

TECHNIQUES D'ETUDE MOLECULAIRE DES ISOLATS : PRINCIPES ET FIABILITE*

H-J. Boulouis¹, Nadia Haddad² et R. Maillard³

RESUME : Les techniques de biologie moléculaire présentent l'avantage de permettre l'étude du génome indépendamment de la culture des microorganismes. Elles s'appuient sur quatre outils de base : des enzymes (polymérase, enzymes de restriction), l'hybridation moléculaire, l'électrophorèse et le séquençage. La combinaison de ces outils conduit à des techniques autorisant soit l'approche globale du génome (électrophorèse en champ pulsé, amplification en chaîne aléatoire,...) soit l'étude de séquences spécifiques (amplification en chaîne, RFLP et southern blotting, séquençage et puce à ADN,...) des clones. Les propriétés de ces techniques, dépendant aussi du matériel biologique étudié, assignent leur emploi soit à une comparaison locale d'isolats soit à une étude plus globale (spatio-temporelle) de souches. Elles constituent pour nombre d'entre elles un outil remarquable de diagnose et de typage, en complément du phénotypage. Leur apport à l'épidémiologie est donc considérable, mais elles présentent aussi diverses limites qu'il convient de connaître au mieux pour adapter cet outil à l'objectif recherché.

SUMMARY : Molecular biology techniques can analyse the genome without needing a cell culture of the targeted microorganisms. These techniques are based on four principles : enzymes (polymerases, restriction enzymes...), molecular hybridisation, electrophoresis and sequencing. These tools can either target the whole genome (pulsed field electrophoresis, RAPD,...) or specific DNA sequences (PCR, Southern blotting, RFLP, micro-arrays...) from the clones. Depending on the biological material these techniques can compare different strains or evaluate time- and space variations for the same strain. These techniques give information for the diagnostic and strain typing complementing informations given by phenotypical studies. These techniques bring a new dimension for epidemiological studies but technical limitations should be assessed to adapt these tools to individual scientific goals.



Les techniques d'études des isolats se sont considérablement développées et diversifiées avec la mise à portée de nombreux laboratoires des outils de la biologie moléculaire. Leur diversité a notamment résulté de la multiplicité des outils disponibles et de la nécessité d'adapter les techniques existantes aux particularités génétiques des micro-organismes étudiés. Elles présentent en effet l'avantage de permettre l'étude du génome et non pas de ses produits d'expression et sont donc, contrairement au phénotypage, indépendantes des conditions de culture et, parfois, de la culture même des micro-organismes.

Cependant, les évidents avantages de ces techniques ne doivent pas en masquer les limites qui restreignent parfois leur utilisation à certains germes ou à certains domaines d'études épidémiologiques.

Le texte qui suit a pour objectif de présenter dans leurs grandes lignes les principales techniques de biologie moléculaire mises en œuvre en épidémiologie et ne saurait être exhaustif. En particulier, nombre d'exemples d'applications se retrouveront au fil des autres textes de ce volume.

* Texte de l'exposé présenté lors de la Journée AEEMA, 17 mai 2001

1 ENVA – Microbiologie, 94704 Maisons-Alfort, France

2 ENVA – UPMC, 94704 Maisons-Alfort, France

3 ENVA – Pathologie du bétail, 94704 Maisons-Alfort, France

I - RAPPELS SUR LES OUTILS DE LA BIOLOGIE MOLECULAIRE

Les techniques de biologie moléculaire reposent sur l'emploi de quatre outils de base :

- Des enzymes telles que:
 - Polymérase capables de copier une molécule d'acide nucléique à partir d'une matrice monobrin (ADN ou ARN) ;
 - Enzymes de restriction qui reconnaissent spécifiquement des séquences du génome et clivent ce génome. La séquence et la taille des sites de reconnaissance (de quelques bases à plusieurs dizaines) définissent le nombre de sites potentiels de coupure pour un germe donné : enzymes à sites fréquents de coupure et enzymes à sites peu fréquents. La séquence reconnue est spécifique d'une enzyme. De ce fait, toute modification d'une base d'un site de reconnaissance se traduit par la perte du site de reconnaissance et de coupure de cette enzyme.

