira d'une généralisation des lois de Mendel, généralisation qui ne prend une forme simple que lorsque la population, anciennement établie, peut être considérée comme ayant atteint un état d'équilibre. On peut alors montrer que, dans ce cas, les fréquences des génotypes sont entièrement déterminées par les fréquences géniques aux divers loci, que ceux-ci soient ou ne soient pas en liaison. Les fréquences génotypiques sont déterminées par les lois de l'association au hasard des allèles d'un même gène et des allèles de tous les gènes en présence et sont données par des expressions symboliques semblables à celles dont nous avons déjà fait usage dans le cas d'un seul couple d'allèles. Pour deux couples A , a de fréquences p et q et B , b de fréquences u et v les neuf génotypes et leurs fréquences sont donnés par le développement du produit :

$$(p^2 AA + 2pq Aa + q^2 aa) (u^2 BB + 2uv Bb + v^2 bb)$$

ou de façon plus condensée par celui du produit :

$$(pA + qa)^2(uB + vb)^2$$

Lorsqu'il s'agit de gènes dominants, le produit peut s'écrire en désignant comme plus haut les dominants, homozygotes ou hétérozygotes, par A et B et les récessifs par a et b :

$$[(1 - q^2) A + q^2 a] [(1 - v^2) B + v^2 b]$$

Il y aura, pour un plus grand nombre de gènes, autant de termes dans le produit que de couples d'allèles en présence, soit quatre dans le cas des cinq types structuraux majeurs des Sphéromes.

Il est inutile de dresser le tableau relatif aux 81 génotypes, mais pour illustrer la façon dont peuvent être calculées les fréquences de tel ou tel phénotype ou génotype, nous donnerons ci-dessous, sous une forme condensée, les formules correspondant aux 16 phénotypes

dont nous avons reconnu l'existence. Pour diminuer autant que possible les complications d'écriture, nous utiliserons la même lettre pour désigner un allèle récessif et sa fréquence, ce qui, dans les calculs que nous aurons à faire ne peut prêter à confusion. Le tableau s'écrit alors :

<i>albicans</i> :	d^2	d	$ $	l^2	1	$ $	o^2	o	$ $	s^2	s
<i>discretum</i> :	$1-d^2$	D	$ $	l^2	1	$ $	o^2	o	$ $	s^2	s
<i>lunulatum</i> :	d^2	d	$ $	$1-l^2$	L	$ $	o^2	o	$ $	s^2	s
<i>ornatum</i> :	$1-d^2$	D	$ $	$1-l^2$	L	$ $	$1-o^2$	O	$ $	s^2	s
<i>signatum</i> :	d^2	d	$ $	l^2	1	$ $	$1-o^2$	O	$ $	$1-s^2$	S

et l'on voit par exemple que les fréquences du double dominant $sO1D$ et du triple dominant $SooLd$ sont respectivement $(1-d^2) l^2(1-o^2) s^2$ et $d^2(1-l^2) o^2(1-s^2)$.

Les résultats les plus utiles sont ceux qui se rapportent aux cinq phénotypes majeurs. Les formules sont particulièrement simples dans ce cas :

A	<i>dd ll oo ss</i>	d^2	l^2	o^2	s^2
D	<i>D d ll oo ss</i>	$(1-d^2)$	l^2	o^2	s^2
L	<i>L L oo ss</i>	$(1-l^2)$	o^2	s^2	
O	<i>O ss</i>	$(1-o^2)$	s^2		
S	<i>S</i>	$(1-s^2)$			

En additionnant, à partir de la base, le nombre des individus appartenant à chacune des catégories phénotypiques classées hiérarchiquement, on obtient le tableau des fréquences suivant :

A	d^2	l^2	o^2	s^2
$A + D$		l^2	o^2	s^2
$A + D + L$			o^2	s^2
$A + D + L + O$				s^2
$A + D + L + O + S$				1

