

Principes et analyse des données de l'électrophorèse des protéines enzymatiques.

Jean François Agnès

ORSTOM, Centre de Recherches Océanologiques, BP V18, Abidjan, Côte d'Ivoire.

Principes généraux

L'électrophorèse, qu'elle soit appliquée aux protéines enzymatiques ou non, permet de séparer des molécules chargées en solution sous l'action d'un champ électrique. Les protéines sont des molécules constituées d'une ou plusieurs chaînes polypeptidiques, elles-même formées d'une succession d'acides aminés. Certains acides aminés possèdent un radical qui peut s'ioniser et être chargé électriquement. La charge électrique (+/-) de chaque acide aminé dépend du pH du milieu qui l'environne. Chaque acide aminé est caractérisé par un pH isoélectrique (pHi), qui est le pH pour lequel sa charge est nulle. En deça de ce pH, l'acide aminé est chargé positivement, au delà, il est chargé négativement. L'ensemble de ces acides aminés donne à la protéine en solution une certaine charge électrique qui provoque sa migration lorsqu'elle est soumise à un champ électrique. De la même manière que précédemment, la protéine est caractérisée par un pHi qui est la résultante de tous les pHi des acides aminés qui la composent. Chaque acide aminé est codé par une séquence de trois nucléotides au sein de la molécule d'ADN qui est tout d'abord transcrite en ARN avant d'être traduite (figure 1). Une mutation (changement d'un nucléotide par un autre, délétion, insertion, etc...) intervenant sur la molécule d'ADN, sera répercutée au niveau de l'ARN messenger et l'acide aminé résultant de la traduction sera changé (exception faite des mutations dites silencieuses qui ne changent pas le sens du codon). Si ce nouvel acide aminé a une charge différente de celui qu'il remplace, la protéine aura alors une charge globale différente et ses potentialités de migration dans un champ électrique seront modifiées. Cette nouvelle protéine pourra être distinguée de la précédente sur le gel d'électrophorèse car elle ne se trouvera pas au même niveau à l'issue de la migration.

De cette manière, en accédant à la variabilité des protéines, nous avons également accès à celle de l'ADN, donc du génome. Il faut cependant émettre quelques restrictions: une substitution de nucléotide n'entraîne pas forcément un changement d'acide aminé (plusieurs triplets de nucléotides codent pour le même acide aminé), un changement d'acide aminé n'entraîne pas forcément un changement de charge (plusieurs acides aminés ont la même charge). D'une manière générale, on admet que le polymorphisme révélé par l'électrophorèse enzymatique représente environ le tiers du polymorphisme de la molécule protéique.

Lorsque l'on fait migrer l'ensemble des protéines d'un organe, muscle, foie, dans un gel qui sert de support à la migration, comment rendre visible une seule d'entre elles sans du même coup visualiser l'ensemble des protéines en solution dans l'homogénat étudié?