- L'hybridation qui permet l'appariement de molécules d'acides nucléiques monobrin. Cette hybridation repose sur la complémentarité des brins qui peut être totale ou partielle. Dans ce dernier cas, les conditions d'hybridation (stringence, température) sont essentielles pour obtenir l'hybridation.
- L'électrophorèse en gel : elle permet la séparation des molécules d'acides nucléiques en fonction de leur poids moléculaire (et donc de leur taille), la charge de ces molécules étant uniforme.
- Le séquençage des acides nucléiques.

Ces techniques sont parfois employées seules ou, le plus souvent, combinées entre elles pour étudier tout ou partie du génome des germes à comparer. Selon qu'est mise en jeu une reconnaissance spécifique de séquence cible (par hybridation par exemple) ou non, on peut regrouper ces techniques en deux catégories, celles qui permettent une étude globale du génome, et celles qui s'intéressent à l'étude de séquences spécifiques.

II - ETUDE GLOBALE DU GENOME

Les principales approches permettant d'identifier des clones en s'appuyant sur des techniques d'étude de séquences non spécifiques sont les techniques séparatives après restriction enzymatique et les techniques d'amplification aléatoire.

1. TECHNIQUES SEPARATIVES APRES RESTRICTION ENZYMATIQUE

Il s'agit de l'électrophorèse en champ pulsé et des techniques de RFLP (Restriction Fragment Length Polymorphism) toutes deux fondées sur la coupure du génome par des enzymes de restriction, suivie de la séparation par électrophorèse des fragments obtenus. Ces deux catégories de méthodes diffèrent par le type de coupure (macrorestriction ou microrestriction) et par le procédé d'électrophorèse.

1.1. ELECTROPHORESE EN CHAMP PULSE (ECP)

L'électrophorèse en champ pulsé est le résultat de la combinaison d'une digestion par des enzymes de restriction à faible nombre de sites de coupure et d'une électrophorèse adaptée à la grande taille des produits de digestion. Le résultat est un profil de restriction intéressant tout le génome du microorganisme.

La réalisation pratique s'effectue en trois étapes :

- multiplication des germes à étudier, la quantité d'ADN nécessaire étant élevée ;

- extraction de l'ADN. Cette étape est la plus délicate car le génome ne doit pas être accidentellement fragmenté. Elle s'effectue donc après incorporation des germes dans une gélose. C'est dans cette gélose que l'ADN est libéré chimiquement puis digéré par une ou deux enzymes de restriction. Le choix des enzymes à utiliser est fonction du micro-organisme étudié : les enzymes doivent couper le génome en un nombre de fragments suffisamment grand pour être discriminant et suffisamment réduit pour que le profil soit exploitable ;
- électrophorèse : les fragments d'ADN obtenus ont en général une taille comprise entre 10 kb et 800 kb. La séparation nécessite l'emploi d'une technique particulière d'électrophorèse dite en champ pulsé : elle utilise deux champs électriques activés alternativement et situés perpendiculairement l'un par rapport à l'autre. Sous l'effet de ces champs alternatifs, les molécules d'ADN progressent dans l'agar plus rapidement que dans un champ linéaire et sont séparées selon leur taille (figures 1 et 2).

Le résultat, obtenu en 24 à 48 heures après révélation de l'ADN, est un profil de macro-restriction définissant un pulsotype.

