

# SYSTÉMATIQUE BIOCHIMIQUE ET ÉVOLUTION

## par

M I . Cariou

Laboratoire de Biologie et Génétique évolutives  
C.N.R.S.91190 Gif-sur-Yvette

Dans l'esprit de bon nombre de biologistes, la systématique est considérée comme synonyme de taxonomie. Cependant, après Simpson (1944) et Mayr (1969), on tend à admettre une acceptation beaucoup plus large du terme. Actuellement, un certain nombre de biologistes admettent que la systématique couvre à la fois l'étude de la diversité des organismes et celle de leurs affinités évolutives. On parle d'ailleurs communément de systématique évolutive.

En systématique, l'unité de base est l'espèce dans le sens du concept défini par Mayr (1963) qui implique l'interfécondité des individus et leur aptitude à produire une descendance fertile. Dans le domaine de l'évolution et de la génétique des populations, l'unité de base est la population, c'est-à-dire une collection d'individus qui, bien que tous différents les uns des autres, possèdent en commun un ensemble de caractères qui les distinguent des autres organismes présents dans leur environnement. Une espèce va comporter un nombre plus ou moins grand de populations dont les caractéristiques sont suffisamment cohérentes pour qu'elle soit reconnue comme une entité parfaitement définie.

Les populations et les espèces ne constituent pas des entités figées dont les caractéristiques sont fixées une fois pour toutes : elles évoluent. Elles subissent des variations génétiques : — par mutations ponctuelles; — sous l'action des pressions sélectives exercées par le milieu au sens large du terme; — par dérive génique (variations aléatoires).

De ce fait, au cours du temps, facteur essentiel, des différences génétiques vont s'accumuler.

Cette évolution graduelle pourrait ainsi conduire à la formation d'espèces nouvelles selon un processus d'anagenèse. Cependant, le mode de spéciation le plus répandu reste celui de la spéciation allopatrique selon lequel, deux populations isolées d'une même espèce, qui n'échangent plus de gènes, vont évoluer indépendamment et acquérir au cours du temps un isolement reproductif. On aura alors deux nouvelles espèces qui ont une origine commune. Ceci a conduit les auteurs à chercher si l'acquisition d'un statut spécifique s'accompagnait d'un niveau de différenciation génétique déterminé; autrement dit, à partir de quel niveau de différenciation peut-on parler d'espèces différentes? Ce débat est maintenant un peu dépassé et il s'est établi un consensus pour reconnaître que les différences biochimiques, des protéines ou des alloenzymes, ne représentent en aucune façon des facteurs primaires de spéciation mais un sous-produit de celle-ci. Les mécanismes responsables de l'isolement reproductif sont de toute autre nature.

Ce qu'il faut retenir, c'est qu'à partir d'une cladogenèse, les différences génétiques biochimiques vont s'accumuler indépendamment dans les deux lignées ainsi créées et leur accumulation sera globalement liée au temps écoulé depuis la divergence. Cela ouvre des perspectives qui dépassent la simple identification des espèces et je me propose ici non pas de faire une revue bibliographique des travaux réalisés sur les Crustacés dans ce domaine (voir pour revue Hedgecock *et al.*, 1982) mais d'examiner les différentes techniques dont nous disposons et de voir jusqu'où peut nous conduire cette approche.

## LES TECHNIQUES ÉLECTROPHORÉTIQUES

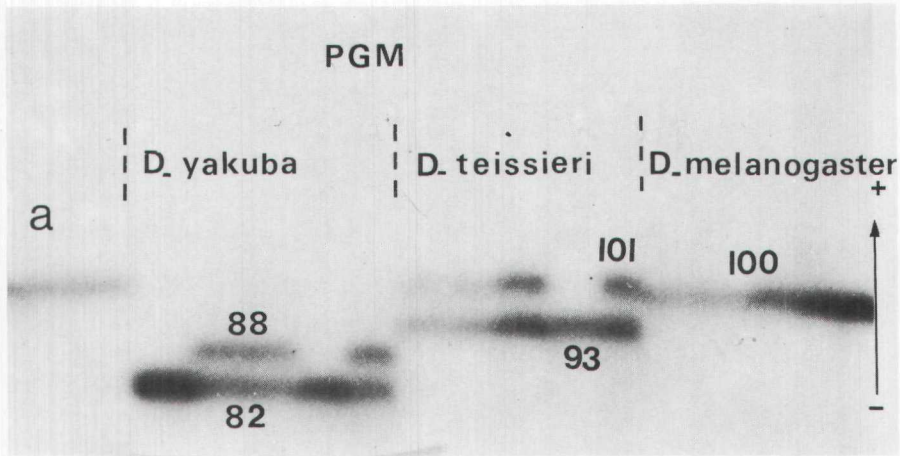
Elles reposent sur la possibilité de faire migrer les protéines dans un champ électrique au travers de supports variés. Les protéines sont ainsi séparées d'abord en fonction de leur charge électrique mais aussi, dans certaines conditions, en fonction de leur poids moléculaire et de leur conformation. Au terme d'une électrophorèse, on peut révéler soit une activité enzymatique spécifique, soit des protéines totales. Les diverses techniques électrophorétiques n'ont pas toutes le même pouvoir résolutif. A partir d'un échantillon de plasma de Vertébrés, contenant 100 à 200 protéines, l'électrophorèse sur acétate de cellulose permet de révéler 5 bandes, sur gel d'amidon et d'acrylamide 15 à 19 bandes respectivement; le pouvoir de résolution est notablement accru par l'isofocalisation (30-40 bandes) et plus encore en électrophorèse bidimensionnelle puisque plus de 100 protéines peuvent alors être identifiées (Anderson et Anderson, 1977).