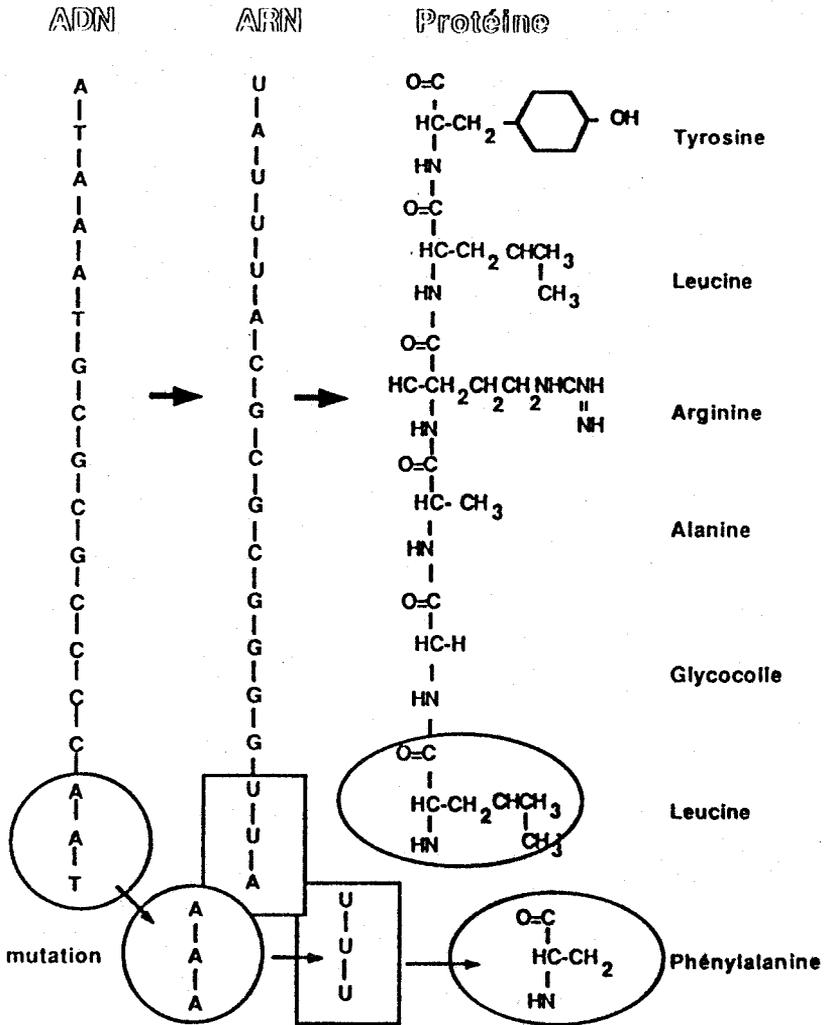


Figure 1. Schéma montrant le codage d'une protéine à partir d'un brin d'ADN et les implications d'une mutation ponctuelle sur ce brin.

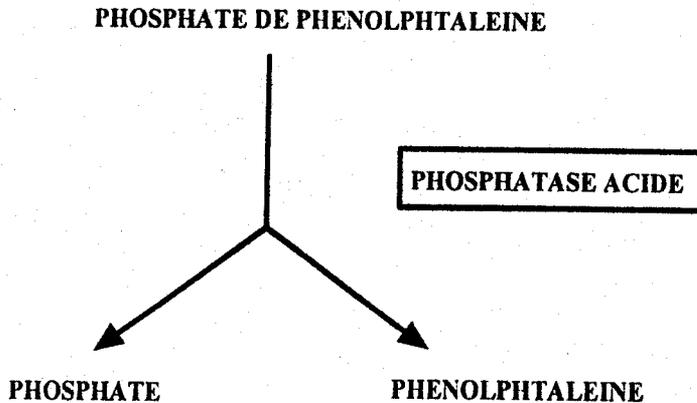


Figure 2. Schéma de la réaction catalysée par la phosphatase acide. La phenolphtaleine est colorée et permet de voir où la réaction a eu lieu dans le gel.

On utilise pour cela les propriétés des protéines ayant une fonction enzymatique (les enzymes sont des catalyseurs biologiques grâce auxquels peuvent s'effectuer les réactions chimiques indispensables au fonctionnement de la cellule). Les enzymes, transforment leur substrat (un alcool, un acide, un lipide, une autre protéine, etc...) à l'endroit même où elles se trouvent dans le gel. Si un des composés de la réaction enzymatique est coloré ou colorable, on met ainsi en évidence l'endroit où ont migré les protéines.

La technique de révélation consistera à faire tremper le gel dans une solution contenant le substrat de l'enzyme que l'on veut révéler pour la faire fonctionner. On ajoutera également des molécules susceptibles de se colorer grâce et à l'endroit même de cette réaction (figure 2).

Le déroulement schématique d'une séance d'électrophorèse des protéines enzymatiques s'établit comme suit (Figure 3): (1) prélèvement d'un morceau de foie et de muscle, (2) broyage et centrifugation afin d'obtenir des extraits liquides, (3) conservation au congélateur (-30 C) ou dans l'azote liquide, (4) et (5) décongélation pour exécuter un prélèvement puis recongélation, (5) prélèvement d'une petite quantité d'extrait par imbibition d'un papier filtre, (6) disposition des papiers filtres dans le gel d'amidon, (7) mise sous tension du gel entraînant la migration de toutes les protéines, (8) la tranche d'amidon est plongée dans une solution contenant les réactifs permettant à un enzyme donné de fonctionner, ce fonctionnement devant provoquer sa visualisation sur le gel, (9) interprétation du gel.

L'électrophorèse des protéines enzymatiques permet d'analyser un grand nombre d'individus ainsi que plusieurs dizaines d'enzymes pour chaque individu. Elle reste une méthode simple, rapide, et relativement peu onéreuse de la génétique biochimique des populations.

