

Aux origines du règne animal

01/07/11

La phylogénie a pour but d'établir les liens de parenté entre les organismes de notre planète et de dresser l'arbre du vivant. Depuis quelques années, une variante « haut débit » de cette discipline a vu le jour : la phylogénomique. Grâce au séquençage complet du génome de nombreux organismes, les scientifiques sont aujourd'hui capables de comparer une multitude de gènes chez diverses espèces. Si la phylogénomique a déjà permis de grandes avancées dans la reconstruction de l'arbre du vivant, il arrive aussi qu'elle parvienne à des conclusions erronées, comme le souligne Denis Baurain dans un article publié en collaboration dans *PLoS Biology*.

À quoi ressemblait le premier animal apparu sur Terre ? Question simple à laquelle il reste pourtant difficile de donner une réponse. Diverses disciplines scientifiques, chacune avec leurs moyens, tentent de remonter la trace des êtres vivants jusqu'au premier animal. La plus connue de ces disciplines est la paléontologie. Les paléontologues reconstituent l'histoire de la vie terrestre à partir de vestiges du passé, les fossiles. En effet, la multitude de créatures qui se sont succédé sur Terre y ont laissé des traces que les paléontologues recherchent et interprètent. Il arrive fréquemment que les journaux relatent la découverte d'un nouveau fossile témoignant de l'existence passée de l'une ou l'autre espèce.



Fossile de Dickinsonia, emblématique de la faune d'Ediacara.
Les fossiles de la période édiacarienne (de 635 à 541 millions d'années) conservent la trace des premiers organismes macroscopiques à morphologie complexe. Précédant immédiatement l'explosion de diversité observée au Cambrien, certains d'entre eux sont vraisemblablement voisins des ancêtres d'animaux contemporains. Ainsi, Dickinsonia aurait été capable de contraction musculaire et de déplacement sur le fond de la mer, ce qui l'a fait classer entre autres parmi les cnidaires (méduses et affiliés) et les polychètes (grand groupe de vers annélides). Pour d'autres auteurs, c'est un « pré-animal » dont il n'existe aucun descendant aujourd'hui, voire un lichen (association d'une algue et d'un champignon) ! L'échelle en bas à gauche de l'image (en cm) donne une idée de la taille de l'organisme. Source de l'image : Wikipedia.

Moins connue du grand public, une autre discipline scientifique s'attèle à remonter le temps du vivant : la phylogénie. « *La phylogénie est la science qui cherche à identifier les relations de parenté entre les êtres vivants avec pour but ultime de reconstruire l'arbre du vivant depuis l'origine de la vie* », explique Denis Baurain, chercheur à l'Unité de Génomique Animale du GIGA et à la Faculté de Médecine Vétérinaire de l'ULg. « Dans le cas des animaux, nous aimerions découvrir à quoi ressemblait notre dernier ancêtre commun et savoir s'il était simple ou complexe », poursuit le chercheur pour qui la phylogénie est une véritable passion.

