

Chapitre 13

Analyser les arbres phylogénétiques pour construire des scénarios évolutifs

Cours

PLAN DU CHAPITRE

- 1 L'établissement de l'arbre du monde vivant
- 2 L'arbre phylogénétique des eucaryotes et son interprétation
- 3 Les virus ont-ils une place dans l'arbre du vivant ?

ZOOM

- 1 Caractéristiques comparées des trois grands domaines du vivant
- 2 Les apomorphies des 9 clades majeurs d'eucaryotes
- 3 Origine et évolution des mitochondries
- 4 Diversité ultrastructurale et biochimique des plastides des eucaryotes

INTRODUCTION

La paléontologie fournit des preuves des changements qu'ont connus les organismes au cours des temps passés, et l'étude de la biodiversité actuelle montre que celle-ci continue de varier, par disparition de certains taxons et apparition d'autres. Nous disposons ainsi d'un enregistrement d'une histoire évolutive qui se poursuit encore maintenant. Malheureusement, cet enregistrement est fragmentaire et les scénarios qui se sont joués dans le passé ne peuvent pas être observés. Afin de contourner cet écueil, nous nous appuyons sur la comparaison d'un nombre toujours plus important de **séquences homologues** d'ADN, ARN, et protéines ; l'analyse des caractères moléculaires observés permet la reconstruction des relations de parenté entre espèces. Les **arbres phylogénétiques** obtenus nous renseignent sur les scénarios d'évolution de différents caractères, par exemple morpho-anatomiques, géographiques, cytologiques, chromosomiques, ou encore génétiques.

- Quels sont les grands clades du monde vivant ?
- Comment s'est construite la cellule eucaryote ?
- Que nous apprennent les phylogénies quant aux scénarios évolutifs et modalités de changement des organismes cellulaires et des virus ?

1 L'établissement de l'arbre du monde vivant

1.1 Les premiers arbres : une représentation hiérarchique de la biodiversité

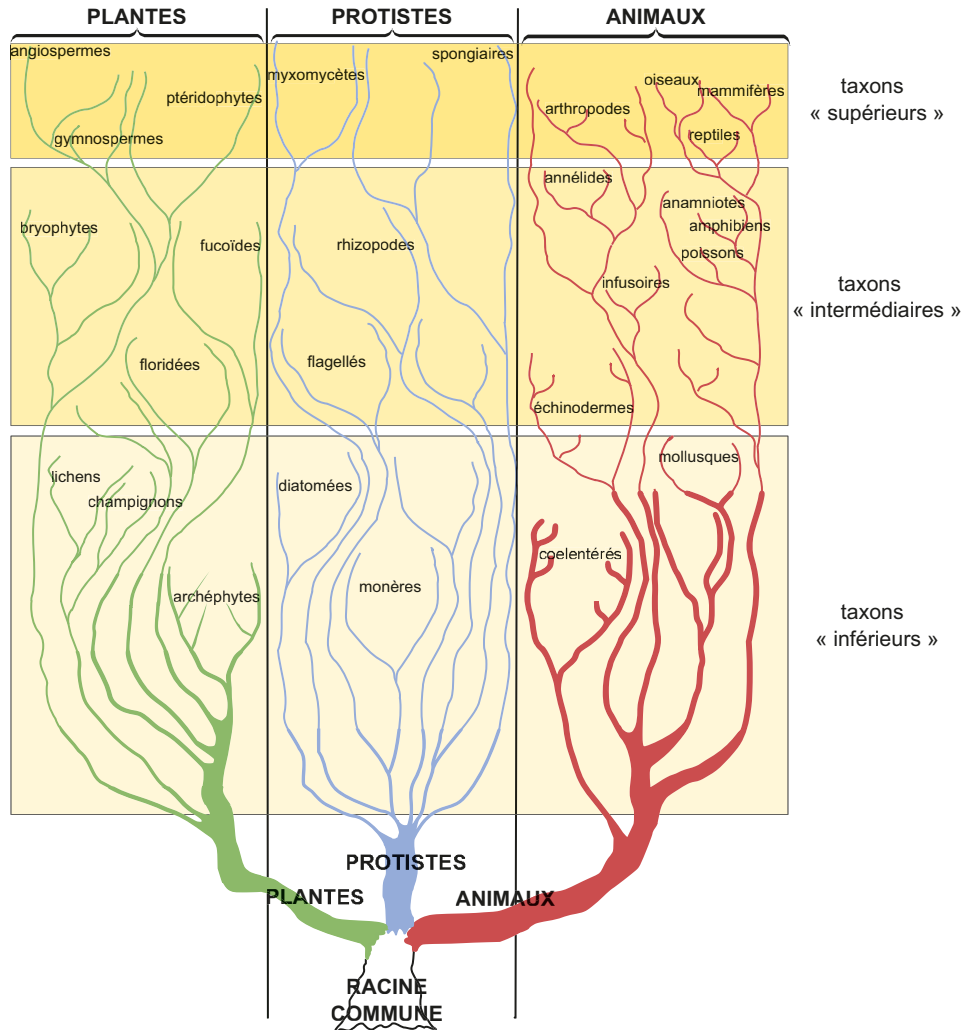


Figure 13.1 Représentation simplifiée de l'arbre de classification d'Ernst Haeckel (1866).

Seuls quelques groupes dont la dénomination a disparu dans les classifications actuelles et qui correspondent à des taxons rencontrés dans le programme de BCPST sont mentionnés ci-après : archéphytes : groupe de chlorobiontes ; floridées : groupe de rhodobiontes ; ficoïdes : groupe de straménopiles ; infusoires : groupe d'alvéolobiontes (ciliés) ; monères : ensemble regroupant des bactéries et des eucaryotes unicellulaires (terme abandonné de nos jours).

Dans son *Systema Naturae* (1758), Linné subdivise le monde vivant en deux règnes : les animaux et les végétaux. Les naturalistes y rajoutent le règne des protistes vers le milieu du XIX^e siècle, en raison de la difficulté à classer certains organismes tels les euglènes qui sont à la fois mobiles et photosynthétiques. Si, dans cette classification, les animaux ne posent que peu de problèmes de définition, le cas des végétaux apparaît plus complexe. En effet, la cellule végétale possède la faculté de synthétiser de la matière organique à partir de matière minérale en utilisant l'énergie

lumineuse via la photosynthèse. Cependant, la photoautotrophie n'est pas l'apanage des végétaux car d'autres organismes tels des unicellulaires eucaryotes ou des bactéries en sont capables. À l'inverse, certains végétaux ont perdu secondairement leur chlorophylle, telles la plupart des plantes parasites qui ne font plus de photosynthèse. Les champignons, à l'époque classés comme des végétaux, sont quant à eux hétérotrophes et leur mode de nutrition, par absorption de matière organique (absorbotrophie), est assez proche de celui des animaux.

À la fin du XIX^e siècle, dans la foulée des travaux de Charles DARWIN et Alfred Russel WALLACE, une première classification des plantes, protistes et animaux sous forme d'arbre a été proposée par Ernst HAECKEL. Rejetant la conception d'un monde vivant qui serait organisé selon une stricte hiérarchie linéaire — « l'Échelle de la Nature » —, il conçoit l'**évolution** comme un processus divergent et donc arborescent. En 1866, il représente l'histoire des êtres vivants sous la forme d'un arbre évoquant une généalogie (figure 13.1).

Cette construction est qualifiée de phylogénie mais HAECKEL donne à ce mot un sens bien différent de l'actuel : il s'agit pour lui de montrer la succession des étapes évolutives dans une lignée donnée. Il assimile arbitrairement le principe de hiérarchisation (ou subordination) des caractères, sur lequel est fondée la hiérarchie du système linnéen, à une chronologie évolutive tournée vers une complexification croissante et progressiste des organismes. L'arbre basé sur la hiérarchie de la distribution des caractères devient alors une représentation des **liens de parenté entre les organismes vivants**.

Sur l'**arbre de Haeckel**, le monde vivant est séparé en trois règnes équilibrés : les plantes, les animaux, et les protistes. Ce troisième règne comprend tous les organismes qui ne sont ni des plantes, ni des animaux : il inclut ainsi divers unicellulaires, qu'ils soient procaryotes ou eucaryotes, cette dernière distinction n'étant à l'époque pas encore établie. La hiérarchie des caractères induit une hiérarchie des organismes, qui se retrouve dans la représentation de l'arbre, avec des strates (du bas vers le haut) de taxons considérés comme inférieurs (car plus simples), intermédiaires, puis supérieurs (car plus complexes). L'évolution ne procédant pas nécessairement par plus de complexité, cette représentation a été abandonnée, et les arbres phylogénétiques modernes placent tous les taxons sur un même pied d'égalité.

1.2 La révolution moléculaire et l'arbre des ARN ribosomiques

a) Les apports des caractères moléculaires

À l'heure actuelle, seules des approches moléculaires qui s'appuient sur l'analyse d'un très grand nombre de gènes et protéines peuvent permettre d'appréhender la complexité du monde vivant. Ainsi, depuis le début des années 1980, l'arbre de la vie a été révolutionné par la possibilité de comparer des séquences homologues d'ADN, d'ARN, et de protéines. Le succès de ces caractères moléculaires résulte de quatre propriétés essentielles.

- **Ils sont ubiquistes.** Alors que les mêmes caractères morpho-anatomiques ne sont plus comparables dès que les plans d'organisation des organismes deviennent trop différents les uns des autres, les gènes et protéines impliqués dans les fonctions fondamentales des cellules (expression génétique, métabolisme) restent comparables, même à l'échelle des différents règnes du vivant. Le seul point à surveiller est l'utilisation de séquences d'autant plus conservées que les relations de parenté à reconstruire sont anciennes. En effet, des séquences à évolution rapide seraient trop différentes les unes des autres pour pouvoir en extraire de l'information phylogénétique.
- **Ils sont objectifs.** Différents naturalistes à qui l'on fournit les mêmes réactifs et matériels de séquençage obtiendront les mêmes séquences d'ADN en partant d'un même échantillon. En revanche, ces mêmes naturalistes ne percevront pas exactement les mêmes caractères morpho-anatomiques s'ils observent un même squelette : l'expérience et l'habitude de chacun(e) pourront conduire à une certaine subjectivité de description des caractères.

Voir ouvrage
de 1^{re} année,
chapitre 18, § 1

Voir TP 9, § 1

- **Ils sont abondants.** Le moindre gène ou la moindre protéine comprend des centaines voire des milliers de nucléotides ou d'acides aminés. Comparées chez différents organismes, ces séquences livrent au biologiste autant de caractères moléculaires analysables. Et à l'heure du séquençage massif des génomes, transcriptomes, et protéomes, les caractères moléculaires disponibles pour la phylogénie peuvent se compter désormais en centaines de milliers voire en millions. Or, lorsqu'il s'agit de reconstituer l'histoire évolutive des organismes, le fait de disposer d'un plus grand nombre de caractères offre l'opportunité de fournir un nombre de synapomorphies qui excédera le nombre des homoplasies existantes (convergences et réversions) et contribuera ainsi à une résolution plus fiable des phylogénies.
- **Ils se prêtent bien à la modélisation** des processus évolutifs. Estimer fidèlement la phylogénie des organismes passe nécessairement par la compréhension des modalités évolutives des caractères qui sont utilisés comme source d'information. Alors que les caractères morpho-anatomiques sont très souvent bien différents les uns des autres, les quatre nucléotides de l'ADN et de l'ARN et les 20 acides aminés fondamentaux des protéines ont des modalités évolutives parfaitement comparables, tant d'une séquence à l'autre que d'un taxon à l'autre. Par exemple, de l'ADN viral de coronavirus, de l'ADN chloroplastique d'angiosperme, de l'ADN nucléaire de champignon ou de mammifère, et de l'ADN bactérien ou archéen subissent tous un excès de mutations de type transitions (échanges entre deux purines ou entre deux pyrimidines) par rapport aux transversions (échanges entre une purine et une pyrimidine, et réciproquement).

b) Les trois domaines du monde vivant

Les premiers travaux de phylogénie moléculaire appliqués à l'identification des grands clades de l'arbre de la vie ont été initiés par Carl Woese et coll. et reposent sur l'analyse des séquences homologues des ARN de la petite sous-unité des ribosomes (ARNr 16S chez les bactéries et ARNr 18S chez les eucaryotes). Du fait de son ubiquité dans le monde vivant, de son importante conservation de séquence, ainsi que des quantités cellulaires disponibles pour son extraction, cet ARN ribosomique est un excellent marqueur pour les analyses phylogénétiques à grande échelle évolutive. C'est là l'origine d'une découverte qui bouleverse notre conception de l'origine de la vie sur Terre (figure 13.2). Le monde « procaryote » se révélait extrêmement hétérogène et scindé en deux sous-ensembles majeurs, à l'époque dénommés les eubactéries (cyanobactéries, bactéries Gram + et -) et les archéobactéries (bactéries méthanogènes, halophiles, thermoacidophiles ou hyperthermophiles).

ZOOM 1
Caractéristiques comparées des trois grands domaines du vivant

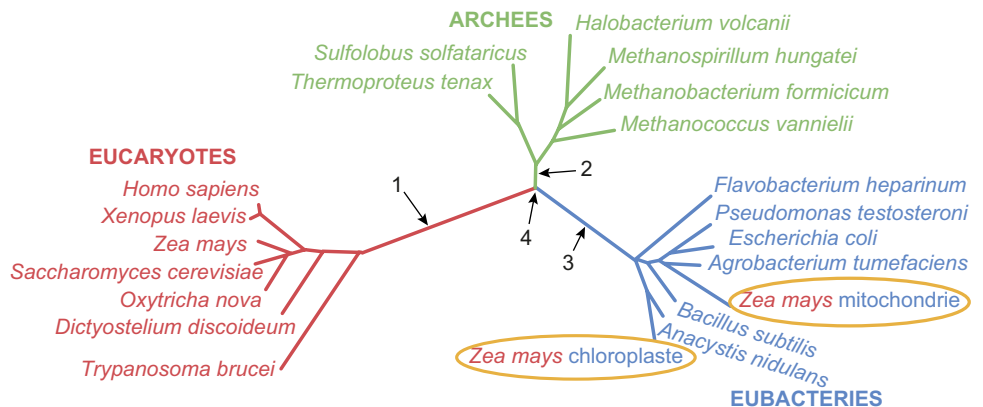


Figure 13.2 Arbre phylogénétique des ARN ribosomiques 16S/18S.

Cet arbre montre les trois domaines du monde vivant (eucaryotes, archées et bactéries) ainsi que leurs importantes divergences moléculaires (modifié d'après Pace et coll., 1986). Les affinités respectivement bactériennes et cyanobactériennes des ARNr mitochondriaux et chloroplastiques sont indiquées en orange. Les flèches indiquent 4 positions possibles pour la racine du monde vivant.



En 1986, Woese et Olsen intègrent à cette phylogénie du vivant les séquences homologues des ARNr 16S de mitochondries et de (chloro)plastides, et montrent qu'elles ne se placent pas au sein même des séquences eucaryotes mais, au contraire, à l'intérieur des eubactéries (figure 13.2). Cet arbre constitue l'un des arguments en faveur d'une origine bactérienne des mitochondries et des plastides.

Notons que l'arbre de la vie n'est pas raciné car aucun groupe externe au monde vivant ne possède d'ARN ribosomiques et d'autres techniques d'enracinement s'avèrent ici inopérantes.

c) Quelle est la position de la racine du vivant ?

Actuellement **cinq possibilités d'enracinement** de l'arbre phylogénétique du vivant peuvent être proposées. Les trois premières enracent l'arbre du vivant le long de chacune des trois branches conduisant à chacun des trois domaines, de manière à ce que chacun des trois domaines soit monophylétique. Si la racine se situe le long de la branche ancestrale des eucaryotes (figure 13.2 : flèche 1), les archées et les bactéries formeraient le groupe traditionnellement reconnu des procaryotes. Si la racine se situe le long de la branche ancestrale des archées (figure 13.2 : flèche 2), les bactéries et les eucaryotes formeraient un groupe monophylétique. Si la racine se situe le long de la branche ancestrale des bactéries (figure 13.2 : flèche 3), les archées et les eucaryotes formeraient alors un clade. Par ailleurs, certains auteurs proposent que dans le contexte d'un réseau du vivant — plutôt que d'un arbre du vivant — cela n'a pas de sens de chercher lequel des trois domaines est le plus proche parent de l'autre, et qu'au contraire, il faut concevoir une racine du vivant ayant donné naissance simultanément aux trois domaines (figure 13.2 : flèche 4). Depuis le milieu des années 2010, une cinquième possibilité apparaît dans la littérature scientifique, celle du fait que les eucaryotes se nichent au sein des archées, rendant paraphylétiques ces dernières. En d'autres termes, les eucaryotes seraient plus proches de certains clades d'archées que les archées ne le seraient entre elles. Aucune des racines proposées ne fait l'objet de consensus.

1.3 L'arbre de la vie ou le réseau du vivant ?

Le concept même d'arbre du vivant est aujourd'hui débattu. Cet arbre est-il l'image de l'évolution des espèces ? Il retrace avant tout l'évolution de l'ARNr de la petite sous-unité ribosomique au sein du vivant. Et à partir d'une phylogénie de gènes ou de protéines, est-il possible d'en déduire une phylogénie d'espèces ?

a) L'évolution réticulée

En 1970, Lynn Margulis revisite une hypothèse ancienne proposée au début du XX^e siècle : l'origine endosymbiotique des plastides et des mitochondries. La cellule eucaryote résulterait d'interactions symbiotiques entre cellules d'origines phylogénétiques complètement distinctes. Bien que certains aspects de cette théorie soient encore controversés, par exemple l'origine du noyau, nous verrons que de nombreux arguments accréditent ce point de vue (voir § 2.3). Ainsi, l'intégralité des relations évolutives entre les organismes vivants ne peut-elle probablement pas être figurée uniquement sous forme d'un arbre.