L'électrophorèse conventionnelle est pratiquée très largement pour les études de génétique des populations, depuis près de 20 ans maintenant. Il s'agit essentiellement des électrophorèses sur gel d'amidon et d'acrylamide, au terme desquelles des activités enzymatiques spécifiques sont le plus souvent révélées. Ainsi, les électrophorégrammes comportent un petit nombre de bandes qui permettent une interprétation génétique selon des principes simples : une seule bande correspond à un individu homozygote pour un allèle déterminé; l'observation de bandes multiples indique un individu hétérozygote et le nombre de bandes dépendra de la structure de l'enzyme (Pl. I). Il n'est pas rare que plusieurs systèmes enzymatiques à commande génétique indépendante soient révélés sur un même gel (isozymes), notamment pour les estérases, les phosphatases, les leucine-aminopeptidases. Dans ce cas, la distinction des allèles appartenant à l'un ou l'autre système est parfois délicate. Une autre difficulté réside dans l'identification d'allèles codant pour des allozymes de mobilité électrophorétique très proche.

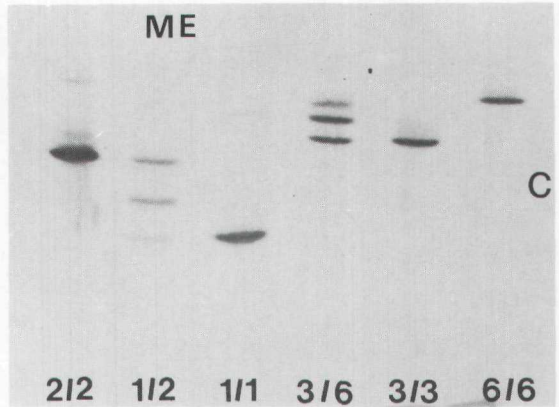
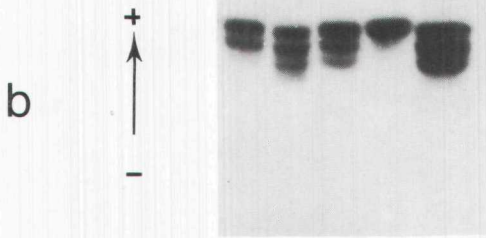
L'isofocalisation est pratiquée en gel d'acrylamide ou d'agarose à larges pores auquel est incorporé un mélange d'ampholytes porteurs couvrant une gamme de pH plus ou moins large. Dans un champ électrique, ces ampholytes ont la particularité de constituer un gradient de pH stable et les protéines vont migrer jusqu'à ce qu'elles atteignent leur point isoélectrique. Elles sont ainsi concentrées en bandes très étroites sur le gel. Cette technique possède un pouvoir de résolution très élevé puisqu'elle permet de séparer des protéines qui diffèrent seulement de 0,01 unité de pI et permet de révéler soit des protéines totales (images bien connues peu favorables à une interprétation génétique) soit des activités enzymatiques spécifiques (PL I).

### PLANCHE I

Electrophorégrammes obtenus par diverses techniques, a : révélation de l'activité de la phosphoglucomutase chez plusieurs espèces du genre *Drosophila*; les hétérozygotes comportent deux bandes, b : 6-phosphogluconate déshydrogénase du parasite *Trypanosoma cruzi* révélée sur gel d'acétate de cellulose. c : enzyme malique d'amibes du genre *Naegleria* révélée après isoélectrofocalisation. Dans les deux derniers cas les souches hétérozygotes présentent trois bandes, d : protéinogramme obtenu à partir d'une drosophile après électrophorèse bidimensionnelle.

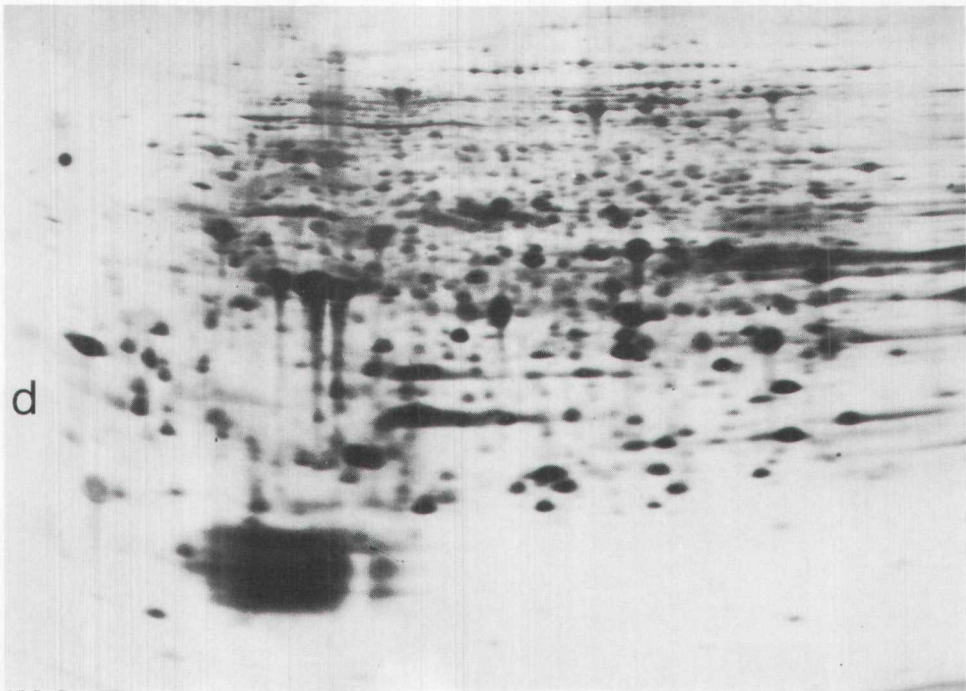


6. PGD



**Génotypes**

3/0    3/4    3/3    3/4



M.-L. CARIOU

Attrayante par son haut pouvoir de résolution, l'isofocalisation reste relativement coûteuse et a l'inconvénient de produire parfois des artéfacts (dédoublés de bandes) que l'on attribue à la fixation des ampholytes sur certaines protéines.