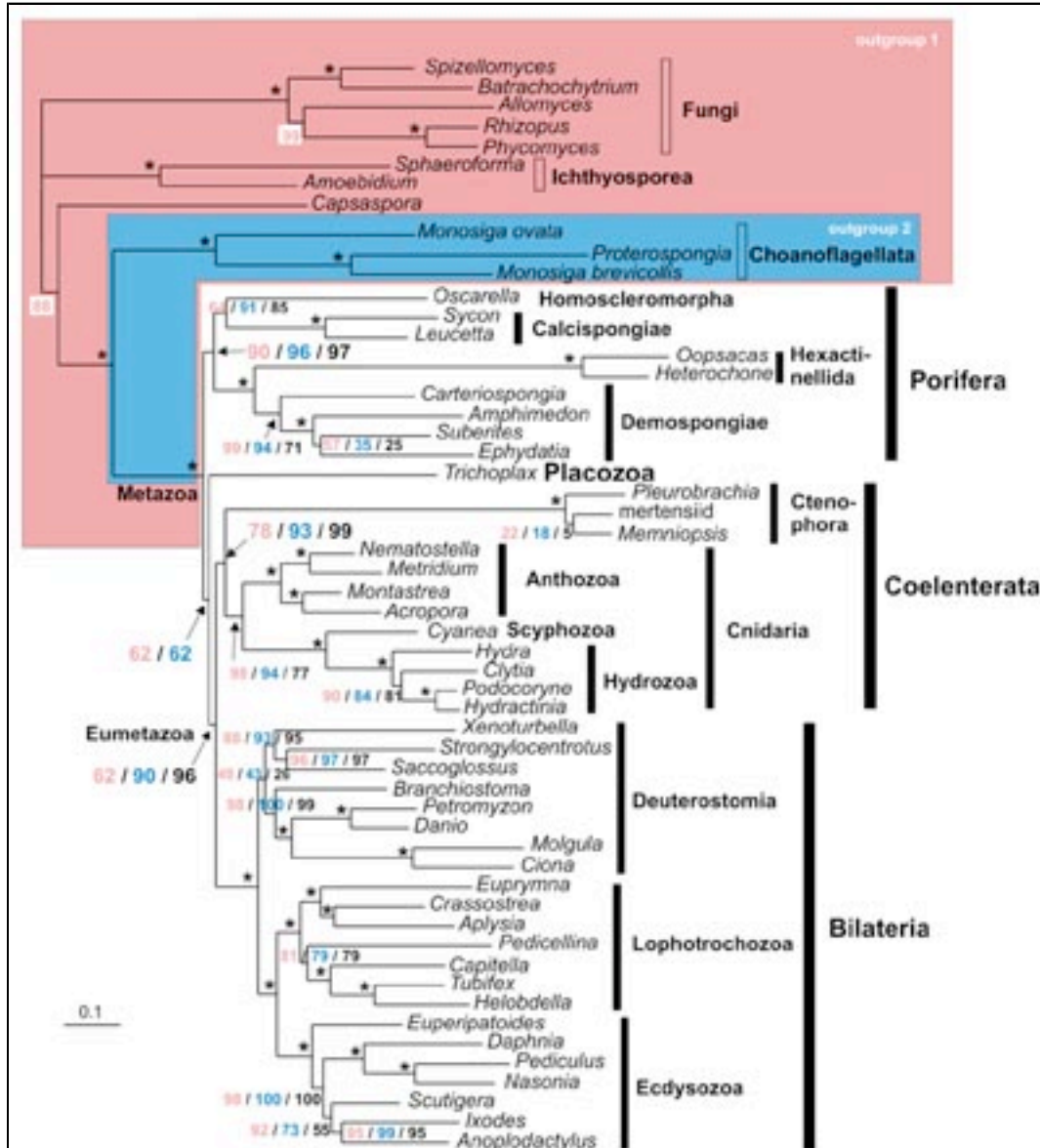
Les relations de parenté, établies au fur et à mesure de la progression des connaissances scientifiques, sont représentées par un arbre phylogénétique. Sur cet arbre, un peu à l'image d'un arbre généalogique,

les organismes actuels occupent les feuilles terminales tandis que chacun des noeuds internes représente l'ancêtre éteint des branches en dérivant et qui symbolisent ses descendants. Le nom d'un noeud est celui du **clade** formé par cet ancêtre et tous ses descendants et non celui de l'ancêtre en question, qui est impossible à identifier car il y a peu de chances que ce soit une espèce fossile en particulier.

La phylogénie en évolution

La phylogénie est née avec la théorie de l'évolution de **Darwin**. « *Au départ, l'étude des liens de parenté entre les êtres vivants se basait exclusivement sur la comparaison des caractéristiques morphologiques des animaux, présents ou fossiles* », indique Denis Baurain. Les scientifiques comparaient ainsi la structure externe des organismes pour juger de leurs affinités évolutives. A la fin des années 60, la phylogénie moléculaire a pris le relais. En effet, les spécialistes en phylogénie utilisent désormais les macromolécules biologiques telles que l'**ADN**, l'**ARN** ou les **protéines**, composants fondamentaux de tous les organismes vivants, pour établir leurs liens de parenté.