De plus, selon les gènes utilisés, les estimations phylogénétiques sont parfois contradictoires car les inférences sont affectées par l'un des mécanismes fondamentaux de variations génétiques, les **transferts horizontaux de gènes** (THG). En effet, les THG sont un moteur prépondérant de l'évolution des procaryotes. Ils affectent également les génomes eucaryotes. Or, l'analyse de l'évolution génique par des méthodes phylogénétiques classiques s'appuie sur une transmission verticale de l'information génétique qui correspond à la descendance avec modifications. Cette démarche est donc délicate à mettre en œuvre pour établir l'arbre du vivant car l'évolution liée aux THG ne correspond pas à une structure en arbre mais bien plus à des échanges d'informations complexes organisés en réseau.

Voir TP 9, § 1

Voir chapitre 12,
figure 12.16

L'existence d'événements de THG et de symbiose — ainsi que les hybridations entre taxons distincts — fait que l'évolution du monde vivant est parfois réticulée : sa représentation la plus adéquate doit alors faire appel à un **réseau** plutôt qu'un arbre.

! Attention !

Dans un arbre, il n'existe qu'un seul trajet le long des branches pour passer d'un nœud à l'autre. Dans un réseau, il y a plusieurs trajets pour passer d'un nœud à l'autre. Alors que la transmission verticale du matériel génétique de génération en génération se prête bien à une représentation arborescente, seule une représentation réticulée de l'évolution du monde vivant permet de figurer les THG (dont nous savons qu'ils sont fréquents dans certains groupes taxonomiques comme les actinobactéries et gamma-protéobactéries) ainsi que les événements d'hybridation.

b) L'arbre du 1 % des gènes codants

Une étude pionnière conduite par Ciccarelli et coll. (2006) montre qu'il est possible d'extraire de 191 génomes, balayant les trois domaines du monde vivant, un sous-ensemble de 31 protéines homologues dont les séquences codantes n'ont pas été affectées par des événements de THG. La concaténation de cette trentaine de protéines permet alors d'estimer un arbre du vivant qui ne reflète pas uniquement la seule information des ARN ribosomiques. Cependant, considérant qu'un génome bactérien code environ 3 000 protéines différentes, il apparaît que cet arbre du vivant est fondé sur 1 % du nombre moyen de protéines présentes dans une cellule de type procaryote. La topologie proposée n'est donc pas le reflet de l'arbre du vivant mais plutôt celle de l'arbre du 1 % des gènes codants ! C'est probablement insuffisant pour inférer des conclusions générales sur l'évolution de la vie.

2 L'arbre phylogénétique des eucaryotes et son interprétation

2.1 Le pouvoir résolutif des analyses phylogénomiques

Voir ouvrage
de 1^{re} année,
chapitre 13, § 1.3

Les programmes de séquençage complet de génomes fournissent les séquences brutes des gènes, que le séquençage des transcriptomes permet d'annoter. Les analyses phylogénétiques sont ainsi entrées dans l'ère de la **phylogénomique**. Cette approche est nommée ainsi dans la mesure où l'inférence phylogénétique bénéficie ici de l'apport d'une quantité d'information d'ampleur génomique. Elle consiste à combiner l'information apportée par plusieurs dizaines, centaines, voire milliers de séquences, notamment des séquences codantes peu ou pas du tout soumises à des THG. Il est courant de penser, à juste titre, que la concaténation de nombreux gènes ou de nombreuses protéines permet d'additionner les homologies, c'est-à-dire le signal phylogénétique présent dans chacun des marqueurs moléculaires pris indépendamment. Nous oublions souvent que le bruit de fond évolutif — c'est-à-dire les homoplasies (convergences et réversions) — n'est pas le même d'un gène ou d'une protéine à l'autre : une convergence conduisant par hasard aux mêmes changements de nucléotides ou d'acides aminés chez deux taxons qui n'ont aucun lien phylogénétique l'un avec l'autre n'a qu'extrêmement peu de chance de survenir dans un autre gène ou une autre protéine exactement chez les deux mêmes taxons ! La combinaison de nombreux marqueurs va donc diluer le bruit de fond évolutif, là où le signal évolutif — qui existe dans chaque marqueur car ces derniers sont liés par la même histoire évolutive des taxons — va s'additionner jusqu'à fournir suffisamment de synapomorphies pour identifier avec confiance de nombreux clades. Le bilan net de l'approche phylogénomique est l'augmentation du signal et la dilution du bruit de fond, ouvrant ainsi la perspective de résoudre de nombreux problèmes de relations de parenté, en tout cas si l'évolution des taxons considérée est plus arborescente que réticulée (car peu affectée par des

hybridations ou des THG), et en gardant toujours à l'esprit que seules des méthodes statistiques performantes et sophistiquées sont en mesure de séparer le bon grain (les homologies) de l'ivraie (les homoplasies).

2.2 La diversité phylogénétique des eucaryotes

a) L'identification moléculaire de 9 principaux clades

Durant les deux premières décennies du XXI^e siècle, en se fondant sur un consensus de différentes topologies obtenues par les approches moléculaires d'ampleur phylogénomique, les phylogénéticiens ont identifié 9 principaux clades d'eucaryotes : straménopiles (= hétérocontes), alvéolobiontes, rhizarias, haptophytes, plantae, cryptophytes, excavata, opisthocontes, et amoebozoaires.

Des hypothèses phylogénétiques sont alors proposées quant aux relations de parenté entre ces 9 clades (figure 13.3). L'ensemble des **unicontes** regroupe des organismes dont les cellules possèdent à l'origine un unique flagelle en position postérieure ; il comprend les amoebozoaires (amibes) et les **opisthocontes** (champignons, animaux, et choanoflagellés). Les autres eucaryotes constituent le groupe des **bicontes**, dont les cellules sont fondamentalement biflagellées. Au sein des bicontes, les straménopiles, alvéolobiontes, et rhizarias forment le clade « S-A-R ». La plupart des organismes étudiés en cours ou en TP au cours des deux années de BCPST appartiennent soit aux opisthocontes, soit à deux clades parmi les bicontes : les **plantae** et les **straménopiles**.

b) Quelques enseignements à tirer de l'arbre des eucaryotes

L'inventaire de la diversité phylogénétique des eucaryotes permet de souligner plusieurs points importants.

- Le terme de « **végétal** » classiquement utilisé pour désigner des organismes photosynthétiques ne correspond pas à un clade mais à de très nombreux organismes dispersés dans l'arbre des eucaryotes. Les « végétaux » constituent ainsi un **groupe polyphylétique**.
- Les organismes photoautotrophes à thalles, vivant fréquemment dans des zones humides, sont qualifiés usuellement par le terme « **algues** ». Cette dénomination désigne des êtres vivants aquatiques réalisant une photosynthèse oxygénique, et les oppose aux « végétaux » terrestres. Pourtant, les « algues » constituent elles aussi un **groupe polyphylétique** car elles se distribuent dans des groupes phylogénétiquement très différents les uns des autres, allant même des eucaryotes (phéophycées, rhodophytes, chlorophytes, euglénophytes) jusqu'aux bactéries (les cyanobactéries ont autrefois été appelées « algues bleues ») !
- Le terme de « **champignons** » désigne aussi un groupe polyphylétique. En effet, des organismes hétérotrophes filamenteux sont observés au sein du clade des **eumycètes** mais également au sein des **oomycètes** qui sont des straménopiles. Les oomycètes (dont fait partie l'agent du mildiou de la luzerne) sont hétérotrophes, parasites ou saprophytes, et ont perdu secondairement leur plaste, mais se distinguent des eumycètes par différents caractères dont l'absence de chitine dans la paroi.
- Le terme courant « **amibe** » désigne des unicellulaires eucaryotes appartenant à divers taxons non apparentés (rhizarias, amoebozoaires) et qui ont comme similitude le fait d'être des hétérotrophes phagotrophes.

Dans un contexte de biologie évolutive, de systématique et de phylogénie, les termes « végétal », « algue », « champignon », et « amibe » sont donc à proscrire car ils correspondent à des ensembles polyphylétiques. Les caractéristiques définissant les « végétaux », « algues », « champignons », et « amibes » sont apparues plusieurs fois indépendamment dans l'histoire des eucaryotes : ce sont des **convergences évolutives**.

ZOOM 2

Les apomorphies des 9 clades majeurs d'eucaryotes

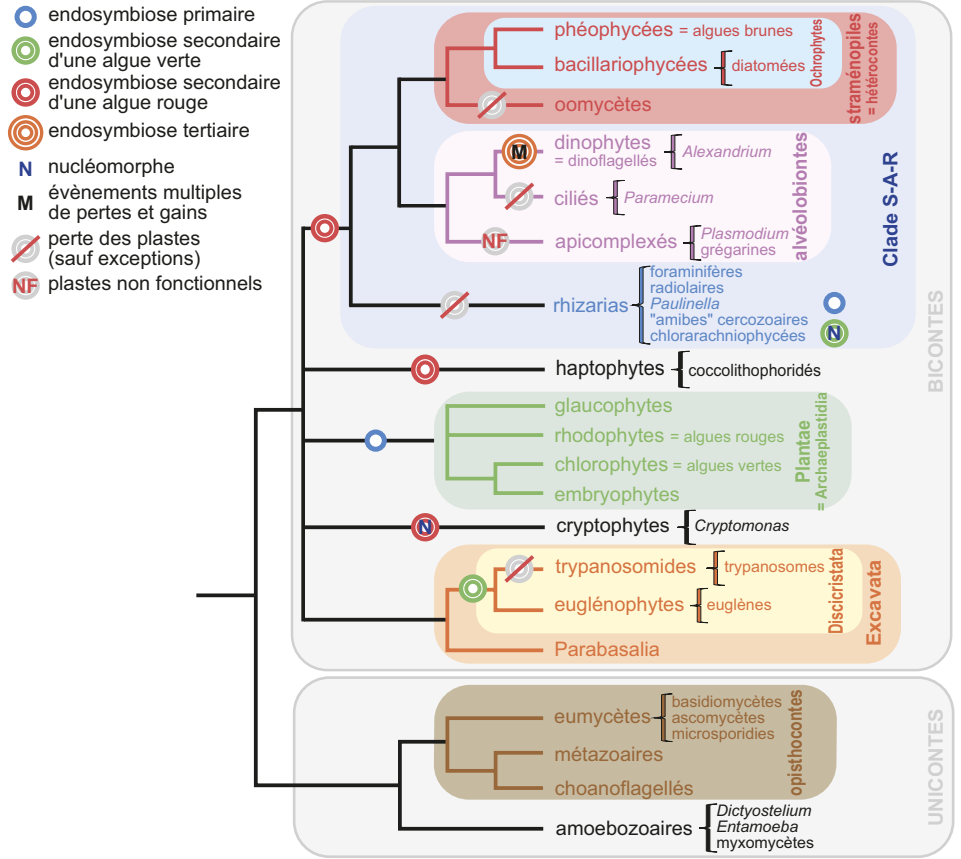


Figure 13.3 Arbre phylogénétique consensuel des eucaryotes fondé sur l'analyse phylogénomique de 320 protéines nucléaires homologues concaténées.

L'étonnante variété des plastes ainsi que leur distribution phylogénétique s'explique par des scénarios d'endosymbioses qui sont figurés par les différents symboles. La construction de ces scénarios évolutifs repose sur la robustesse et la fiabilité de la topologie consensuelle analysée ici. Foraminifères, radiolaires et coccolithophoridés ont une grande importance géologique. [Modifié d'après Strassert J. F. H., Irisarri I., Williams T. A. & Burki F. (2021) « A molecular timescale for eukaryote evolution with implications for the origin of red algal-derived plastids ». *Nature Communications* 12:1879. [<https://doi.org/10.1038/s41467-021-22044-z>].

c) La racine des eucaryotes et l'artéfact d'attraction des longues branches

Les difficultés rencontrées au sujet de l'enracinement de l'arbre du vivant (voir § 1.2c) se transposent à la question du positionnement de la racine de l'arbre des eucaryotes. Une position couramment admise pour la racine des eucaryotes implique une dichotomie entre unicontes et bicontes, mais aussi une grande incertitude sur l'ordre de branchement des lignées de bicontes, d'où la multifurcation impliquant cinq clades : les straménopiles + alvéolobiontes + rhizarias (clade « S-A-R »), les haptophytes, les plantae, les cryptophytes, et les excavata (figure 13.3). À noter que derrière cette multifurcation se cachent 105 phylogénies alternatives qui permettraient de connecter ces cinq clades uniquement avec des dichotomies (et ce problème serait encore plus complexe si la monophylie respective des bicontes et des unicontes était remise en cause). Trouver la position de la racine de la phylogénie des eucaryotes — ou des bicontes — suppose l'utilisation d'un groupe externe. Or les groupes externes indubitables les plus proches pour les

eucaryotes sont soit les archées, soit les bactéries. Mais ces deux domaines du monde vivant sont évolutivement très éloignés des eucaryotes, et il est connu que des groupes externes trop éloignés ont tendance à enraciner les groupes internes de manière artificielle, notamment en plaçant la racine le long des branches les plus longues, c'est-à-dire celles correspondants aux taxons à évolution la plus rapide. Cette anomalie méthodologique — à laquelle notamment la cladistique avec maximum de parcimonie est très sensible — porte le nom d'artéfact d'**attraction des longues branches**. Pour en comprendre la cause, il faut se souvenir que des taxons à évolution plus rapide que les autres voient leurs caractères changer plus fréquemment d'état, ce qui génère un plus grand nombre d'homoplasies sous la forme de convergences et de réversions. Or, si des taxons non apparentés d'un point de vue phylogénétique présentent une fréquence élevée de convergences et réversions, les biologistes risquent d'interpréter ces similitudes de caractères comme des synapomorphies potentielles qui auraient alors tendance à les rapprocher de manière artificielle. Et malheureusement, plus les taxons concernés évoluent rapidement — ou plus ils sont anciens, comme c'est le cas des groupes externes —, plus le nombre d'homoplasies accumulées augmente, et plus le nombre de « fausses synapomorphies » détectées augmente lui aussi, renforçant notre confiance dans un ensemble taxonomique artéfactuel, qui ne fait que regrouper les taxons sur la base de leur vitesse d'évolution et non sur leurs véritables relations de parenté.

2.3 L'origine endosymbiotique de la cellule eucaryote

La diversité des eucaryotes va de pair avec une évolution complexe, profondément marquée par des interactions symbiotiques entre micro-organismes, ayant conduit à l'apparition des mitochondries et des plastes. Les génomes de ces organites semi-autonomes caractéristiques de la cellule eucaryote s'apparentent aux génomes bactériens. Mitochondries et plastes possèdent des ribosomes qui ont une composition similaire à celle des ribosomes bactériens. La séquence de l'ARNr 16S est fortement conservée au sein des plastes, mitochondries et bactéries et elle est à ce titre utilisée comme marqueur phylogénétique. Nous avons ainsi vu (figure 13.2) que les phylogénies moléculaires placent les ARNr 16S des mitochondries et des plastes non pas au sein même des séquences eucaryotes mais à l'intérieur des bactéries. C'est un argument majeur en faveur de l'origine endosymbiotique de la cellule eucaryote. D'un point de vue sémantique, le terme « endosymbiose », ou plus exactement « **endocytobiose** », fait référence à la symbiose entre deux cellules dont l'une est internalisée dans l'autre.

Quels ont été les mécanismes évolutifs qui ont conditionné ces associations symbiotiques et leur évolution chez les eucaryotes ?

a) Une origine unique des mitochondries mais plusieurs pertes secondaires

En 1890, le biologiste allemand Richard Altmann postule que les mitochondries ont une origine bactérienne. Le séquençage de l'ADN mitochondrial (ADNmt) a largement confirmé cette hypothèse ancienne basée sur une simple considération cytologique. Certains résultats suggèrent une parenté des mitochondries avec les bactéries alpha-pourpres (protéobactéries non sulfureuses), proches du groupe des *Rickettsia* (*R. typhi* est l'agent du typhus). Nombre d'espèces de *Rickettsia* et de genres apparentés sont parasites ou symbiotes de cellules eucaryotes ce qui tend à accréditer l'hypothèse d'une endocytobiose. D'autres résultats suggèrent que les mitochondries pourraient constituer le groupe-frère de toutes les alpha-protéobactéries. Dans tous les cas, l'acquisition des mitochondries résulte probablement d'un **événement unique** car la plupart des gènes mitochondriaux fondamentaux sont présents chez tous les eucaryotes. En revanche, en raison de l'importance des transferts géniques — endosymbiotiques ou horizontaux — ayant eu lieu entre les génomes mitochondriaux et nucléaires, il est

ZOOM 3

Origine et évolution
des mitochondries

difficile de dresser le portrait-robot de la première cellule eucaryote chez laquelle l'endocytobiose mitochondriale s'est produite.

Si la mitochondrie a été acquise une seule fois au cours de l'histoire des eucaryotes, son évolution ultérieure est en revanche complexe et marquée par de nombreux événements indépendants de perte de gènes et de simplification structurale et fonctionnelle. L'évolution des cellules eucaryotes a donc été marquée par des phases de complexification mais aussi de simplification.

b) L'endosymbiose primaire d'une cyanobactérie est à l'origine des plastides des plantae

L'endosymbiose primaire d'une cyanobactérie

De nombreux arguments ultrastructuraux et moléculaires ont conduit à formuler l'hypothèse d'une homologie entre le plaste des plantae et la cellule cyanobactérienne. Cette hypothèse est étayée par trois arguments phylogénétiques principaux.