L'électrophorèse bidimensionnelle a été proposée en 1975 par O'Farrell; elle associe deux techniques d'électrophorèse :

— La première dimension est une isofocalisation en boudin d'acrylamide qui sépare les protéines selon leur point isoélectrique.

— La seconde dimension est une électrophorèse dans un gel à gradient de concentration d'acrylamide dans des conditions dénaturantes, c'est-à-dire en présence d'un détergent (SDS, Sodium Dodecyl Sulfate), pour séparer les protéines en fonction de leur poids moléculaire.

Les polypeptides, et non plus les enzymes, sont révélés par marquage radioactif ou coloration à l'argent.

On obtient ainsi des plaques comportant plusieurs centaines de spots, jusqu'à 1.000, correspondant à autant de composants protéiques d'un même individu (PI. I). Devant la complexité des images, il devient alors difficile de proposer une interprétation génétique et l'établissement des homologues protéiques entre deux individus est une entreprise délicate.

### **Avantages et Inconvénients des différentes techniques**

On peut grouper ces techniques en fonction de leur cible : enzymes, ou protéines totales. Dans le premier cas, l'électrophorèse conventionnelle et l'isofocalisation seront appropriées, dans le second cas, l'électrophorèse en double dimension sera adéquate.

— L'étude des activités enzymatiques a l'avantage indéniable de permettre de déduire directement du phénotype observé le génotype de l'individu pour chaque locus. Pour l'ensemble des individus d'une population il est alors possible de déterminer les fréquences alléliques qui la caractérisent. On peut cependant opposer trois critiques principales à cette approche :

1. Elle ne permet pas un échantillonnage au hasard du génome. Il s'avère que pour la plupart des organismes, les mêmes enzymes sont révélées de manière satisfaisante. On trouve donc régulièrement le même jeu de base d'une vingtaine d'enzymes dans toutes les études de polymorphisme. Cet inconvénient est compensé par la faculté d'établir des comparaisons entre des caractères homologues.

2. Les 20, 30 ou même 50 gènes étudiés dans les cas les plus favorables ne représentent qu'une infime partie de l'ensemble des gènes structuraux de l'organisme. On estime en effet à environ 50.000 le nombre de gènes de structure d'une drosophile.

3. Elle ne permet guère de détecter plus de 25 à 30 p. 100 des substitutions d'acides aminés.

Ce résultat est notablement amélioré par l'utilisation de l'électrophorèse séquentielle ou analytique qui associe plusieurs techniques et notamment l'électrophorèse en présence d'urée ou d'un détergent (SDS) qui ont la propriété de solubiliser certaines catégories de protéines. En outre, des gels de concentration variable en acrylamide et la dénaturation thermique des extraits permettent de

révéler des électromorphes supplémentaires. C'est en appliquant ces techniques que Lewontin et ses collaborateurs ont mis en évidence 37 allèles au locus de la Xdh alors que les méthodes conventionnelles n'en détectaient que 6 chez *Drosophila pseudoobscura* (Singh *et al.*, 1976).

Plus récemment, des résultats comparables ont été obtenus pour d'autres locus, l'aldehyde oxydase et l'estérase-6.

Lewontin (comm. orale) a estimé que 90% de la variabilité totale est détectée par l'électrophorèse séquentielle. Les résultats de Ramshaw *et al.* (1979) montrent que 17 des 20 mutations de l'hémoglobine identifiées par séquençage sont ainsi détectées.

— L'analyse des protéines totales en double dimension augmente considérablement le nombre de gènes de structure que l'on peut étudier. La grande difficulté réside dans la lecture des gels et en particulier l'identification des protéines homologues. Elle nécessite presque obligatoirement un dispositif de lecture automatique couplé à une infrastructure informatique sérieuse. Les programmes de traitement de tels gels sont actuellement en plein développement (Tarroux *et al.*, 1985).

Cette approche exclut toute interprétation génétique, à l'exception de quelques cas particuliers quand le nombre de spots est limité. Racine et Langley (1980) ont

TABLEAU I

Variabilité détectée par différentes méthodes d'électrophorèse.  $ne^*$  : nombre effectif d'allèles révélés par électrophorèse conventionnelle;  $ne$  : nombre effectif d'allèles révélés par les techniques alternatives. 2D : électrophorèse bidimensionnelle. ( ) : nombre de locus. (D'après Selander et Whittam, 1983).

| Espèce                  | $ne^*$       | $ne^P$       | $ne'/ne$ | Technique                          | Référence                        |
|-------------------------|--------------|--------------|----------|------------------------------------|----------------------------------|
| <i>D. pseudoobscura</i> | 1,38<br>(13) | 1,65<br>(13) | 1,12     | Electrophorèse<br>séquentielle     | Ayala (1982)                     |
| <i>D. subobscura</i>    | 1,83<br>(8)  | 2,42<br>(8)  | 1,25     | Conventionnel +<br>dénat. à urée   | Loukas <i>et al.</i><br>(1981)   |
| <i>D. melanogaster</i>  | 1,73         | 2,06         | 1,18     | Conventionnel +<br>dénat thermique | Ayala (1982)                     |
|                         | 1,20<br>(21) | 1,04<br>(54) | 0,87     | 2D                                 | Leigh Brown et<br>Langley (1979) |
| <i>D. simulons</i>      | 1,06<br>(21) | 1,00<br>(70) | 0,94     | 2D                                 | Ohnishi <i>et al.</i><br>(1983)  |
| <i>Mus musculus</i>     | 1,06<br>(22) |              |          |                                    | Berry et Petters<br>(1977)       |
|                         |              | 1,02<br>(72) | 0,96     | 2D                                 | Racine et Langley<br>(1980)      |
| <i>Homo sapiens</i>     | 1,06<br>(87) |              | —        |                                    | Harris <i>et al.</i><br>(1977)   |
|                         |              | 1,00<br>(83) | 0,94     | 2D                                 | Smith <i>et al.</i><br>(1980)    |
|                         | —            | 1,01<br>(61) | 0,95     | 2D                                 | Comings (1979)                   |