FIGURE 1
Principe de l'électrophorèse en champ pulsé

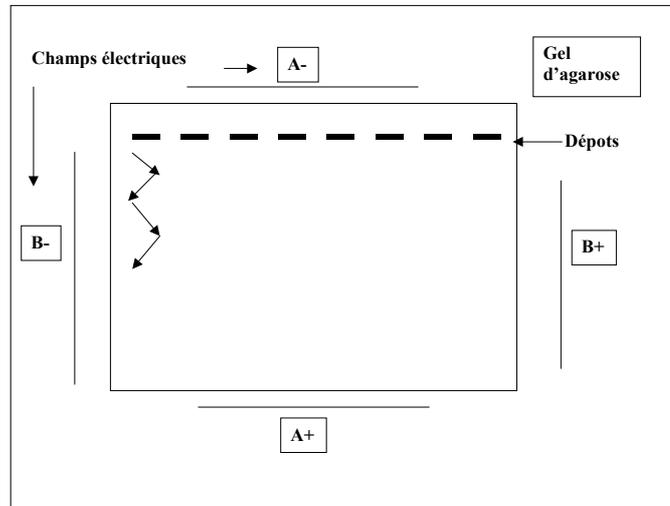
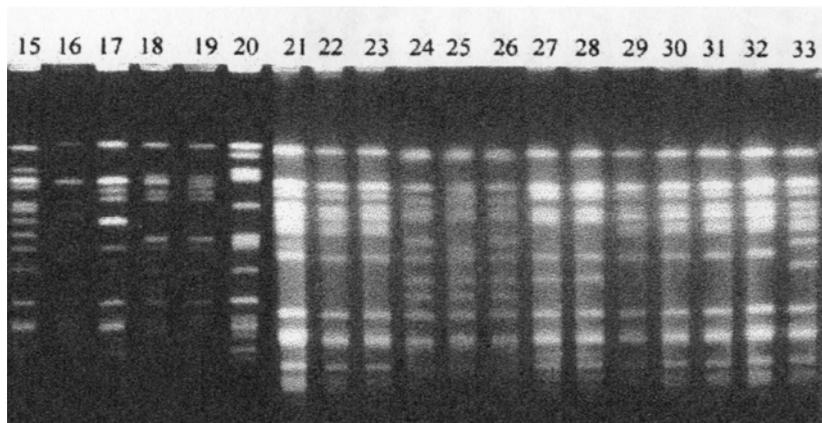


FIGURE 2
Exemple de pulsotypes obtenus avec différentes souches de *Bartonella henselae* [In Maruyama S. et al.]



L'interprétation repose sur le nombre et la taille des fragments, et obéit à différents critères : l'analyse ne peut porter que sur des pulsotypes de plus de dix fragments et sur des espèces ayant une certaine diversité génétique naturelle. Dans ces conditions, la similitude des profils ou la différence sur une bande (mutation ponctuelle) est en faveur de l'origine clonale des souches étudiées. A l'opposé, trois différences ou plus (c'est-à-dire trois mutations ou plus) plaident en faveur de souches génotypiquement différentes.

L'intérêt majeur de cette technique réside dans le fait qu'elle permet d'étudier l'ensemble du génome et présente donc un pouvoir discriminant potentiellement important.

Cependant, elle demande encore un degré certain d'expertise et sa mise au point peut s'avérer longue. Elle est plus souvent employée pour la comparaison de souches au sein d'un laboratoire, compte tenu, parfois, de la faible reproductibilité inter laboratoires des résultats. Néanmoins, pour certains germes, des efforts de standardisation permettent une comparaison inter laboratoires (c'est le cas pour l'étude des *Listeria* – voir dans ce numéro l'article d'Anne Brisabois).

1.2. RFLP

La technique de RFLP est le résultat de la combinaison d'une digestion par des enzymes de restriction à nombre élevé de sites de coupure et d'une électrophorèse simple. En bactériologie, cette technique en tant que telle est abandonnée, le très

grand nombre de bandes rendant très difficile la différenciation des profils obtenus (cf. figure 3). En revanche, elle reste très utilisée lorsqu'elle est couplée à une autre technique de révélation (cf. III.2).

