

## De la génétique des populations bactériennes à l'épidémiologie des maladies infectieuses

Erick Denamur  
Bertrand Picard

Les notions acquises sur *Escherichia coli* sont trop souvent généralisées et ne reflètent pas la réalité de l'extrême diversité de l'ensemble du monde bactérien. La détermination de la structure génétique d'une population bactérienne semble être un préalable pour l'étude épidémiologique des maladies infectieuses. Les groupes bactériens majeurs, définis par une approche phylogénétique fondée sur l'étude de l'ARN ribosomique, ont clairement une évolution verticale. L'espèce bactérienne, qu'elle soit définie en termes biologique ou phénotypique, correspond bien à une réalité. A l'intérieur d'une espèce, la structure de la population peut varier de la forme clonale à la panmixie selon l'importance des transferts horizontaux de matériel génétique.

### Remerciements

Nous remercions Jacques Élion et Bernard Grandchamp pour les nombreuses discussions que nous avons eues sur le sujet ainsi que pour la relecture du manuscrit.

### ADRESSE

E. Denamur : maître de conférences des universités, praticien hospitalier. Laboratoire de biochimie génétique et Inserm U. 120, hôpital Robert-Debré, 48, boulevard Sérurier, 75019 Paris, France. B. Picard : professeur des universités, praticien hospitalier. Laboratoire de microbiologie, hôpital Morvan, 29609 Brest Cedex, France.

### TIRÉS À PART

E. Denamur.

L'utilisation du typage de l'ADN en médecine légale fait couler beaucoup d'encre. Le problème majeur est celui-ci : quelle est la probabilité que par coïncidence, l'ADN d'un suspect soit le même que celui trouvé sur les lieux du crime (et donc appartenant au criminel) [1] ? A l'écart de ce débat houleux et médiatique, la structure génétique des populations microbiennes est l'objet d'un intérêt croissant depuis une quinzaine d'années. Les microbes ne siègent pas dans les tribunaux, et pourtant, *Neisseria meningitidis*, *Mycobacterium tuberculosis* ou *Plasmodium falciparum*, par exemple, sont responsables d'un nombre de morts bien supérieur à celui dû au pire meurtrier.

Le monde bactérien, trop souvent assimilé à *Escherichia coli* par facilité, est en fait extrêmement divers. Les bactéries peuvent être trouvées à peu près partout, dans n'importe quelle niche écologique, et établir des relations symbiotiques avec d'autres organismes. Cette remarquable ubiquité réside dans leur capacité de produire de l'énergie et d'utiliser des nutriments à partir de sources très diverses [2]. De plus, de nombreuses espèces restent à découvrir (on estime que moins de 20 % des bactéries existantes sont identifiées) car, pendant près d'un siècle, les bactériologistes ont considéré que les micro-organismes que l'on pouvait cultiver constituaient la quasi-totalité des bactéries existantes. Une nouvelle bactériologie, qui vise à isoler en première

## RÉFÉRENCES

1. Lenski RE. Assessing the genetic structure of microbial populations. *Proc Natl Acad Sci USA* 1993; 90: 4334-6.
2. Hodgson DA. Bacterial diversity : the range of interesting things that bacteria do. In : Hopwood DA, Chater KF, eds. *Genetics of bacterial diversity*. London: Academic Press, 1989 : 3-22.
3. Woese CR. Microbiology in transition. *Proc Natl Acad Sci USA* 1994; 91: 1601-3.
4. Young JPW. The population genetics of bacteria. In : Hopwood DA, Chater KF, eds. *Genetics of bacterial diversity*. London: Academic Press, 1989: 417-38.
5. Chater KF, Hopwood DA. Diversity of bacterial genetics. In : Hopwood DA, Chater KF, eds. *Genetics of bacterial diversity*. London: Academic Press, 1989: 23-52.
6. Woese CR. Bacterial evolution. *Microbiol Rev* 1987; 51: 221-71.
7. Qu LH, Hardman N, Gill L, Chappell L, Nicoloso M, Bachelier JP. Phylogeny of helminths determined by rRNA sequence comparison. *Mol Biochem Parasitol* 1986; 20: 132-8.
8. Cocito C. Archaeobactéries et évolution. *médecine/sciences* 1986; 2: 436-44.
9. Barinaga M. Archaea and eukaryotes grow closer. *Science* 1994; 264: 1251.
10. Rowlands T, Baumann P, Jackson SP. The TATA-binding protein : a general transcription factor in eukaryotes and archaeobacteria. *Science* 1994; 264: 1326-9.

intention des gènes bactériens (en particulier ceux codant pour les ARN ribosomiques) et non plus des bactéries, révèle une diversité supplémentaire insoupçonnée jusqu'alors [3].

La génétique des populations bactériennes est au confluent de la génétique, de l'écologie et de la systématique. Elle étudie l'importance et la nature de la variation génétique, sa distribution dans le temps et dans l'espace, les forces qui l'affectent et sa signification écologique et médicale. La génétique des populations bactériennes est donc extrêmement liée à l'évolution [4]. Comprendre la structure des populations bactériennes nécessite d'avoir une approche phylogénétique avec, d'une part, caractérisation de grands groupes bactériens et, d'autre part, utilisation de critères précis de définition de l'espèce bactérienne. C'est à partir de ces données que l'on peut commencer à étudier la structure génétique au niveau de l'espèce, indispensable à la compréhension de l'épidémiologie et de la physiopathologie des maladies dont sont responsables les bactéries. Tant que les microbiologistes n'ont eu à leur disposition que des caractères phénotypiques trop simples ou trop variables (forme de la cellule, mobilité, morphologie de la colonie, variation antigénique, sensibilité aux bactériophages, capacité métabolique), la nature des relations entre les procaryotes est restée obscure. Ce n'est qu'avec l'émergence de critères moléculaires qu'une approche phylogénétique a pu être menée, révélant parfois une absence de corrélation avec les taxinomies antérieures.