- Les plantae ont un ancêtre commun unique. L'ensemble des taxons formés par les glaucophytes, les rhodophytes, les chlorophytes et les embryophytes forme un clade bien soutenu (figure 13.3).
- Les plastides des plantae sont eux-mêmes monophylétiques. Cela suggère qu'il n'y a pas eu dans ce groupe d'acquisition indépendante de plastides.
- La plupart des gènes plastidiaux des plantae forme un groupe frère avec les gènes de cyanobactéries (cf. le clade chloroplaste + *Anacystis nidulans* sur la figure 13.2). Les gènes plastidiaux sont donc phylogénétiquement indépendants des gènes eucaryotes. Ces analyses phylogénétiques sont cependant difficiles à conduire en raison du « dialogue » symbiotique entre les gènes nucléaires et les gènes bactériens (devenus plastidiaux) qui, après la mise en place de la symbiose, a progressivement modifié chacun des deux génomes.

Des arguments supplémentaires proviennent des nombreuses associations durables qui existent entre bactéries et organismes eucaryotes, comme les cas de rétention transitoire de plastides ingérés dans le cytoplasme et qui sont utilisés temporairement par la cellule hôte.

Aujourd'hui, il est donc bien établi que les plastides à deux membranes ou plastides primaires ont probablement une origine unique. La membrane interne du plaste est homologue de la membrane de la cyanobactérie endocytée ; la membrane externe correspond à celle de la vacuole d'endocytose de l'hôte (figure 13.4A). La comparaison de la composition biochimique des membranes constitue un élément important de cette analyse. Les glaucophytes possèdent entre les deux membranes plastidiales une concentration élevée de peptidoglycane, principal constituant de la paroi cyanobactérienne (figure 13.4C2). Cette enveloppe vestigiale de peptidoglycane fait défaut chez les autres plantes. Ceci suggère que l'endocytobiose s'est accompagnée de modifications profondes des structures du symbiote bactérien (= plaste primaire) (figure 13.4A). Un unique événement d'endosymbiose a ainsi conduit à la mise en place de l'autotrophie au sein des plantae.

L'hôte de la cyanobactérie symbiotique

L'ancêtre commun des plantae était probablement un **unicellulaire hétérotrophe flagellé**, présentant des mitochondries et capable d'un mécanisme de type « phagocytose ». Après l'ingestion d'une cyanobactérie dans une vacuole de phagocytose, le flagellé ne l'a pas digérée et a établi une relation symbiotique avec elle. Puisque les algues rouges sont dépourvues d'appareil flagellaire, nous pouvons supposer que la totalité de cette structure (y compris les centrioles) a été secondairement perdue dans ce clade. L'établissement de la symbiose a ensuite progressivement engendré un dialogue noyau-mitochondries-plastes lié à la redondance d'informations génétiques dont une partie transite par une structure commune : le cytoplasme. Ce point sera repris plus loin.

ZOOM 4

Diversité
ultrastructurale et
biochimique des
plastides des eucaryotes

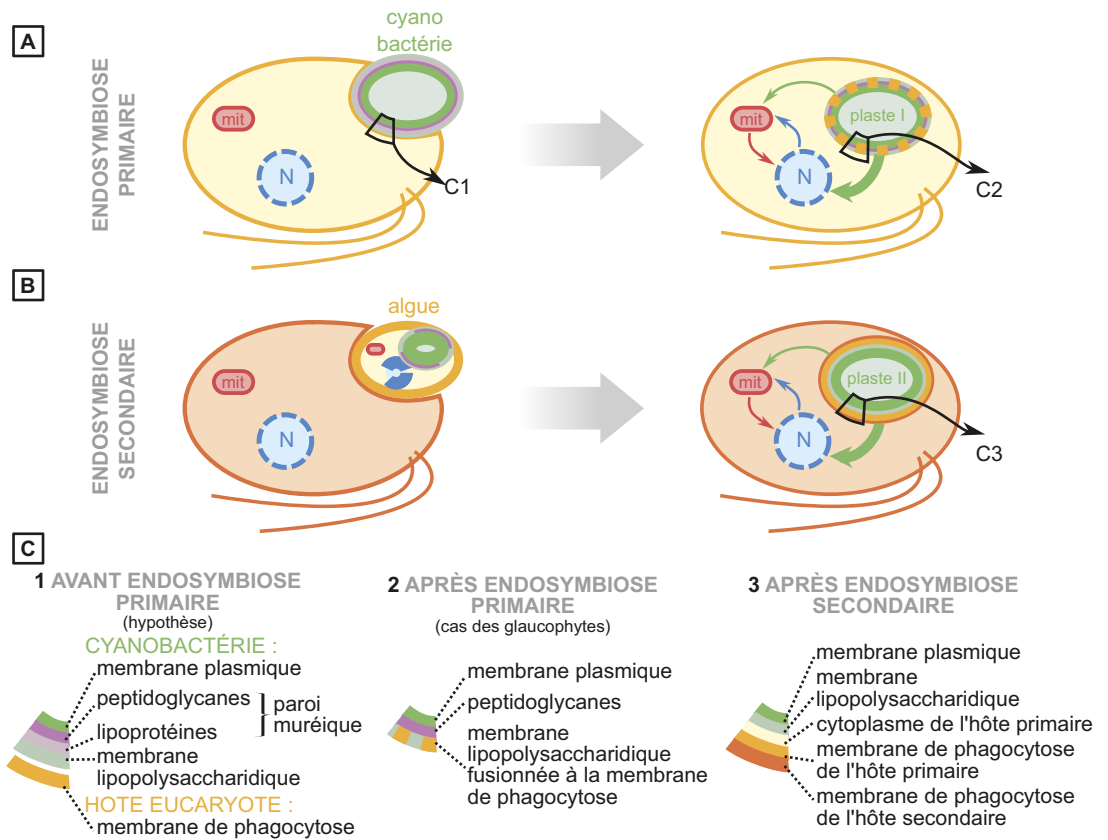


Figure 13.4 Origine des plastes par endosymbiose primaire et secondaire.

Représentation schématique des mécanismes d'endosymbiose primaire (A) d'une cyanobactérie par un eucaryote unicellulaire non-phototrophe, et d'endosymbiose secondaire (B) d'une algue eucaryote phototrophe par un eucaryote non-phototrophe. Les flèches bleues, rouges et vertes indiquent les transferts géniques entre noyau, mitochondrie et plastes ; leur épaisseur souligne l'importance du transfert. Le noyau (N) et la mitochondrie (mit) sont indiqués. (C) Organisation ultrastructurale schématisée des feuillets membranaires avant endosymbiose (C1), chez un glaucophyte (*Cyanophora paradoxa*) (les lipoprotéines sont ici perdues) (C2) et après endosymbiose secondaire (C3).

c) Divers évènements d'endocytobiose secondaire à l'origine de la diversité des plastes

Des plastes à quatre membranes ont été décrits chez les straménopiles (fucus, diatomées) ainsi que dans d'autres clades (alvéolobiontes, haptophytes, cryptophytes). Il existe aussi des plastes à trois membranes chez les euglénophytes ou les dinophytes. Ceci est en faveur d'une évolution indépendante des plastes dans ces groupes.

La théorie de l'endosymbiose secondaire

Sarah Gibbs, en 1978, propose que les plastes à quatre membranes dérivent de l'endocytobiose par un eucaryote d'une algue contenant déjà une cyanobactérie symbiotique. On connaît actuellement de telles associations endosymbiotiques entre des algues et des organismes hétérotrophes : par exemple les xanthelles (dinophytes) vivent en symbiose avec les coraux et d'autres animaux marins. Cette hypothèse s'appuie par ailleurs sur des données ultrastructurales, et elle est étayée par des arguments phylogénétiques. La membrane la plus externe des plastes des algues brunes

ZOOM 4

Diversité ultrastructurale et biochimique des plastes des eucaryotes

(phéophycées) correspondrait à une membrane endocellulaire de la cellule hôte. La membrane située immédiatement en dessous appartient à l'algue eucaryote d'origine symbiotique et les deux plus internes correspondraient à celles du plaste primaire de l'algue eucaryote (figure 13.4-B). Nous pouvons donc supposer que tous les constituants de l'algue symbiotique, à l'exception de sa membrane et de son plaste ont été perdus par évolution simplificatrice (régressive).

Des évènements d'endosymbiose secondaires indépendants

L'arbre des eucaryotes (figure 13.3) montre que des endocytobioses secondaires ont eu lieu de manière récurrente dans différents taxons. L'existence de plastes à trois membranes chez les euglénophytes ou les dinophytes suggère une origine endosymbiotique secondaire de ces plastes, suivie d'une régression de l'une des membranes, probablement la membrane plasmique de l'algue symbiotique internalisée. De nombreux dinophytes hétérotrophes peuvent en effet phagocyter des micro-eucaryotes sans consommer leur membrane plasmique qui est rejetée. L'organisme ingéré dans une vacuole reste fréquemment non digéré, au moins pendant un temps, ce qui pourrait expliquer la mise en place d'une endocytobiose. Lorsque la proie est photosynthétique, ces dinophytes détournent les plastes phagocytés à leur profit. Ainsi, des représentants de différents clades eucaryotes ont sans doute été « ingérés » à de multiples reprises par des dinophytes ou d'autres unicellulaires ; ceci tend à démontrer le caractère polyphylétique de l'acquisition des plastes secondaires.

Un degré supplémentaire de complexification cytologique est même atteint chez certains dinophytes avec des plastes qui résultent d'une endosymbiose tertiaire, c'est-à-dire d'un troisième événement d'endocytobiose.

Une seconde endosymbiose primaire chez les eucaryotes ?

Il a longtemps été considéré que l'endosymbiose primaire était un événement unique qui caractérisait les plantae. La découverte du genre *Paulinella*, petits eucaryotes amœboïdes hétérotrophes du groupe des rhizarias, est venue déstabiliser ce dogme. En effet, chaque cellule de *P. chromatophora* contient deux structures réniformes, nommées « chromatophores » qui dérivent d'une cyanobactérie symbiotique. Les phylogénies moléculaires ont démontré que ce symbiote est différent des cyanobactéries à l'origine du plaste des plantae. Bien qu'ils soient très modifiés, ces chromatophores sont bien considérés comme des plastes et dérivant d'une endosymbiose primaire. Plantae et rhizarias étant distants phylogénétiquement, l'acquisition de l'autotrophie par endocytobiose d'un plaste primaire pourrait avoir eu lieu de manière convergente, c'est-à-dire deux fois indépendamment au cours de l'évolution des eucaryotes.

d) La coévolution entre la cellule eucaryote et ses organites

Les exemples précédents montrent la grande diversité et surtout la grande fréquence des associations symbiotiques au sein des eucaryotes. Ces associations variées entre organites et hôtes sous-tendent une diversité de relations génétiques entre les partenaires. Lors des endosymbioses mitochondriale et plastidiale, il s'est produit des échanges de gènes entre les génomes nucléaires de l'hôte et ceux, mitochondriaux ou plastidiaux, des symbiotes.

La taille des génomes mitochondriaux est très variable : chez *Plasmodium falciparum*, le génome mitochondrial est extrêmement réduit (6 kb) et ne comprend que trois gènes codants des protéines de la chaîne respiratoire. L'endosymbiose des mitochondries a ainsi conduit à la perte quasi-totale du génome mitochondrial. En revanche, chez la plupart des plantes, les ADNmt dépassent 200 kb, pouvant atteindre 11 000 kb, ce qui excède de loin la taille moyenne des génomes bactériens. Ainsi des gains de gènes se sont également produits dans différents clades, en particulier dans celui des plantes. Cette augmentation de taille du génome résulte des transferts de gènes endosymbiotiques provenant essentiellement du noyau. Pertes et gains de caractères sont des traits aléatoires, non orientés, de l'évolution des eucaryotes et qui ont dû dépendre du contexte phylogénétique et de l'environnement.

Pour les plastes, l'endocytobiose s'accompagne systématiquement d'une extrême réduction de leur génome. Une grande part des gènes plastidiaux « perdus » est relocalisée et codée par le génome nucléaire de l'hôte. Ainsi le génome nucléaire des bicondes peut être considéré comme une *mosaïque génique*, associant des gènes archéens, sans doute ancestraux, à des gènes bactériens qui ont été acquis à maintes reprises en lien avec des endocytobioses multiples, conditionnant ainsi l'évolution de l'autotrophie.

e) La perte secondaire des plastes

La régression du génome des plastes s'accompagne fréquemment de modifications structurales parfois importantes qui peuvent aller jusqu'à la disparition physique de cet organite. La perte des plastes s'argumente plus ou moins facilement en s'appuyant sur la phylogénie des eucaryotes (figure 13.3). Au sein du clade S-A-R, l'endosymbiose secondaire est probablement un événement unique réalisé principalement via une algue rouge. Postérieurement à cet événement, chez de nombreux alvéolobiontes, les plastes ont fortement régressé. Tout d'abord, un organite dérivé d'un plaste à trois membranes est présent chez certains apicomplexés, tels *Plasmodium*, l'agent du paludisme, mais cette structure n'effectue plus la photosynthèse. Ensuite, l'existence de gènes plastidiaux dans le génome des ciliés (*Paramecium*) suggère la présence puis la disparition totale du plaste dans ce groupe. Par ailleurs, au sein des dinophytes, près de la moitié des espèces sont secondairement hétérotrophes mais diverses ré-acquisitions indépendantes du plaste sont argumentées chez ces organismes. D'autres pertes indépendantes sont enfin reconnues chez les straménopiles, les oomycètes, et dans le groupe des rhizarias et des excavata (trypanosomes). Gains et pertes des plastes sont donc des événements récurrents, indépendants, qui soulignent l'absence d'orientation des mécanismes évolutifs ; ainsi, l'évolution n'a pas de but.

2.4 La pluricellularité : un ensemble de caractères convergents associés à des fonctions similaires

Plusieurs lignées d'eucaryotes pluricellulaires (métazoaires, plantes, différents groupes « d'algues » ou de « champignons ») ont une origine phylogénétiquement indépendante (figure 13.3). Si la **pluricellularité** était un état ancestral des eucaryotes, il faudrait supposer que le passage à l'unicellularité se serait produit un nombre de fois plus élevé que la transition réciproque (de l'unicellularité à la pluricellularité). Le raisonnement par maximum de parcimonie favorise ainsi l'hypothèse de l'unicellularité comme état ancestral chez les eucaryotes. La pluricellularité est donc un état dérivé, apparu plus de 20 fois par convergence, et qui constitue une étape importante de l'évolution eucaryote, marquée par une transition entre une cellule isolée et autonome et un édifice complexe caractérisé par une coopération intercellulaire. Cette transition peut être illustrée par différents clades d'unicellulaires tels les ciliés et les diatomées qui comportent des organismes unicellulaires coloniaux.

a) Une diversité d'acquisitions de la matrice extracellulaire

Lorsque l'on considère l'ensemble des lignées pluricellulaires eucaryotes, nous constatons que les constituants squelettiques (rigides) des matrices extracellulaires (MEC) sont à prédominance polysaccharidique et de nature très variée, la plus classique étant la cellulose. Seule la MEC animale est pour l'essentiel de structure protéique (collagène). Les éléments matriciels (flexibles) sont également diversifiés mais en général polysaccharidiques : pectines (embryophytes), galactanes (rhodophytes), alginates (phéophycées), à l'exception des animaux chez lesquels la matrice est surtout protéique (protéoglycanes). À la différence des métazoaires, les eumycètes possèdent une paroi en chitine (β 1-4 N-acétylglucosamine), c'est-à-dire polysaccharidique. Ces différences majeures supposent une batterie d'enzymes très différentes dans chacune des lignées. Dans la plupart des cas, la pluricellularité est associée à une immobilité cellulaire ; la seule véritable exception est constituée par l'organisation des métazoaires chez qui les cellules sont mobiles.

Voir ouvrage
de 1^{re} année,
chapitre 5, zoom 2

b) De la juxtaposition des cellules à la pluricellularité

Si la pluricellularité a émergé à une vingtaine de reprises chez les eucaryotes, il existe également de nombreux cas d'association de cellules sous forme de filaments (eumycètes, oomycètes) ou d'agrégats ou colonies de cellules (algues volvocales, choanoflagellés). Parmi les organismes modèles des microbiologistes, la levure *Saccharomyces cerevisiae* (ascomycète) est fondamentalement unicellulaire. Cependant des travaux récents ont montré qu'une transition adaptative vers un état pluricellulaire clonal peut être obtenue expérimentalement chez cet organisme. Comme dans beaucoup d'autres cas, la formation d'agrégats cellulaires résulte de l'absence de séparation des cellules-filles après la division cellulaire. À l'inverse, la spécialisation cellulaire liée à la pluricellularité revêtant des caractéristiques bien différentes selon les groupes d'eucaryotes, rien n'empêche une perte de cette organisation complexe avec réversion vers l'unicellularité.

Tous ces éléments soulignent une grande diversité de l'organisation cellulaire et génétique des eucaryotes pluricellulaires. La vie sous forme de colonies de cellules, dont la séparation des rôles est plus ou moins nette et intégrée, a été inventée de nombreuses fois chez les eucaryotes. Les mécanismes de communication entre cellules sont, eux aussi, convergents à grande échelle évolutive.