1.3. AMPLIFICATION ALEATOIRE

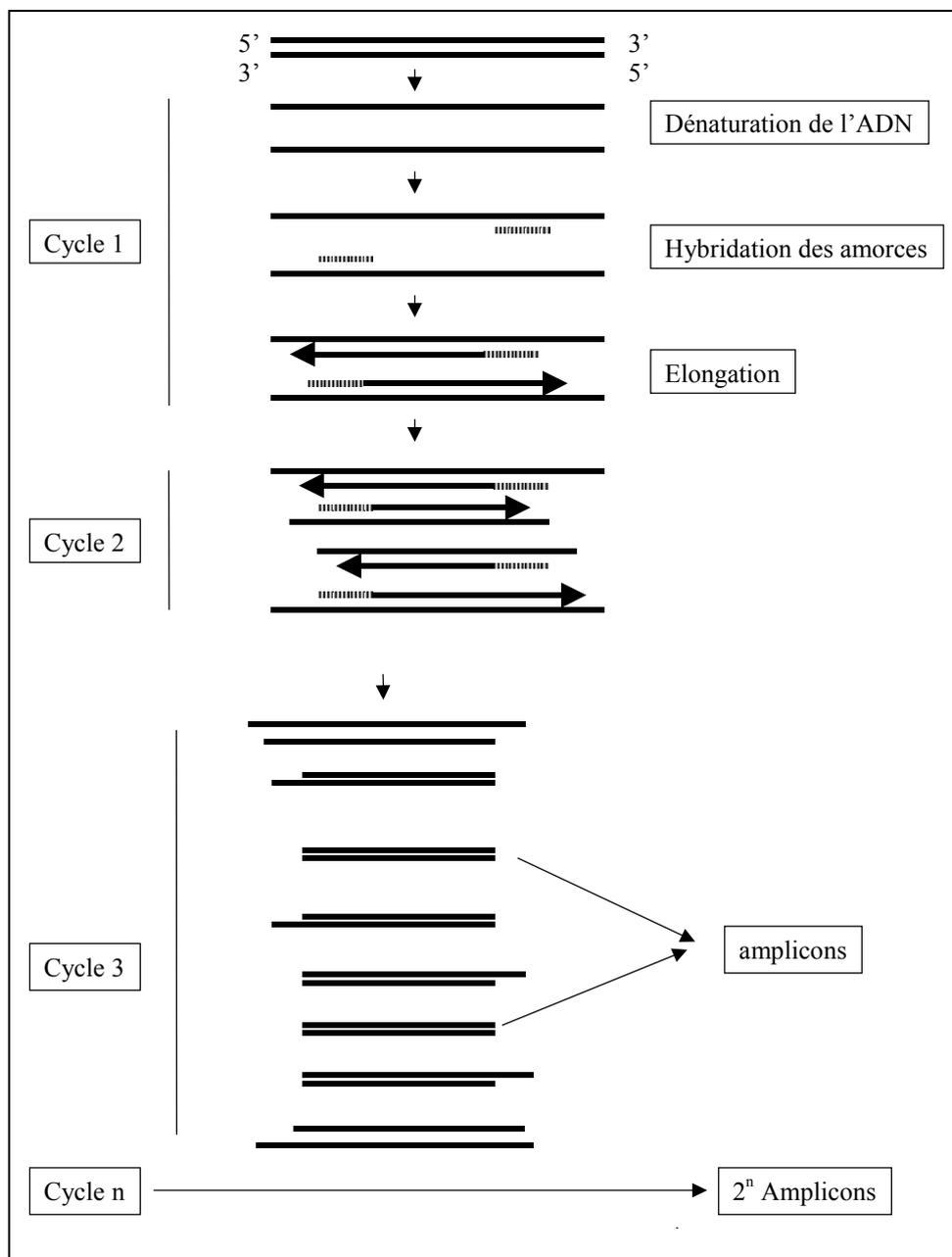
Ces techniques s'appuient sur le principe de l'amplification en chaîne (Cf. encadré : PCR). Dans les

techniques d'amplification aléatoire, (la ou) les amorces ne sont pas spécifiques d'un germe ni d'un gène. L'une des techniques employées en épidémiologie est la « Random Amplified Polymorphic DNA » (RAPD).

ENCADRE 1

Polymerase Chain Réaction (PCR/amplification en chaîne) [d'après Bogart et Lamoril.]

Cette technique consiste à amplifier une partie de génome, grâce à une polymérase et des amorces encadrant la région à amplifier. Le résultat de l'amplification est la production d'un amplicon/amplifiat en quantité telle qu'il est macroscopiquement visible. Cette technique est aussi applicable aux ARN : elle nécessite alors une étape intermédiaire de RT-PCR (Reverse transcriptase-PCR) permettant la transcription de l'ARN en ADN avant l'amplification de ce dernier.



Dans cette technique, les séquences des amorces utilisées sont choisies au hasard (taille variable -de 5 à plus de 30 bases- et séquence pouvant être définie par le GC% du génome étudié s'il est connu). La taille des amorces conditionne la taille des amplicons. Les températures d'hybridation, basses, permettent un accrochage non stringent de l'amorce. Lorsque l'amorce se fixe sur les deux brins complémentaires à

une distance de moins de 5000 bases on peut obtenir un amplicon. Cette technique est assez peu reproductible car les profils obtenus sont très dépendants des conditions opératoires de l'amplification (nature et concentration de la polymérase, températures, thermocycleur,...) et ne peut servir en général qu'à des comparaisons de souches au sein d'un laboratoire.

III - ETUDE DE SEQUENCES SPECIFIQUES

Les techniques s'attachant à mettre en évidence des séquences spécifiques sont très variées et parfois combinatoires. Il est donc impossible de donner des exemples pour chaque situation.