Les bactéries diffèrent des eucaryotes en bien des points: elles ont un génome haploïde, une reproduction clonale, des tailles de populations importantes et un temps de reproduction court [5]. Malgré cette apparente simplicité, la structure des populations bactériennes se caractérise par une grande complexité. En terme de génétique des populations, la caractéristique majeure chez les bactéries est la fréquence et la diversité des mécanismes d'échange génétique entre organismes qui incluent transformation, transduction et conjugaison. L'importance de ces phénomènes conditionne la structure de la population [4]. Comme nous le verrons, cette caractéristique est

encore mal connue et il semble maintenant évident qu'elle varie d'une espèce bactérienne à l'autre.

Le but de cette synthèse est de montrer comment les techniques moléculaires actuelles permettent d'établir les bases d'une génétique d'une population bactérienne. De plus, quelques exemples illustreront en quoi la connaissance de la structure génétique des procaryotes contribue à la compréhension de l'épidémiologie des maladies infectieuses.

### **L'étude de l'ARN ribosomique a permis une approche phylogénétique des groupes majeurs du vivant**

Le séquençage des protéines et des acides nucléiques a permis une approche évolutionniste particulièrement efficace en terme d'« horloge moléculaire ». Le pourcentage de différence entre deux séquences permet d'estimer le temps écoulé depuis que les organismes ont divergé à partir d'un ancêtre commun [6]. Les molécules d'ARN de chacune des deux sous-unités ribosomiques (ARNr) constituent un indicateur particulièrement riche pour aborder les problèmes de l'évolution. Ces molécules sont présentes dans toutes les espèces; elles ont gardé une fonction générale équivalente, celle d'assurer la traduction. Les ARNr sont des molécules qui contiennent une alternance de régions extrêmement conservées, qui peuvent servir à étudier les relations phylogénétiques les plus distantes (interrègnes), et de régions dont la vitesse de variation est beaucoup plus rapide et qui peuvent être utilisées pour mesurer des distances évolutives relativement courtes. De plus, la taille des molécules d'ARNr est assez élevée pour permettre une validité statistique des résultats [7].

La surprise majeure à l'analyse des résultats de la comparaison des séquences des ARNr a été la mise en évidence d'une diversité phylogénétique jusque-là insoupçonnée des procaryotes sans commune mesure avec celle, en comparaison fort restreinte, des plantes et des animaux (*m/s n° 7, vol. 6, p. 710*). Deux groupes ont été caractérisés parmi les

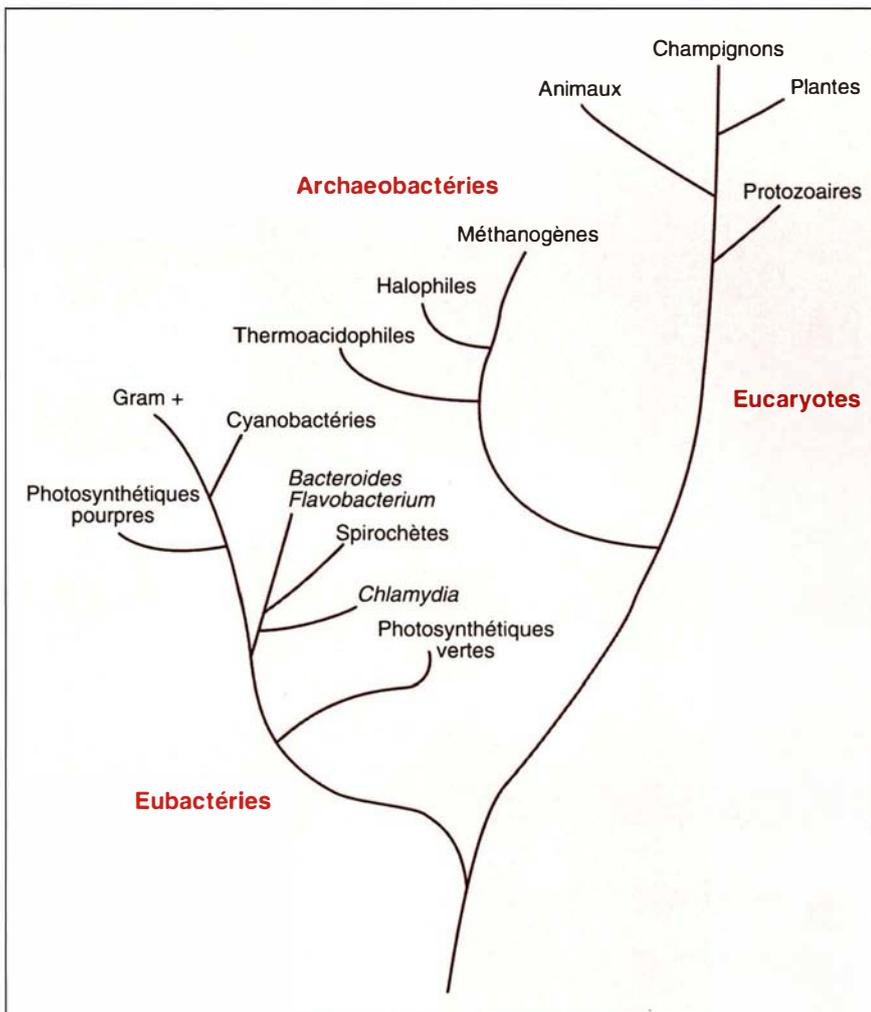


Figure 1. **Arbre phylogénétique déterminé à partir des comparaisons de séquences de l'ARNr** [6]. Le branchement des archaeobactéries repose sur la similitude des appareils transcriptionnels chez les eucaryotes et les archaeobactéries [9, 10]. Pour les eubactéries, l'importance des branches varie de quelques espèces (Bacteroides-Flavobacterium, Spirochetes, Chlamydiae) à de très nombreuses espèces (Gram positif, photosynthétiques pourpres). Le groupe des Gram positif est divisé en deux sous-groupes (haut et bas, selon leur pourcentage de G + C) tandis que celui des photosynthétiques pourpres l'est en quatre ( $\alpha$ ,  $\beta$ ,  $\gamma$  et  $\delta$ ). A titre d'exemple, les mycobactéries appartiennent au sous-groupe haut, selon leur pourcentage de G + C alors que les streptocoques et les staphylocoques appartiennent au sous-groupe bas, selon leur pourcentage de G + C. Les entérobactéries et *Pseudomonas aeruginosa* sont membres de la sous-division  $\gamma$  des bactéries photosynthétiques pourpres.

procaryotes, les eubactéries et les archaeobactéries qui ne sont pas plus reliés entre eux qu'ils ne le sont au groupe des eucaryotes [3, 6, 8].