$ne$  : réciproque de la somme des carrés des fréquences des allèles.

ainsi détecté des variations alléliques chez la souris et montré que les protéines détectées par électrophorèse bidimensionnelle sont moins polymorphes que les protéines enzymatiques puisque le taux d'hétérozygotie est de 0,02 dans le premier cas et 0,17 dans le second. Des constatations analogues ont été faites pour d'autres organismes (Tabl. 1) (Selander et Whittam, 1983; Avise et Aquadro, 1983). Cependant, chez les végétaux, des résultats récents montrent qu'une modification de la technique d'électrophorèse bidimensionnelle d'O'Farrell permet d'augmenter la résolution analytique dans un rapport de 2,5 (Damerval *et al.*, 1986). De plus l'analyse de cinq lignées de maïs a révélé une variabilité qualitative très importante (de Vienne et Damerval, comm. pers.) dont l'origine, allélique ou non, n'est pas encore établie.

### LA VARIABILITÉ INTRASPÉCIFIQUE

L'étude des protéines et plus particulièrement des alloenzymes conduit à déterminer la structure génétique des populations, chacune étant caractérisée par la fréquence des allèles qui y ségrègent. La première constatation qui s'impose à partir de l'ensemble des données maintenant disponibles dans la littérature, est l'existence d'une variabilité génétique considérable dans les populations naturelles des organismes les plus divers.

Si la majorité des Crustacés a une hétérozygotie inférieure à 10%, certaines espèces, dont l'Isopode *Jaera albifrons*, se situent parmi les organismes les plus variables avec une hétérozygotie comprise entre 20 et 25% (Cariou, 1977).

Une telle ampleur du polymorphisme indique clairement que l'analyse de la variabilité intraspécifique constitue un préalable indispensable à toute comparaison interspécifique.

Des variations des protéines liées aux caractéristiques biologiques des individus ou à leur origine géographique sont connues dans de nombreux cas :

— Variations liées à l'âge des individus; certaines protéines ne s'expriment qu'à des stades particuliers du développement.

Quelques enzymes sont codées par des gènes liés au sexe. C'est le cas d'une forme d'amylase chez *Drosophila hydei* (Doane *et al.*, 1975), de la G6PDH et de la phosphoglycérate kinase chez l'Homme (Cooper *et al.*, 1975) et de l'Est-5 chez *Drosophila pseudoobscura*.

— D'autres alloenzymes sont synthétisés dans certains tissus seulement; le cas le plus connu est la lactate déshydrogénase de Vertébré qui se caractérise par une forme cardiaque, une forme musculaire et une forme rétinienne que l'on observe aussi dans les testicules des oiseaux et des Mammifères. Le cas des  $\alpha$ -amylases est particulièrement complexe. Chez *Drosophila melanogaster*, le locus structural est dupliqué, fortement polymorphe et il existe des gènes de régulation agissant soit sur le niveau de la transcription (sous l'influence des sucres), soit sur l'expression tissulaire dans le tube digestif (Bahn, 1967; Doane, 1965, 1969).

— L'analyse de la variation géographique, notamment pour les espèces à vaste répartition, a montré l'existence de variations clinales latitudinales, altitudinales, etc. des fréquences de certains allèles et dans d'autres cas une forte hétérogénéité génétique des populations conspécifiques.

Il est bien évident que la validité des comparaisons que l'on peut établir par les enzymes dépend étroitement :

- du nombre d'individus étudiés par population;
- du nombre de populations analysées;
- du nombre de locus enzymatiques considérés.

Les exigences qu'il convient d'avoir vis-à-vis de ces paramètres dépendent du polymorphisme des locus. Plus un locus est polymorphe, c'est-à-dire comporte un grand nombre d'allèles, plus grand sera l'échantillon étudié pour établir les fréquences alléliques de la population. Quant au nombre de locus, le plus grand nombre est souhaitable. Nei (1978) estime qu'il est nécessaire de considérer une cinquantaine de locus, ce qui reste inaccessible dans la plupart des cas. On peut admettre qu'une trentaine de locus constitue déjà un échantillon satisfaisant.

### LA DIFFÉRENCIATION INTERSPÉCIFIQUE

La simple comparaison locus par locus peut révéler des situations intéressantes : si une collection d'individus, que rien ne différencie morphologiquement, est examinée pour un locus particulier et révèle deux catégories d'individus homozygotes pour des allèles différents en l'absence totale des hétérozygotes correspondants, on peut légitimement conclure qu'ils n'engagent pas de gènes et qu'il s'agit donc de deux espèces jumelles.

Il existe dans la littérature de nombreux exemples d'espèces jumelles identifiées grâce aux alloenzymes, tant chez les Vertébrés que les Invertébrés mais l'un des premiers, et aussi l'un des plus convaincants, est celui des six espèces sympatriques de *Capitella* (Annélides) qui n'ont aucun allèle en commun au locus de la PHI (Grassle et Grassle, 1976). Il est bien évident que les locus enzymatiques diagnostiques sont particulièrement précieux pour les organismes qui présentent peu ou pas de différenciation morphologique, c'est le cas de nombreux protozoaires, parasites, etc.