3 Les virus ont-ils une place dans l'arbre du vivant ?

3.1 Quelques caractéristiques des virus

Les découvertes de virus géants (« *Giant Virus* » ou Girus) ont nettement complexifié notre vision de l'évolution. Un virus est traditionnellement défini comme un parasite intracellulaire obligatoire, infectieux, capable d'autoréplication, et doté d'un génome isolé du milieu extérieur par une capsid protéique, éventuellement entourée d'une enveloppe lipoprotéique. Cependant, derrière cette définition des virus se cache une extraordinaire diversité qualitative — près de 10 000 espèces sont recensées — et surtout quantitative, — les eaux marines par exemple contiennent au moins dix fois plus de virus que de cellules procaryotes —, au point que les cellules semblent vivre dans un « océan de virus ».

Les génomes viraux sont circulaires ou linéaires, simples ou doubles brins (db), constitués d'ARN ou d'ADN, et pour certains de taille plus importante que certains génomes de procaryotes symbiotiques (800 kb pour les *Mimivirus* et 2,5 Mb pour *Pandoravirus*, contre 160 kb pour la bactérie *Carsonella ruddii* ou 500 kb pour l'archée *Nanoarchaeum equitans*). Ils sont caractérisés par différents types d'**évolution non-verticale** de l'information génique : des THG par intégration des fragments d'ADN de leur hôte, des recombinaisons homologues, et enfin, une mosaïque de gènes qui correspond à des génomes chimériques produits par recombinaisons non-homologues entre deux génomes viraux co-infectant un même hôte.

Le point commun de tous les virus est qu'ils détournent systématiquement une partie de la machinerie de leur hôte pour compléter leur propre cycle de réplication. Les étudier et se pencher sur leur place dans le monde vivant a des vertus pédagogiques : cela nous permet de nous poser un certain nombre de questions sur les virus, et plus généralement sur la vie et l'évolution biologique.

3.2 Les virus sont-ils vivants ?

Définir si les virus sont ou non vivants passe d'abord par définir ce qu'est la vie. Une définition qui s'appuierait sur les propriétés d'auto-organisation et d'auto-entretien des êtres vivants ne s'applique pas ici aux virus, ces entités étant dépourvues de métabolisme énergétique. La principale propriété du monde vivant qui pourrait donc s'y appliquer est basée sur la capacité à évoluer. Cependant, cette définition ne permet pas de faire la distinction entre les virus biologiques et

Voir ouvrage
de 1^{re} année,
chapitre 13, § 4

les virus informatiques. Ces derniers sont conçus pour se propager avec de légers changements dans leur code (c'est-à-dire muter), et ils subissent une sélection (par leur capacité différentielle à infecter de nouveaux ordinateurs) tout en résistant aux logiciels antivirus, sans pour autant qu'ils soient considérés comme vivants. Force est donc de constater que nous ne pouvons pas proposer une définition univoque de la vie permettant d'y inclure les virus. Notons enfin que les virus sont des parasites moléculaires, ne se répliquant ni n'évoluant par eux-mêmes : ce sont les cellules qui les font évoluer. Sans cellules, il n'y a pas d'évolution virale. Les virus ne sont donc pas « auto-reproduits » mais simplement « produits ». Autrement dit, les virus ne sont pas des entités vivantes, mais des **entités vécues**, c'est-à-dire ayant subi l'expérience de la vie cellulaire.

3.3 Tous les virus présentent-ils une continuité évolutive ?

a) Possèdent-ils une continuité structurale ?

Un attribut universel des cellules et, par conséquent, des êtres vivants, est la possession de membranes. Ces dernières se constituent en divisant des membranes préexistantes qui sont ensuite enrichies en nouveaux lipides et nouvelles protéines membranaires. Ceci implique une continuité structurale depuis les cellules primordiales jusqu'aux cellules contemporaines. Chez les virus en revanche, tous les constituants viraux sont synthétisés *de novo* à chaque cycle d'infection virale par les machineries cellulaires asservies. Quant aux enveloppes lipidiques qui caractérisent certains virus, elles sont néoformées : elles ne présentent donc pas de continuité structurale avec les membranes virales précédentes. Alors que l'existence de membranes héritées est une preuve de la monophylie de toutes les cellules modernes, leur absence chez les virus suggère l'origine polyphylétique de ces derniers.

b) Peuvent-ils être replacés dans une même phylogénie ?

La reconstruction des phylogénies requiert de comparer simultanément tous les caractères, en particulier moléculaires, chez les organismes dont les relations de parenté doivent être estimées. Il est ainsi possible d'inférer des arbres phylogénétiques contenant des ensembles de virus échantillonnés à des niveaux taxonomiques suffisamment réduits pour qu'ils présentent des portions génomiques en commun. Par exemple, comme l'actualité sanitaire avec la pandémie de la COVID-19 nous l'a montré, la phylogénie des souches virales de SARS-CoV2 peut être reconstruite à partir de la comparaison de leurs génomes (figure 13.5). Un tel arbre phylogénétique peut être utilisé pour estimer l'âge de l'émergence du virus, caractériser ses modalités de propagation géographique, reconstruire sa dynamique épidémiologique, et détecter les sites nucléotidiques qui sont sous sélection positive dans ses gènes.

Si nous souhaitons maintenant replacer toute la diversité des virus connus dans l'arbre du vivant, un obstacle majeur survient : il apparaît que les virus dans leur intégralité ne partagent pas de caractéristiques homologues avec les cellules-hôtes, et nous ne connaissons pas de gènes qui soient partagés par l'ensemble des virus. Par conséquent, les virus connus ne peuvent être directement comparés avec les organismes cellulaires : il n'est pas possible de les positionner dans l'arbre du vivant cellulaire. Néanmoins, l'existence de motifs structuraux communs aux protéines de capsid de lignées virales distantes (par exemple, les phages entérobactériens et les adénovirus eucaryotes) suggérerait que malgré l'extrême divergence des séquences protéiques primaires, tous les virus connus possèdent une origine unique et pourraient être replacés dans une même phylogénie. Cependant, il existe deux hypothèses alternatives pour expliquer la présence de motifs protéiques similaires dans des lignées virales distinctes.

- En raison de fortes contraintes subies par leurs structures tertiaires protéiques et afin de s'adapter aux contraintes physico-chimiques des environnements cellulaires, les capsides virales pourraient n'adopter qu'un petit nombre d'architectures géométriques simples, conduisant ainsi à des phénomènes de **convergence structurale**.



Voir TP 9, § 1

- La **fréquence élevée des THG** entre cellules et virus lors de l'évolution pourrait expliquer que différents virus non apparentés partagent certains gènes, et donc certaines protéines. Les THG, couplés à des taux élevés de recombinaison, tendent à rendre plus similaires qu'elles ne le sont réellement des lignées virales évolutivement distantes. En effet, après un THG, l'organisme receveur va posséder dans son génome une séquence très similaire à celle qu'un donneur lui aura transféré : d'un point de vue phylogénétique, le receveur apparaîtra alors comme un parent anormalement proche du donneur. Pour les virus en particulier, mais aussi pour tous les organismes en général, les THG constituent donc un défi permanent pour une inférence fiable des arbres phylogénétiques.

Tous ces éléments nous montrent qu'il n'est pas possible de replacer l'intégralité des virus connus dans un même arbre phylogénétique. Ceci amène deux conséquences : d'une part il n'est pas possible de prouver la monophylie de toutes les lignées virales, et d'autre part il n'est pas possible de les positionner dans l'arbre du vivant cellulaire. Et quand bien même nous pourrions estimer un arbre contenant tous les organismes cellulaires ainsi que les virus, le taux d'évolution très élevé de ces derniers ferait que leur position phylogénétique serait probablement très fragile en raison de problèmes d'artéfacts d'attraction des longues branches. Par ailleurs, en raison de la plasticité génomique virale, il est possible de reconstruire l'histoire de l'évolution de chaque gène individuel d'une lignée virale et d'en déduire les événements de THG, mais ces histoires ne refléteront pas l'évolution des virus dans leur ensemble.

3.4 Les virus sont-ils des entités très anciennes ?

a) Recherche de coévolution entre les virus et leurs hôtes

Le fait que certaines lignées virales infectent des hôtes phylogénétiquement éloignés peut être interprété comme un argument en faveur de leur origine très ancienne. Cette hypothèse repose sur l'existence d'une **coévolution** entre les virus et leurs hôtes : la spéciation des virus résulterait de la spéciation de leurs hôtes. Démontrer la coévolution des virus, et de manière plus générale celle des parasites, et de leurs hôtes nécessite de prouver que les phylogénies respectives des hôtes et de leurs parasites sont congruentes, c'est-à-dire que leurs topologies — les relations de parenté — et leurs âges de divergence doivent être similaires. Cependant, il est extrêmement difficile d'appliquer ce type de raisonnement aux parasites, notamment viraux, pour lesquels le changement d'hôte — c'est-à-dire la possibilité pour un parasite d'infecter des hôtes phylogénétiquement non apparentés — est fréquent.

Ainsi, des virus qui infectent des hôtes phylogénétiquement (très) distants peuvent tout simplement constituer des entités virales qui ont la capacité de sauter d'un hôte initial à un hôte (très) différent, sans que cela ne reflète ni une quelconque ancienneté (les sauts d'hôtes peuvent avoir eu lieu très récemment), ni une coévolution.

b) L'argument de la simplicité

La vision anthropocentrique de l'évolution — qui ne serait qu'un processus de progression vers une complexité toujours accrue — explique la tendance à considérer que les virus doivent être extrêmement anciens parce qu'ils sont extrêmement simples. Pourtant, l'évolution n'est pas un processus directionnel, et il existe de nombreux exemples d'évolution simplificatrice. Ainsi, la simplicité de structure des organismes n'est une preuve ni de leur ancienneté, ni du fait qu'ils cumulent des états de caractères ancestraux. L'**évolution simplificatrice** (appelée aussi de manière un peu péjorative « évolution régressive ») est un processus fondamental dans l'évolution des parasites. Même si certaines lignées virales, comme les virus géants, sont caractérisées par une augmentation de taille du génome et de complexité, la règle générale est que les virus sont soumis à une forte pression sélective pour conserver une taille de génome minimale, ceci leur garantissant une vitesse élevée de multiplication. Dans un contexte de course à l'armement face aux cellules-hôtes, cette

Voir chapitre 12,
§ 4.1f

contrainte évolutive majeure limite leur complexification : les virus présentent donc très généralement une simplicité d'organisation qui n'est que la conséquence du parasitisme.

3.5 Bilan : des virus jusqu'à la biodiversité

En conclusion, les virus sont des entités organiques caractérisées par une incapacité à s'auto-entretenir et à s'auto-répliquer, par une origine polyphylétique, par une remarquable plasticité de leur génome au cours des temps évolutifs, et par l'impossibilité de les incorporer dans l'arbre de la vie. Cela n'implique pas pour autant qu'ils n'ont pas eu, et qu'ils n'ont pas un rôle important dans l'évolution de la vie sur Terre. Au contraire, leur abondance couplée aux pressions sélectives qu'ils exercent sur les populations de nombreux organismes cellulaires fait que les virus constituent une source et un moyen importants de maintenir la biodiversité. Comme les éléments génétiques mobiles, tels que les transposons, qui peuvent s'intégrer au génome d'un hôte, ils contribuent à la production de la variabilité génétique. De plus, en tant qu'éléments génétiques ayant une phase extracellulaire, ils constituent des véhicules pour les transferts génétiques horizontaux, y compris entre hôtes phylogénétiquement très éloignés.

ZOOM 1

Caractéristiques comparées des trois grands domaines du vivant

Chacun des trois domaines du monde vivant — bactéries, eucaryotes, et archées — est généralement considéré comme constituant un groupe monophylétique. Ainsi, un certain nombre de caractères sont classiquement mis en avant afin de définir chaque domaine. L'observation comparée d'électronographies montre tout d'abord des différences assez fondamentales entre l'organisation cellulaire de type procaryote (bactéries, archées) et eucaryote. Elle révèle par ailleurs l'existence d'organites, d'éléments du cytosquelette, et de compartiments cellulaires — c'est-à-dire des structures intracellulaires limitées par une bicouche lipidique, avec un contenu protéique qui fournit un environnement dédié à la séquestration et à l'exécution de réactions biochimiques — dans des cellules appartenant aux trois domaines du vivant. L'organisation cellulaire des bactéries et des archées est donc certainement plus complexe que ce qui est suggéré par la division classique entre procaryotes et eucaryotes.

Les bactéries

Elles sont en général des organismes unicellulaires dont les cellules sont **dépourvues de noyau**. Leur taille, majoritairement de quelques micromètres, peut néanmoins varier de 50 nm (pour les nanobactéries) à 600 µm. L'**ADN, circulaire**, se trouve directement dans le cytosol. Les bactéries possèdent généralement une paroi, faite de **peptidoglycanes**. La paroi est épaisse pour les bactéries Gram+ ; elle est fine, et doublée d'une membrane externe chez les bactéries Gram-. La distinction de ces deux groupes repose sur un test de coloration mais n'a pas de valeur phylogénétique. Certaines bactéries présentent des membranes cytoplasmiques très invaginées et des compartiments cellulaires (c'est le cas des thylacoïdes chez les cyanobactéries).

Les eucaryotes

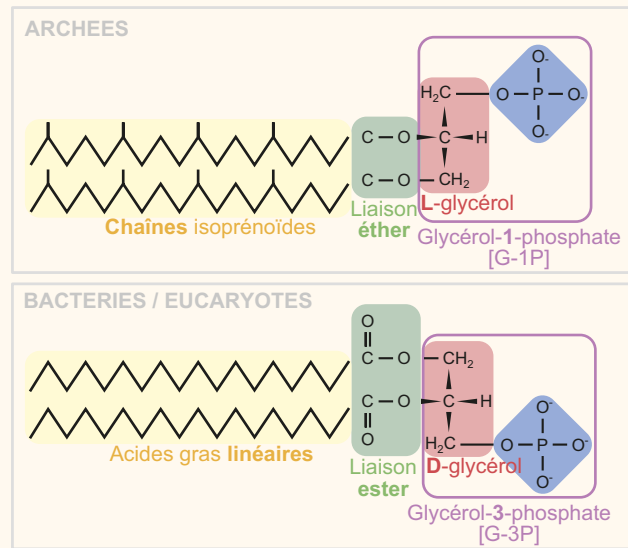
Ils sont caractérisés par la présence dans leurs cellules de **nombreuses endomembranes**, délimitant des compartiments variés, dont le nucléoplasme. Cette organisation suppose un fonctionnement cellulaire bien différent de celui d'une bactérie et donc des propriétés biochimiques particulières. De nombreuses molécules impliquées dans l'adhésion ou les mouvements cellulaires (moteurs moléculaires, partie axiale des cils ou des flagelles) semblent propres aux eucaryotes. L'ADN des eucaryotes est linéaire et associé à des **protéines histones** qui structurent un génome morcelé par des introns, et dans lequel les opérons font généralement défaut (certains animaux comme les nématodes en présentent cependant).

Voir ouvrage
de 1^{re} année,
chapitre 6, § 2

La transcription eucaryote est quant à elle suivie d'un mécanisme de maturation des ARNm qui précède dans le temps l'initiation de la traduction : ce caractère pourrait être une signature de la monophylie des eucaryotes. Le dernier ancêtre commun aux eucaryotes (LECA, « Last Eukaryotic Common Ancestor ») aurait ainsi acquis, parmi d'autres caractéristiques, un épissosome (= splicéosome), un système endomembranaire complexe, et des machineries associées à la méiose et à la phagotrophie.

Les archées

Elles présentent une organisation générale (taille et forme) proche de celle des bactéries. Mais ces microorganismes partagent de nombreux points communs avec les cellules eucaryotes : leurs gènes codant les enzymes de la réplication et de la transcription sont homologues de ceux des eucaryotes ; leur ADN est associé à des **histones** ; leurs **gènes sont morcelés**. D'autres caractéristiques paraissent les relier aux bactéries telles l'organisation des chromosomes et la présence d'opérons. Les archées possèdent aussi des caractéristiques propres, par exemple des phospholipides membranaires particuliers constitués de polymères d'isoprénoïdes associés au glycérol-1-phosphate (G-1P) par une **liaison éther**, à la différence des eucaryotes et bactéries dont les membranes plasmiques sont formées de polymères d'acides gras linéaires associés au glycérol-3-phosphate (G-3P) par une liaison ester.



Structure comparée des phospholipides membranaires des archées par rapport à ceux des bactéries et des eucaryotes.

Le métabolisme très particulier des archées, parfois même inédit (cf. la méthanogenèse), a conduit à les classer initialement selon des critères purement métaboliques (thermophiles, méthanogènes, halophiles, sulfo-dépendantes) ; toutefois ce point de vue est largement abandonné aujourd'hui au profit d'une classification phylogénétique basée sur des caractères moléculaires, et qui reconnaît plusieurs phylums archéens (par exemple les euryarchées, thaumarchées, et crenarchées). La monophylie des archées est faiblement soutenue par l'analyse phylogénétique de marqueurs ribosomiaux. Les travaux de phylogénomique basés sur la comparaison de protéines très conservées proposent même que les eucaryotes sont plus apparentés à certains phylums archéens que ces derniers ne le sont entre eux, ce qui rendrait paraphylétique le groupe des archées.

ZOOM 2

Les apomorphies des 9 clades majeurs d'eucaryotes

Trois clades majeurs d'eucaryotes auxquels appartiennent des organismes vus en TP sont recensés dans le tableau ci-dessous avec les quelques apomorphies cellulaires qui les caractérisent.