La figure 1 propose un arbre phylogénétique universel des trois règnes primaires fondé sur les séquences des ARNr. C'est en 1977 que les archaeobactéries ont été individualisées en règne primaire mais ce n'est que très récemment que l'on a été capable de situer leur branchement dans l'arbre principal. Il semble en effet que les archaeobactéries et les eucaryotes descendent d'un tronc commun duquel avaient antérieurement divergé les eubactéries, comme le montre la similitude entre les appareils transcriptionnels des archaeobactéries et des eucaryotes qui diffèrent de celui des eubactéries [9, 10]. La réalisation d'un tel arbre pose comme postulat

que les transferts latéraux d'ADN entre espèces éloignées ne jouent pas un rôle majeur et que, par conséquent, l'évolution bactérienne se fait sous forme d'un arbre plus que d'un réseau. Cela semble vrai car des arbres phylogénétiques obtenus par la comparaison d'autres parties du génome (cytochrome c, nitrogénase) sont concordants dans le monde bactérien avec celui des ARNr [4].

### Qu'est-ce qu'une espèce bactérienne ?

L'espèce biologique est définie comme un ensemble d'organismes se reproduisant entre eux mais pas avec ceux d'une autre espèce. Ce concept a été largement appliqué en systématique chez les plantes et les animaux parce que l'échange de matériel gé-

nétique est une puissante force de cohésion de ces groupes [11]. Mais chez les bactéries, la reproduction et l'échange de gènes chromosomiques sont des fonctions indépendantes, à l'inverse de la majorité des animaux et des plantes où ces deux fonctions sont liées [12]. Ainsi, la reproduction est asexuée chez l'ensemble des bactéries et les recombinaisons entre souches sont dues à des transferts horizontaux. En conséquence, l'espèce bactérienne a-t-elle une réalité et, dans ce contexte, comment la définir ?

Bien qu'ayant une reproduction non sexuée, les procaryotes échangent du matériel génétique et l'on peut donc essayer d'appliquer, sur cette base, la notion d'espèce biologique aux bactéries. Il faut alors concevoir qu'il existe un taux de recombinaison suf-

## RÉFÉRENCES

11. Cohan FM. Genetic exchange and evolutionary divergence in prokaryotes. *Trends Ecol Evol* 1994; 9: 175-80.
12. Dykhuizen DE, Green L. Recombination in *Escherichia coli* and the definition of biological species. *J Bacteriol* 1991; 173: 7257-68.
13. Wayne LG, Brenner DJ, Colwell RR, Grimont PAD, Kandler O, Krichevsky MI, Moore LH, Moore WEC, Murray RGE, Stackebrandt E, Starr MP, Trüper HG. Report of the ad hoc committee on reconciliation of approaches to bacterial systematics. *Int J Syst Bacteriol* 1987; 37: 463-4.
14. Grimont PAD. Use of DNA reassociation in bacterial classification. *Can J Microbiol* 1988; 34: 541-6.
15. Rayssiguier C, Thaler DS, Radman M. The barrier to recombination between *Escherichia coli* and *Salmonella typhimurium* is disrupted in mismatch-repair mutants. *Nature* 1989; 342: 396-400.
16. Radman M, Taddei F, Halliday J. Correction des erreurs dans l'ADN : de la génétique bactérienne aux mécanismes de prédisposition héréditaire aux cancers chez l'homme. *médecine/sciences* 1994; 10: 1024-30.
17. Matic I, Rayssiguier C, Radman M. Interspecies gene exchange in bacteria : the role of SOS and mismatch repair systems in evolution of species. *Cell* 1995; 80: 507-15.
18. Maynard Smith J, Dowson CG, Spratt BG. Localized sex in bacteria. *Nature* 1991; 349: 29-31.
19. Siebold C, Henricksen J, Koening A, Martin C, Chalkley L, Hakenbeck R. Mosaic *pbpX* genes of major clones of penicillin resistant *Streptococcus pneumoniae* have evolved from *pbpX* genes of penicillin sensitive *Streptococcus oralis*. *Mol Microbiol* 1994; 12: 1013-23.
20. Dowson CG, Coffey TJ, Spratt BG. Origin and molecular epidemiology of penicillin-binding-protein-mediated resistance to  $\beta$ -lactam antibiotics. *Trends Microbiol* 1994; 2: 361-6.