**La comparaison de la structure génétique globale** et l'évaluation de la divergence entre des espèces ou des entités taxonomiques de rang plus élevé s'est rapidement avérée nécessaire. Le problème n'est plus d'attribuer une signification taxonomique à un quelconque degré de divergence puisqu'il est bien établi qu'une spéciation peut se produire sans entraîner de divergence génétique biochimique notable ou, au contraire, être accompagnée d'une forte différenciation de génomes. Il n'en reste pas moins que l'analyse du polymorphisme des populations naturelles pour un grand nombre de caractères nécessite une approche synthétique et quantitative. Pour cela, deux voies différentes sont possibles :

— La première est l'utilisation des estimateurs de distance et de nombreuses «distances», basées sur des concepts différents, ont été proposées. La plus populaire est la distance de Nei (1972, 1975) qui estime le nombre moyen de substitutions de codons par locus à partir de la séparation de deux taxons. On lui reproche cependant d'être dépendante de l'hétérozygotie, donc du nombre d'allèles présents à un locus, d'admettre un taux de substitution constant pour tous les locus et de présenter une mauvaise linéarité dans le temps. Sur ce dernier point en particulier, il semble que d'autres distances se comportent mieux (Gregorius, 1984). Il est clair que le choix de l'indice est un problème important auquel

s'ajoutent ceux des limites de validité des différents indices, de la nécessité d'en utiliser un seul ou plusieurs. Tous ces problèmes sont tout à fait d'actualité (Katz, 1986; Capy et Cariou, 1986).

— La seconde est l'application des analyses multidimensionnelles aux fréquences des alloenzymes. L'analyse des Données Centrées (Lefebvre, 1976) s'avère la mieux adaptée à ce type de données (Cariou et Lefebvre, 1982) et ne nécessite aucun postulat quant à la vitesse d'évolution des gènes et la structure des populations.

La comparaison d'espèces bien différenciées ne soulève généralement que peu de difficultés. Les problèmes de l'échantillonnage des espèces, c'est-à-dire du nombre d'individus qu'il convient d'examiner pour les caractériser, du choix de l'indice de distance ou de l'utilisation des analyses multidimensionnelles ne se posent guère quand la différenciation génétique est importante. A la limite l'analyse d'un unique individu pour un petit nombre de locus suffit et, quelle que soit la méthode utilisée, les résultats sont concordants. La situation est beaucoup plus délicate lorsque les espèces sont étroitement apparentées et que leur divergence est faible. Il devient alors nécessaire d'examiner plusieurs dizaines d'individus pour déterminer avec précision les fréquences des alloenzymes et le plus grand nombre possible de locus pour obtenir un bon échantillon de gènes.

**Le cas des Jaera** : Le complexe *Jaera albifrons* avec ses cinq espèces étroitement apparentées illustrera cette situation. Morphologiquement très proches puisque le diagnose spécifique repose exclusivement sur les variants sexuels secondaires des périopodes des mâles, les femelles restant indifférenciées, les espèces de *Jaera* sont génétiquement étroitement apparentées. L'analyse bio-

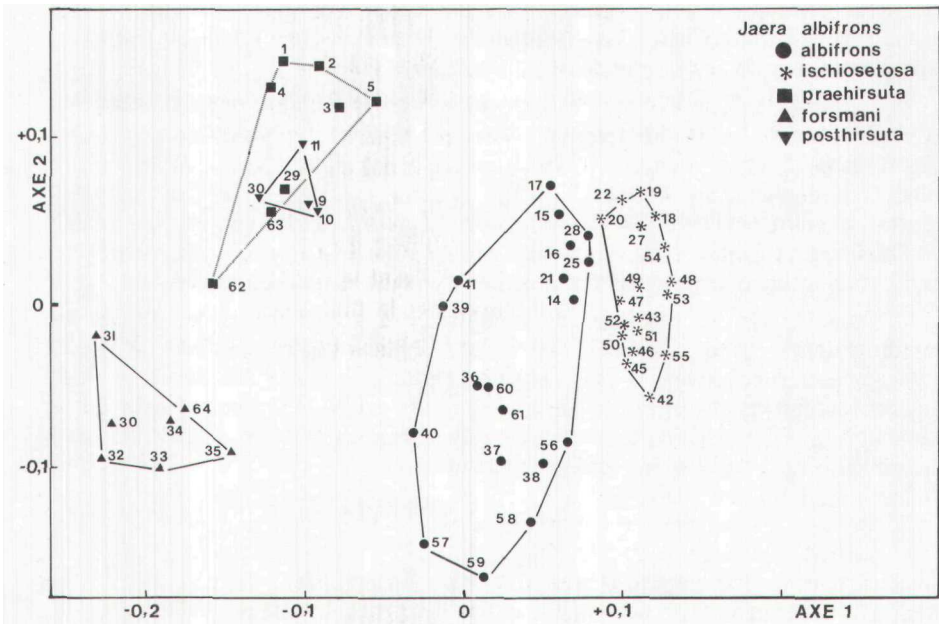


FIG. 1

Projection des populations de *Jaera albifrons* (Crustacé, Isopode) dans le plan défini par les deux premières composantes de l'analyse des Données Centrées appliquée aux fréquences des alloenzymes. Quatre espèces sont ainsi parfaitement discriminées : la cinquième espèce est discriminée sur l'axe 3.

chimique est basée sur 18 locus enzymatiques analysés dans une cinquantaine de populations sur un échantillon de 20 à plus de 100 individus. L'ampleur de la différenciation génétique (distance de Nei) varie sensiblement selon les couples d'espèces considérées. La divergence est maximale ( $D = 0,60$ ) entre *forsmani*, espèce strictement européenne et *posthirsuta*, exclusivement américaine, mais aussi *ischioetosa*, espèce présente de part et d'autre de l'Océan Atlantique. La divergence est minimale entre les espèces *forsmani* et *praehirsuta* ( $D = 0,20$ ), résultat concordant avec les données morphologiques et biogéographiques antérieures (Cariou, 1977). Il est intéressant de noter qu'entre ces deux espèces, la distance génétique est du même ordre de grandeur ou même inférieure à celles qui séparent les populations de l'espèce la plus hétérogène, *albifrons* (distance moyenne intraspécifique : 0,23; Cariou, 1985).

