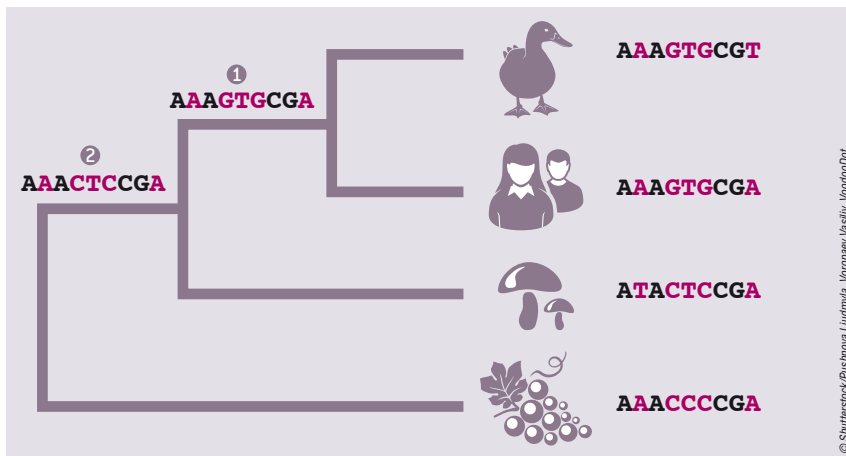




© Shutterstock/Pushnova Ludmila, Voropaev Vasily, VoodooDot

**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 AAAGTGC GA ou AAAGT GCG T. 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 AAAC TCC GA.

## 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 ATACTC CGA dans la vigne, AAAC CCG A chez le cèpe, AAAGTGC GA chez le canard et AAAGT GCG T 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