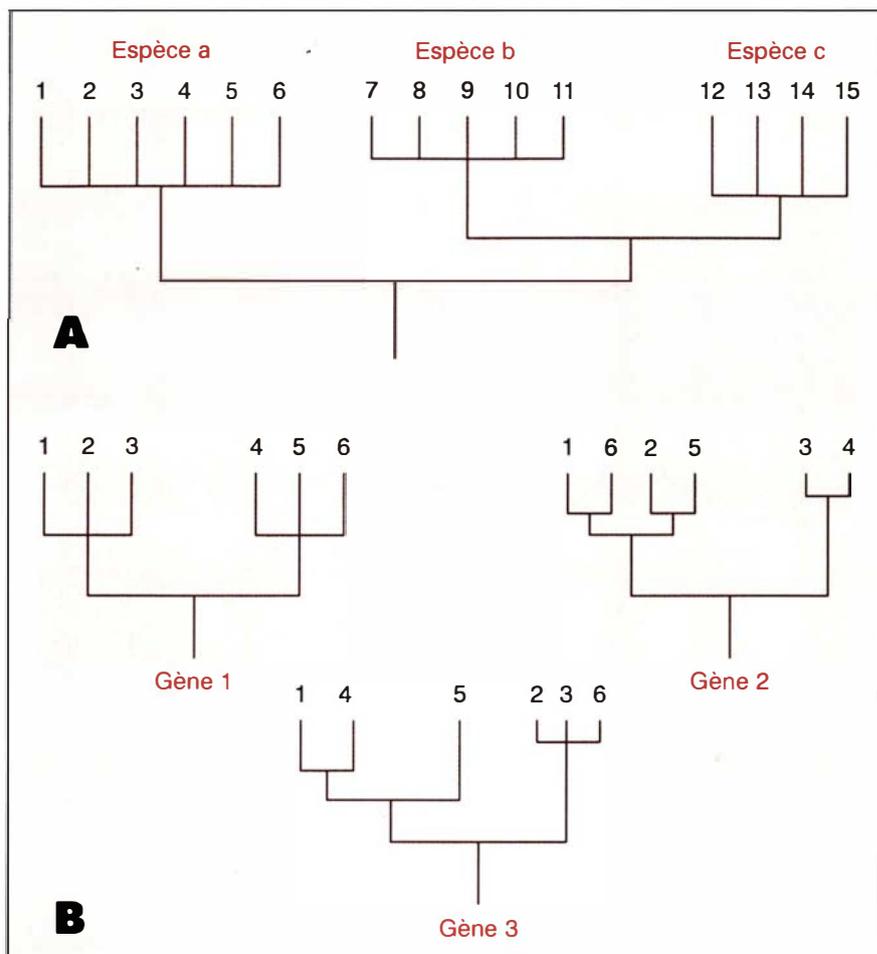


Figure 2. **Meilleur arbre phylogénétique obtenu pour 15 souches hypothétiques à partir d'un groupe d'arbres phylogénétiques, chacun dérivé des données du séquençage nucléotidique d'un gène (A).** Le groupement des souches 1 à 6 est différent pour différents gènes (B) mais ces souches sont toujours regroupées et distinctes des autres souches lorsque l'on considère les arbres phylogénétiques de tous les gènes (A). Ces souches représentent donc une espèce (espèce a). Il en est de même pour les souches 7 à 11 (espèce b) et 12 à 15 (espèce c). Les espèces B et C sont plus proches l'une de l'autre qu'elles ne le sont de l'espèce a. Leur divergence dans le temps est moins ancienne que ne l'est la divergence espèce a/espèces b et c.

fisant à l'intérieur d'une espèce et relativement peu important entre les espèces. Cette définition implique que les phylogénies de divers gènes établies sur des individus de la même espèce devraient être significativement différentes les unes des autres alors que les phylogénies des mêmes gènes établies sur des individus appartenant à des espèces distinctes ne devraient pas l'être [12]. Dykhuizen et Green [12] viennent récemment de proposer d'utiliser ce critère pour la

définition de l'espèce. Considérons quinze souches bactériennes isolées de la nature parmi lesquelles on veut délimiter des espèces. Des portions de quelques gènes sont séquencées chez chaque souche et des arbres phylogénétiques fondés sur le pourcentage de divergence des séquences nucléotidiques sont construits. Ces différents arbres sont comparés et l'on trouve certains groupes de souches à l'intérieur desquels les arbres obtenus pour les différents

gènes ne sont pas corrélés alors que ces souches sont toujours groupées ensemble par rapport aux autres souches sur les arbres de tous les gènes étudiés. Ces groupes de souches peuvent alors être considérés comme constituant autant d'espèces bactériennes (figure 2). Ces auteurs ont validé ce concept en séquençant des gènes ubiquistes tels que les gènes *gnd* (6-phosphogluconate déshydrogénase), *phoA* (phosphatase alcaline) ainsi que la région de l'opéron *trp* (tryptophane) chez différentes souches de *E. coli* et de *Salmonella typhimurium*. Ils ont montré qu'à l'intérieur de *E. coli*, les arbres phylogénétiques individuels obtenus pour chaque gène étudié étaient différents [12]. La concordance retrouvée entre les arbres phylogénétiques des individus appartenant à des espèces distinctes peut être attribuée à l'évolution verticale ayant pour conséquence une modification importante des caractères entre les espèces, l'absence de recombinaison n'étant alors qu'une conséquence de la faible homologie d'ADN entre les souches des différentes espèces.

La définition génotypique de l'espèce repose également sur une approche phylogénétique. Elle est fondée sur des expériences d'hybridation en solution entre l'ADN d'une souche de référence et l'ADN de la souche que l'on veut étudier. Avec cette technique, outre le pourcentage d'hybridation entre deux souches (il est de 100 % pour la souche avec elle-même), il est possible de mesurer la différence de stabilité thermique des hybrides. Cette valeur est appelée  $\Delta T_{m(e)}$  (thermal elution midpoint) et est exprimée en degré C, 1 °C de  $\Delta T_{m(e)}$  étant supposé correspondre à 1 % de divergence dans les séquences d'ADN. A partir de l'expérience acquise avec les *Enterobacteriaceae*, l'espèce a été définie comme un ensemble de souches ayant au moins 70 % d'homologie de l'ADN à la température optimale de renaturation avec un  $\Delta T_{m(e)}$  inférieur ou égal à 5 °C [13]. La valeur de 5 °C a été choisie à partir des résultats d'un histogramme des valeurs de  $\Delta T_{m(e)}$  obtenues avec les *Enterobacteriaceae* qui montre clairement deux pics ou catégories de divergences séparés par un intervalle (figure 3). Un premier pic avec un maximum à 3,5-4 °C correspond à

l'espèce, un second pic au-dessus de 7 °C avec un maximum à 9-9,5 °C et des ondulations jusqu'à 20 °C correspond à des espèces différentes. Entre 5,5 et 7 °C, on est dans un *no bacterial's land* avec très peu de souches [14].