Les **gènes** utilisés pour faire la comparaison doivent être choisis avec soin car les différentes régions de l'ADN des organismes n'évoluent pas toutes à la même vitesse. Pour étudier les relations entre les êtres vivants à une large échelle de temps, il est nécessaire d'utiliser des gènes pour lesquels le taux de mutation observé d'une espèce à l'autre est faible. Le plus souvent, il s'agit des gènes assurant des fonctions vitales pour la cellule et sur lesquels la pression de sélection est importante, telle que la synthèse des protéines.



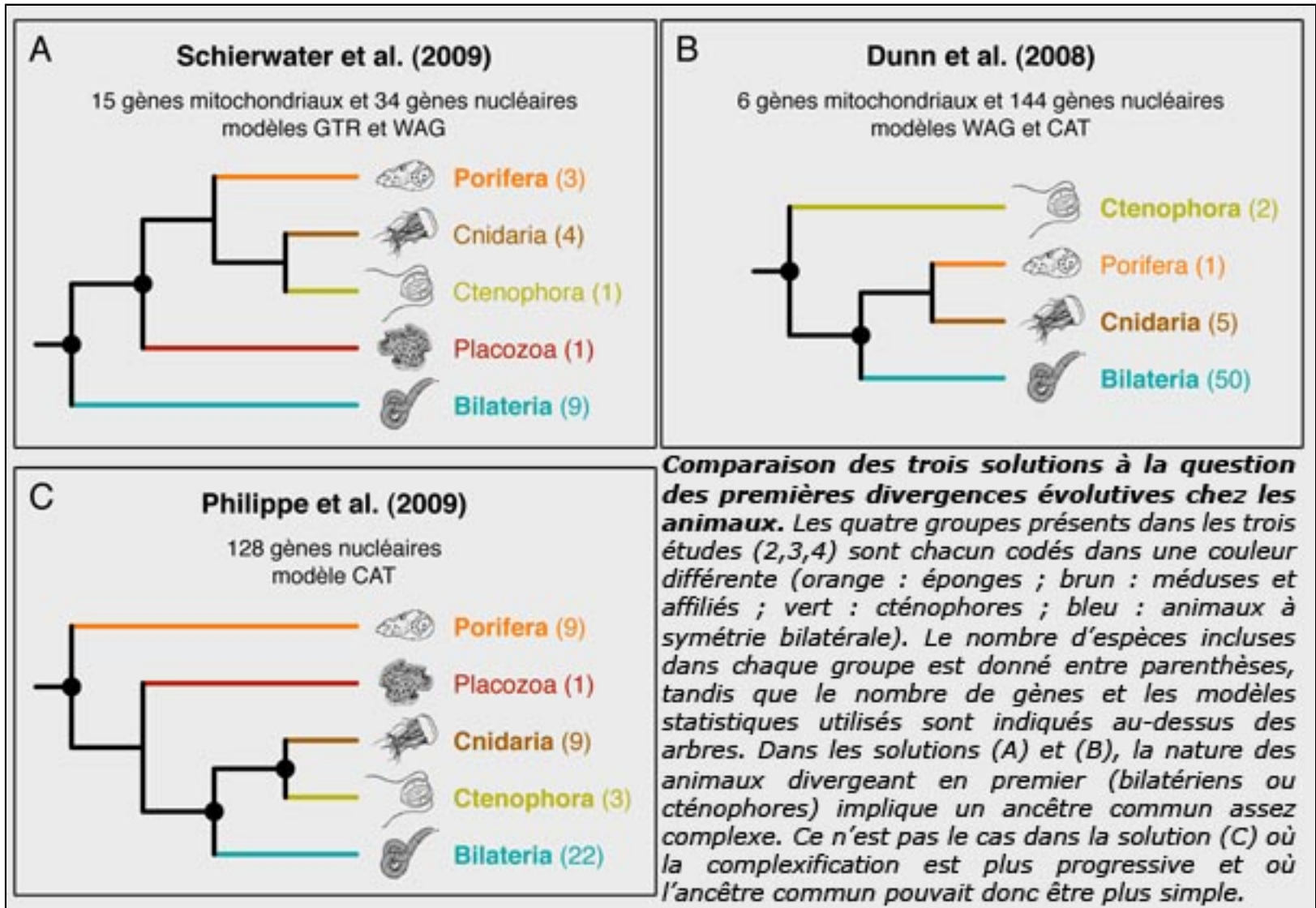
Phylogénie des animaux se concentrant sur les divergences les plus anciennes. Il s'agit de l'arbre phylogénomique correspondant à la « bonne solution » (voir Fig.suivante tableau C). Il a été calculé sur 128 gènes nucléaires en utilisant une approche bayésienne et le modèle CAT. Les animaux sont sur fond blanc (un seul vertébré, le poisson zèbre *Danio rerio*), tandis que les groupes en bleu et rose correspondent aux organismes actuels les plus proches des animaux (dont les choanoflagellés et les champignons). Les noms des différents clades sont donnés à droite de l'arbre. La fiabilité de la reconstruction à chaque noeud interne est indiquée sous forme de trois pourcentages (ou d'une étoile lorsque ces trois valeurs sont maximales).