L'application de l'analyse des Données Centrées aux mêmes données de base, les fréquences des alloenzymes, conduit à une image globalement compatible avec la précédente mais néanmoins différente à certains égards. L'examen de la figure 1 montre en effet que la discrimination est bien maximale entre *ischioetosa* et *forsmani*; elle est encore forte, mais moins marquée, entre *posthirsuta* et *ischioetosa*. La position relative de *forsmani* et *praehirsuta* est plus inattendue; ces deux espèces que l'on sait étroitement apparentées sont fortement discriminées sur l'axe 2. Ceci s'explique simplement par le fait que l'axe 2 rend compte de la différenciation géographique Europe-Amérique du Nord qui s'exprime au niveau des espèces, (rappelons que *forsmani* est exclusivement européenne et *posthirsuta* américaine) mais aussi au niveau des populations d'une même espèce : les populations américaines *d'albifrons* et *d'ischioetosa* sont positives sur l'axe 2, les populations européennes, négatives sur ce même axe.

L'analyse des Données Centrées permet donc de mettre en évidence l'ampleur de la différenciation génétique entre les populations et les espèces et de plus d'analyser la différenciation en plusieurs composantes. C'est donc une approche complémentaire de l'utilisation des distances génétiques.

L'obtention d'une matrice de distances génétiques conduit tout naturellement à la construction de dendrogrammes. Il existe plusieurs méthodes qui produisent des réseaux (Fitch et Margoliash, 1967; Farris, 1972; Tateno *et al.*, 1982) ou des cladogrammes (Sneath et Sokal, 1973). La plus populaire et la plus simple à mettre en œuvre est l'UPGMA (Unweighted pair-group method analysis) de Sneath et Sokal qui procède en regroupant d'abord les espèces (ou populations) les plus semblables génétiquement puis une à une chacune des autres successivement.

Dans le cas des espèces du complexe *Jaera albifrons*, on aboutit au dendrogramme suivant (Fig. 2) : les cinq espèces actuelles résultent de quatre cladogenèses successives, la plus ancienne divisant la lignée ancestrale en deux lignées évolutives divergentes. La cladogenèse la plus récente est celle qui a conduit à la différenciation *de forsmani*; cette dernière espèce et *praehirsuta* étant indiscutablement les plus étroitement apparentées.

### L'ESTIMATION DU TEMPS ÉVOLUTIF

Moyennant certaines hypothèses, la divergence estimée par les protéines reflètera non seulement la phylogénie mais aussi le temps écoulé depuis la première cladogenèse.

Ainsi, si l'on accepte les postulats suivants : taux de substitution par année constant; taux de mutation, pression de sélection et effets de la dérive comparables dans les populations, la distance génétique de Nei (D) est directement proportionnelle au temps écoulé depuis la divergence des deux entités.

Cette relation s'exprime de la manière suivante :

$$D = 2 \alpha T$$

ou  $\alpha$  est le taux de substitution des codons détectables par électrophorèse, estimé à  $10^{-7}$  par Nei (1975) ainsi,

$$T = 5 \times 10^6 D$$

Autrement dit, une distance génétique de 1 correspondra à un temps de divergence de 5 millions d'années. Les protéines apparaissent donc comme une horloge moléculaire (evolutionary clock) permettant de dater les événements évolutifs.

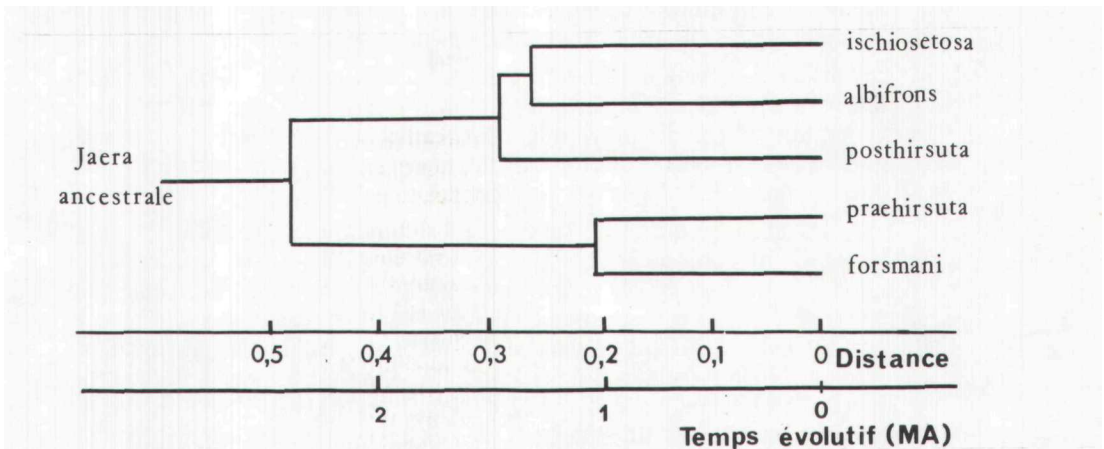


FIG. 2

Le dendrogramme obtenu à partir de la matrice des distances génétiques établit les parentés évolutives des cinq espèces affines du complexe *Jaera albifrons*. *Forsmani*, limitée à une portion des côtes européennes, est l'espèce différenciée le plus récemment; sa divergence à partir de *praehirsuta* (ou d'une «*praehirsuta-like*») remonterait à environ 1 MA. La divergence d'*ischiosetosa* et *albifrons* serait plus ancienne; la cladogenèse la plus ancienne remonterait à près de 2,5 MA.