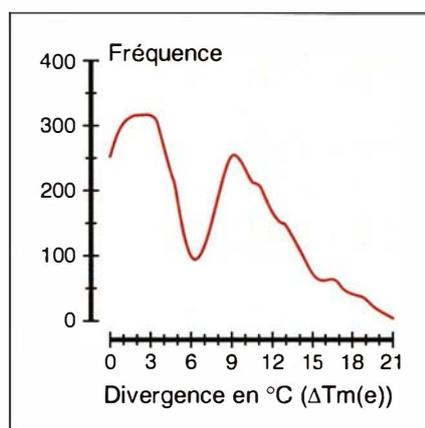


Figure 3. **Distribution de 2 030 valeurs de divergence ( $\Delta T_{m(e)}$ ) entre des souches bactériennes appartenant à la famille des *Enterobacteriaceae* montrant clairement deux pics ou catégories de divergences séparés par un intervalle.** L'espèce est définie comme un ensemble de souches ayant des ADN avec au moins 70 % d'analogie et une différence de température optimale de renaturation  $\Delta T_{m(e)}$  inférieure ou égale à 5 °C. Le premier pic correspond à l'espèce bactérienne, le second à des espèces bactériennes différentes. (D'après [14]).

Il semble bien que, quels que soient les critères retenus pour définir une espèce, la notion de groupes séparés entre eux, que l'on nomme espèces, soit une réalité. Il n'existe pas un continuum entre les différents individus du monde bactérien et les groupes bactériens au niveau de l'espèce se comportent en terme d'évolution plus comme les branches d'un arbre que comme un réseau. Il est peu probable que la discontinuité observée entre les espèces soit due à

un simple problème d'échantillonnage. La recombinaison *in vivo* nécessite une homologie très importante entre les ADN; elle est en effet abolie par une divergence de séquence de 10 % à 20 %. C'est le système de réparation des mésappariements qui est à la base de cette barrière à la recombinaison interspécies. Il a en effet été montré qu'entre *E. coli* et *Salmonella enterica*, deux espèces bactériennes ayant divergé il y a à peu près 150 millions d'années et qui ont 15 % à 20 % de différence de séquences d'ADN, la barrière de la recombinaison intergénérique est levée lorsque les souches présentent des mutations dans les gènes *mutL*, *mutS* et *mutH* impliqués dans la réparation des mésappariements [15, 16]. Ainsi, c'est un équilibre entre le système SOS impliqué dans la recombinaison et le système de réparation des mésappariements qui est à la base du processus de spéciation [16, 17]. Il semble toutefois que cette barrière de recombinaison entre espèces ne soit pas absolue car il a été rapporté que différentes espèces de *Neisseria*, dont les séquences d'ADN présentent entre elles jusqu'à 23 % de différence, avaient échangé des fragments du gène *PBP2B*. Les PBP (penicillin binding protein) sont les protéines qui lient la pénicilline et dont les modifications constituent un des mécanismes de résistance à la famille des  $\beta$  lactamines [18]. La formation d'un gène *PBP2B* mosaïque par échanges horizontaux est susceptible de faire apparaître chez la bactérie une résistance à la pénicilline; ce type de transfert est donc très fortement sélectionné. De même, la résistance du pneumocoque à la pénicilline pourrait être due à la formation de gènes *PBP* mosaïques par transmission horizontale de gènes *PBP* de souches de streptocoques appartenant à des espèces nettement éloignées comme *Streptococcus oralis* [19, 20]. Plus d'études sont nécessaires pour savoir si de tels transferts sont ou non exceptionnels.

De telles définitions phylogénétiques de l'espèce peuvent se trouver en porte à faux avec une logique de bactériologie médicale fondée en premier lieu sur des données cliniques et des caractères phénotypiques, biochimiques ou antigéniques plutôt que sur des données moléculaires trop

## RÉFÉRENCES

21. Bercovier H, Mollaret HH, Alonso JM, Braut J, Fanning GR, Steigerwalt AG, Brenner DJ. Intra- and interspecies relatedness of *Yersinia pestis* by DNA hybridization and its relationship to *Yersinia pseudotuberculosis*. *Curr Microbiol* 1980; 4: 225-9.
22. Judicial Commission of the International Committee on Systematic Bacteriology. Opinion 60. Rejection of the name *Yersinia pseudotuberculosis* subsp. *pestis* (van Loghem) Bercovier *et al.* 1981 and conservation of the name *Yersinia pestis* (Lehmann and Neumann) van Loghem 1944 for the plague bacillus. *Int J Syst Bacteriol* 1985; 35: 540.
23. Selander RK, Levin BR. Gene diversity and structure in *Escherichia coli* populations. *Science* 1980; 210: 545-7.
24. Selander RK, Caugant DA, Ochman H, Musser JM, Gilmour MN, Whittman TS. Methods of multilocus enzyme electrophoresis for bacterial population genetics and systematics. *Appl Environ Microbiol* 1986; 51: 873-84.
25. Selander RK, Caugant DA, Whittam TS. Genetic structure and variation in natural population of *Escherichia coli*. In : Neidhardt FC, Ingragam JL, Low KB, Magasanik B, Schaechter M, Umberger HE, eds. *Escherichia coli and Salmonella typhimurium: cellular and molecular biology*. Washington, DC: American Society of Microbiology, 1987; 2: 1625-48.
26. Goulet P, Picard B. Comparative electrophoretic polymorphism of esterases and other enzymes in *Escherichia coli*. *J Gen Microbiol* 1989; 135: 135-43.
27. Brunham RC, Plummer FA, Stephens RS. Bacterial antigenic variation, host immune response, and pathogen-host coevolution. *Infect Immun* 1993; 61: 2273-6.
28. Maynard Smith J, Smith NH, O'Rourke M, Spratt BG. How clonal are bacteria? *Proc Natl Acad Sci USA* 1993; 90: 4384-8.
29. Martinez JL. Do bacteria have sex? *Nature* 1991; 352: 288.
30. Caugant DA, Frøholm LO, Bøvre K, Holten E, Frasch CE, Mocca IF, Zollinger WD, Selander RK. Intercontinental spread of a genetically distinctive complex of clones of *Neisseria meningitidis* causing epidemic disease. *Proc Natl Acad Sci USA* 1986; 83: 4927-31.

