

longtemps. Une multitude de génomes entiers de bactéries et d'archées ont été séquencés. Quand on a comparé le génome d'une espèce bactérienne à celui de nombreux clones provenant de celle-ci, on s'est aperçu qu'il existe beaucoup plus de gènes au niveau de l'espèce qu'au niveau de l'organisme. Tout se passe comme si une bactérie comportait d'une part des gènes essentiels – des gènes informationnels clés pour la réplication, la transcription et la traduction – et, d'autre part, des gènes qui changent d'un individu à l'autre, agissant plus sur le métabolisme et la pathogénicité.

Ainsi est apparue une notion ahurissante, qui n'existe pas chez les eucaryotes *a priori* : le pangénome – l'ensemble de ces gènes changeants (voir l'article de Bianca Sclavi et John Herrick, pages 34 à 41). Cette notion contient celle de transfert horizontal, c'est-à-dire d'échanges de gènes entre organismes. L'idée existait déjà pour les chloroplastes et les mitochondries, mais tout à coup, on se rendait compte que ce phénomène se produit peut-être en permanence. Comment se font ces transferts horizontaux ? Via les virus, probablement, ou au moins en grande partie. Les virus sont ainsi de retour dans l'histoire du vivant...

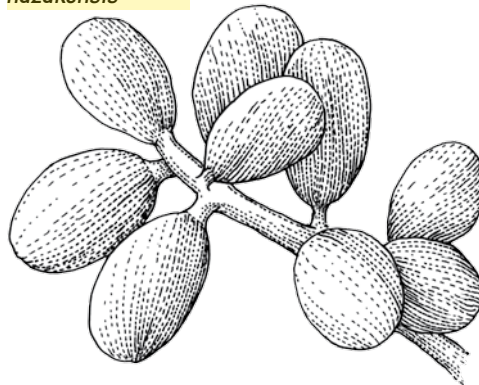
La notion de transfert horizontal a amené à penser le vivant en termes non plus de phylogénie, mais de réseau. On a commencé à comprendre pourquoi trouver une racine à l'arbre du vivant de Woese était impossible : on imaginait que les gènes étaient obtenus par descendance, par transfert vertical, alors que sans doute beaucoup provenaient de transferts horizontaux. L'arbre du vivant cachait en fait une forêt.

PLS

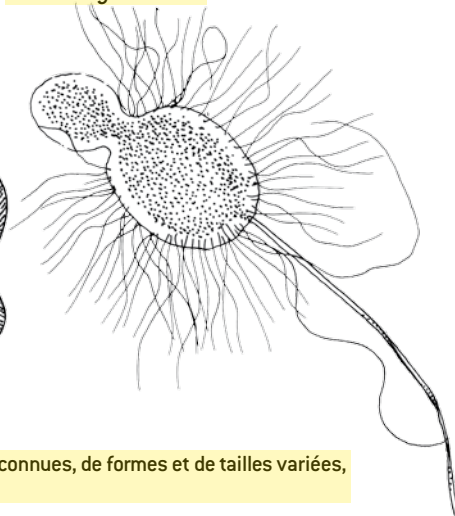
Si l'arbre du vivant est un réseau, peut-on encore l'étudier ?

H. L. G. : Il y a deux écoles. La première considère qu'il est toujours possible de faire une phylogénie des archées et des bactéries : l'idée est que si l'on est capable de repérer les gènes transférés horizontalement, il suffit de les éliminer et de ne garder que les gènes informationnels, qui ont beaucoup plus de chance d'être transmis verticalement. C'est alors ceux-là que l'on utilise pour faire une phylogénie, sans être brouillé par les gènes transmis horizontalement. C'est ce que font par exemple Céline Brochier-Armanet à l'université Lyon-1 ou Simonetta Gribaldo à l'institut Pasteur.

Thermosporothrix hazakensis



Planctomyces maris



LE GROUPE DES BACTÉRIES compte 16 760 espèces connues, de formes et de tailles variées, et de voies métaboliques très diversifiées.

L'autre école est celle de Bill Martin, de l'université de Düsseldorf, en Allemagne, que résume bien le titre d'un article qu'il a publié en 2006, *The tree of one percent* (L'arbre de 1%). Si vous enlevez tous les gènes que vous pensez transmis horizontalement, vous ne gardez que 1 % du génome et vous construisez votre phylogénie avec 1 % des gènes. Est-ce vraiment raisonnable ? L'idée est de dire que les phylogénies ne tiennent pas la route. Les arbres que l'on construit sur les archées ou les bactéries servent à faire une classification, mais pas une phylogénie. Ils ne reflètent pas l'histoire puisque celle-ci est plus compliquée. Il faut utiliser les réseaux.

PLS

Concrètement, comment étudie-t-on les réseaux ?

H. L. G. : Ce champ d'étude mêle trois domaines : la biologie moléculaire pour collecter des génomes à comparer, la bio-informatique pour la gestion des données et les mathématiques pour l'interprétation. Il a explosé avec le séquençage à haut débit. La biologie moléculaire a engrangé beaucoup de séquences de gènes et d'ARN à la fin des années 1980 et au début des années 1990, mais à l'époque, tout était encore fait à la main. Puis, dans les années 1990, on a commencé à avoir des génomes entiers, notamment le génome humain à la fin des années 1990 et au début des années 2000.

Aujourd'hui, obtenir le génome d'une bactérie ne coûte que 1 000 euros. D'énormes quantités de séquences sont ainsi engrangées. Dans notre

laboratoire, par exemple, Philippe Lopez et Éric Baptiste traitent ainsi des millions de séquences. Il y a quelques années, le goulet d'étranglement était l'obtention des séquences, maintenant c'est le traitement de l'information, tant son volume est grand.

L'interprétation des données, quant à elle, n'est pas plus simple. Elle fait appel à l'étude des réseaux d'un point de vue mathématique, notamment à la théorie des graphes. Il s'agit d'un domaine extrêmement ardu des mathématiques, aussi les biologistes travaillent-ils avec des mathématiciens purs. Le domaine n'avance pas vite, mais il rassemble une communauté de plus en plus importante de chercheurs, notamment canadiens, américains, irlandais, anglais et allemands.

PLS

Qu'a apporté cette notion de réseau ?

H. L. G. : Elle a montré que la vie est beaucoup plus complexe qu'on ne le pensait. Le point clé est que les transferts horizontaux continuent toujours à se produire. Cela change complètement la vision des microorganismes. Si vous imaginez des archées dans les sédiments marins ou ailleurs, il faut les penser avec des échanges de gènes. Les chercheurs essaient ainsi de tirer des enseignements de ces échanges actuels pour reconstituer une histoire des bactéries et des archées. L'idée est la suivante : si les transferts horizontaux sont aussi importants aujourd'hui, ils l'étaient certainement davantage auparavant. Cette idée a permis de clarifier notre compréhension des eucaryotes.