Si l'on considère à nouveau le dendrogramme de la figure 2, on peut dire, avec une certaine approximation, que la divergence la plus ancienne chez les *Jaera* se situe entre 2 et 3 millions d'années. La dernière cladogenèse, celle qui a donné naissance à *forsmani* à partir de *praehirsuta* ou d'une «*praehirsuta-like*», remonterait à environ 1 million d'années.

On sait cependant que le taux d'évolution des protéines n'est pas constant et Sarich (1977) reconnaît deux catégories de protéines, celles qui évoluent lentement ( $T = 30 \times 10^6 D$ ) et celles qui évoluent rapidement ( $T = 2,4 \times 10^6 D$ ). L'estimation de Nei est donc plutôt une estimation basse. Plus récemment, Avise et Aquadro (1983) ont souligné que les unités de temps utilisées par les différents auteurs sont extrêmement variables puisque à une distance génétique égale à l'unité correspondent des temps variant de 0,7 à 33 millions d'années. Ils considèrent que le taux d'évolution moléculaire varie non seulement d'une protéine à l'autre mais aussi d'un taxon à un autre, notamment chez les Vertébrés. Pour les Drosophiles de Hawaii, par exemple, Carson (1976) retient 2 MA pour une unité de distance génétique, soit une vitesse d'évolution 15 fois plus rapide que celle proposée par

Yang (1974) pour les lézards. Quoi qu'il en soit, ces estimations du temps évolutif demeurent précieuses malgré les incertitudes qui les entourent, en particulier pour les nombreux organismes pour lesquels on ne dispose d'aucun document paléontologique. La tendance actuelle est de considérer les arguments paléontologiques, géologiques, biogéographiques pour, en quelque sorte, étalonner l'horloge moléculaire.

### CONCLUSION

Indépendamment de la signification, adaptative ou non, des polymorphismes biochimiques, les alloenzymes constituent des marqueurs génétiques nucléaires qui restent infiniment précieux pour l'étude des populations et des espèces.

Les distances génétiques, dans certaines limites, pas trop faibles car elles sont entachées d'une erreur importante, pas trop élevées car elles perdent toute signification, restent l'un des moyens les plus commodes pour la quantification de la divergence entre des entités différentes. Dans ces limites, la plupart des indices de distances donnent des estimations compatibles.

L'un des aboutissements de l'approche biochimique est l'établissement de phylogénies. De nombreuses méthodes de construction d'arbres ou de réseaux ont été proposées et les discussions sur leurs avantages ou inconvénients, compte tenu de la philosophie qui les sous-tend, font l'objet de nombreuses publications dans la littérature internationale. Ceci d'autant plus qu'elles sont appliquées aux résultats des études sur le DNA, mitochondrial ou nucléaire, en plein essor actuellement. Ces nouvelles techniques moléculaires constituent, sans aucun doute, des outils puissants pour analyser la différenciation des espèces et établir les phylogénies. Elles soulèvent cependant quelquefois des problèmes inattendus. Comment, en effet, expliquer que les deux espèces *Drosophila yakuba* et *D. teissieri* du sous-groupe *melanogaster* soient identiques par leur DNA mitochondrial, ce qui suggère une divergence récente (Solignac *et al.*, 1986), alors que leur différenciation pour l'ensemble du génome nucléaire (caractères morphologiques, chromosomiques, biochimiques, etc.) indiquerait au contraire une divergence ancienne.

Ce que je crois, c'est qu'il ne faut négliger ou privilégier aucune approche particulière. Il est au contraire tout à fait nécessaire de confronter les approches les plus diversifiées dans le but de dégager un consensus susceptible de rendre compte de la plupart des résultats. A cet égard, l'étude consacrée récemment à la phylogénie du Panda (O'Brien *et al.*, 1985) me paraît tout à fait exemplaire.

### INDEX BIBLIOGRAPHIQUE

- ANDERSON, L. and ANDERSON, NG., 1977. — High resolution two dimensional electrophoresis of human plasma proteins. *Proc. Nat. Acad. Sci. U.S.A.*, 74, pp. 5421-5425.
- AVISE, J.C and AQUADRO, CF., 1983. — A comparative summary of genetic distances in the Vertebrates. *Evol. Biol.*, 15, pp. 151-185.