On peut alors écrire, en désignant par (A), (D), (L), etc., le nombre des *albicans*, *discretum*, *lunulatum*, etc. :

$$\begin{aligned}
 d^2 &= (A)/(A + D) \\
 l^2 &= (A + D)/(A + D + L) \\
 o^2 &= (A + D + L)/(A + D + L + O) \\
 s^2 &= (A + D + L + O)/(A + D + L + O + S)
 \end{aligned}$$

On peut démontrer par des raisonnements moins élémentaires que ceux dont nous avons fait usage, que les estimations données par ces formules sont les meilleures qui puissent être extraites d'un échantillon classé en cinq types structuraux majeurs. Les fréquences géniques et leurs écarts-types s'en déduisent, comme il a été expliqué, par application de la loi de Hardy-Weinberg, la correction de Haldane pouvant être utilement faite pour les gènes particulièrement rares.

III. PRÉCISION DES ESTIMATIONS

Les formules que nous venons de justifier sont les seules qui puissent être utilisées dans le travail courant et nous n'en employons pas d'autres depuis le début de notre travail. Elles résoudraient entièrement les problèmes posés si l'on pouvait admettre avec une entière sécurité que les populations de *Sphaeroma serratum* sont panmictiques et que tous leurs génotypes structuraux ont la même valeur sélective. Mais il est peu vraisemblable que ces deux conditions soient toujours exactement remplies et il nous faut examiner les conséquences éventuelles d'un écart aux situations génétiques idéales que nous avons postulées.

Nous remarquerons en premier lieu que, dans nos formules, la valeur de d^2 ne dépend que de (A) et de (D). Il ne pouvait en être autrement puisque, en raison des faits d'épistasie, les *lunulatum*, les *ornatum* et les *signatum* ne peuvent fournir aucun renseignement sur la fréquence des allèles *D* et *d*, et que l'on peut seulement supposer que les *D* et les *d* s'y rencontrent dans les mêmes proportions que dans l'ensemble des *albicans* et des *discretum*, comme nous l'avons fait en admettant que les différents gènes s'associent au hasard. Les mêmes remarques sont valables pour les formules donnant l^2 et o^2 , si nous ne reconnaissons que cinq phénotypes majeurs, mais le problème ne se pose pas pour s^2 , dont le calcul fait intervenir tous les individus dénombrés.

On voit bien, par ce qui précède, que, tant que l'on ne distingue que cinq types structuraux, aucune justification expérimentale des calculs ne peut être donnée, puisque le nombre des paramètres à estimer est égal à celui des variables indépendantes. Il en irait autrement si les *ornatum* pouvaient être classés régulièrement en O1 et O L et les *signatum* en S O1, S O L, S o L et S O L. Disposant alors de huit données indépendantes pour calculer quatre paramètres, on pourrait améliorer l'estimation de certaines fréquences géniques, et, en même temps, contrôler dans une certaine mesure, la validité des formules utilisées.

Aucune correction ne peut évidemment être apportée à l'estimation de la fréquence des récessifs *dd*, puisque l'action du gène *discretum* est toujours complètement masquée par celle des trois autres gènes, ni à celle des récessifs *ss* puisque, au contraire, l'action du gène *signatum* n'est jamais masquée. Il n'en est pas de même pour *ll* qui peut, au moins en principe, être reconnu, chez les *ornatum* sous le phénotype O1 et chez les *signatum* sous les phénotypes S O1 et S O L ; ni pour *oo*, qui peut être retrouvé chez les *signatum* sous les formes S o L et S o L. D'où les deux formules :

$$l^2 = (A + D + o1 + S O1 + S O L) / (A + D + L + O + S)$$

$$o^2 = (A + D + L + S o1 + S o L) / (A + D + L + O + S)$$

relations qui, comme celles qui donnent s^2 , ont le grand mérite de ne rien préjuger sur la nature des règles qui président à l'association des différents gènes.