Jusqu'il y a peu, la phylogénie moléculaire se limitait à comparer l'une ou l'autre région de l'ADN de plusieurs espèces afin de pouvoir les positionner les unes par rapport aux autres sur un arbre phylogénétique. « Depuis 8 ou 9 ans, une nouvelle

*variante de la phylogénie a fait son apparition et est de plus en plus utilisée pour dresser l'arbre du vivant. Il s'agit de la phylogénomique », précise Denis Baurain. Les nouvelles méthodes de **séquençage de génomes** ont permis d'augmenter radicalement le nombre de gènes disponibles pour réaliser les études de phylogénie moléculaire. Aujourd'hui, la disponibilité de génomes complets, mais surtout de transcriptomes (ensemble des ARN messagers) obtenus à un coût dérisoire, rend possible l'étude simultanée de centaines de gènes pour lesquels il existe une **homologie** d'une espèce à l'autre. « *En principe, comparer plusieurs gènes en parallèle chez plusieurs espèces est une technique d'autant plus fiable que le nombre de gènes comparés et que l'échantillon d'espèces considéré sont grands* », continue le scientifique. Bien entendu, ces grands jeux de données requièrent des capacités de calcul très importantes, telles que celles offertes par NIC3, l'ordinateur parallèle de puissance de l'ULg.*

Des analyses plus poussées mais pas sans failles

Si la phylogénomique a certes permis de faire un bond en avant dans la compréhension des liens de parenté entre espèces, elle ne constitue cependant pas une solution miracle. « *Cette approche présente aussi des désavantages. Il ne suffit pas d'augmenter le nombre de séquences de gènes à comparer pour obtenir un arbre fiable* », explique Denis Baurain. « *Encore faut-il utiliser des données bien contrôlées et des méthodes d'analyse sophistiquées* », insiste le chercheur. C'est précisément cela que Denis Baurain, en collaboration avec une équipe canadienne de l'Université de Montréal, souligne dans un article publié dans *PLoS Biology* (1). A quoi ressemblait l'ancêtre commun de tous les animaux en existence ? C'est-à-dire avant l'explosion cambrienne au cours de laquelle la plupart des grands embranchements actuels ont vu le jour. C'est à cette question que trois études phylogénomiques concurrentes (2,3,4) ont tenté de répondre. « Or, ces dernières sont arrivées à trois conclusions différentes dont une seule est correcte (Fig. ci-dessous) », ajoute Denis Baurain. Ce dernier et ses collègues d'Outre-Atlantique ont alors pris l'initiative de ré-examiner critiquement les données et les méthodes d'analyse utilisées dans les deux études aux conclusions erronées. « *Nous y avons soit corrigé les données, soit utilisé une méthode d'analyse plus appropriée et, dans chaque cas, nous sommes parvenus à la conclusion correcte* », poursuit Denis Baurain. « *Le problème de ces deux études venait d'une accumulation d'erreurs dans les données combinée à de mauvaises techniques d'analyse* ».

