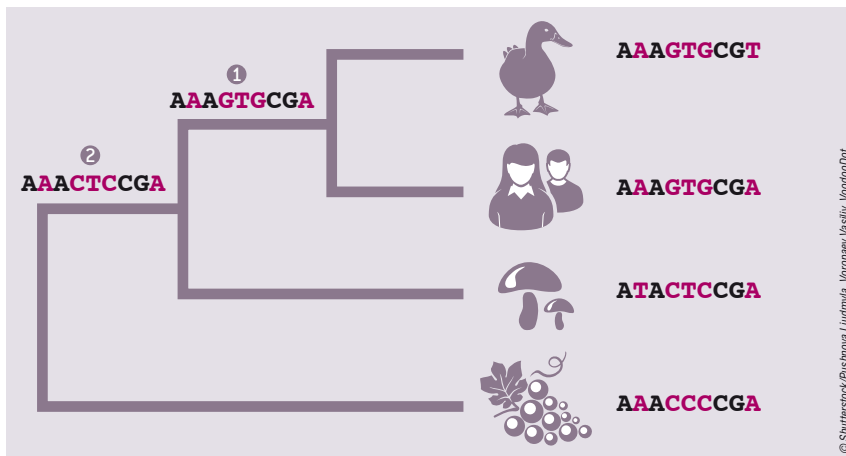




**2. SUR CET ARBRE PHYLOGÉNÉTIQUE SIMPLIFIÉ**, on peut reconstruire, à partir des séquences d'un même gène repérées chez le canard, l'humain, le cèpe et le raisin, les séquences du même gène de l'ancêtre commun de l'humain et du canard d'une part (1), et de l'ancêtre des animaux et des champignons d'autre part (2). On suppose que les ressemblances sont dues à un héritage commun plutôt qu'à des événements indépendants. Les séquences les plus proches sont celles du canard et de l'humain (une seule mutation). La séquence du même gène chez leur ancêtre commun sera donc AAAGTGCGA ou AAAGTGCGT. Les séquences du cèpe et du raisin portant un A à la fin, la première option est sans doute la bonne. La séquence de l'ancêtre commun du cèpe, du canard et de l'humain est celle qui minimise le nombre de mutations entre cet ancêtre et le dernier parent dans l'arbre, le raisin, soit AAATCCGA.

## BIBLIOGRAPHIE

G. J. Szöllösi *et al.*, Lateral gene transfer from the dead, *Systematic Biology*, vol. 62, pp. 386-397, 2013.

J.-G. Dumas *et al.*, Théorie des codes, Dunod, 2013.

G. J. Szöllösi *et al.*, Phylogenetic modeling of lateral gene transfer reconstructs the pattern and relative timing of speciations, *PNAS*, vol. 109, pp. 17513-17518, 2012.

S. Abby *et al.*, Lateral gene transfer as a support for the tree of life, *PNAS*, vol. 109, pp. 4962-4967, 2012.

B. Boussau *et al.*, Parallel adaptations to high temperatures in the Archaean eon, *Nature*, vol. 456, pp. 942-945, 2008.

E. Zuckerkandl et L. Pauling, Molecules as documents of evolutionary history, *Journal of Theoretical Biology*, vol. 8, pp. 357-366, 1965.

sens ou dans l'autre, alors sa fréquence augmente ou diminue dans une population au fil des générations. D'autre part, il n'est pas systématique qu'une mutation induise un avantage ou un inconvénient pour la reproduction. Dans certains génomes, par exemple ceux des mammifères ou de certaines plantes qui sont de longs textes verbeux, la plupart des mutations sont sans effet immédiat. Dans ce cas, la fréquence des génomes qui portent une mutation suit une marche aléatoire sans pression de l'environnement. C'est le principe de la dérive génétique.

Que cette variation soit influencée ou non par la sélection naturelle, la fréquence finit toujours par atteindre 0 ou 1. Si elle atteint 0, la mutation a disparu de la population ; il n'en reste pas de trace dans les génomes. Si elle atteint 1, tous les génomes de la population sont porteurs de la mutation. Dès lors, il ne reste pas de trace dans les génomes de leur état ancestral, avant la mutation.

Avec un nombre réaliste de mutations par génération (une centaine), la dérive remplace la plupart des lettres du texte génomique en quelques millions de générations. Ainsi, on ne peut retrouver des traces que de génomes ayant au plus quelques millions d'années – ce qui représente une

partie infime de l'histoire du monde vivant. En revanche, la sélection permet une meilleure conservation de l'information : on retrouve dans les génomes actuels des traces de l'origine du vivant, il y a presque quatre milliards d'années.