Les fréquences géniques une fois estimées par le procédé qui vient d'être indiqué, il est possible de vérifier si l'association des gènes *lunulatum*, *ornatum* et *signatum* se fait ou non au hasard, en comparant aux effectifs observés les effectifs calculés en admettant la panmixie. Les *discretum* et les *albicans* doivent être réunis en une même catégorie, puisque l'effet des allèles *D* ou *d* est entièrement masqué par tout autre gène dominant. Il y a quatre degrés de liberté et les fréquences théoriques sont données par les formules suivantes :

<i>albicans</i> + <i>discretum</i>	A + D	l^2	o^2	s^2
<i>lunulatum</i>	L	$(1-l^2)$	o^2	s^2
<i>signatum</i>	0 L	l^2	$(1-o^2)$	s^2
<i>ornatum</i>	S 0 L	l^2	o^2	$(1-s^2)$
	S 0 L	l^2	$(1-o^2)$	$(1-s^2)$
	S O L	$(1-l^2)$	o^2	$(1-s^2)$
	S O L	$(1-l^2)$	$(1-o^2)$	$(1-s^2)$

La méthode de contrôle qui vient d'être exposée, très satisfaisante en théorie, l'est très peu en pratique, les dominants doubles ou triples étant presque toujours beaucoup trop rares et trop difficiles à classer pour que l'on puisse en tirer normalement parti dans l'étude de la structure génétique de la population. Dans le travail courant, il faudra presque toujours se contenter des formules générales obtenues en distinguant seulement les cinq phénotypes majeurs. La question qui se pose à leur propos est de savoir si la marge d'incertitude que comportent normalement les estimations qu'elles permettent, est ou non supérieure aux erreurs qui peuvent résulter de l'emploi d'une formule imparfaitement adaptée aux structures génétiques de la population étudiée.

Les deux hypothèses qui sont à la base de nos calculs, panmixie et égalité des valeurs sélectives des divers génotypes, sont indépendantes l'une de l'autre, mais conditionnent l'une et l'autre les fréquences des phénotypes qui restent en présence lors du dénombrement. Il en résulte qu'un test de conformité portant sur un échantillon classé en neuf catégories peut permettre de constater que les fréquences observées diffèrent de façon significative des fréquences calculées, mais qu'il ne peut apporter aucune information sur les causes de ces écarts, qui ne pourraient être décelées que par des recherches expérimentales spécialement organisées à cet effet. Il est cependant possible de trouver une limite supérieure vraisemblable aux erreurs qui peuvent résulter d'une réalisation imparfaite des conditions théoriques que nous avons placées à la base de nos calculs, en ne faisant appel qu'aux caractéristiques les plus évidentes des populations étudiées.

Nous illustrerons cette proposition en étudiant le cas relativement simple du gène *S*. Nous savons que les *signatum* sont presque toujours rares et que, dans le plus grand nombre de populations, leur fréquence est bien inférieure à 10 pour 100. Nous savons aussi que presque tous les *signatum* dont la descendance a pu être étudiée, étaient hétérozygotes. Ce résultat est conforme à la théorie, puisque, si la fréquence du gène *S* est p , celle des homozygotes *SS* doit être p^2 , et

celle des hétérozygotes Ss , $2 pq$: pour $S = 0,05$, on devrait trouver 1 SS pour 38 Ss et 361 ss , et pour $S = 0,01$, 1 SS pour 198 Ss et 9.801 ss .

Il pourrait cependant se faire que, nos hypothèses de départ n'étant pas absolument exactes, la fréquence réelle de SS diffère assez notablement de p^2 . On peut sans doute écarter l'idée que les croisements consanguins sont chez les Sphéromes, favorisés par rapport aux autres, mais il n'est pas sûr que, chez ces animaux comme chez beaucoup d'autres, il n'existe pas un certain degré d'homogamie ou un système de croisements préférentiels. S'il en est ainsi, les *signatum* se croisant plus volontiers entre eux qu'avec des individus appartenant à d'autres phénotypes, le nombre des homozygotes formés lors de la fécondation sera supérieur au nombre théorique. Il est vraisemblable, en revanche, si le mécanisme responsable de l'équilibre de S et de s est conforme au schéma classique, que les homozygotes SS sont moins viables que les hétérozygotes Ss , ce qui doit avoir pour conséquence d'en diminuer la fréquence relative et pourrait même, à la limite, les faire disparaître complètement.