1. AMPLIFICATION EN CHAINE (POLYMERASE CHAIN REACTION/PCR)

L'essor de cette technique en microbiologie et en particulier en épidémiologie tient à son extrême adaptabilité du fait :

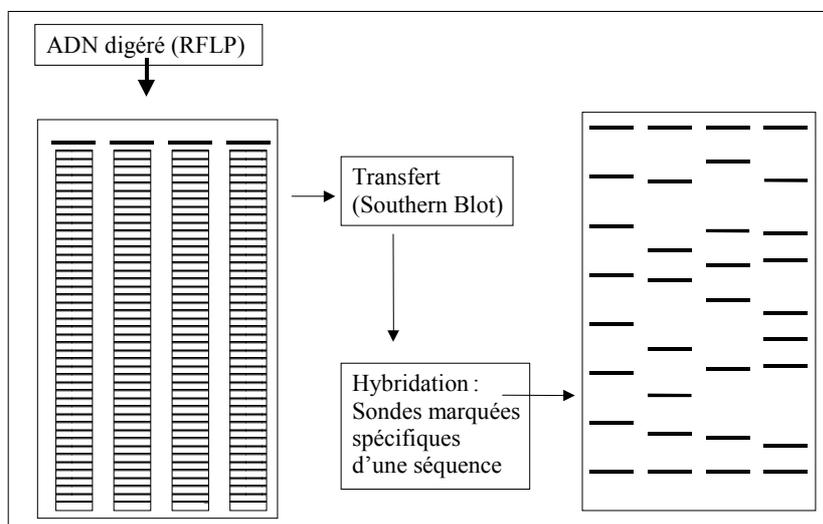
- de la nature des amorces : en fonction des objectifs, les amorces choisies pourront être, soit spécifiques d'un gène (ou portion de gène) propre à un micro-organisme, soit spécifiques d'un gène (ou portion de gène) universellement retrouvé mais dont certaines séquences sont très variables (par exemple gène codant pour l'ARN ribosomal 16S des bactéries) ; elles peuvent aussi être non spécifiques (comme dans le cas de la RAPD, cf. supra). Selon le cas, il s'agira donc d'un outil de diagnose (famille, genre, espèce) ou de typage ;
- du traitement de l'amplicon/amplifiat :

- électrophorèse simple de l'amplicon,
- digestion par enzyme de restriction : profil de micro-restriction (RFLP),
- PCR nichée (emboîtée/ Nested PCR) : deuxième PCR utilisant l'amplicon initial comme matrice,
- séquençage.

2. DIGESTION PUIS HYBRIDATION (RFLP AVEC SOUTHERN BLOTTING)

La combinaison de ces deux techniques s'appuie sur le principe suivant : le génome est digéré par une enzyme de restriction à nombreux sites de coupure. La localisation de la/des séquence(s) d'intérêt s'effectue, après séparation préparatoire électrophorétique, par hybridation avec une sonde spécifique marquée. L'objectif est la différenciation des isolats en fonction de plusieurs critères : présence/absence de la séquence cible, nombre variable de copies, taille du/des segment(s) de restriction qui porte(nt) cette séquence (figure 3).

FIGURE 3
RFLP et Southern blotting



3. SÉQUENÇAGE

Toutes les méthodes précédentes permettent d'approcher la structure primaire du génome. Le séquençage, qui fournit cette structure, n'est plus une technique du ressort de laboratoires fondamentaux ou de recherche. Il devient, grâce à son automatisation, à la portée de laboratoires de routine (soit sous forme d'automate de séquençage soit sous forme de prestation de services) et peut s'avérer moins onéreux que certaines techniques décrites plus haut. Les techniques de séquençage, du fait de leur accessibilité, sont amenées dans un avenir proche à supplanter la plupart des autres techniques.

Une fois la/les séquences établie(s), il est possible de construire un dendrogramme ou un arbre phylogénique qui matérialise la comparaison de souches.

4. PUCE A ADN

Dans ce domaine, les puces à ADN semblent être le point ultime de l'évolution des techniques de biologie moléculaire appliquées à la microbiologie.

