



Point de vue

Biodiversité microbienne et référence : le défi de la génomique

S. Brisse, Institut Pasteur, Plate-forme Genotypage des Pathogènes et Santé Publique (France)

S. Brisse. (2010). Biodiversité microbienne et référence : le défi de la génomique, EuroReference, No. 4, ER04-10P01. <http://www.anses.fr/euroreference/numero4/PN2001.htm>

Il est devenu banal de dire que la biodiversité microbienne, définie comme la diversité génétique, phénotypique et fonctionnelle des procaryotes, virus ou eucaryotes unicellulaires, est considérable. Mais ceux qui l'étudient de près préfèrent la qualifier de « vertigineuse ».



Il est assez largement accepté que le monde bactérien, par exemple, est composé d'une centaine de divisions majeures (phylum), alors qu'on en connaissait seulement une douzaine il y a 30 ans. Ces divisions profondes ont été révélées dans le sillage de la révolution de l'ARN ribosomique 16S apporté par Carl Woese, qui a découvert le troisième domaine du vivant, les Archées, qui sont aussi différents des bactéries que des eucaryotes (et dont, curieusement, aucun représentant pathogène n'est connu). On sait aussi que le corps humain comporte environ 10 fois plus de cellules procaryotes (Bactéries et Archées) que de cellules humaines, et que ces bactéries appartiennent à plus de 1000 espèces distinctes, dont l'inventaire a été initié récemment par de grands consortiums internationaux (Qin *et al.*, 2010). Mais les estimations du nombre total d'espèces existantes dépassent largement le million, voire le milliard (Dykhuizen, 1998). Au niveau fonctionnel, l'inventaire des familles de gènes présents dans l'ensemble du domaine bactérien en est à ses balbutiements, avec « seulement » 60 000 familles de gènes découvertes à l'heure actuelle (Wu *et al.*, 2009). On estime entre 0,1 % et 1 % la fraction des espèces que l'on parvient à cultiver en laboratoire. L'analyse génomique de cette « matière sombre » phylogénétique est maintenant possible grâce aux avancées des technologies de séquençage de l'ADN, applicable à des cellules bactériennes uniques isolées physiquement dans des micro-canaux.

En quoi les laboratoires de référence sont-ils concernés par la biodiversité, puisqu'ils sont en général spécialisés sur une ou un petit nombre d'espèces précises ? La raison majeure est qu'au niveau de l'espèce aussi, la diversité microbienne est vertigineuse. D'une part, les souches d'une espèce bactérienne typique peuvent différer entre elles, au niveau de leurs séquences orthologues (héritées de leur ancêtre commun), de plus de 5 % en moyenne. En comparaison, l'Homme diffère du chimpanzé, son plus proche voisin, de 1,2 % en moyenne, alors que deux personnes, toutes origines confondues, partagent plus de 99,9 % de leurs séquences génomiques. Du point de vue bactérien, l'espèce *Homo sapiens*, qui appartient dans la taxonomie des mammifères à l'ordre des Primates, est donc une « souche » au sein de « l'espèce Primates », qui inclut aussi les grands singes, les singes du Nouveau Monde et les lémuriers. De plus, le contenu en gènes (présence/absence) est très variable d'une souche bactérienne à l'autre. Par exemple, les trois premières souches séquencées de *Escherichia coli* ne partageaient qu'un tiers du total de leurs gènes en commun, et sur 20 génomes de cette espèce

comparés plus récemment, environ 2 000 gènes sont communs à tous les génomes, mais plus de 11 000 gènes distincts ont été répertoriés au total (Touchon *et al.*, 2009). Des extrapolations montrent que la somme des gènes présents dans l'ensemble des souches d'une même espèce (le « pan-génome ») peut donc être « considérable ». Les régions génomiques variables sont souvent responsables des caractéristiques médicalement importantes des souches, portant par exemple des îlots de virulence ou de résistance (Fournier *et al.*, 2006), rendant leur identification cruciale pour le laboratoire de référence.

Comment est organisée la diversité génétique au sein des espèces ? Des dizaines de familles clonales sont typiquement décrites au sein des espèces bactériennes pathogènes (Ragon *et al.*, 2008 ; Brisse *et al.*, 2009 ; Diancourt *et al.*, 2010). Comment délimiter, pour pouvoir ensuite les identifier, les clones les plus menaçants pour un patient ou la Santé publique ? Comment démontrer l'origine commune de cas groupés et déterminer la source de contamination ? Comment caractériser rapidement les nouvelles souches émergentes ? Ces questions sont au cœur de l'activité des laboratoires de référence. Elles sont aussi centrales en génétique des populations microbiennes, discipline qui a pour objet d'étudier la « biodiversité des souches » et l'évolution de leurs caractéristiques (Brisse, 2008). Même pour les espèces les plus étudiées, comme illustré récemment pour *E. coli* (Walk *et al.*, 2009), l'inventaire de la diversité des souches et les frontières de l'espèce évoluent constamment.

De façon regrettable, les activités de référence et la biologie des populations se sont côtoyées trop longtemps, sans beaucoup interagir. Alors que la microbiologie n'est quasiment pas enseignée dans les spécialisations de biologie évolutive, les bases mêmes de la science de l'évolution sont virtuellement absentes de la formation des microbiologistes, pharmaciens, médecins biologistes et vétérinaires. Alors que le laboratoire de référence utilise des méthodes de typage faciles d'emploi et hautement discriminantes pour répondre rapidement aux questions épidémiologiques locales, le généticien des populations applique des méthodes plus standardisées, comme le séquençotypage multilocus (www.pasteur.fr/mlst), qui ont un pouvoir discriminant plus limité mais approprié pour l'analyse de collections de souches plus étendues dans le temps et dans l'espace. Tandis que le laboratoire de référence cherche à connaître le contenu en gènes d'intérêt médical (toxines, gènes de résistance), le généticien des populations ne cache pas sa préférence pour les gènes phénotypiquement « neutres », qui seuls permettent de caractériser la diversité et



Point de vue

les relations entre souches de manière fiable.