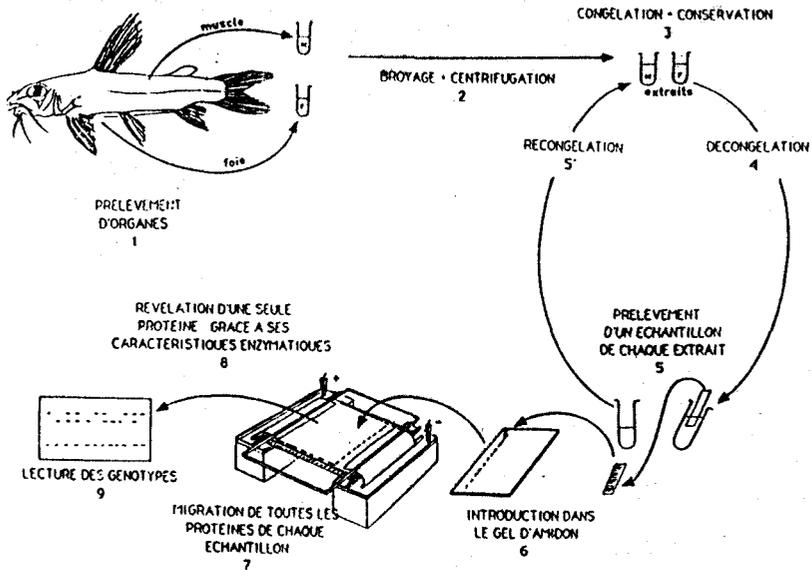


Figure 3. Déroulement schématique d'une séance d'électrophorèse des protéines enzymatiques, (1) prélèvement d'un morceau de foie et de muscle, (2) broyage et centrifugation afin d'obtenir des extraits liquides, (3) conservation au congélateur (-30 C) ou dans l'azote liquide, (4) et (5) décongélation pour exécuter un prélèvement puis recongélation, (5) prélèvement d'une petite quantité d'extrait par imbibition d'un papier filtre, (6) disposition des papiers filtres dans le gel d'amidon, (7) mise sous tension du gel entraînant la migration de toutes les protéines, (8) la tranche d'amidon est plongée dans une solution contenant les réactifs permettant à un enzyme donné de fonctionner, ce fonctionnement devant provoquer sa visualisation sur le gel, (9) interprétation du gel.

L'étude du polymorphisme enzymatique

L'électrophorèse nous fournit des résultats sous forme de génotypes, pour chaque individu et pour chaque enzyme étudié. Chaque individu diploïde, possède deux copies de chaque gène (deux allèles), chacune lui provenant d'un de ses parents. Si ces deux copies sont identiques, on dira que l'individu est homozygote pour l'allèle en question. Si elles sont différentes (à cause d'une mutation ayant affecté une des copies), on dira que l'individu est hétérozygote. Le premier travail consistera à transformer ces données génotypiques en fréquences alléliques caractéristiques de la population à laquelle appartiennent les individus étudiés. Ensuite, on calcule ce que devraient être les proportions des différents génotypes dans la population que l'échantillon représente.

Pour cela, on dispose d'un théorème ou "loi de Hardy-Weinberg" qui stipule que si dans une population deux allèles A et B sont présents à des fréquences respectives p et q ($q=1-p$), alors, la fréquence théorique des génotypes homozygotes sera p^2 (AA) et q^2 (BB), et la fréquence théorique du génotype hétérozygote sera $2pq$ (AB).

Dans le cas où les proportions observées seraient significativement différentes des proportions attendues, il est nécessaire de vérifier si l'on a pas fait d'erreur dans l'interprétation des zymogrammes ou dans l'échantillonnage (présence de plusieurs espèces dans le même échantillon), puis de tenter d'en expliquer la cause (absence de panmixie, effets sélectifs).

On peut ensuite calculer des paramètres statistiques qui caractériseront chaque population, ou encore les populations les unes par rapport aux autres. Dans le premier cas, il s'agira de mesurer la variabilité génétique, dans le second, la divergence génétique.

MESURE DE LA VARIABILITE GENETIQUE

On dispose pour cela de trois indices: a) Le taux de polymorphisme P qui correspond au nombre de locus polymorphes par rapport au nombre de locus étudiés. On peut considérer qu'un locus est polymorphe lorsque l'allèle le plus commun a une fréquence inférieure ou égale à 99% ($P_{99\%}$), ou à 95% ($P_{95\%}$); b) La diversité allélique A , est le nombre moyen d'allèles par locus; c) la diversité génétique moyenne H_i (encore appelée hétérozygotie théorique), concerne les locus au sein de chaque population. H_i est fonction du nombre d'allèles présents à chaque locus et de leurs fréquences.