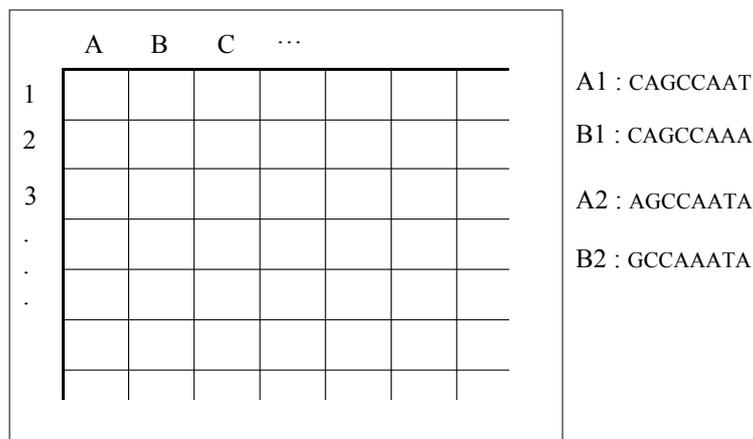
Le principe des puces à ADN repose sur la miniaturisation de l'hybridation permettant d'identifier simultanément des milliers de molécules différentes d'acides nucléiques d'un micro-organisme.

Une puce à ADN est fabriquée par dépôt ou synthèse, sur un support, d'oligonucléotides de séquences différentes selon un plan préétabli (figure 4). Chaque oligo ou cible a donc une position connue sur la puce. Il est ainsi possible de disposer selon les techniques de fabrication, de 10^4 à 10^5 séquences par cm^2 . L'incubation de cette puce avec de l'ADN ou de l'ARN à étudier qui a été préalablement marqué, conduit dans des conditions de stringence élevée, à leur hybridation avec les cibles correspondantes. Cette hybridation est détectée et localisée. L'ensemble des signaux produits fait l'objet d'un traitement informatique. Il est donc possible d'obtenir des profils différents d'hybridation (reflétant des différences de séquences), voire la séquence d'un gène ou d'une portion de gène, en fonction des oligonucléotides hybridés. Le choix des oligonucléotides synthétisés sur la puce sera fonction du gène ou du génome étudié.

FIGURE 4

Principe des puces à ADN

A1, B1, A2, B2, ... : Zone de fixation sur la puce pour chacun des oligonucléotides susceptibles de s'hybrider avec des séquences homologues des isolats à étudier.



5. APPLICATIONS

Une sélection des principales applications de l'utilisation d'une technique ou de plusieurs techniques combinées est présentée dans le tableau I.

Trois illustrations parmi d'autres semblent intéressantes à développer pour leur caractère exemplaire.

□ Le séquençage automatique, qui inclut des étapes PCR, est sans doute la technique de choix en épidémiologie des infections virales, ne serait ce que

parce que le nombre de gènes et la taille du génome de ces agents infectieux est compatible techniquement et économiquement avec ce type d'approche (cf. exposé de Stéphan Zientara).

□ La combinaison d'une PCR et d'une électrophorèse, très employée en bactériologie pour l'étude des séquences VNTR (Variable Number Tandem Repeats) de différentes bactéries, des séquences ERIC (Enterobacterial repetitive intergenic consensus) des Entérobactéries ou des séquence IS (Séquence d'Insertion) des Mycobactéries [cf. exposé

de Nadia Haddad et B. Durand]. Ces séquences sont constituées d'unités de base monomorphes, simples ou répétées, uniques ou multiples dans le génome, insérées aléatoirement ou dans des sites privilégiés. Les amorces sont constituées de tout ou partie de ces séquences (dans le cas de séquences répétées inversées, une seule amorce suffit). L'amplification aboutit à des amplicons dont la taille ou le nombre sont variables d'un souche à l'autre. (Exemple : figure 5 et tableau II).

- L'association RFLP et Southern blotting pour le ribotypage des bactéries. Elle utilise des sondes spécifiques des gènes codant pour les ARN ribosomiaux (16S ou 23S) ou des séquences intergéniques. Le ribotypage est suffisamment reproductible pour avoir donné naissance à des banques de données internationales.