Dans cette dernière hypothèse, la fréquence de S serait égale à $(S)/(A + D + L + O + S)$; c'est-à-dire à la moitié de la fréquence des *signatum*, l'écart à la valeur donnée par la relation de Hardy-Weinberg étant $-p^2/2$. Si, par suite des croisements préférentiels, le nombre des homozygotes devenait deux fois plus grand que celui qui doit s'observer dans le cas de la panmixie, l'erreur commise serait la même, mais en sens inverse. Il est facile de calculer que cette erreur reste inférieure à l'écart-type de l'estimation donnée par nos formules tant que la fréquence de S est inférieure à 0,12 pour un échantillon de 1.000 individus et à 0,10 pour un échantillon de 2.000, effectif rarement atteint dans nos dénombremens. Si le nombre des homozygotes était quadruple ou quintuple de celui auquel on pouvait s'attendre, l'erreur commise ne serait supérieure à l'écart-type que si la fréquence de S dépassait 0,06 ou 0,05 dans un échantillon de 1.000 et 0,05 ou 0,04 dans un échantillon de 2.000.

Des calculs semblables à ceux qui précèdent peuvent être faits pour les *ornatum* et les *lunulatum* lorsque les Sphéromes étudiés ont été classés en neuf phénotypes. Le raisonnement est un peu plus compliqué si l'on ne distingue que cinq phénotypes, l'incertitude étant alors double. Si l'on peut, comme toujours, commettre une erreur dans l'estimation du rapport du nombre des homozygotes aux hétérozygotes chez les dominants, il se peut aussi, en même temps, que les fréquences des *signatum-ornatum*, des *signatum-lunulatum*, des *signatum-ornatum-lunulatum* et des *ornatum-lunulatum*, ne soit pas exactement égales aux fréquences calculées. Ces deux causes d'erreur sont indépendantes, la première étant, comme on l'a vu, de l'ordre de O^2 ou de L^2 , et la deuxième de l'ordre de $O.S$ ou de $L.S + L.O$. Elles seront presque toujours négligeables, O et L étant, comme S , généralement petits. On voit par là que, dans le plus grand nombre des cas, l'estimation des fréquences de S , O et L données par nos formules peut légitimement être tenue pour satisfaisante. Il en est certainement de même pour les *rubrum R* et, plus généralement, pour tous les gènes de couleur, qui sont en même temps rares et reconnaissables dans tous les phénotypes structuraux.

Le cas de *discretum* est différent, l'allèle dominant *D* ayant une fréquence qui atteint ou dépasse 0,50 dans beaucoup de populations. S'il est fort improbable que les valeurs sélectives des trois génotypes *DD*, *Dd* et *dd* soient très différentes l'une de l'autre, on ne peut exclure *a priori* la possibilité de certains écarts à la panmixie, qui, sans être très importants, affecteraient sensiblement la précision de nos estimations. Cette cause d'erreur s'ajoute à celle que nous avons souvent signalée, en rappelant que la distinction des *albicans* et des *discretum* est parfois malaisée.

(*Laboratoire de Zoologie, Sorbonne.*)

INDEX BIBLIOGRAPHIQUE

- BOCQUET, C., LÉVI, C. et TEISSIER, G., 1951. — Recherches sur le polychromatisme de *Sphaeroma serratum* (F.). *Arch. Zool. Exp. Gen.*, 87, pp. 245-297.
- BOCQUET, C. et TEISSIER, G., 1960. — Génétique des populations de *Sphaeroma serratum* (F.). I Stabilité du polychromatisme. *Cah. Biol. Mar.* 1, n° 1, pp. 103-111.
- HALDANE, J.B.S., 1956. — Almost unbiased estimates of functions of frequencies. *Sankhyā* 17 Part 3, pp. 201-208.