Les trois clades majeurs des eucaryotes

	(i) straménopiles (= hétérocontes)	(ii) plantae (= lignée verte)	(iii) opisthocontes
exemples de taxons vus en TP	phéophycées (« algues » brunes) ; diatomées ; oomycètes (autrefois classés dans les champignons)	algues vertes (ou chlorophytes) ; algues rouges (ou rhodophytes) ; embryophytes (dont fougères et angiospermes)	eumycètes et métazoaires
type trophique	autotrophie pour le carbone, ou hétérotrophie ; plastes à quatre membranes (= plastes secondaires) ; chlorophylles a et c	autotrophie pour le carbone ; plastes à deux membranes (= plastes primaires) ; chlorophylles a et b	hétérotrophie pour le carbone
polyosides	β1-3 glucane (réserve)	amidon (réserve)	glycogène (réserve) ; chitine (structure)
flagelles des cellules	deux flagelles différents	deux à quatre flagelles, perdus chez les rhodophytes et les angiospermes	un flagelle unique et propulseur

Les six autres clades identifiés par les analyses phylogénétiques de caractères moléculaires sont les suivants.

(iv) Les **rhizarias** : unicellulaires amoéboïdes et/ou flagellés. Certains possèdent des plastes et sont photosynthétiques. On ne leur connaît aucune synapomorphie cytologique. Les données moléculaires les rapprochent des straménopiles. Exemples : globigérines (foraminifères), radiolaires.

(v) Les **alvéolobiontes** : unicellulaires planctoniques ou parasites dont beaucoup ont perdu secondairement leurs plastes, comme le suggère l'existence de gènes plastidiaux dans leur noyau. La présence de saccules (dont ils tirent leur nom) sous la membrane plasmique pourrait constituer une synapomorphie cytologique. Exemples : *Paramecium*, *Plasmodium*.

(vi) Les **haptophytes** : algues unicellulaires biflagellées possédant un haptonème (appendice filamenteux constitué de microtubules) impliqué dans l'adhérence et la prédation. Leurs chloroplastes possèdent une membrane externe en continuité avec la membrane nucléaire externe, ainsi que des thylacoïdes groupés par trois. Exemples : coccolithophoridés.

(vii) Les **cryptophytes** : unicellulaires généralement photosynthétiques possédant deux flagelles. Leurs chloroplastes présentent quatre membranes – avec un vestige de noyau, le nucléomorphe, localisé entre les deuxièmes et troisièmes membranes – et leurs thylacoïdes sont appariés. Exemples : *Cryptomonas*, *Guillardia* (taxons non étudiés selon le programme de BCPST).

(viii) Les **excavata** : ils regroupent de nombreux taxons, dont les discicristata d'une part, et les parabasalia d'autre part. Les premiers forment un clade d'unicellulaires libres (euglènes) ou parasites (trypanosomes), dont la monophylie est soutenue par la présence de mitochondries

Voir chapitre 1, § 1 et 2

Voir chapitre 1, § 3.2b

particulières à crêtes discoïdes (en forme de raquette de ping-pong). Les seconds contiennent des symbiotes du tube digestif de métazoaires et des parasites pathogènes (*Trichomonas* est responsable de maladies sexuellement transmissibles). Certains des excavata ont perdu secondairement leurs mitochondries.

(ix) Les **amoébozoaires** : unicellulaires couramment qualifiés d'« amibes » qui se déplacent par vagues cytoplasmiques internes. Ils sont vraisemblablement le groupe-frère des opisthontes avec lesquels ils partagent des synapomorphies : des mitochondries à crêtes aplaties, un flagelle unique et propulseur, et différents caractères biochimiques (possession de glyco-gène et de chitine).

ZOOM 3

Origine et évolution des mitochondries

Il existe **plusieurs hypothèses** quant à la nature de l'ancêtre-hôte eucaryote ayant réalisé l'endocytobiose avec une bactérie pour conduire à la mitochondrie. L'une des hypothèses proposées implique un « archézoaire » qui serait un hypothétique eucaryote privé de mitochondries, tandis que les autres font appel à des scénarios d'endocytobiose entre une bactérie et une archée qui aurait secondairement évolué en cellule compartimentée et nucléée, c'est-à-dire une cellule eucaryote.

L'hypothèse de l'archézoaire

Le scénario de l'archézoaire s'appuie sur des analyses phylogénétiques de la petite sous-unité d'ARNr. Les premiers arbres ainsi obtenus en utilisant l'approche cladistique ont identifié un groupe particulier d'eucaryotes désignés par le terme « d'archézoaires ». C'est un ensemble de parasites unicellulaires anaérobies — comme les trichomonadines ou les microsporidies — qui présentent deux particularités : (i) ils apparaissent comme les plus anciens rameaux eucaryotes connus ; et (ii) ils sont dépourvus de mitochondries. L'absence de ces organites a initialement été considérée comme plésiomorphe (i.e., ancestrale). Ultérieurement, les arbres reconstruits à partir de la combinaison de nombreuses protéines nucléaires indiquent sans ambiguïté d'une part que le groupe des archézoaires est polyphylétique (figure 13.3) et d'autre part que leur position phylogénétique très ancienne est un artéfact d'attraction des longues branches. L'absence des mitochondries chez les trichomonadines et microsporidies est donc vraisemblablement liée à des pertes secondaires et indépendantes chez ces taxons.

L'hypothèse symbiogénétique

L'un des scénarios dit « symbiogénétique » postule que la complexité de la cellule eucaryote serait la conséquence de l'endosymbiose mitochondriale ; ainsi, l'origine de la mitochondrie serait directement associée à celle de la cellule eucaryote. Cette hypothèse s'appuie sur le double héritage, archéen et bactérien, d'une grande partie du génome eucaryote. L'hypothèse propose également que, dans certains clades eucaryotes, la respiration aérobie aurait été perdue secondairement. La mitochondrie aurait alors été modifiée en structures variées correspondant, par exemple, aux hydrogénosomes des ciliés. En effet, ces organites produisent de l'ATP et du CO₂ mais ils ne possèdent pas de cytochromes et ils sont le siège d'un cycle des acides tricarboxyliques (ou cycle de Krebs) incomplet : ils s'apparentent donc bien aux mitochondries. Par ailleurs, on a longtemps pensé que l'amibe *Entamoeba histolytica* (amoébozoaire : voir Zoom 2), organisme anaérobie, était dépourvue de mitochondrie mais les analyses phylogénétiques ont permis de montrer que cet organisme possède un mitosome, structure à deux membranes, ayant totalement perdu l'ADNmt, dont la seule fonction est d'associer le fer et le soufre en complexe Fe-S. Les mitosomes, qui possèdent des protéines parentes de celles des mitochondries, pourraient donc en dériver.

L'évolution simplificatrice des mitochondries

Même si les biologistes rencontrent des difficultés pour établir le portrait-robot des premières cellules eucaryotes et de leurs organites, tout se passe ainsi comme si les mitochondries et les génomes mitochondriaux que nous observons dans la nature actuelle constituent une photographie instantanée d'une évolution globalement simplificatrice, qui s'est enclenchée postérieurement à l'établissement de l'endosymbiose mitochondriale. Cette évolution conduit à des pertes de gènes dans le génome mitochondrial (mitogénome), à des transferts de gènes vers le génome nucléaire, et parfois au recrutement de nouveaux gènes nucléaires pour produire des protéines importées dans la mitochondrie. Cette évolution s'est néanmoins produite à des vitesses très contrastées selon les clades eucaryotes considérés, ce qui fait que les biologistes observent aujourd'hui des situations très différentes, avec certains eucaryotes possédant des mitochondries complexes hébergeant des génomes possédant une centaine de gènes (cas de l'eucaryote excavata *Reclinomonas*) jusqu'à certains autres possédant des mitochondries simplifiées hébergeant des mitogénomes extrêmement réduits (trois gènes chez l'alvéolobionte *Plasmodium*) voire absents au sein de « fantômes » ou « reliques » de mitochondries (les mitosomes).

ZOOM 4

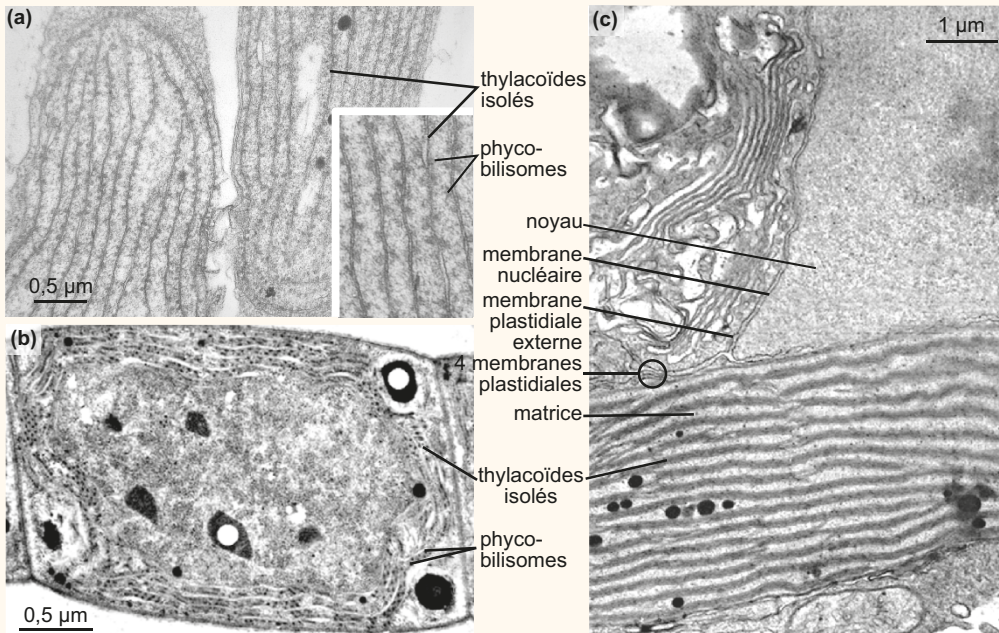
Diversité ultrastructurale et biochimique des plastes des eucaryotes

La structure et le fonctionnement des plastes et de leurs pigments sont très diversifiés chez les bicontes photosynthétiques (tableau) : les plastes des « algues » présentent par exemple la particularité d'être ceinturés par deux, trois ou quatre membranes. Cette étonnante variété des plastes est le résultat d'événements primaires et secondaires d'endocytobiose.

Les plastes primaires des plantae

En microscopie électronique à transmission (MET), les plastes des plantae présentent des structures relativement variées. Au sein de cette diversité, les plastes d'algues rouges (rhodophytes) se distinguent par leur surprenante ressemblance avec les cellules des cyanobactéries. Entourés d'une double membrane, ils présentent des thylacoïdes, organisés en structures parallèles jamais accolées les unes aux autres. Les membranes des thylacoïdes assurent la photosynthèse grâce à la chlorophylle a et aux phycobilines (protéines), localisées dans des structures sphériques particulières, les **phycobilisomes** (figure a et b). Les cyanobactéries possèdent des repliements membranaires, qualifiés également de thylacoïdes, qui comportent eux aussi des phycobilisomes. Ces structures font défaut chez d'autres bactéries photosynthétiques (bactéries pourpres ou vertes). Ces observations s'accordent bien avec ce qu'on peut attendre dans l'hypothèse où une cellule hôte « végétale » héberge une cyanobactérie : la membrane interne du plaste correspond à la membrane plasmique bactérienne, et la membrane externe à celle de la vacuole dans laquelle elle a été internalisée lors de l'endocytobiose. Mais cette situation « nette » est relativement rare en dehors des algues rouges et, dans bien des cas, comme chez certaines algues vertes, telle *Dunaliella sp.* (chlorophyte), la membrane externe du plaste est en continuité avec le réticulum endoplasmique.

Il existe d'autres similitudes entre cyanobactéries et plastes : la chlorophylle a, caractéristique de tous les plastes eucaryotes, est, au sein des bactéries, uniquement présente chez les cyanobactéries (tableau). De même, plastes et cyanobactéries possèdent deux photosystèmes alors que les autres bactéries n'en ont qu'un seul. Le gain de la chlorophylle b et la perte des phycobiliprotéines ont sans doute accompagné l'endocytobiose chez les chlorophytes et les embryophytes. Soulignons enfin que la composition des lipides polaires (cardiolipides) des membranes de cyanobactéries est identique à celle des chloroplastes.



Ultrastructure des plastes en microscopie électronique à transmission.

Le terme de « plante » est utilisé ici plutôt que celui plus restrictif de « chloroplaste » qui est spécifique des plantae (= lignée verte). **(a)** Plaste de l'algue rouge *Polysiphonia sp.* (plantae : rhodophytes) ; **(b)** Comparaison avec la cyanobactérie *Oscillatoria splendida* ; **(c)** Plaste de l'algue brune *Chorda filum* (ochrophytes : phaeophytes). (Clichés fournis par (a) : C. Lichtlé, École Normale Supérieure, Paris, (b) : J.-C. Thomas, École Normale Supérieure, Paris et (c) : modifié d'après Bouck, © 1965 Rockefeller University Press. Initialement publié dans *Journal of Cell Biology* 26 :523-537).

Comparaison de différentes caractéristiques cytologiques et biochimiques des cyanobactéries et des plastes des eucaryotes.

Phylum	Membranes	Thylacoïdes	Chlorophylles	Phycobilines	Caroténoïdes	Localisation
cyanobactéries	—	non accolés	a	PC, APC, PE	—	—
eucaryotes						
rhodophytes	2	non accolés	a (et d)	PC, APC, PE	—	cytoplasme
glaucophytes	2	non accolés	a	PC, APC	—	cytoplasme
chlorophytes	2	par 2 à 6 (grana)	et b	—	zéaxanthine	cytoplasme
embryophytes	2	associés, grana	a et b	—	zéaxanthine	cytoplasme
dinophytes	3	par 3	a et c	—	péridinine	cytoplasme
euglénophytes	3	par 3	a et b	—	diadinoxanthine	cytoplasme
phéophycées	4	par 3	a et c	—	fucoxanthine	lumière du RER
haptophytes	4	par 3	a et c	—	fucoxanthine	lumière du RER
cryptophytes	4	par 2	a et c	PE, (PC)	alloxanthine	lumière du RER

Abréviations : APC = allophycocyanine, PC = phycocyanine, PE = phycoérythrine. Modifié d'après Kim E. & Archibald J.M. (2008). *Diversity and evolution of plastids and their genomes*. *Plant Cell Monographs*. Springer, Berlin. [https://doi.org/10.1007/7089_2008_17]

Les plastes secondaires à quatre membranes

Chez les plastes des algues brunes (phéophycées), la membrane la plus externe est pourvue de ribosomes et se situe en continuité avec l'enveloppe nucléaire (figure c). Dans l'hypothèse d'une endosymbiose secondaire, cette membrane doit correspondre à une enveloppe endocellulaire de la cellule hôte. La membrane située immédiatement en dessous appartient à l'algue eucaryote d'origine symbiotique et les deux plus internes correspondraient à celles du plaste primaire de l'algue eucaryote. Un **compartiment périplastidial** situé entre la deuxième et la troisième membrane a été maintes fois observé au MET et atteste d'un vestige probable du compartiment cytoplasmique de l'algue symbiotique contenant des ribosomes.

De plus, chez certaines algues (cryptophytes et chlorarachniophycées, figure 13.3), il est possible d'observer dans l'espace périplastidial un élément ressemblant à un noyau très modifié et nommé **nucléomorphe**. Chez les algues brunes, le compartiment périplastidial et le nucléomorphe sont inexistant. L'existence d'un nucléomorphe chez les cryptophytes est une précieuse source de caractères moléculaires : leur analyse montre que ce génome est proche de celui du noyau des algues rouges. Quant au génome du plaste, il s'apparente au génome cyanobactérien ou à celui des plastes des algues rouges ou vertes. Par ailleurs, le génome du nucléomorphe varie énormément selon les clades étudiés. Ce constat tend à démontrer que son évolution en lien avec la symbiose entraîne une compaction du génome du symbiote, dont l'intensité varie selon l'organisation génétique de son hôte.

Réviser

Résumé

Les progrès des techniques de séquençage des génomes, transcriptomes et protéomes ont révolutionné l'arbre phylogénétique du vivant en permettant l'analyse de dizaines ou de centaines de milliers de caractères moléculaires. Celui qui a été établi à partir de la comparaison des séquences des ARN ribosomiques présente trois domaines (archées, eucaryotes, et bactéries) ; il n'est pas raciné de manière consensuelle. Une telle représentation en arbre ne permet pas de rendre compte de l'évolution réticulée qui résulte des transferts horizontaux de gènes, des symbioses ou des hybridations.

L'arbre des eucaryotes comprend deux grands clades : les unicontes, parmi lesquels se trouvent les opisthocontes (dont les animaux et champignons) et les bicontes (dont les plantae et les straménopiles). La racine de l'arbre des eucaryotes est elle aussi encore sujette à discussion. Certains groupes définis par leurs ressemblances écologiques apparaissent polyphylétiques : ils témoignent de convergences évolutives, comme les végétaux (ensemble d'organismes réalisant la photosynthèse oxygénique), les champignons (hétérotrophes filamenteux), ou encore les algues (photoautotrophes à thalles). L'origine endosymbiotique de la cellule eucaryote est notamment argumentée par le fait que les phylogénies moléculaires placent les ARNr 16S des mitochondries et des plastes au sein des bactéries. Il est admis que les mitochondries ont une origine unique. En revanche, plusieurs événements indépendants d'endocytobiose sont à l'origine des plastes des algues, notamment l'endocytobiose primaire d'une cyanobactérie à l'origine des plastes à deux membranes des plantae, ainsi que l'endocytobiose secondaire d'une algue rouge à l'origine des plastes à quatre membranes des straménopiles. L'état pluricellulaire est un état dérivé apparu à une vingtaine de reprises. La possession de plastes ou la pluricellularité ont été perdus par certains taxons (évolution régressive). L'évolution peut donc conduire à la complexification mais aussi à la simplification des organismes.