TABLEAU I

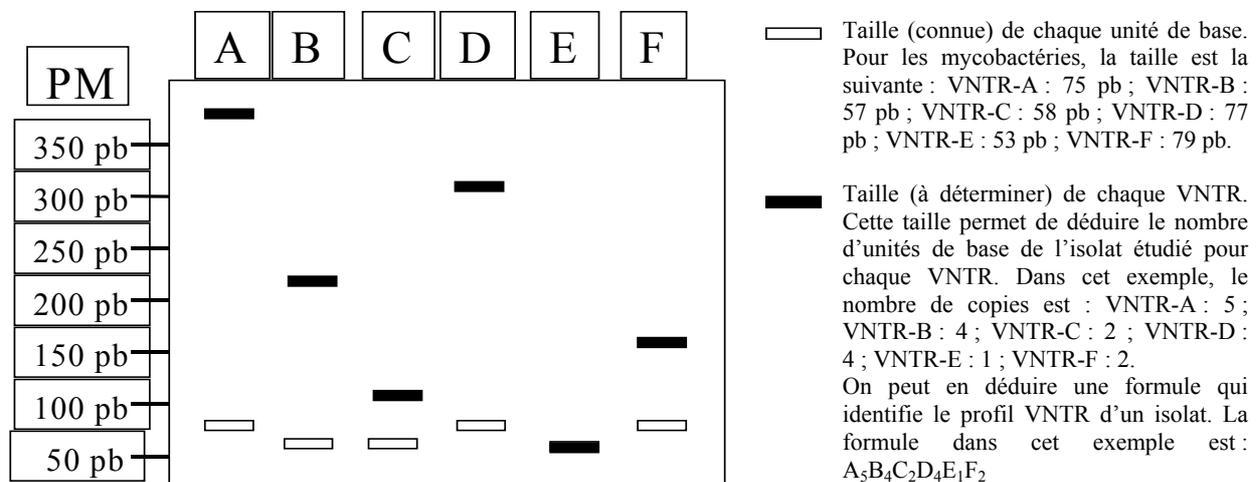
Applications utilisant un test ou la combinaison de plusieurs tests à des fins de typage moléculaire

X : sans objet
 ITS : séquence intergénique
 VNTR : Variable Number Tandem Repeat
 ARN16S : gène codant pour l'ARN ribosomal 16S
 PCR : Polymerase Chain Reaction
 RFLP : Restriction Fragment Length Polymorphism
 + à +++ : fréquence d'utilisation des tests ou de leur combinaison pour le typage moléculaire

1 ^{er} test \ 2 ^{ème} test	PCR	Electrophorèse	Hybridation
Electrophorèse	+++ (ARN16S, ITS, VNTR, ...)	X	X
Hybridation (RFLP + southern ou puce à ADN)	++ Confirmation par hybridation de la spécificité de la séquence amplifiée	+ Confirmation par hybridation de la spécificité de la séquence ayant la taille attendue	+++ Technique à part entière : (Très utilisée, notamment en bactériologie)
Séquençage	+++ (Surtout en virologie). Utilisation systématique croissante	+ Confirmation par hybridation de la spécificité de la séquence ayant la taille attendue	X

FIGURE 5

Typage des séquences VNTR chez les mycobactéries



IV - INTERET ET LIMITES DES TECHNIQUES DE BIOLOGIE MOLECULAIRE DANS LE TYPAGE D'ISOLATS

Les avantages de la biologie moléculaire pour le typage des souches sont évidents :

- toutes les souches d'une espèce peuvent faire l'objet d'un typage moléculaire (typabilité voisine de 100%), puisque ces techniques libèrent de l'inconstance de l'expression des gènes ;
- certaines techniques rendent possible l'analyse directe du prélèvement : cela ne dispense pas de l'isolement mais permet dans certains cas un gain de temps appréciable lorsque le germe étudié est à croissance lente ou d'isolement délicat ou les prélèvements difficilement exploitables pour la culture. Et les techniques sont utilisables même si les germes sont non revivifiables (sauf dans le cas de nécessité d'une grande quantité d'ADN : ECP ou RFLP) ;
- pour une espèce donnée, l'analyse de différents gènes peut fournir différents marqueurs possibles dont la combinaison accroît le pouvoir discriminant c'est-à-dire la capacité à distinguer comme différents des isolats qui le sont réellement (= sensibilité) ;
- certains marqueurs fournissent des résultats numérisables pouvant être rassemblés dans des banques de données accessibles.

Les limites du typage de souches par des techniques de biologie moléculaire tiennent :

- aux techniques employées, dont les performances doivent être évaluées en terme de :

- lisibilité et maniabilité des résultats,
- reproductibilité,
- pouvoir discriminant,
- coût du matériel et des réactifs,
- lourdeur de mise en œuvre ;

Ces limites tendent à assigner chaque technique à l'un des deux grands domaines de la comparaison de souches en épidémiologie (tableaux II et III) :

- la comparaison locale ou intra-laboratoire (techniques lourdes ou peu reproductibles) ;
- la comparaison inter-laboratoires lors d'études épidémiologiques spatio-temporelles ;
- au matériel biologique étudié (Horloge moléculaire et Variabilité subclonale).