MESURE DE LA DIVERGENCE GENETIQUE

Toute une gamme d'indices permet d'estimer les distances génétiques entre populations ou taxons (Cavalli-Sforza et Edwards, 1967; Farris, 1972; Felsenstein, 1973; Fitch, 1971; Nei, 1972, 1978; Sneath et Sokal, 1973; Autem et Bonhomme, 1980). Chaque indice possède des propriétés biologiques et mathématiques qui lui sont propres.

Cette divergence génétique étant mesurée, il va falloir traduire les relations entre les distributions d'allèles, les fréquences alléliques, ou les génotypes d'une part et les taxons (genres, espèces ou populations conspécifiques) d'autre part grâce à des analyses de type phylogénétique.

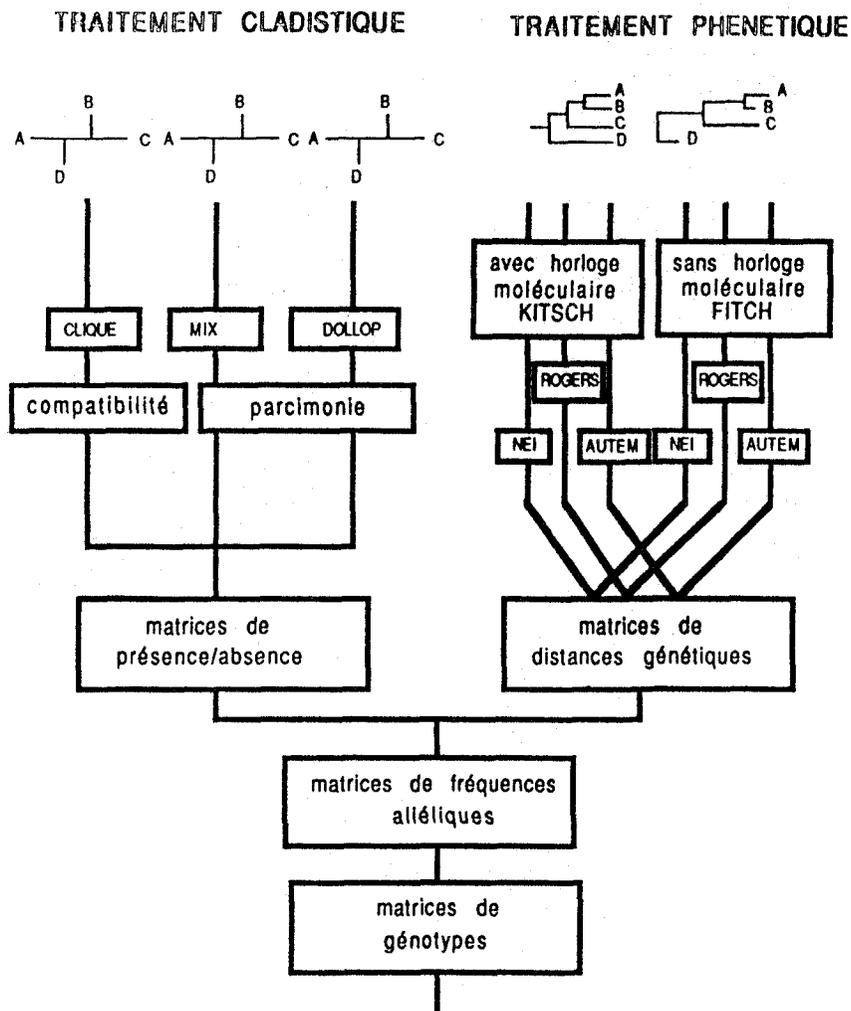


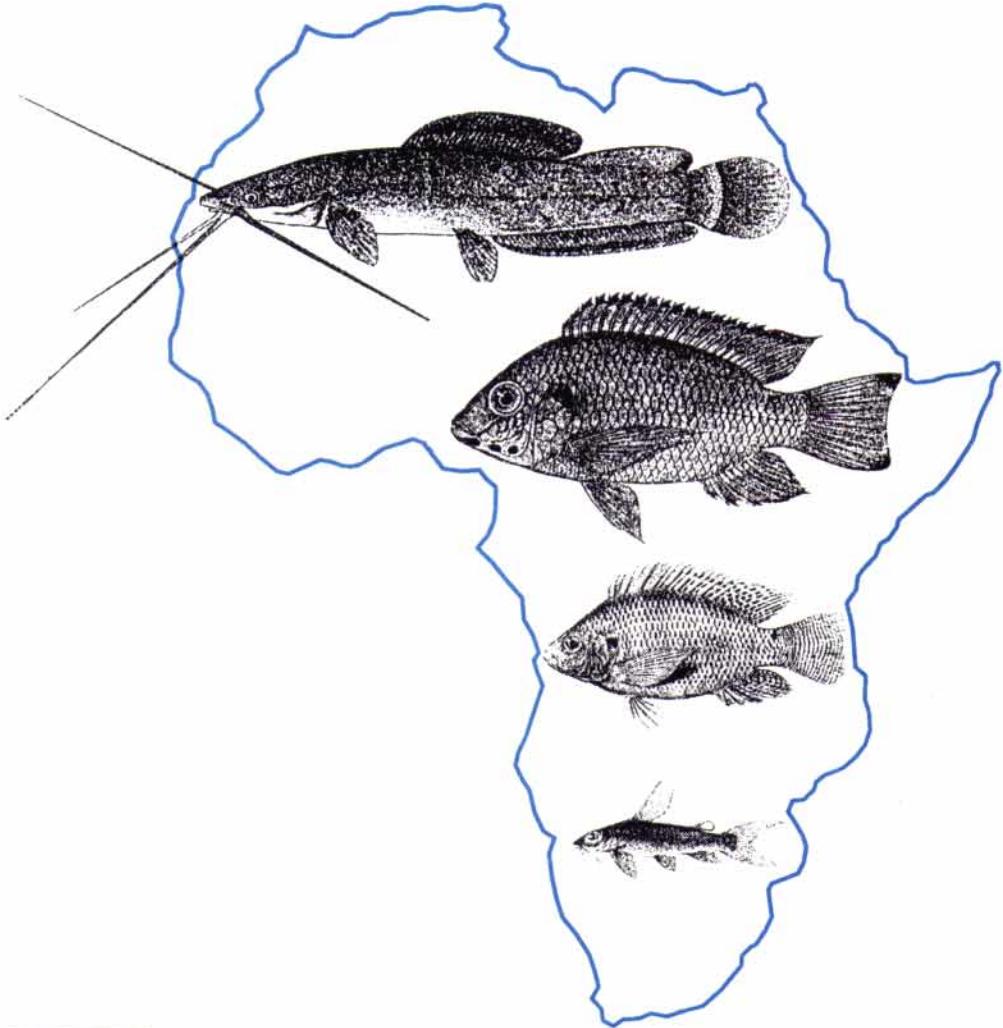
Figure 4. Les données de l'étude du polymorphisme enzymatique sont recueillies sous forme de génotypes. Ces matrices de génotypes sont transformées en matrices de fréquences alléliques qui sont caractéristiques de chaque échantillon. Ces matrices peuvent à leur tour être transformées en matrices de distances génétiques pour un traitement phénétique, ou en matrices de présence/absence d'allèles (0/1) pour un traitement cladistique. Pour le traitement phénétique, les matrices de distances génétiques (indices de Nei, 1972; Rogers, 1972; ou de Autem et Bonhomme, 1980) peuvent être