difficiles ou trop lourdes à obtenir en routine. Ainsi, par exemple, le bacille de Yersin, responsable de la peste, a longtemps constitué une espèce distincte à l'intérieur du genre *Yersinia*. Pourtant, par hybridation ADN/ADN, son appartenance à l'espèce *Y. pseudotuberculosis* fut clairement établie [21] et l'appellation proposée par les taxinomistes fut alors *Y. pseudotuberculosis* subspecies *pestis*. Néanmoins, le pouvoir pathogène particulier du bacille pesteux conduisit des bactériologistes à contester cette appellation jugée trop minimaliste et à demander la conservation de l'appellation *Y. pestis*, plus conforme à la réalité médicale [22].

### Diversité intraspécifique : de la structure clonale à la panmixie

C'est au début des années 1980 que les premiers travaux sur la structure génétique à l'intérieur de l'espèce ont été rapportés. Ces travaux, menés en particulier par Selander *et al.* [23-25], reposent sur l'analyse du polymorphisme électrophorétique d'enzymes métaboliques et ont permis de proposer la théorie de l'organisation clonale du monde bactérien. A cette époque, la fréquence des phénomènes de transferts horizontaux était nettement sous-estimée *in vivo*, ce type d'échange étant surtout considéré comme survenant *in vitro* au laboratoire. L'étude électrophorétique du polymorphisme enzymatique est une technique simple qui permet de détecter des substitutions d'acides aminés, reflets des variations nucléotidiques du gène correspondant. Chaque souche se caractérise par une combinaison de variants de mobilité pour le nombre d'enzymes (donc de *loci*) étudiés. Des isolats ayant des allèles identiques à tous les *loci* examinés sont dits appartenir au même type électrophorétique. Des milliers de souches de *E. coli* ont maintenant été étudiées sur une quarantaine de *loci* chromosomiques [25, 26]. L'espèce *E. coli* présente un niveau de diversité génétique des régions étudiées d'ADN codant bien plus important que la majorité des espèces eucaryotes. Plus de 90 % des *loci* sont polymorphes (30 % chez les eucaryotes) et la diversité à chaque *locus* est en moyenne trois fois plus élevée que

chez les eucaryotes. Une organisation clonale implique l'existence d'une forte association d'allèles entre les *loci* (déséquilibre de liaison) et la mise en évidence de mêmes génotypes (correspondant, par hypothèse, aux types électrophorétiques) à travers le monde sur une large période de temps (une centaine d'années). En effet, malgré l'impressionnante variation allélique retrouvée, le nombre de types électrophorétiques détectés est bien moins élevé que ce que l'on aurait pu attendre car certains allèles sont quasiment toujours trouvés associés à d'autres. Par exemple, deux génotypes AB et ab résultant de deux *loci* sont très communs alors que les génotypes Ab et aB sont absents ou extrêmement rares. A côté de *E. coli*, de nombreuses autres espèces bactériennes ont maintenant été étudiées par cette approche (essentiellement des bactéries d'intérêt médical) et cette notion de clonalité, fondée sur des caractères protéiques, donc indirects, semblait pouvoir s'appliquer à l'ensemble du monde bactérien.

En fait, les choses sont apparues bien plus complexes lorsque les méthodes directes de détermination de séquences nucléotidiques ont pu être appliquées plus aisément à la comparaison d'un nombre appréciable d'individus. Ainsi, les données de séquences nucléotidiques accumulées depuis quelques années ont clairement montré l'existence de gènes mosaïques formés à partir de fragments issus de transferts horizontaux. Cela a été mis en évidence, le plus souvent, au niveau de gènes impliqués dans la virulence, gènes de membrane externe (*porA* de *Neisseria meningitidis*), gènes de la piline de *N. gonorrhoeae* ou gènes de résistance aux antibiotiques (gènes codant pour les PBP) [18, 27]. Dans ces cas, la souche recombinante a clairement acquis un avantage sélectif qui lui permet d'échapper aux défenses de l'hôte ou d'être plus résistante aux antibiotiques. Mais il est maintenant évident que de tels transferts surviennent également pour des gènes ubiquistes au polymorphisme considéré comme neutre. Il est à noter que de tels gènes mosaïques sont observés aussi bien chez des espèces naturellement transformables (*Streptococcus*, *Bacillus*, *Haemophilus*, *Neisseria*) que chez des