Les germes font l'objet d'une microévolution dite subclonale qui dépend de la structure génétique du micro-organisme étudié et de sa plasticité génotypique/génomique. Cette variabilité subclonale peut se manifester au cours des repiquages (pouvant affecter la reproductibilité des résultats) ou au cours de l'infection. Lorsqu'elle est connue, cette plasticité peut conduire à choisir un marqueur ou un autre en fonction de ses propres objectifs. On peut être amené à écarter un marqueur connu pour sa variabilité trop importante : en absence de stabilité, le marqueur ne permettra pas d'affirmer l'identité clonale au cours d'une épidémiologie. Cependant, les études spatio-temporelles peuvent utiliser cette plasticité dans la mesure où elle peut être considérée comme une horloge moléculaire (tableau III).

TABLEAU II
Performances des marqueurs épidémiologiques

ECP : Electrophorèse en champs pulsés ; PCR : Polymerase Chain Reaction ; RFLP : Restriction Fragment Length Polymorphism.

¹ : Selon le type de germe.

² : Appliqué aux comparaisons d'isolats.

³ : Appliqué aux études globales.

Marqueurs	Typabilité	Pouvoir discriminant	Niveau analytique
ECP	Excellente	Variable ¹	Comparatif ²
Ribotypage	Excellente	Passable	Définitif ³
PCR+RFLP	Excellente	Passable	(Définitif) ¹
Olygotype	Excellente	Bon	Définitif
Séquençage	Excellente	Optimal	Définitif

TABLEAU III
Applications des techniques moléculaires de typage d'isolat

+/- et +++ : Importance du caractère considéré (polymorphisme et stabilité) en fonction de l'application.

Comparaison d'isolats	Etude plus globale d'isolats
<ul style="list-style-type: none">• Foyers (cheptel)• Transmission interspèces• Individus (double infection)• Laboratoire	<ul style="list-style-type: none">• Géographie génotypique (spatiale et/ou temporelle)<ul style="list-style-type: none">▪ dans un pays▪ entre pays• Liens phylogéniques• Nouvelles sous espèces
(Identité ou différence ?) Polymorphisme +++ Stabilité +/-	Polymorphisme +/- Stabilité +++

V - CONCLUSION

La biologie moléculaire offre à l'épidémiologiste de nombreux outils dont le pouvoir discriminant et la nouveauté ne doivent pas masquer les limites. Pour être idéal, un marqueur épidémiologique devrait présenter un polymorphisme et une stabilité « optimale », une mise en œuvre simple et économique et donner des résultats reproductibles. La biologie moléculaire ne fournit pas ce type de marqueur. Cependant, une analyse fine des besoins et des moyens doit permettre à l'épidémiologiste d'effectuer son choix en fonction

d'objectifs (comparaison intra ou inter-laboratoires,...) et/ou de critères pratiques (rapidité d'exécution, simplicité, lisibilité, coût, polyvalence, ...). Afin de minimiser ces limites, une association de techniques moléculaires semble raisonnable. En tout état de cause, ces techniques de génotypage ne remplacent pas celles de phénotypage qui restent performantes dans certains cas. Le choix de l'une ou l'autre des approches dépend de la nature du travail à effectuer.

VI - BIBLIOGRAPHIE

Bogard M. et Lamoril J. ~ Biologie Moléculaire en biologie clinique. I. Méthodes. Collection Option Bio. Elsevier, 1998

Freney J, Renaud F, Hansen W et Bollet C. ~ Précis de bactériologie clinique. Ed Eska/Alexandre Lacassagne, 2000.

Lee P.J. et Hudson T.J. ~ La puce à ADN en médecine et en science. *Médecine/Sciences*, 2000, **16** : 43-49.

Maruyama S., Kasten R., Boulouis H.J., Gurfield N., Katsube Y. et Chomel B. ~ Genomic diversity of Bartonella henselae isolates from domestic cats from Japan, The USA and France by Pulse field gel Electrophoresis. *Vet. Microbiol.*, 2000, **2074** : 1-13.