analysées en supposant l'existence d'une horloge moléculaire (programme KITSCH), ou l'inverse (programme FITCH). Dans le dernier cas, on obtient un dendrogramme non raciné. En ce qui concerne le traitement cladistique, les matrices de présence/absence d'allèles peuvent être traitées selon le principe de compatibilité (programme CLIQUE) ou de parcimonie (programme MIX et DOLLOP). Dans les deux cas, on obtient un réseau.

Ce type d'analyses (fig. 4) permet non seulement de classer les objets analysés en fonction de leurs ressemblances, mais aussi et surtout en fonction de leurs liens de parenté, et ainsi de rendre compte des évènements évolutifs au sein des familles, genres ou espèces. En pratique, les dendrogrammes ou arbre phylogénétiques ou encore réseaux phylogénétiques proposés sont construits soit à partir d'algorithmes de classification hiérarchique appliqués à des matrices de distances génétiques (traitement phénétique) soit à partir d'autres techniques utilisant les données brutes sous forme de fréquences géniques ou présence/absence d'allèles (traitement cladistique).

ATELIER

BIODIVERSITÉ ET AQUACULTURE EN AFRIQUE

ABIDJAN 21/25 NOVEMBRE 1994



cro
CENTRE DE RECHERCHES
OCÉANOGRAPHIQUES
ABIDJAN



UNION EUROPÉENNE

ORSTOM

INSTITUT FRANÇAIS DE RECHERCHE SCIENTIFIQUE
POUR LE DÉVELOPPEMENT EN COOPÉRATION