espèces non naturellement transformables (*Salmonella*, *E. coli*) [18]. Le problème est de savoir jusqu'à quel point ces transferts horizontaux peuvent obscurcir une structure clonale. Maynard Smith *et al.* [28] ont réanalysé les résultats de différents travaux fondés sur l'étude du polymorphisme enzymatique en utilisant un test statistique capable de détecter les associations entre les gènes à différents loci. Il ressort de ce travail que les populations bactériennes ne sont pas invariablement clonales mais qu'elles occupent un spectre allant de la structure fortement sexuée (terme un peu abusivement emprunté à la reproduction des eucaryotes et qui indique une très grande proportion d'échanges horizontaux entre souches [29]) à la structure clonale quasiment stricte. Certaines espèces telles que *Salmonella enterica* sont trouvées clonales (figure 4A). A l'autre extrémité, *N. gonorrhoeae* est panmictique avec une association au hasard entre les loci. Des simulations de populations suggèrent que pour produire une population panmictique, la recombinaison localisée touchant un gène doit être deux à quatre fois plus fréquente que les mutations ponctuelles [26]. Deux types de structure de population intermédiaires ont été mis en évidence. Chez une population de *Rhizobium meliloti*, le déséquilibre de liaison existe seulement parce que l'analyse porte sur des isolats appartenant à des groupes génétiquement isolés. Si l'on s'intéresse à des souches à l'intérieur de chaque groupe génétique, l'association entre les loci est quasiment au hasard (figure 4B). A l'inverse, *N. meningitidis* présente ce qui est appelé une structure « épidémique » (figure 4C). Il existe une association significative entre les loci mais elle est due à une augmentation récente et explosive de types électrophorétiques particuliers. Un exemple tout à fait significatif de ce type de phénomène est celui d'une pandémie de méningite cérébrospinale, survenue dans les années 1980, due à un clone de *N. meningitidis* caractérisé par un type électrophorétique d'enzymes. Ce clone avait vraisemblablement acquis un avantage sélectif important car, au sein de la population panmictique de *N. meningitidis*, il a émergé puis s'est disséminé sur trois continents. Les souches

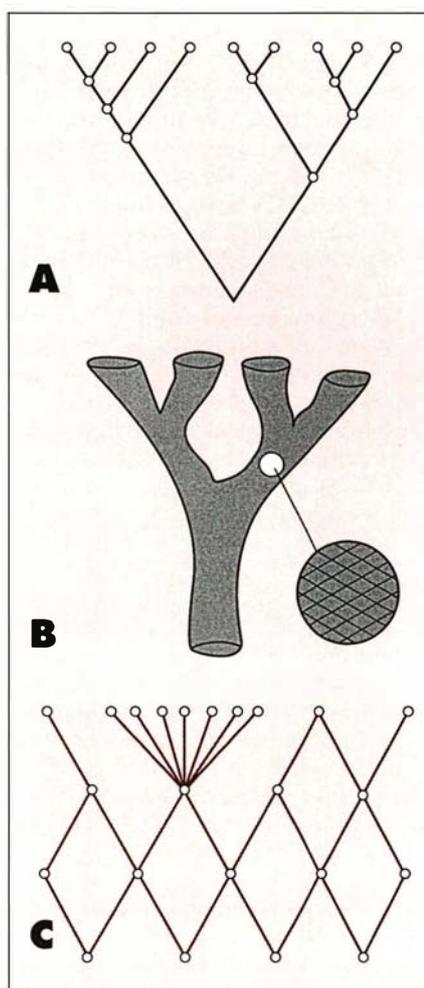


Figure 4. **Représentation des différents types de structure de population bactérienne.** A et B. Les populations consistent en des souches qui sont divisées en deux groupes phylogénétiques majeurs. La structure de la population dans A est clonale à tous les niveaux. Aucune recombinaison ne survient, que ce soit entre des individus de la même ou de différentes branches de l'arbre. Dans B, il n'y a pas de recombinaison entre les deux branches mais à l'intérieur de chacune d'elles, la structure ressemble plus à un filet qu'à un arbre, comme cela est représenté dans l'agrandissement. C Structure épidémique dans laquelle les recombinaisons sont fréquentes entre tous les membres de la population avec une structure en filet. Toutefois, occasionnellement, une souche ayant un avantage sélectif important apparaît et se développe rapidement pour produire un clone épidémique. (D'après [27]).

de ce clone apparurent au Danemark en 1976 puis en Finlande en 1979, aux îles Féroé en 1981 et en Islande en 1983. On les retrouva en Écosse, en Hollande, en Allemagne et en Espagne au début des années 1980. En 1983 elles prédominèrent en Angleterre et furent retrouvées en France. C'est également au début des années 1980 qu'elles furent isolées en Afrique du Sud dans la ville du Cap et à Cuba où elles causèrent de sévères épidémies sans que l'on pût expliquer leur transmission à partir des foyers européens. En revanche, on relia l'épidémie survenue à Miami à la même époque à l'arrivée des réfugiés cubains. Enfin, ce clone fut aussi retrouvé lors d'une épidémie survenue au Chili en 1985 [30].

Un autre moyen d'apprécier le degré de clonalité d'une population bactérienne est d'étudier plusieurs marqueurs génétiques et de les comparer entre eux. Cette approche a permis de montrer chez des isolats humains de *Pseudomonas aeruginosa* qu'il existait un taux important de recombinaison avec disparition de la structure clonale [31, 32]. En effet, une absence de corrélation a été mise en évidence entre, d'une part, les arbres obtenus à partir de l'étude du polymorphisme électrophorétique des estérases et, d'autre part, les arbres obtenus à partir du polymorphisme de restriction (RFLP : *restriction fragment length polymorphism*) de l'ADNr. De même, la comparaison des résultats de RFLP de l'ADNr et du polymorphisme d'amplification au hasard de l'ADN par PCR (RAPD : *random amplified polymorphic DNA*) [33, 34] à ceux du polymorphisme électrophorétique des enzymes a montré que *E. coli* avait une structure de base clonale mais qu'il existait aussi, à l'intérieur de l'espèce, des variations dans le degré de clonalité. Les souches du groupe B2 causant des infections extra-intestinales chez l'homme ont une organisation fortement clonale alors que les souches retrouvées chez les mammifères non primates ont une structure clonale plus relâchée [35].