Mais la situation est en train de changer, et à une vitesse « vertigineuse ». Technologiquement, la possibilité de séquencer les génomes de nombreuses souches bactériennes, virales ou eucaryotes (parasites, champignons/levures) est à notre portée. En recherche, les génomes complets apportent déjà de nouveaux éclairages sur la diversification clonale et la dynamique des épidémies (Holt *et al.*, 2008). Pour les activités de référence, ces technologies signifient la possibilité concrète d'obtenir la caractérisation ultime d'un pathogène en quelques jours, évitant les approches ciblées et permettant de révéler des caractéristiques génomiques inattendues pouvant déboucher sur de nouveaux diagnostics. Ces nouvelles approches prennent tout leur sens pour des pathogènes émergents (nouvelles souches) ou pour le typage à résolution maximale pour la dissection génétique des épidémies (Beres *et al.*, 2010; Harris *et al.* 2010). Cette nouvelle situation provoque de facto un rapprochement des laboratoires de référence et des spécialistes de biodiversité et évolution. La génération des séquences n'étant plus le facteur limitant, une importance grandissante est portée à l'échantillonnage des souches, que le laboratoire de référence est le mieux placé pour définir. Mais ces possibilités apportent des défis nouveaux : quantité massive de données difficilement gérable, engorgement des serveurs de calcul et de stockage. Surtout, de nouvelles expertises deviennent limitantes : traitement bioinformatique des données et interprétation biologique des séquences. Le développement de la génomique des populations microbiennes souligne l'importance de rapprocher les laboratoires de référence des disciplines de l'évolution et de génomique, tant au niveau des collaborations scientifiques que de la formation. Les laboratoires français de référence peuvent s'appuyer sur une forte communauté française en microbiologie et biologie évolutive et sur un réseau national de plates-formes de génomique performant.

Références bibliographiques

Beres SB, Carroll RK, Shea PR, Sitkiewicz I, Martinez-Gutierrez JC, Low DE, McGeer A, Willey BM, Green K, Tyrrell GJ, Goldman TD, Feldgarden M, Birren BW, Fofanov Y, Boos J, Wheaton WD, Honisch C, Musser JM. 2010. Molecular complexity of successive bacterial epidemics deconvoluted by comparative pathogenomics. *Proceedings of the National Academy of Sciences*, 107(9): 4371-6.

Brisse, S. 2008. L'espèce bactérienne : nécessaire, mais pas suffisante. *Bulletin de la Société Française de Microbiologie*, 23(3): 164-174.

Brisse S, Fevre C, Passet V, Issenhuth-Jeanjean S, Tournebize R, Diancourt L, Grimont P. 2009. Virulent clones of *Klebsiella pneumoniae*: identification and evolutionary scenario based on genomic and phenotypic characterization. *PLoS One*, 4(3): e4982.

Diancourt L, Passet V, Nemec A, Dijkshoorn L, Brisse S. 2010. The population structure of *Acinetobacter baumannii*: expanding multiresistant clones from an ancestral susceptible genetic pool. *PLoS One*, 5(4): e10034.

Dykhuizen, D. E. 1998. Santa Rosalia revisited: why are there so many species of bacteria? *Antonie Van Leeuwenhoek*, 73(1): 25-33.

Fournier P-E, Vallenet D, Barbe V, Audic S, Ogata H, Poirel L, Richet H, Robert C, Mangenot S, Abergel C, Nordmann P, Weissenbach J, Raoult D, Claverie J-M. 2006. Comparative genomics of multidrug resistance in *Acinetobacter baumannii*. *PLoS Genetics*, 2(1): e7.

Harris SR, Feil EJ, Holden MTG, Quail MA, Nickerson EK, Chantratita N, Gardete S, Tavares A, Day N, Lindsay JA, Edgeworth JD, de Lencastre H, Parkhill J, Peacock SJ, Bentley SD. 2010. Evolution of MRSA during hospital transmission and intercontinental spread. *Science*, 327(5964): 469-74.

Holt KE, Parkhill J, Mazzoni CJ, Roumagnac P, Weill F-X, Goodhead I, Rance R, Baker S, Maskell DJ, Wain J, Dolecek C, Achtman M, Dougan G. 2008. High-throughput sequencing provides insights into genome variation and evolution in *Salmonella* Typhi. *Nature Genetics*, 40(8): 987-93.

Qin, J., R. Li, J. Raes, *et al.* 2010 A human gut microbial gene catalogue established by metagenomic sequencing. *Nature*, 464(7285): 59-65.

Ragon M, Wirth T, Hollandt F, Lavenir R, Lecuit M, Le Monnier A, Brisse S. 2008. A new perspective on *Listeria monocytogenes* evolution. *PLoS Pathogens*, 4(9): e1000146.

Touchon, M., C. Hoede, O. Tenaillon, *et al.* 2009. Organised genome dynamics in the *Escherichia coli* species results in highly diverse adaptive paths. *PLoS Genetics*, 5(1): e1000344.

Walk ST, Alm EW, Gordon DM, Ram JL, Toranzos GA, Tiedje JM, Whittam TS. 2009. Cryptic lineages of the genus *Escherichia*. *Applied and Environmental Microbiology*, 75(20): 6534-44.

Wu, D., P. Hugenholtz, K. Mavromatis, *et al.* 2009. A phylogeny-driven genomic encyclopaedia of Bacteria and Archaea. *Nature*, 462(7276): 1056-60.