Les virus ne sont pas vivants : ce sont des parasites moléculaires incapables de se répliquer et d'évoluer par eux-mêmes ; leurs constituants sont synthétisés de novo à chaque cycle d'infection au lieu d'être en partie hérités d'un autre virus ; ils ne peuvent pas être placés dans l'arbre du vivant cellulaire.

Attention

- Sachez argumenter les nombreuses évolutions convergentes au sein des eucaryotes.
- Ne confondez pas l'existence supposée d'une structure (plaste, mitochondrie) et celle de leur fonction (autotrophie, respiration).
- Comprenez la notion d'évolution réticulée : les transferts horizontaux de gènes ont un rôle tout aussi important dans l'évolution que la reproduction sexuée (transferts verticaux d'informations).

S'entraîner

QCM de connaissances

- 1 Identifiez parmi les propositions ci-dessous celles qui expliquent que les relations de parenté entre les cinq clades majeurs du groupe des bicontes ne soient pas résolues, ce qui conduit à la multifurcation observée sur l'arbre des eucaryotes (figure 13.3).
 - a. Des séquences d'ADN ou d'ARN ont été analysées.
 - b. Des séquences de protéines ont été analysées.
 - c. Un seul gène a été analysé.
 - d. De nombreux gènes ont été analysés.
 - e. La cladistique avec le principe de maximum de parcimonie a été utilisée.
 - f. L'approche probabiliste avec maximum de vraisemblance a été utilisée.
 - g. Les cladogenèses des bicontes sont très espacées dans le temps.
 - h. Les cladogenèses des bicontes sont séparées par de très courts laps de temps.
- 2 Identifiez les propositions exactes parmi les suivantes.
 - a. Tous les virus sont petits et simples.
 - b. Tout organisme qui est petit et simple est ancestral.
 - c. Les virus sont les premières entités biologiques à être apparues sur Terre.
 - d. Les virus peuvent être replacés dans la phylogénie du monde vivant.
 - e. Les virus contribuent aux transferts horizontaux de gènes (THG).
 - f. Les virus sont des entités vivantes.
 - g. Les virus présentent des membranes héréditaires.

Question de synthèse courte

Origine endosymbiotique de la cellule eucaryote

Arguments pour comprendre les mécanismes de l'évolution

Activités pratiques

PLAN DU CHAPITRE

- 1 Analyser la structure génétique de populations
- 2 Utiliser des données pour argumenter sur la coévolution

INTRODUCTION

L'étude de la structure génétique des populations peut permettre de mettre en évidence les effets des mécanismes de l'évolution et de mieux connaître les caractéristiques biologiques (système de reproduction) et écologiques des populations considérées. En étudiant des espèces aux traits contrastés, il est possible de comprendre quels sont les principaux facteurs qui gouvernent la structure génétique des populations.

1 Analyser la structure génétique de populations

L'étude de la structure génotypique des populations suppose la comparaison avec la structure théorique attendue sous régime de reproduction panmictique. En cas d'écart significatif, des explications peuvent être proposées quant au régime de reproduction et/ou aux effets de forces évolutives. Parallèlement, les variations de structure allélique dans le temps et/ou la comparaison de différentes populations plus ou moins isolées et/ou soumises à diverses contraintes permettent d'aborder les mécanismes de l'évolution.

1.1 Comparer des données d'observation de populations naturelles avec celles attendues selon le modèle nul

L'escargot des haies, *Cepaea nemoralis*, largement distribué en Europe, présente un important polymorphisme de couleur. La limnée tronquée (*Galba truncatula*) est un petit escargot qui vit dans des mares et ruisseaux temporaires; elle est très étudiée car elle est vectrice d'un parasite d'importance médicale et vétérinaire, la grande douve du foie.

La structure des populations de ces mollusques hermaphrodites, qui peuvent chacun se reproduire par autofécondation ou allofécondation est étudiée au moyen de marqueurs neutres (non-soumis à sélection). Les deux loci étudiés dans le tableau sont représentatifs de l'ensemble des marqueurs neutres étudiés pour chacune de ces deux populations. Dans les deux cas, l'échantillonnage a été réalisé sur une très petite surface. Le [tableau TP8.1](#) résume les données acquises.

Voir chapitre 12,
figure 12.1

Tableau TP8.1 Effectifs génotypiques observés pour une population de *C. nemoralis* et une de *G. truncatula*.

modèle biologique	<i>C. nemoralis</i>	<i>G. truncatula</i>
site d'étude	prairie en Virginie (USA)	ruisseau temporaire en Occitanie (France)
locus et allèles	gène codant pour une isoenzyme allèles <i>A1</i> et <i>A2</i>	gène codant pour la leucine aminopeptidase allèles <i>B1</i> et <i>B2</i>
effectifs génotypiques observés	<i>A1A1</i> : 19 <i>A1A2</i> : 50 <i>A2A2</i> : 22	<i>B1B1</i> : 10 <i>B1B2</i> : 6 <i>B2B2</i> : 56
effectif total (N)	91	72

Modifié d'après Brussard et McCracken (1974) -Heredity- et Hurtrez-Boussès et al. (diverses sources).

On cherche à savoir si la structure génotypique observée s'approche de celle du modèle de Hardy-Weinberg.

Pour cela, on commence par utiliser les effectifs génotypiques observés pour calculer les fréquences alléliques. Puis, on calcule les effectifs génotypiques attendus sous modèle de Hardy-Weinberg : Np^2 et Nq^2 pour les homozygotes et $2pq$ pour les hétérozygotes.

Les résultats des calculs sont récapitulés dans le [tableau TP8.2](#).

Tableau TP8.2 Test de l'équilibre de Hardy-Weinberg pour deux populations étudiées à un locus diallélique neutre.

modèle biologique	<i>C. nemoralis</i>	<i>G. truncatula</i>
calcul des fréquences alléliques	$f(A1) = p = (2 \times 19 + 50) / (2 \times 91) = \mathbf{0,48}$ $f(A2) = q = (2 \times 22 + 50) / (2 \times 91) = \mathbf{0,52}$	$f(B1) = p = (2 \times 10 + 6) / (2 \times 72) = \mathbf{0,18}$ $f(B2) = q = (2 \times 56 + 6) / (2 \times 72) = \mathbf{0,82}$
calcul des effectifs génotypiques attendus sous l'équilibre de Hardy-Weinberg	<i>A1A1</i> : $Np^2 = 91 \times 0,484^2 = \mathbf{21}$ <i>A1A2</i> : $N2pq = 91 \times 2 \times 0,484 \times 0,516 = \mathbf{46}$ <i>A2A2</i> : $Nq^2 = 91 \times 0,516^2 = \mathbf{24}$	<i>B1B1</i> : $Np^2 = 72 \times 0,181^2 = \mathbf{2}$, <i>B1B2</i> : $N2pq = 72 \times 2 \times 0,181 \times 0,819 = \mathbf{22}$ <i>B2B2</i> : $Nq^2 = 72 \times 0,819^2 = \mathbf{48}$
conclusion	La population peut être considérée comme en équilibre de Hardy-Weinberg	La population n'est pas en équilibre de Hardy-Weinberg . Par rapport aux effectifs génotypiques attendus selon Hardy-Weinberg, on observe ici un déficit en hétérozygotes .

Les effectifs observés sont, pour *C. nemoralis*, très proches de ceux attendus, tandis que pour *G. truncatula* les effectifs observés sont très différents de ceux attendus dans le modèle de Hardy-Weinberg. Un test statistique adapté (test du Chi², non développé ici) permet de confirmer qu'il n'y a pas de différence significative pour *C. nemoralis*. En revanche, dans le cas de *G. truncatula*, les effectifs s'écartent significativement du modèle de Hardy-Weinberg : on observe un déficit en hétérozygotes par rapport à ce qui est attendu.

Voir chapitre 12,
zoom 6

La population d'escargots des haies a une structure génotypique proche de celle attendue sous panmixie : on en déduit que ce mollusque hermaphrodite se reproduit essentiellement par fécondation croisée et les croisements se font majoritairement au hasard.

Dans la population de limnées tronquées, on observe un fort déficit en hétérozygotes par rapport à ce qui serait attendu sous panmixie. La zone d'échantillonnage étant peu étendue, on peut probablement exclure un effet Wahlund. Ces limnées ont probablement un régime de reproduction fermé (fortement consanguin, voire autofécondant) – ce qui peut par ailleurs être confirmé expérimentalement. Sachant qu'elles vivent dans des habitats très temporaires (soumis à des assèchements périodiques), un effet de la dérive génétique (goulots d'étranglement fréquents) est également très probable.

1.2 Exploiter des données pour déterminer une modalité de sélection

Chez *C. nemoralis*, la couleur de la coquille est gouvernée par un gène principal. On s'intéresse ici à des populations où seuls deux allèles sont recensés : *R*, pour rose et *J* pour jaune. L'allèle *R* est dominant. Les fréquences alléliques ont été étudiées dans plusieurs populations en Europe (tableau TP8.3).

a) Émettre des hypothèses quant aux mécanismes de l'évolution responsables de la différenciation entre populations

Tableau TP8.3 Effectifs phénotypiques (escargots à coquille jaune et à coquille rose) et fréquences alléliques (*J* et *R*) dans 4 populations en Europe.

	nombre d'escargots		fréquence allélique	
	jaunes	roses	allèle <i>J</i>	allèle <i>R</i>
Stockholm (Suède)	23	33	0,64	0,36
Dilbeek (Belgique)	22	24	0,69	0,31
Niederbronn (France, Alsace)	50	15	0,82	0,18
Montségur (France, Ariège)	60	8	0,94	0,06

D'après les données du tableau, la proportion d'escargots à coquille jaune augmente avec la latitude. Deux hypothèses peuvent être proposées pour expliquer cela.

Hypothèse 1 : expansion récente de l'aire de distribution, avec une colonisation progressive du nord au sud (ou du sud au nord), avec une meilleure aptitude à la dispersion des escargots à coquille jaune (respectivement, à coquille rose). Cela conduirait à une différenciation des populations par la distance.

Hypothèse 2 : sélection en faveur des escargots jaunes aux plus hautes latitudes. La pression de sélection pourrait être de nature abiotique (augmentation des températures du nord au sud) ou biotique (par exemple, une ressource alimentaire plus favorable aux escargots jaunes ou un prédateur qui préférerait les escargots à coquille jaune pourraient être d'autant plus abondant qu'on va vers le sud).

b) Proposer des hypothèses quant aux pressions de sélection

Cain et Shepard (1954) ont étudié, dans un site en Angleterre, la répartition des morphes dans deux types d'habitats, prairies ou forêts (figure TP 8.1). Silvertown et coll. (2011) ont, quant à eux, compilé des données sur les proportions, toutes régions confondues, de morphe jaune dans deux types d'habitats, forêts et dunes (tableau TP8.4).

De plus, des études de physiologie ont montré que le pouvoir de réflexion des rayons solaires par la coquille est d'autant plus important que celle-ci est claire et que les escargots à coquille claire accumulent moins la chaleur que ceux à coquille plus sombre.

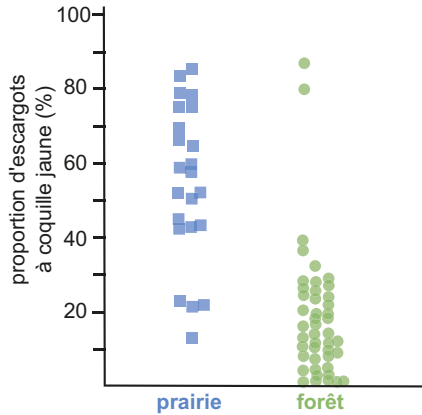


Figure TP 8.1 Proportion par population d'escargots à coquille jaune dans deux types de milieux : prairie (carrés bleus) et forêt (cercles verts). Chaque point représente une population.

Modifié d'après Cain et Sheppard (1954) - Genetics-

Tableau TP8.4 Fréquence d'escargots des haies à coquille jaune, en 1950 et 2010, dans deux types d'habitats. Modifié d'après Silvertown et al. (2011) - PLoS-ONE-

année	1950	2010
forêts	0,38	0,44
dunes	0,53	0,67

Les escargots à coquille jaune sont plus abondants dans les milieux ouverts (prairies, dunes), où l'ensoleillement est fort, que dans les milieux fermés (forêts), où l'ombre est plus importante. En cohérence avec l'effet de la latitude, il semble donc que **l'ensoleillement soit une pression de sélection** sur les populations de *C. nemoralis* : un fort ensoleillement sélectionnerait les escargots à coquille jaune. Deux hypothèses, non-exclusives, peuvent permettre d'expliquer cela :

- **effet thermique** : les escargots à coquille jaune, qui réfléchit mieux les rayonnements, résisteraient mieux à la dessiccation que leurs congénères à coquille rose. Ceci est cohérent avec les études physiologiques citées plus haut ;
- **détection par les prédateurs** : les escargots jaunes pourraient être moins repérables par les prédateurs lorsqu'ils sont dans des milieux ouverts, tandis qu'en milieu ombragé les escargots roses, plus sombres, seraient avantagés car moins détectables.

c) Proposer un scénario évolutif dans un contexte de changement climatique

D'après le [tableau TP8.4](#), en 60 ans, la proportion d'escargots à coquille jaune a augmenté, dans les forêts et dans les dunes. Cette augmentation est plus marquée dans les dunes.

Compte tenu de ce qui a été dit plus haut, et sachant que les températures moyennes ont augmenté de 0,6 °C durant la période étudiée, la pression de sélection exercée par les températures serait de plus en plus importante et le morphe jaune serait de plus en plus favorisé. La pression de sélection est plus importante dans les milieux très ouverts que sont les dunes, que dans les forêts où des refuges plus frais peuvent être trouvés.

D'après les analyses ci-dessus, la **pression de sélection exercée par les températures et l'ensoleillement** semble être un **moteur majeur dans l'évolution du polymorphisme de couleur** chez l'escargot des haies.

Voir TP 17, § 1

1.3 Repérer des arguments en faveur d'un effet fondateur

Des populations de limnée tronquée se retrouvent en différents endroits du monde (Europe, Afrique du Nord, Amérique du Sud). Une vaste étude a permis d'analyser la diversité allélique à des marqueurs neutres (6 loci microsatellites) dans 13 sites de l'Altiplano bolivien, où *Galba truncatula* a été observée pour la première fois en 1975, et des populations en Europe et Afrique du Nord. Les résultats sont récapitulés dans le [tableau TP 8.5](#). D'autres études ont montré que les populations boliviennes sont génétiquement plus proches des populations portugaises que de n'importe quelle autre population.

Tableau 8.5 Diversité allélique dans des populations de *Galba truncatula*.

Populations	M	F 1	F 2	E 1	E 2	E 3	E 4	P	AB
Nombre moyen d'allèles/locus	1,7	1,7	1,2	1,5	2,2	2	1,7	1,8	1

Populations du Maroc (M), France (F, 2 populations), Espagne (E, 4 populations), Portugal (P) et Altiplano bolivien (AB, 13 populations confondues). Modifié d'après Meunier et al. (2001) -Molecular Ecology-.

D'après le tableau, les populations d'Europe et d'Afrique du Nord présentent une **faible diversité allélique**. Celle-ci est cohérente avec l'hypothèse évoquée plus haut d'un fort effet de la **dérive génétique** chez ce mollusque qui vit dans des milieux temporaires et subit donc des goulots d'étranglement répétés. Sur l'Altiplano, il n'y a absolument **pas de diversité génétique** aux loci étudiés : les 13 populations sont entièrement monomorphes. Cette absence de diversité pourrait s'expliquer par un **effet fondateur** récent ; ceci est cohérent avec l'arrivée de la limnée tronquée sur l'Altiplano remontant à seulement quelques décennies. Cette hypothèse est soutenue par les similitudes entre populations boliviennes et portugaises : quelques individus, issus d'une population de la péninsule ibérique, auraient été importés sur l'Altiplano. Les capacités d'auto-fécondation de la limnée tronquée lui auraient permis une colonisation très rapide des sites, sans que la diversification (par mutations puis sélection) ait eu le temps de se faire. La dérive génétique semble donc être le moteur majeur dans l'évolution des populations de *G. truncatula*.

2

Utiliser des données pour argumenter sur la coévolution

Lapeirousia anceps est une plante de la famille des Iridacea de la région du Cap en Afrique du Sud. La corolle de sa fleur est caractérisée par un tube de plusieurs cm de long, au fond duquel se trouvent les nectaires. L'insecte diptère *Moegistorhynchus longirostris* possède une trompe (proboscis) longue de plusieurs cm également, qui lui permet de consommer le nectar de cette plante.

2.1 Analyser des données biologiques pour caractériser une interaction

La [figure TP8.2](#) décrit les résultats d'une expérience visant à caractériser la nature de l'interaction entre la plante et l'insecte. Pour cela, des bourgeons floraux de *L. anceps* ont subi une ablation des anthères ; au moment de la floraison, la longueur du tube de chaque fleur est mesurée et la quantité de nectar (visible par transparence) est notée. Chaque fleur est surveillée : dès qu'un *M. longirostris* vient la visiter, il est capturé et la longueur de sa trompe est mesurée. La fleur visitée est prélevée : le niveau de nectar est à nouveau mesuré (la différence entre les deux mesures correspond à ce qui a été prélevé par l'insecte) et les grains de pollens déposés sur le stigmate sont comptés.