L'ensemble de ces données a des retombées très pratiques pour l'étude des phénomènes épidémiques. En effet, la connaissance du caractère clonal ou recombinant d'une population bactérienne dont un groupe est

## RÉFÉRENCES

31. Denamur E, Picard B, Decoux G, Denis JB, Elion J. The absence of correlation between allozyme and *rml* RFLP analysis indicates a high gene flow rate within human clinical. *Pseudomonas aeruginosa* isolates. *FEMS Microbiol Lett* 1993; 110: 275-80.
32. Picard B, Denamur E, Barakat A, Elion J, Goulet P. Genetic heterogeneity of *Pseudomonas aeruginosa* clinical isolates revealed by esterase electrophoretic polymorphism and restriction fragment length polymorphism of the ribosomal RNA gene region. *J Med Microbiol* 1994; 40: 313-22.
33. Williams JGK, Kubelik AR, Livak KJ, Rafalski JA, Tingey SV. DNA polymorphisms amplified by arbitrary primers are useful as genetic markers. *Nucleic Acids Res* 1990; 18: 6531-5.
34. Welsh J, McClelland M. Genomic fingerprinting using arbitrarily primed PCR and a matrix of pairwise combinations of primers. *Nucleic Acids Res* 1991; 19: 5275-9.
35. Desjardins P, Picard B, Kaltenböck B, Elion J, Denamur E. Sex in *Escherichia coli* does not disrupt the clonal structure of the population: evidence from RAPD and *rml* RFLP data. *J Mol Evol* 1995 (sous presse).
36. Kreiswirth B, Kornblum J, Arbeit RD, Eisner W, Maslow JN, McGeer A, Low DE, Novick RP. Evidence for a clonal origin of methicillin resistance in *Staphylococcus aureus*. *Science* 1993; 259: 227-30.
37. Pitt TL, Livermore DM, Pitcher D, Vatoopoulos AC, Legakis NJ. Multiresistant serotype O 12. *Pseudomonas aeruginosa*: evidence for a common strain in Europe. *Epidem Infect* 1989; 103: 565-76.
38. Musser JM, Selander RK. Brazilian purpuric fever: evolutionary genetic relationships of the case clone of *Haemophilus influenzae* biogroup aegyptius to encapsulated strains of *Haemophilus influenzae*. *J Infect Dis* 1990; 161: 130-3.
39. Mariani-Kurkdjian P, Denamur E, Milon A, Picard B, Cavé H, Lambert-Zechovsky N, Loirat C, Goulet P, Sansonetti PJ, Elion J. Identification of a clone of *Escherichia coli* O103:H2 as a potential agent of hemolytic-uremic syndrome in France. *J Clin Microbiol* 1993; 31: 296-301.
40. Lenski RE, Travisano M. Dynamics of adaptation and diversification: A 10,000-generation experiment with bacterial populations. *Proc Natl Acad Sci USA* 1994; 91: 6808-14.

impliqué dans un processus épidémique va permettre d'orienter la stratégie de choix des méthodes de typage des souches. Avec une population ayant un taux important de recombinaison, l'absence de corrélation entre les résultats des typages fait qu'il suffit d'en combiner un nombre limité (en général deux) pour rapidement atteindre un degré important de différenciation, ce qui rend l'étude de ce type de population relativement aisée. A l'inverse, l'étude épidémiologique des populations clonales est plus délicate, particulièrement lorsqu'un clone a été sélectionné par des caractères de résistance aux antibiotiques comme les souches de *Staphylococcus aureus* résistantes à la méthicilline [36] ou de *P. aeruginosa* du sérotype O12 [37] ou par des caractères de virulence, comme les souches de *Haemophilus influenzae* responsables de la fièvre brésilienne purpurique [38]. Les techniques de typage au pouvoir discriminant le plus élevé devront alors être mises en œuvre pour tenter de différencier les souches à l'intérieur de la population. Dans des cas extrêmes, le caractère clonal de certaines populations bactériennes serait tel qu'il rendrait vaine toute tentative de différenciation d'un groupe de souches impliqué dans un processus épidémique. L'appartenance des souches à ce groupe constituerait alors le seul marqueur épidémiologique utilisable, comme dans le cas du clone de *E. coli* O103:H2 responsable de syndromes hémolytiques et urémiques en France entre 1987 et 1989 [39].

## Conclusions et perspectives

Ainsi, il semble bien que l'importance des phénomènes de recombinaison génétique entre les bactéries soit le principal déterminant du type de la structure d'une population bactérienne. Les groupes bactériens majeurs ont clairement une évolution verticale. A l'intérieur des espèces, le degré de recombinaison est variable; il n'a été étudié que chez très peu d'espèces et il reste donc un effort majeur à faire dans ce sens. La détermination de la structure génétique d'une population bactérienne semble être un préalable de plus en plus im-

portant à l'étude épidémiologique des maladies infectieuses.

Dans cette revue, l'importance des transposons, éléments accessoires utilisant la recombinaison illégitime, qui participent à l'échange génétique entre les bactéries n'a pas été discutée car les conséquences du transfert d'éléments accessoires ont été surtout étudiées en terme d'adaptation bactérienne: différents souches, espèces, genres dans un habitat peuvent échanger des gènes sur des éléments accessoires, permettant à la population d'exploiter l'expérience génétique de l'ensemble d'un écosystème bactérien [4].

Une approche séduisante pour mieux comprendre les mécanismes de l'évolution chez les bactéries vient d'être proposée par Lenski et Travisano sous le nom de « paléontologie expérimentale » [40]. Ils ont, en effet, étudié les dynamiques d'adaptation et de diversification chez douze populations de *E. coli* sur 10 000 générations (soit une période d'à peu près cinq ans) dans un environnement identique. Bien qu'encore loin de la complexité d'une population bactérienne évoluant dans un milieu extérieur, cette approche méthodologique s'est avérée être un bon modèle d'évolution [38]. ■

## Summary

### From population genetics of bacteria to epidemiology of infectious diseases

The bacterial world has been too often reduced to *Escherichia coli*. In fact, there is a tremendous bacterial diversity. Knowledge of the structure of bacterial populations is a prerequisite to the understanding of epidemiology of infectious diseases. The major bacterial groups defined by a phylogenetic approach based on ribosomal RNA studies clearly have a vertical evolution. Bacterial species, defined as biologic or phenetic entities, really exist. Within a species, bacteria occupy a spectrum of population structures ranging from highly sexual to almost strictly clonal according to the level of horizontal exchange.