- BAHN, E., 1967. — Crossing over in the chromosomal region determining amylase isozymes in *Drosophila melanogaster*. *Hereditas*, 58, pp. 1-12.
- CARIOU, Mi., 1977. — Recherches sur le polymorphisme enzymatique du complexe *Jaera albifrons*, Leach (Crustacé, Isopode). Thèse Doct. d'Etat; Univ. Pierre et Marie Curie (Paris VI).
- CARIOU, M.L., 1985. — Distances génétiques et analyses multidimensionnelles : différenciation génétique des espèces du complexe *Jaera albifrons* (Crustacés, Isopodes). *Bull. Soc. Zool. Fr.*, vol. 42, pp. 281-295.
- CARIOU, MX. et LEFEBVRE, J., 1982. — Le miniordinateur, outil préférentiel pour le calcul des distances génétiques biochimiques chez le Crustacé *Jaera albifrons*. Actes X<sup>e</sup> Colloque «Informatique et Biosphère», Paris, pp. 110-114.
- CAPY, P. et CARIOU, MX., 1986. — Estimation de distances génétiques entre populations et entre espèces de drosophiles du groupe *obscura*. Actes Colloque "Distance divergence et variabilité génétique". Méribel 86, I.N.R.A. pp 23-32
- CARSON, HX., 1976. — Inference of the time of origin of some *Drosophila* species. *Nature*, 259, pp. 395-396.
- COOPER, D.W., JOHNSON, P.G., MURTAGH, CE, SHARMAN, G.B, VANDERBERG, JX. and POOLE, W.E., 1975. — Sex-linked Isozymes and sex chromosome evolution and interaction in Kangaroos. In : Isozymes, vol. III, Ed. Markert, C.L., Academic Press, London.
- DAMERVAL, c, de VIENNE, D., THIELLEMENT, H. and ZIVY, M., 1986. — Technical improvements in two-dimensional electrophoresis increase the level of genetic variation detected in wheat-siblings proteins. *Electrophoresis*, 7, pp. 52-54.
- DOANE, W.W. ABRAHAM, I., KOLAR, MM, MARTENSON, E. and DEIBLER, G.E. 1975. — Purified *Drosophila* a amylase Isozymes : genetical, biochemical and molecular characterization. In : Isozymes, vol. IV, Ed. Markert, C.L., Academic Press, London.
- DOANE, w.w., 1965. — Genetic control of amylase activity in *Drosophila melanogaster*. *Proc. XII. Int. Congr. Ent.*, London, 1964, p. 234.
- DOANE, w.w., 1969. — *Drosophila* amylases and problems in cellular differentiation. In : RNA in development ed. Hanly, E.W., University of Utah Press, Salt Lake City, Utah, pp. 73-109
- FARRIS, J.S., 1972. — Estimate phylogenetic trees from distance matrices. *Am. Nat.*, 106, pp. 645-668.
- FITCH, WM, and MARGOLIASH, E., 1967. — Construction of phylogenetic trees. *Science*, 155, pp. 279-284.
- GRASSLE, JP. and GRASSLE, J.F., 1976. — Sibling species in the marine pollution indicator *Capitella* (Polychaeta). *Science*, 192, pp. 567-569.
- GREGORIUS, HR, 1984. — An unique genetic distance. *Biometric Journal*, 26, pp. 13-18.
- HEDGECOCK, D, TRACEY, MX and NELSON, K, 1982. — Genetics. *The Biology of Crustacea*, vol. 2, Academic Press, pp. 283-403.
- KATZ, M., 1986. — Etude de quelques indices de distances génétiques et de leurs estimateurs. Thèse Doc, Univ. Paris VII.
- LEFEBVRE, J., 1976. — Introduction aux analyses statistiques multidimensionnelles. Ed. Masson, Paris.
- MAYR, E., 1963. — *Animal species and Evolution*, Harwad University Press.
- MAYR, E., 1969. — *Principles of Systematic Zoology*, McGraw-Hill, London.
- NEI, M, 1972. — Genetic distance between populations. *Am. Nat.*, 106, pp. 283-292.
- NEI, M, 1975. — Molecular population genetics and Evolution. North Holland, Amsterdam.
- NEI, M, 1978. — Estimation of average heterozygosity and genetic distance from a small number of individuals. *Genetics*, 89, pp. 583-590.
- OBRIEN, SJ., NASH, w.G, WILDT, DE, BUSH, M.E. and BENVENISTE, RJE, 1985. — A molecular solution to the riddle of the giant panda's phylogeny. *Nature*, 317, pp. 140-144.
- OFARRELL, PH, 1975. — High resolution two-dimensional electrophoresis of proteins. *J. Biol. Chem.*, 250, pp. 4007-4021.
- RACINE, R.R. and LANGLEY, C.H., 1980. — Genetic heterozygosity in a natural population of *Mus musculus* assessed using wo-dimensional electrophoresis. *Nature*, 282, pp. 855-857.
- RAMSHAW, J.A.M. COYNE, JA. and LEWONTIN, RC., 1979. — The sensitivity of gel electrophoresis as a detector of genetic variation. *Genetics*, 93, pp. 1019-1037.
- SARICH, 1977. — Rates, sample sizes and the neutrality hypothesis for electrophoresis in evolutionary studies. *Nature*, 265, pp. 24-28.
- SELANDER, RK. and WHITAM, T.S., 1983. — Protein polymorphism and the genetic structure of populations. In : Evolution of Genes and Proteins. Sinauer Associates INC., pp. 89-114.
- SIMPSON, GG, 1944. — Tempo and mode in evolution. Columbia, Univ. Press New York.

- SINGH, R.S., LEWONTIN, R.C. and PELTON, A.A., 1976. — Genetic heterogeneity within electrophoretic «alleles» of xanthine dehydrogenase in *Drosophila pseudoobscura*. *Genetics*, 84, pp. 609-629.
- SNEATH, P.H.A. and SOKAL, R.R., 1973. — Numerical Taxonomy. W.H. Freeman and Co. Ed. San Francisco.
- SOLIGNAC, M., MONNEROT, M. and MOUNOLOU, J.C., 1986. — Mitochondrial DNA evolution in the *melanogaster* species subgroup of *Drosophila*. *J. Mol. Evol.*, 23, pp. 31-40.
- TARROUX, p., RABILLOU, T., PENNETIER, J.L. et GODEFROY, o., 1985. — Electrophorese bidimensionnelle et informatique. *Bull. Soc. Zool. Fr.*, 42, pp. 181-203.
- TAIENO, Y., NEI, M. and TAJIMA, F., 1982. — Accuracy of estimated phylogenetic trees from molecular data. I. Distantly related species. *J. Mol. Evol.*, 18, pp. 387-404.
- WILSON, A.C., CARLSON, S.S. and WHITE, T.J., 1977. — Biochemical Evolution. *Annu. Rev. Biochem.*, 46, pp. 573-639.
- YANG, S.Y., SOULE, M. and GORMAN, G.C., 1974. — *Anolis* lizards of the Eastern Carribean : A case study in Evolution. I. Genetic relationships, phylogeny, and colonization sequence of the *roquet* group.