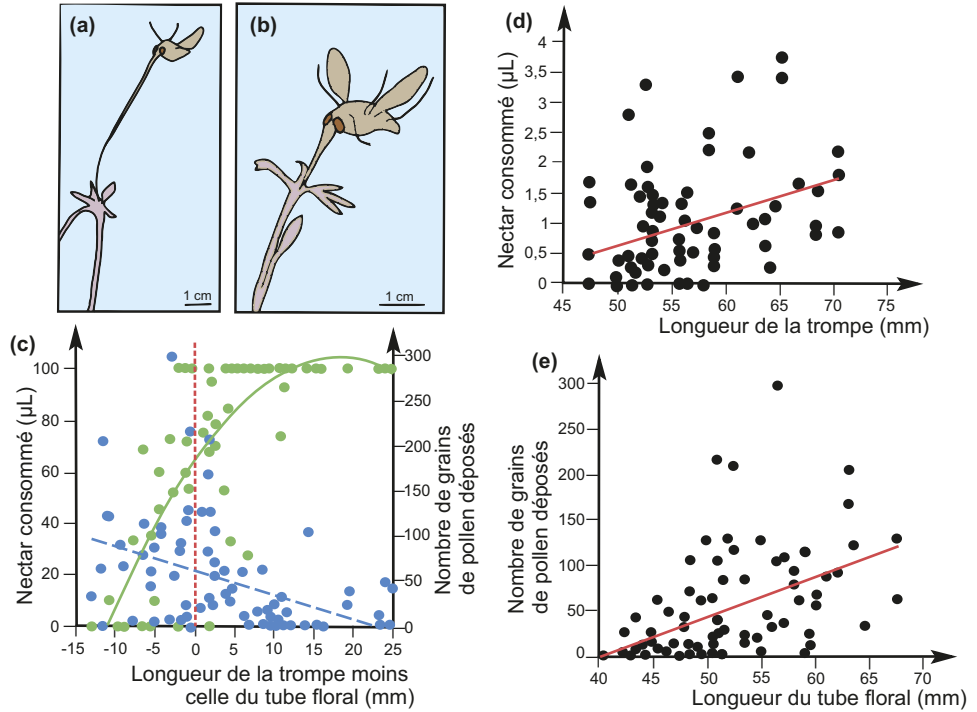


Figure TP 8.2 Interactions entre une plante et son pollinisateur.

(a) et (b) schémas de la fleur de *Lapeirousia anceps* et de l'insecte *Moegistorhynchus longirostris* ;

(c) quantité de nectar consommé (en vert) et nombre de grains de pollen (en bleu) déposés en fonction de la différence de longueur entre la trompe de l'insecte et le tube floral et courbes d'ajustement correspondantes ;

(d) quantité de nectar consommé en fonction de la longueur de la trompe de l'insecte et droite de régression correspondante ;

(e) nombre de grains de pollen déposés en fonction de la longueur du tube floral et droite de régression correspondante.

(Modifié d'après Pauw et al., 2009).

Voir ouvrage
de 1^{re} année,
chapitre 17, § 2.1

La figure montre une relation positive significative entre la longueur de la trompe de l'insecte et la quantité de nectar consommé. La relation pourrait être qualifiée de prédateur (insecte)-proie (plante) et l'insecte présente une adaptation morphologique à l'appropriation de la nourriture. Toutefois, du pollen est aussi déposé lors des visites de l'insecte ; or, la fleur ayant été privée de ses étamines, il ne peut pas s'agir de son propre pollen. *M. longirostris* est donc un pollinisateur de la plante. Il s'agit d'une association à bénéfice réciproque (la plante fournissant le nectar et l'insecte permettant la pollinisation) : cette interaction est donc un **mutualisme**.

Compte tenu des longueurs très importantes du tube de la fleur et de la trompe de l'insecte, on peut penser que la plante a peu, voire pas, d'autres pollinisateurs et que l'insecte n'exploite que peu, voire pas, d'autres plantes. Il y a donc probablement une forte spécificité entre *L. anceps* et *M. longirostris*.

2.2 Repérer les pressions de sélection réciproques

La figure TP8.2e montre une relation significative positive entre la longueur du tube de la plante et le nombre de grains de pollens recueillis. On peut proposer que plus le tube est long, plus l'insecte force et passe du temps pour atteindre les nectaires et donc plus il se frotte contre le

stigmate et y dépose du pollen. Ceci est confirmé par l'analyse des relations entre la différence « longueur de la trompe – longueur du tube de la fleur » et la quantité de nectar consommée d'une part et le nombre de grains de pollen déposé d'autre part (résultats de la [figure TP8.2c](#)). Plus la trompe est longue par rapport au tube floral, plus le prélèvement de nectar par l'insecte est important : les insectes à trompe plus longue que le tube sont avantagés. On s'attend donc à une **sélection directionnelle en faveur d'un allongement de la trompe**. Inversement, plus le tube est long par rapport à la fleur, plus la pollinisation est efficace : les plantes à tube plus long que la trompe ont une meilleure fitness. On s'attend donc à une **sélection directionnelle en faveur d'un allongement du tube floral**.

2.3 Élaborer un scénario coévolutif

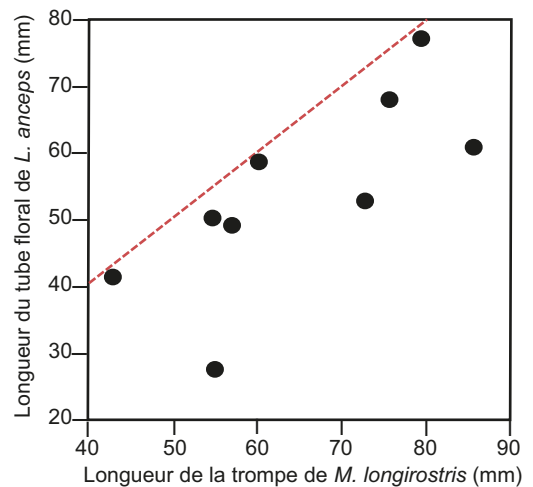
La longueur du tube de la fleur exerce une pression de sélection sur la longueur de la trompe de l'insecte (les insectes qui ont une trompe trop courte sont contre-sélectionnés). Réciproquement, la longueur de la trompe de l'insecte exerce une pression de sélection sur la longueur du tube floral : les plantes qui ont un tube trop court ne sont pas pollinisées et sont donc contre-sélectionnées. Un **insecte « tricheur »**, qui aurait une trompe très longue et prélèverait systématiquement le nectar sans polliniser aurait une fitness supérieure à celle d'un insecte « honnête », puisqu'il gagnerait le bénéfice de l'association (la nourriture) sans investir de coût (l'acte de pollinisation) ; on pourrait donc s'attendre à ce qu'il envahisse la population. Symétriquement, une **plante « tricheuse »** qui attirerait l'insecte et se ferait polliniser sans fournir de nectar (tube anormalement long ou nectaires non-fonctionnels) aurait une fitness supérieure à celle des autres plantes et devrait donc augmenter en fréquence. Quand la quantité d'insectes tricheurs augmente fortement, il n'y a plus assez de pollinisation dans la population de plantes : celle-ci s'effondre. Ce faisant, la quantité de nourriture disponible pour les insectes diminue aussi, entraînant un effondrement de leurs effectifs aussi. De la même manière, s'il y a beaucoup de plantes tricheuses, la population d'insectes voit ses effectifs diminuer, ce qui entraîne la chute des effectifs de plantes. On s'attend donc à des **dynamiques fréquence-dépendantes**.

2.4 Proposer des hypothèses pour expliquer la diversité entre populations

L'étude expérimentale a été complétée par des expériences *in natura* dans plusieurs populations : à chaque visite de plante, l'insecte a été capturé et sa trompe mesurée. La longueur du tube de la fleur a également été mesurée. Les résultats pour l'ensemble des populations sont présentés sur la [figure TP 8.3](#).

Figure TP8.3 Relation entre la longueur de la trompe de l'insecte *Moegistorhynchus longirostris* et la longueur du tube floral de *Lapeirousia anceps* dans un ensemble de populations.

Chaque point représente la moyenne des valeurs pour une population. La droite correspond à des longueurs identiques de la trompe de l'insecte et du tube floral.





Par ailleurs, d'autres travaux plus récents ont montré que l'endroit où le pollen se dépose sur la trompe de l'insecte est variable.

Les résultats par population montrent une relation positive entre longueur de la trompe de l'insecte et longueur du tube floral : cela confirme les prédictions faites ci-dessus quant à la **coévolution**. Toutefois, aucune population n'est au-dessus de la droite qui modélise une correspondance parfaite : les longueurs des trompes sont, dans certaines populations (points le long de la droite) très proches des longueurs des tubes floraux mais sont, dans d'autres populations, plus longs que les tubes.

Les longueurs ne sont pas équivalentes pour les différentes populations : on peut donc proposer que d'**autres contraintes** s'exercent sur les plantes et les insectes, qui font que la sélection directionnelle ne s'exerce pas indéfiniment. Le fait que dans certaines populations la longueur de la trompe soit plus importante que celle du tube floral pourrait s'expliquer par une diversité des modes de pollinisation : si l'accumulation de pollen se fait par exemple, non pas au niveau de la tête de l'insecte, mais à mi-distance sur la trompe, le scénario coévolutif permet de prédire que la longueur de la trompe devrait être deux fois supérieure à celle du tube floral.

Analyser les arbres phylogénétiques pour construire des scénarios évolutifs

Activités pratiques

PLAN DU CHAPITRE

- 1 Identifier un transfert horizontal de gène (THG)
- 2 Comprendre l'histoire évolutive d'un groupe

INTRODUCTION

Les arbres phylogénétiques sont des hypothèses sur les relations de parenté estimées à partir de la comparaison de différents caractères présents chez des organismes d'intérêt. En retour, leur analyse permet de construire des scénarios évolutifs. Deux exemples illustrent les méthodes de raisonnement qui permettent de le faire.

1 Identifier un transfert horizontal de gène (THG)

1.1 Données biologiques : aligner des séquences protéiques chez les bactéries

La phylogénie de 10 bactéries appartenant aux genres *Salmonella*, *Escherichia*, *Yersinia*, *Buchnera*, *Wigglesworthia*, *Haemophilus*, *Pasteurella*, *Vibrio*, et *Pseudomonas* est estimée à l'aide de séquences d'acides aminés de quatre protéines, déduites des génomes : la chaîne A de l'ATP synthétase (ATPB), la biotine synthétase (BIOB), la sous-unité A de l'ADN gyrase (GYRA), et la chaîne alpha de l'ARN polymérase ADN dépendante (RPOA). Trois groupes externes appartenant aux genres *Xanthomonas* et *Xylella* sont utilisés.

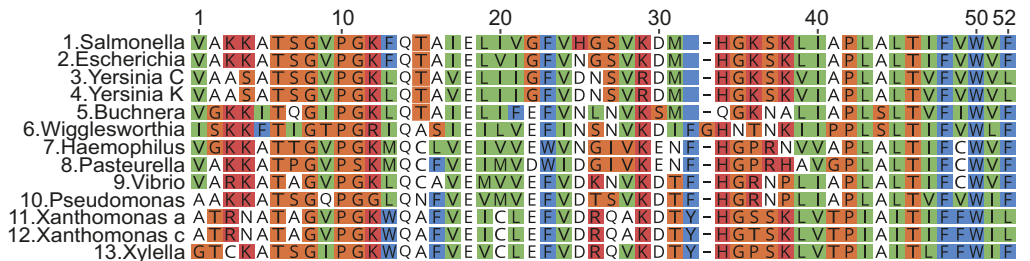


Figure TP 9.1 Extrait d'alignement de séquences de la protéine ATPB.

Sur chaque colonne (numérotée de 1 à 52), les acides aminés sont donnés avec leur code à une lettre, et coloriés en fonction de leurs propriétés biochimiques. Le tiret à la colonne 33 est un « gap » : il matérialise un événement d'insertion d'acide aminé chez *Wigglesworthia* ou de délétion chez toutes les autres bactéries. [D'après Lerat E, Daubin V, Moran NA (2003) From gene trees to organismal phylogeny in Prokaryotes: The case of the gamma-Proteobacteria. PLoS Biology 1:101-109. <https://doi.org/10.1371/journal.pbio.0000019>].

Les séquences homologues des quatre protéines sont alignées, c'est-à-dire que les positions homologues de chaque séquence sont placées les unes en dessous des autres, et représentent une colonne ou un site de l'alignement. Chaque colonne numérotée constitue un caractère moléculaire, et chaque acide aminé constitue un état de caractère moléculaire (figure TP9.1).

Voir ouvrage de
1^{re} année, chapitre 18,
tableau 18.3

Voir chapitre 13,
§ 2.2c

1.2 Comparaison des phylogénies : identifier l'incongruence

À partir des quatre alignements de séquences protéiques, les arbres phylogénétiques ont été estimés par l'approche probabiliste fondée sur le maximum de vraisemblance (figure TP9.2). Cette méthode est en effet moins sensible à l'artéfact d'attraction des longues branches que la cladistique avec principe de maximum de parcimonie.

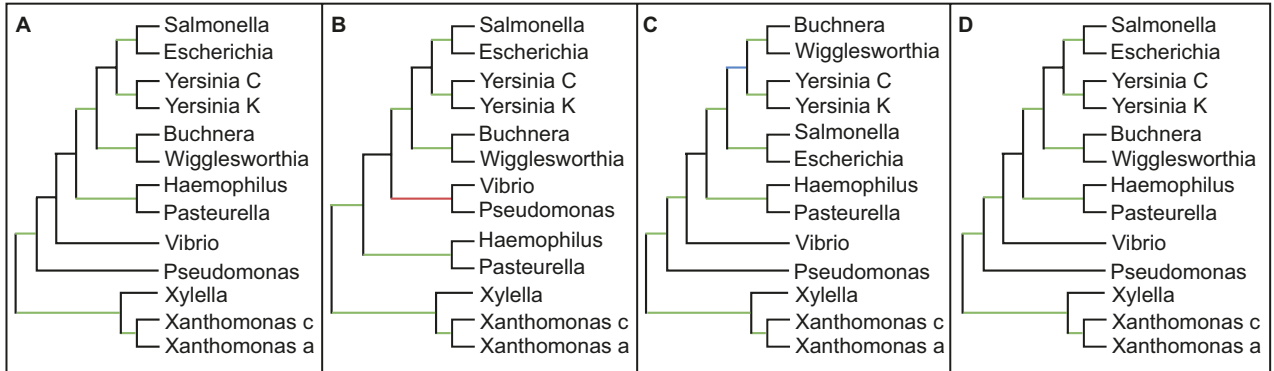


Figure TP9.2 Arbres phylogénétiques estimés pour les mêmes bactéries mais à partir de quatre protéines différentes.

Les branches vertes conduisent à des clades identiques pour les quatre topologies. Les branches bleue et rouge indiquent des incongruences soutenues de manière respectivement faible et forte par une analyse de bootstrap. (A) Arbre d'ATPB (estimé à partir d'un alignement de 298 sites) ; (B) Arbre de BIOB (368 sites) ; (C) Arbre de GYRA (1 042 sites) ; (D) Arbre de RPOA (335 sites).

Pour analyser ces résultats phylogénétiques, il nous faut donc comparer les quatre arbres. Comme dans toute comparaison, cela passe par l'identification des points communs et des différences qui existent entre ces quatre topologies.

- Les points communs, ou **congruences**, s'interprètent par le fait que les gènes qui codent ces protéines sont liés dans un même génome bactérien : ils ont tous subi la même histoire évolutive, et ils conduisent donc tous à inférer la même phylogénie. Les clades définis par les branches vertes se retrouvent ainsi dans tous les arbres :
 - *Salmonella* + *Escherichia* ;
 - les deux souches de *Yersinia* ;
 - *Buchnera* + *Wigglesworthia* ;
 - le clade formé par les 6 bactéries précédemment mentionnées ;
 - *Haemophilus* + *Pasteurella* ;
 - *Xanthomonas* + *Xylella*. Sans surprise, ces trois bactéries sont le groupe-frère de toutes les autres car elles ont été définies comme groupe externe.
- Les différences topologiques, ou **incongruences** sont les suivantes :
 - *Yersinia* est le groupe-frère de *Buchnera* + *Wigglesworthia* pour GYRA (branche bleue de la figure TP9.2C), alors que c'est le groupe-frère de *Salmonella* + *Escherichia* pour les trois autres protéines ;
 - *Vibrio* est le groupe-frère de *Pseudomonas* pour BIOB (branche rouge de la figure TP9.2B), alors que c'est le groupe-frère du clade *Salmonella* + *Escherichia* + *Yersinia* + *Buchnera* + *Wigglesworthia* + *Haemophilus* + *Pasteurella* pour les trois autres protéines.

1.3 Interprétation de l'incongruence entre arbres : poser l'hypothèse d'un THG

Voir ouvrage
de 1^{re} année,
chapitre 18, § 2.2b

Toutes les relations de parenté qui sont congruentes pour les quatre protéines suggèrent qu'elles sont représentatives de la phylogénie des espèces bactériennes correspondantes. En revanche, nous observons deux cas d'incongruence phylogénétique. Lorsque ce phénomène est observé, il y a une question à se poser immédiatement : l'incongruence observée est-elle statistiquement significative ? En d'autres termes, la branche matérialisant l'incongruence topologique est-elle robuste ? La réponse réside dans l'estimation par bootstrap du nombre de synapomorphies qui soutient le clade considéré.

- Pour la **phylogénie de GYRA** : en raison d'une faible valeur de bootstrap, il y a trop peu de signal phylogénétique pour considérer la branche comme solide (sa fragilité est soulignée en bleu). Dans ce cas, il n'y a pas vraiment d'incongruence, mais plutôt un manque d'information phylogénétique.
- Pour la **phylogénie de BIOB** : en raison d'une forte valeur de bootstrap, il y a suffisamment de signal phylogénétique pour considérer la branche comme solide (sa robustesse est soulignée en rouge). Dans ce cas, il y a une forte incongruence, révélatrice d'un événement biologique.