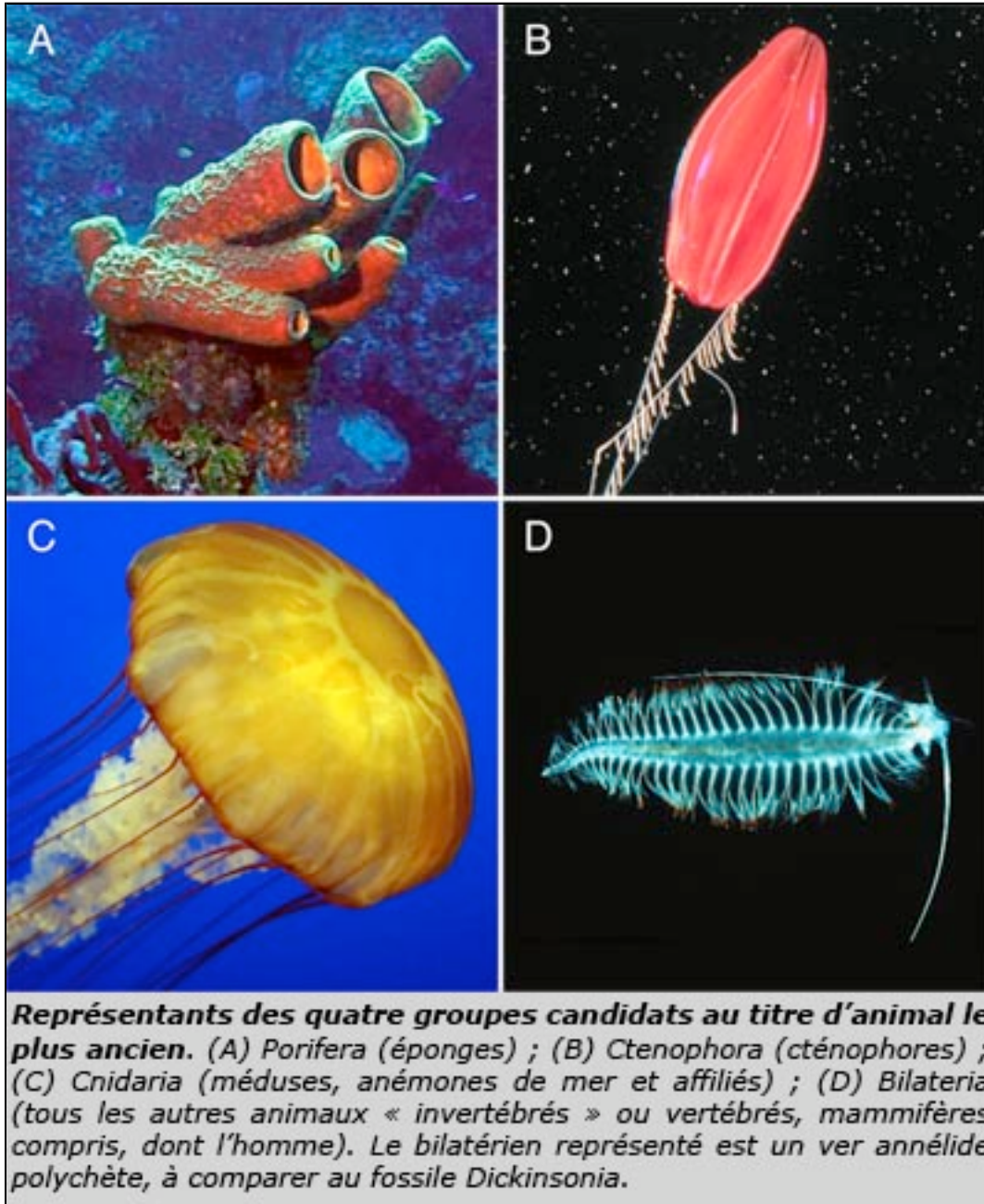


En augmentant la taille de l'échantillon de séquences de gènes, la phylogénomique permet avant tout de réduire l'erreur stochastique, soit l'erreur qui résulte de l'usage d'une quantité trop faible de données. Mais un autre type d'erreur prend alors le dessus : l'erreur systématique. Egalement appelé signal non-phylogénétique, ce type d'erreur est lié aux approximations des modèles statistiques utilisés pour analyser les données. « Quand on modélise la réalité, dans ce cas-ci les processus évolutifs, il faut utiliser des modèles capables de rendre compte de la complexité de ces processus », explique Denis Baurain.