Pour reconstituer l'information ancienne, on compare plusieurs génomes actuels dans un cadre phylogénétique. Par exemple, prenons les génomes d'une variété de vigne, d'un cèpe, d'un canard et d'un humain (voir la figure 2). Les ressemblances entre séquences permettent de reconstituer les relations de parenté entre ces espèces, ce qui se traduit par un arbre phylogénétique, c'est-à-dire une généalogie des espèces. Cette généalogie implique que le canard et l'humain ont un ancêtre commun qui leur est exclusif et qui descend lui-même de l'ancêtre qu'ils partagent avec le cèpe. On identifie alors des morceaux (des gènes) suffisamment similaires pour que l'on suppose qu'ils descendent d'une même séquence ancestrale : on trouve ATACTCCGA dans la vigne, AAATCCGA chez le cèpe, AAAGTGCGA chez le canard et AAAGTGCGT chez l'humain.

À partir de l'analyse de ces séquences actuelles d'un même gène, on reconstitue les séquences du gène chez les ancêtres communs des espèces considérées en supposant que les ressemblances entre espèces sont plus fréquemment dues à leur héritage commun qu'à des événements indépendants qui auraient donné le même résultat : dans notre exemple, cette hypothèse se traduit par un scénario unique où chacune des positions variables (en rouge) n'a changé qu'une seule fois au cours de l'évolution des espèces.

En multipliant le nombre de gènes et de génomes comparés (aujourd'hui on peut avoir accès aux séquences de plusieurs milliers de génomes, chacun contenant des milliers de gènes, voir l'article de G. Perrière dans ce numéro) et avec l'aide de modèles plus réalistes prenant en compte nos connaissances des processus biologiques, on peut ainsi estimer des séquences de génomes ancestraux et décrire l'histoire des mutations.

En pratique, on sait aujourd'hui reconstituer les séquences quasi complètes d'ancêtres récents, par exemple le génome de la bactérie qui fut probablement responsable des grandes épidémies de peste du Moyen Âge, à l'aide du séquençage de ses descendantes, des

souches de *Yersinia pestis* contemporaines. Pour des ancêtres plus éloignés, tel celui des mammifères qui vivait il y a une centaine de millions d'années, une partie de l'information a été perdue par dérive. Cependant, pour les parties du génome contraintes par la sélection, les gènes en particulier, la reconstruction reste tout à fait possible. On peut même connaître le nombre de chromosomes et la façon dont les gènes étaient organisés dans ces génomes. À partir de ces données, on commence à comprendre les modes de vie des espèces disparues (température optimale de croissance, profondeur préférée d'anciens mammifères marins, utilisation de l'oxygène par des bactéries anciennes, alimentation, taille ou longévité de mammifères anciens, histoire des migrations et mélanges humains...), un peu comme si l'on retrouvait le code d'un vieux programme et que l'on reconstituait son fonctionnement à partir de ses versions nouvelles (voir la figure 3).

Cependant, l'application de ces méthodes à l'ensemble du monde vivant

se heurte à une difficulté : nombre d'organismes ont la capacité d'intégrer dans leur génome des gènes de leurs contemporains, proches apparentés ou non. Ces transferts horizontaux violent les hypothèses de reconstruction des génomes ancestraux, qui supposent que l'ancêtre commun des séquences appartient à l'ancêtre commun des espèces.

### Le transfert de gène : obstacle ou indice ?

Il est apparu que chaque gène, quelle que soit sa fonction, pouvait, à un moment ou un autre de son histoire, avoir été transféré. À long terme, aucun gène n'a donc la même histoire que les organismes qui le portent. Ce constat semblait remettre en cause la possibilité théorique de reconstruire l'évolution du vivant à partir des gènes.

Ainsi, le gène était trop « individualiste » pour refléter l'histoire des organismes. N'avait-il pour autant rien à en raconter ? De fait, lorsqu'on observe l'ensemble des

gènes et, pour peu que l'on sache interpréter les singularités que révèle chaque histoire de gènes, on s'aperçoit que les transferts horizontaux contiennent une information inédite sur l'histoire du vivant. D'une part, ils peuvent venir à l'appui d'une hypothèse phylogénétique : si un ancêtre a reçu un gène par transfert, ce gène doit être observé chez ses descendants. Ainsi, la présence du gène peut être un caractère partagé permettant de reconnaître les organismes qui sont proches.