La comparaison des quatre arbres nous montre un cas d'incongruence topologique (pour BIOB) qui implique *Vibrio* et *Pseudomonas*. Comme les trois autres phylogénies de protéines (ATPB, GYRA, et RPOA) indiquent que *Pseudomonas* occupe une position plus profonde dans l'arbre des bactéries, tandis que *Vibrio* est le groupe-frère du clade *Salmonella* + *Escherichia* + *Yersinia* + *Buchnera* + *Wigglesworthia* + *Haemophilus* + *Pasteurella*, nous pouvons supposer que cet ordre de branchement phylogénétique – qui est celui qui est majoritaire – reflète la phylogénie des espèces. Le branchement minoritaire *Vibrio* + *Pseudomonas* suggère quant à lui qu'un événement biologique a affecté la protéine BIOB – et donc le gène qui la code – chez ces deux bactéries.

Chez les bactéries, la cause la plus fréquente d'incongruence entre un « arbre de gène » (la phylogénie estimée à partir de la protéine BIOB) et l'arbre des espèces (la phylogénie estimée à partir des autres protéines) est le **transfert horizontal de gène (THG)**. Ici, la position phylogénétique de *Pseudomonas* est anormalement proche de celle de *Vibrio*. Nous pouvons donc supposer qu'un ancêtre de *Vibrio* a donné une copie du gène codant BIOB à un ancêtre de *Pseudomonas*, remplaçant alors la copie qui était présente chez ce dernier taxon. *Vibrio* apparaît comme le **donneur**, et *Pseudomonas* comme le **receveur**, ce qui conduit à une protéine BIOB de *Pseudomonas* qui se retrouve très apparentée à la protéine BIOB de *Vibrio*. S'il n'y avait pas eu de THG, *Pseudomonas* occuperait très certainement une position phylogénétique éloignée de *Vibrio*, comme c'est le cas pour les arbres d'ATPB, de GYRA, et de RPOA.

En conclusion, pour reconstruire la phylogénie des espèces, il faut d'abord identifier et **éliminer les arbres de gènes et de protéines qui ont été affectés par des événements de THG**. Ensuite, dans le cadre d'une approche phylogénomique, il faut **combinaison tous les alignements non affectés par les THG**, et réanalyser l'alignement du super-gène ou de la super-protéine ainsi obtenue afin d'inférer l'arbre des espèces. Ceci agit comme une **synergie** qui va permettre d'ajouter les homologues, c'est-à-dire le signal phylogénétique présent dans chacun des marqueurs moléculaires pris indépendamment. Par ailleurs, si nous étudions des événements anciens, la reconstruction phylogénétique impose l'emploi de caractères à évolution plus lente, comme c'est le cas des remplacements d'acides aminés, moins fréquents que les éventuels changements synonymes ou non synonymes de nucléotides dans les codons des gènes correspondants. Notons enfin que si la fréquence des THG est trop importante, il sera plus approprié de reconstruire l'évolution des taxons sous la forme d'un réseau phylogénétique plutôt que d'un arbre phylogénétique.

2

Comprendre l'histoire évolutive d'un groupe

Voir chapitre 12
et TP 8

Les mollusques gastéropodes sont des organismes très étudiés en biologie évolutive. Sur la [figure TP9.3](#), la phylogénie de limnées appartenant aux genres *Austropeplea*, *Radix*, *Omphiscola*, *Lymnaea*, *Pseudosuccinea*, *Galba*, et *Stagnicola* est estimée à partir de caractères moléculaires, et enracinée par trois groupes externes appartenant aux genres *Bulinus*, *Biomphalaria*, et *Physa*.

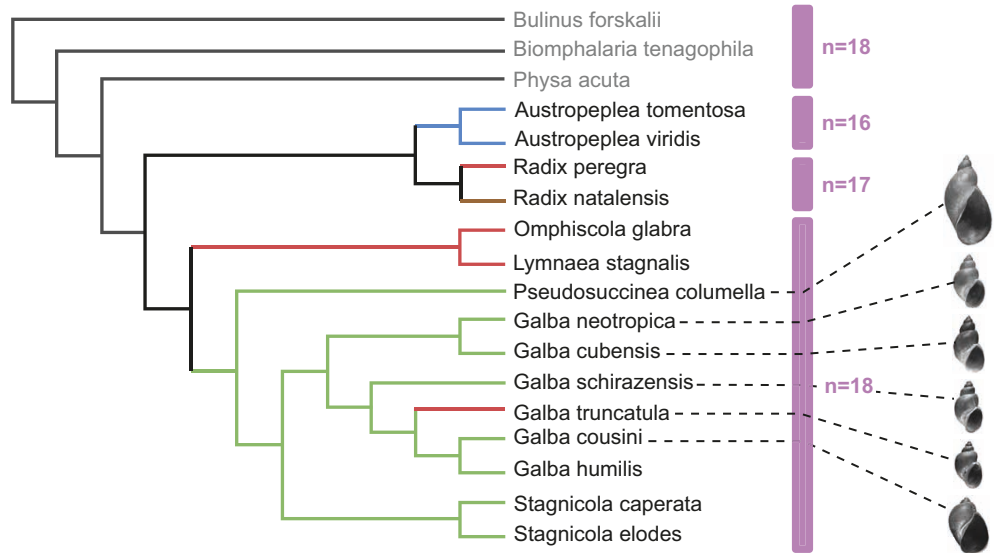


Figure TP9.3 Arbre phylogénétique de 7 genres de limnées et trois groupes externes (en gris) inféré par maximum de vraisemblance à partir de séquences d'ADN mitochondrial (gène de l'ARNr 16S) et nucléaire (espaceurs ribosomiques ITS1 et ITS2).

La morphologie de la coquille est représentée pour 6 taxons (les dimensions des plus petites d'entre elles n'excèdent pas 5 mm en largeur pour 10 mm en hauteur). Le nombre chromosomique haploïde de tous les taxons est donné sur le côté de l'arbre (en magenta). La couleur des branches indique l'origine géographique des limnées : Australasie (bleu), Eurasie (rouge), Afrique (marron), et Amérique (vert). [D'après Correa et al. 2010, BMC Evolutionary Biology 2010, 10:381 — <http://www.biomedcentral.com/1471-2148/10/381> ; Correa et al. 2011, Infection, Genetics and Evolution 11:1978-1988 — <https://doi.org/10.1016/j.meegid.2011.09.003>].

Cet arbre apporte des informations sur la morphologie des limnées, les événements chromosomiques qui ont accompagné leur évolution et la distribution géographique associée.

2.1 Identifier des espèces cryptiques à l'aide de la phylogénie moléculaire

Voir chapitre 12, § 1.1

Les coquilles des limnées du genre *Galba* peuvent être très similaires entre taxons proches ([figure TP9.3](#)). Par ailleurs, chaque espèce présente des gammes de variations morphologiques importantes en fonction des conditions environnementales (plasticité phénotypique). Le seul examen des caractéristiques de la coquille pourra conduire les biologistes à confondre plusieurs espèces (par exemple *Galba neotropica*, *G. cubensis*, *G. schirazensis*, et *G. truncatula*) en une seule : ces espèces sont dites **cryptiques**, c'est-à-dire présentant une importante ressemblance morphologique.

L'arbre de la [figure TP9.3](#) permet de montrer que des espèces phénotypiquement similaires sont **génétiquement très différenciées** car elles occupent des positions phylogénétiques bien distinctes. Les caractères moléculaires des séquences d'ADN mitochondrial et nucléaire permettent

Voir ouvrage de 1^{re} année, chapitre 18, § 1.1b

ainsi d'identifier la plupart des espèces de limnées tropicales américaines. Utiliser des caractères moléculaires appropriés peut donc servir de « codes-barres » moléculaires pour identifier de manière fiable des espèces morphologiquement cryptiques.

Selon la phylogénie de la figure TP9.3, le clade comprenant notamment *Galba neotropica*, *G. cubensis*, *G. schirazensis*, et *G. truncatula* regroupe des individus avec une petite coquille (en forme de tour), à l'exception notable de *Galba cousini* qui s'écarte de cette homogénéité morphologique. Néanmoins, l'inclusion de cette espèce dans le clade précédemment mentionné est sans ambiguïté, ce qui suggère alors que l'ancêtre de tous ces taxons aurait eu une morphologie turriciforme à petite coquille, semblable à celle de *Galba truncatula*, avant que, pour une raison inconnue, un changement de morphologie ait lieu le long de la branche menant à *Galba cousini*.

2.2 Proposer un scénario d'évolution chromosomique

La formule chromosomique a changé durant l'évolution des gastéropodes qui sont présentés ici, avec des taxons possédant 16, 17, ou 18 chromosomes à l'état haploïde. Nous pouvons nous appuyer sur l'arbre phylogénétique (figure TP9.3) pour comparer deux scénarios d'évolution chromosomique expliquant les données observées.

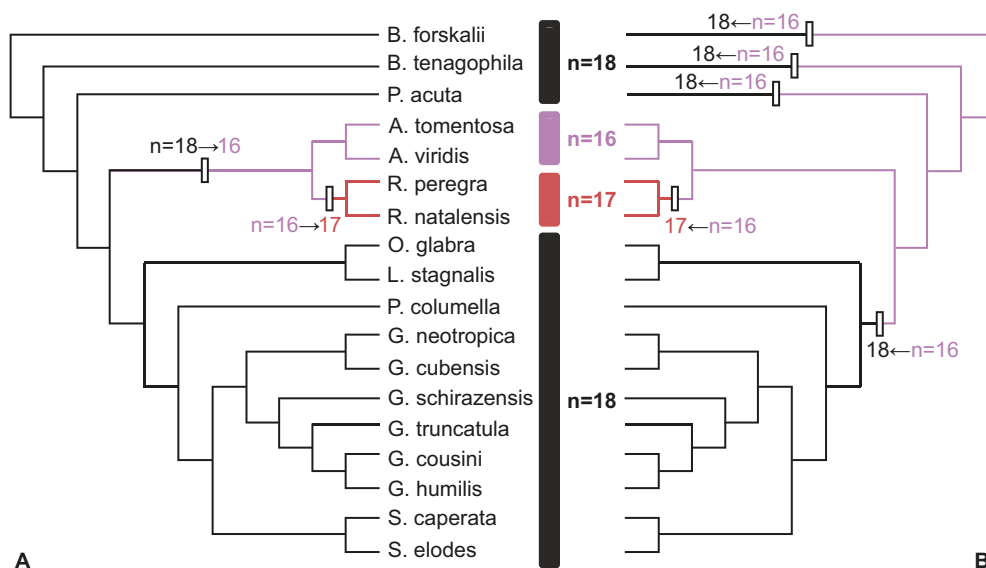


Figure TP9.4 Deux scénarios concurrentiels d'évolution chromosomique.

Les branches sont colorées en fonction des états chromosomiques (n = 16 en rose, n = 17 en rouge, et n = 18 en noir).

(A) Scénario dans lequel n = 18 est ancestral. Expliquer les données chromosomiques suppose un total de deux changements de nombre chromosomique.

(B) Scénario dans lequel n = 16 est ancestral. Expliquer les données chromosomiques suppose un total de cinq changements de nombre chromosomique.

Le scénario (A) est plus parcimonieux que le scénario (B). Dans un contexte d'économie d'hypothèses, c'est celui qui est préféré.

Dans le **premier scénario**, n = 18 est l'état plésiomorphe, et il faut supposer deux changements de formule chromosomique pour expliquer le caryotype des genres *Austropeplea* et *Radix* (figure TP9.4A) : une réduction du nombre haploïde — par exemple suite à deux fusions de deux chromosomes pour passer de n = 18 à n = 16 —, suivie d'un passage de n = 16 à n = 17 l'ancêtre de *Radix*, par exemple suite à une fission chromosomique. Une hypothèse alternative tout aussi parcimonieuse consisterait en une diminution graduelle du nombre de chromosomes,

avec un premier passage de $n = 18$ à $n = 17$ suivi d'un second passage de $n = 17$ à $n = 16$, cette fois-ci uniquement chez l'ancêtre de *Austropeplea*. Dans ces deux cas — ex-aequo du point de vue du nombre total de changements chromosomiques — il faut noter que les groupes externes (*Bulinus*, *Biomphalaria*, et *Physa*) posséderaient l'état ancestral ($n = 18$). Contrairement à une idée répandue, ce n'est pourtant pas toujours le cas : les groupes externes évoluent eux aussi, et ils ne présentent pas toujours leurs caractères dans l'état ancestral !

Dans le **second scénario**, $n = 16$ est l'état plésiomorphe, et il faut alors supposer cinq changements de formule chromosomique pour expliquer le caryotype des différents genres (figure TP9.4B) : quatre passages convergents de $n = 16$ à $n = 18$, et un passage de $n = 16$ à $n = 17$ chez l'ancêtre de *Radix*. Le coût total de ce scénario est de cinq changements, ce qui est moins économe que le scénario précédent qui n'impliquait que deux changements. Bien entendu, d'autres scénarios sont envisageables, mais aucun d'entre eux n'arrivera à atteindre ce nombre minimum de deux changements chromosomiques (figure TP9.4A).

2.3 Retracer un scénario de colonisation géographique

La distribution géographique actuelle des limnées comprend l'Australasie, l'Eurasie, l'Afrique, et le Nouveau Monde (figure TP9.3). Concentrons-nous sur le clade des limnées du Nouveau Monde, au sein duquel vient se brancher une limnée de distribution eurasiennne (*Galba truncatula*). Nous pouvons nous appuyer sur l'arbre phylogénétique pour comparer trois scénarios de colonisation géographique expliquant les distributions observées.

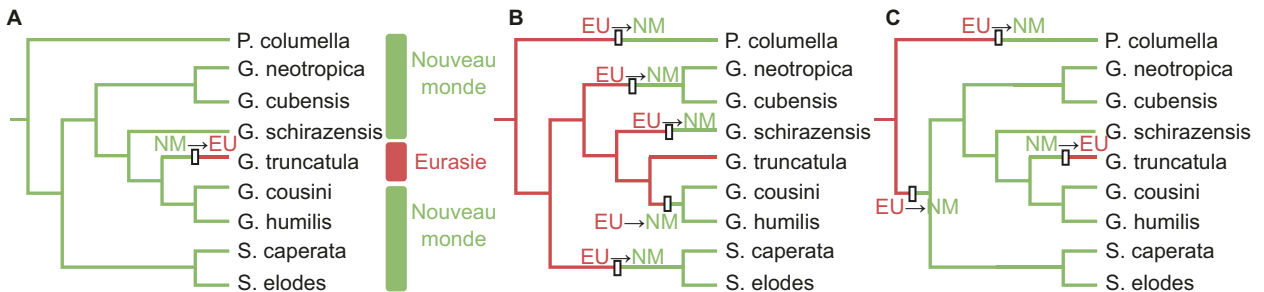


Figure TP9.5 Sous-arbre de limnées dont la distribution géographique couvre le Nouveau Monde (Amérique du Sud et du Nord : en vert) et l'Eurasie (en rouge).

(A) Scénario dans lequel la distribution dans le Nouveau Monde (NM) est ancestrale. Un seul changement d'aire de répartition vers l'Eurasie (EU) permettrait d'expliquer ce qui est observé pour *Galba truncatula*. **(B)** Scénario dans lequel la distribution en Eurasie est ancestrale. Cinq changements convergents d'aire de répartition vers le Nouveau Monde expliqueraient la répartition actuelle des limnées. **(C)** Scénario dans lequel la distribution en Eurasie est ancestrale. Deux changements convergents d'aire de répartition vers le Nouveau Monde et une réversion vers l'Eurasie expliqueraient la répartition des limnées.

Supposer une distribution géographique ancestrale dans le Nouveau Monde pour ces limnées implique un seul changement d'aire géographique pour *Galba truncatula*, ce qui constituerait un événement de migration du Nouveau Monde vers l'Ancien Monde (figure TP9.5A). En revanche, si nous supposons une distribution ancestrale eurasiennne, un scénario évolutif qui n'impliquerait que des changements convergents d'aire géographique coûterait un total de cinq migrations vers le Nouveau Monde (figure TP9.5B). Un scénario alternatif serait alors d'autoriser à la fois des changements convergents d'aire géographique et un retour vers l'aire d'origine — autrement dit une réversion — ce qui ne coûterait plus que trois migrations (figure TP9.5C). Dans tous les cas, le scénario le plus parcimonieux reste le premier, dans lequel *Galba truncatula* a migré une seule fois vers l'Eurasie.

De manière générale, lorsqu'un arbre phylogénétique est utilisé pour construire des scénarios évolutifs, le décompte du nombre d'événements évolutifs nécessaires pour expliquer les données observées implique souvent des convergences et des réversions. S'il n'y a aucune information quant à la probabilité relative des phénomènes de convergence et de réversion pour les caractères analysés, ils peuvent présenter le même score de parcimonie, et il faut accepter que différents scénarios évolutifs soient indiscernables les uns des autres. Pour les hiérarchiser, il faut alors faire appel à des informations supplémentaires. En revanche, si les caractères étudiés sont a priori plus susceptibles d'être sujets soit à convergence ou soit à réversion, alors l'un ou l'autre des scénarios évolutifs reconstruits pourra constituer la meilleure façon d'expliquer les données. Il y a par exemple des situations dans lesquelles les réversions sont improbables : retour à un plan d'organisation ancestral quand des structures morpho-anatomiques ont été perdues, facilité différentielle des fissions par rapport aux fusions chromosomiques selon la forme des chromosomes, ou recolonisation d'une aire géographique donnée lorsque les voies migratoires sont impraticables.