Gare aux pièges de l'histoire évolutive

Lorsqu'un gène chez une espèce ressemble à un gène chez une autre espèce, ces deux gènes sont très probablement issus d'un ancêtre commun et sont alors dits « homologues ». Plus les gènes homologues présentent de différences, plus cet ancêtre est éloigné dans le temps. C'est sur base de ces homologies que les scientifiques parviennent à reconstruire peu à peu l'arbre du vivant. Mais l'histoire évolutive peut jouer

des tours à ceux qui tentent de remonter sa trace. Il arrive en effet que des espèces présentent des gènes similaires qui n'ont pourtant pas été hérités directement d'un ancêtre commun à ces espèces. « *C'est comme le dauphin et le requin* », compare Denis Baurain. « *Ces animaux se ressemblent superficiellement car ils répondent à des contraintes fonctionnelles comparables. Pourtant, ils n'ont rien à voir l'un avec l'autre et n'ont pas d'ancêtre commun récent, le premier étant un mammifère proche des hippopotames, le second un poisson cartilagineux vieux de 400 millions d'années* ». Au niveau moléculaire, ce type d'évolution convergente se produit également. Ainsi, deux gènes homologues éloignés peuvent finir par se ressembler fortement suite à une série de mutations indépendantes. C'est ce qu'on appelle l'**homoplasie**. Plus grave, certains gènes homologues connaissent des problème d'orthologie, c'est-à-dire existent en plusieurs copies dans certains génomes, voire ne représentent pas fidèlement leur génome pour cause de transfert de gènes entre espèces parfois très distantes (OGMs naturels).



On comprend dès lors la confusion que cela peut entraîner pour les analyses phylogénomiques ! Afin d'éviter d'être induit en erreur par ces phénomènes, il est donc crucial de contrôler scrupuleusement l'orthologie des séquences de gènes utilisées et d'avoir recours à une modélisation qui tienne compte de l'homoplasie. « C'est ce que nous avons fait avec les deux études phylogénomiques pour lesquelles nous avons des doutes quant aux conclusions », explique Denis Baurain.

Une fois le tir rectifié, les résultats des trois études sur l'évolution des organismes avant le Cambrien abondent dans le même sens : « ils révèlent que l'histoire évolutive des premiers animaux montre une complexité croissante, depuis les éponges jusqu'aux animaux à symétrie bilatérale en passant par les méduses », indique Denis Baurain. « Cela paraît logique, mais toutes les parties de l'arbre du vivant ne vont pas nécessairement

dans ce sens », poursuit-il. « *La levure du boulanger, par exemple, est dérivée par simplification secondaire d'un champignon qui était à l'origine beaucoup plus complexe qu'elle* ».

Les explosions de biodiversité sous la loupe

« *Nous sommes en quelque sorte des méthodologistes empiristes. Nous identifions les zones les plus floues et compliquées de l'arbre du vivant et nous y attaquons les unes après les autres pour finir par dresser un arbre complet* », indique Denis Baurain. Outre l'homoplasie et les problèmes d'orthologie, les spécialistes en phylogénie sont confrontés à d'autres difficultés. « *A l'aube du Cambrien, voici 550 millions d'années, il y a probablement eu une explosion de la diversité animale. Or, il est parfois difficile de retracer la chronologie d'évènements qui se déroulent de manière si rapide* », continue le chercheur. Les épisodes de diversification brutale donnent en effet du fil à retordre aux spécialistes. Durant ces épisodes les organismes évoluent vite pour rester compétitifs et coloniser de nouvelles niches écologiques. C'est ce qu'on appelle la radiation adaptative. « *Cela a été le cas notamment pour les mammifères modernes, qui se sont diversifiés en très peu de temps* », précise Denis Baurain. De plus, comme la levure, il arrive que les êtres vivants évoluent dans une direction régressive, c'est-à-dire perdent en complexité par rapport à leurs ancêtres. « *C'est pourquoi les parasites sont souvent pris à tort pour des organismes primitifs* ».

Homoplasie, duplication et transferts de gènes, radiation adaptative, régression, l'histoire évolutive est loin d'être un long fleuve tranquille aisé à remonter. Mais, la phylogénomique - « *à condition d'être utilisée convenablement* », souligne Denis Baurain - est très efficace et devrait permettre aux spécialistes d'élucider un à un les mystères de l'arbre du vivant.