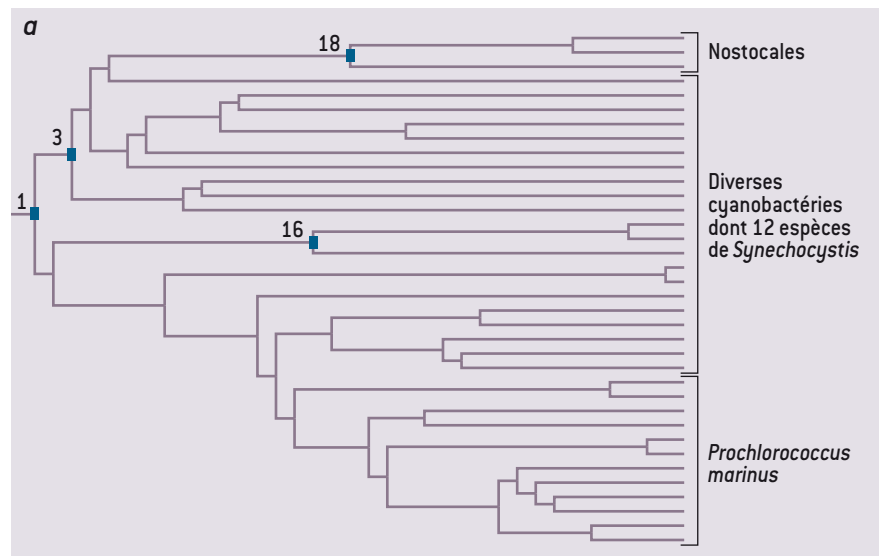
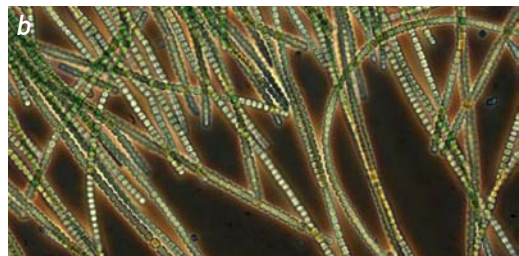
D'autre part, un échange de gènes entre deux organismes ne peut avoir lieu que s'ils sont contemporains. Le transfert horizontal contient donc une information sur les âges relatifs des lignées qui échangent des gènes : on ignore à quelle époque exactement vivaient les organismes anciens, mais, avec des relations de descendance et des échanges entre espèces contemporaines, on peut espérer établir une chronologie de l'histoire de la vie. Or les transferts ont lieu principalement chez les organismes microscopiques (bactéries, virus, champignons unicellulaires, autres

## TRANSFERT DE GÈNES ET DATATION

Par définition, le transfert de gènes entre deux espèces n'est possible que si les deux espèces sont contemporaines. Ce processus, qui transparaît dans le génome des espèces actuelles, donne aux phylogénéticiens de précieuses informations sur les âges relatifs d'espèces anciennes dont les traces fossiles sont rares. En 2012, à partir de 36 génomes d'espèces différentes, les auteurs ont ainsi reconstitué la phylogénie chronologique des cyanobactéries, les premiers organismes à avoir fait de la photosynthèse et produit de l'oxygène en grande quantité (a). Les plantes ne récupèrent cette faculté en intégrant des cyanobactéries au fonctionnement de leurs cellules.

Les nœuds de l'arbre sont positionnés de gauche à droite selon leur chronologie. Le nœud 1, le plus ancien, est parfois daté à 3,8 milliards d'années. Les plantes ne figurent pas sur cet arbre, mais on voit la naissance au nœud 18 du clade des cyanobactéries multicellulaires, les nostocales, qui forment des filaments verts de cellules différenciées (b). Comme ce clade a reçu, par transfert, des gènes d'autres mor-

ceaux de la phylogénie, on estime que cette forme de multicellularité est relativement récente. De même, la présence de nombreux transferts du clade du nœud 3 vers le clade du nœud 16 lui donne une origine relativement récente, alors que les hypothèses habituelles plaçaient cette diversification à la racine des cyanobactéries.



Adapté de G. J. Szallasi et al., PNAS, vol. 109, pp. 17513-17518, 2012