

Sous la direction de

DENIS FAURE, DOMINIQUE JOLY
et SYLVIE SALAMITOU

101 secrets de l'ADN

CNRS EDITIONS

Présentation de l'éditeur



Tout organisme vivant sur Terre contient de l'ADN, support de l'information génétique. L'étudier, c'est étudier le vivant dans toutes les composantes de sa diversité présente et passée, c'est saisir sa fragilité mais aussi sa capacité à s'adapter aux contraintes les plus diverses et les plus extrêmes, c'est comprendre qui nous sommes, nous humains, nos origines et dans quel monde complexe nous vivons.

Les années 1950 ont vu la découverte de la double hélice de cette molécule, puis dans les années 1970 émergent les premières technologies de séquençage qui permettent de décrire très précisément sa composition chimique, son organisation et son fonctionnement. Depuis, les progrès accomplis dans les technologies d'analyse et les recherches sur l'ADN et les génomes dans leur intégralité ont été considérables.

De l'arbre du vivant à la conservation des espèces, en passant par la paléogénomique, la phylogénie, le transfert de gènes, l'ARN, les symbioses, le microbiote, le recyclage des déchets, les *Big Data* appliquées au vivant, cet ouvrage propose 101 notions et objets de recherche au plus près de l'actualité scientifique.

Denis Faure (Institut de Biologie Intégrative de la Cellule - I2BC) et Dominique Joly (Évolution, Génomes, Comportement, Écologie - EGCE) sont biologistes, directeurs de recherche au CNRS.
Sylvie Salamitou est chargée de communication de l'unité EGCE.

101 secrets de l'ADN

Coordonné par Denis Faure,
Dominique Joly et Sylvie Salamitou

CNRS Éditions, Paris, 2019
ISBN : 978-2-271-12743-3

101 secrets de l'ADN

CNRS ÉDITIONS

15, rue Malebranche – 75005 Paris

SOMMAIRE

6 INTRODUCTION

26 PREMIER CHAPITRE
EXPLORATION HAUT DÉBIT DU VIVANT

L'accès à haut débit aux séquences ADN et ARN ouvre de nouvelles opportunités d'exploration et de caractérisation du vivant, dans tous les milieux, sous toutes les latitudes, et en révolutionne la compréhension.

64 DEUXIÈME CHAPITRE
INVENTAIRE ET CONNAISSANCE DE LA BIODIVERSITÉ

Du micro au macroscopique, de nouvelles espèces, y compris les plus rares, sont inventoriées grâce à l'exploration de l'ADN dans tous les écosystèmes. La massification des données ADN propulse l'étude de la biodiversité dans l'ère des *big data*.

112 TROISIÈME CHAPITRE
L'ADN, TÉMOIN MOLÉCULAIRE DE L'ÉVOLUTION DES ORGANISMES

Chaque espèce actuelle est le reflet de l'histoire de ses interactions avec l'environnement et avec les autres espèces. L'information contenue dans les séquences ADN d'un génome permet de retracer cette histoire évolutive avec une précision jusque-là inégalée.

144 QUATRIÈME CHAPITRE
VIVRE SOUS CONTRAINTES

Les organismes doivent s'adapter aux changements, fuir ou disparaître. L'analyse des génomes permet d'identifier les mécanismes de l'adaptation des organismes aux contraintes naturelles et anthropiques.

186 CINQUIÈME CHAPITRE
INTERACTIONS ENTRE ESPÈCES

Les relations entre espèces affectent positivement ou négativement la survie des organismes et leur évolution. Les nouvelles méthodes de séquençage ADN à haut débit permettent d'étudier ces interactions et d'en analyser les règles de fonctionnement.

216 SIXIÈME CHAPITRE
FONCTIONNEMENT ET FRAGILITÉ DES ÉCOSYSTÈMES

Les écosystèmes se distinguent les uns des autres par la diversité des organismes qui les composent et par les relations qui s'établissent entre eux et avec leur environnement. L'étude de la dynamique de ces relations permet de mieux comprendre le fonctionnement des écosystèmes.

256 SEPTIÈME CHAPITRE
PALÉOGÉNOMIQUE ET HISTOIRES ÉVOLUTIVES

L'étude des ADN anciens et des ADN contemporains permet de proposer des scénarios d'évolution et de dispersion des espèces, voire de reconstituer les écosystèmes passés.

290 HUITIÈME CHAPITRE
L'HUMANITÉ ET SON ENVIRONNEMENT

Les humains changent directement ou indirectement la biodiversité et les écosystèmes qui les entourent ; en retour, ces changements modulent à court et long termes le développement et la santé des populations humaines. Comprendre ces interactions, c'est pouvoir proposer de nouvelles pratiques en santé humaine et environnementale.

326 INDEX

330 TABLE DES MATIÈRES, AUTEURS, BIBLIOGRAPHIE



Introduction

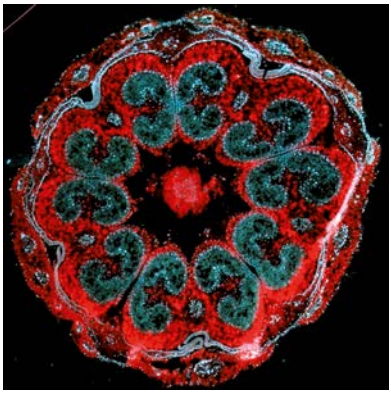
L'ADN révèle les secrets du vivant

Comme un livre auquel il est souvent comparé, l'ADN raconte le vivant, sa diversité, sa fragilité mais aussi sa capacité à s'adapter aux contraintes les plus diverses et les plus extrêmes grâce à des innovations biologiques plus surprenantes les unes que les autres.

Capter l'énergie solaire, utiliser des gaz atmosphériques comme nutriments et sources d'énergie, recycler les déchets, faire des « OGM », changer de sexe et de sexualité, produire des antibiotiques, y résister… voilà quelques exemples de ce que l'ADN révèle des innovations du vivant, dont certaines ont inspiré notre propre espèce, *Homo sapiens*.

Étudier l'ADN c'est comprendre la vie, son histoire, c'est aussi comprendre qui nous sommes, quelles sont nos origines, et dans quel monde complexe nous vivons.

Cet ouvrage invite au voyage dans le monde de l'ADN, un voyage dont les 101 escales révèlent autant de secrets du vivant.



Coupe d'un bourgeon de fleur de tomate. La couleur rouge correspond à la chlorophylle, les parties en bleu sont de l'ADN. Photographie A. Smirnova / GMCM / Unistra / Cemibio / CNRS Photothèque.

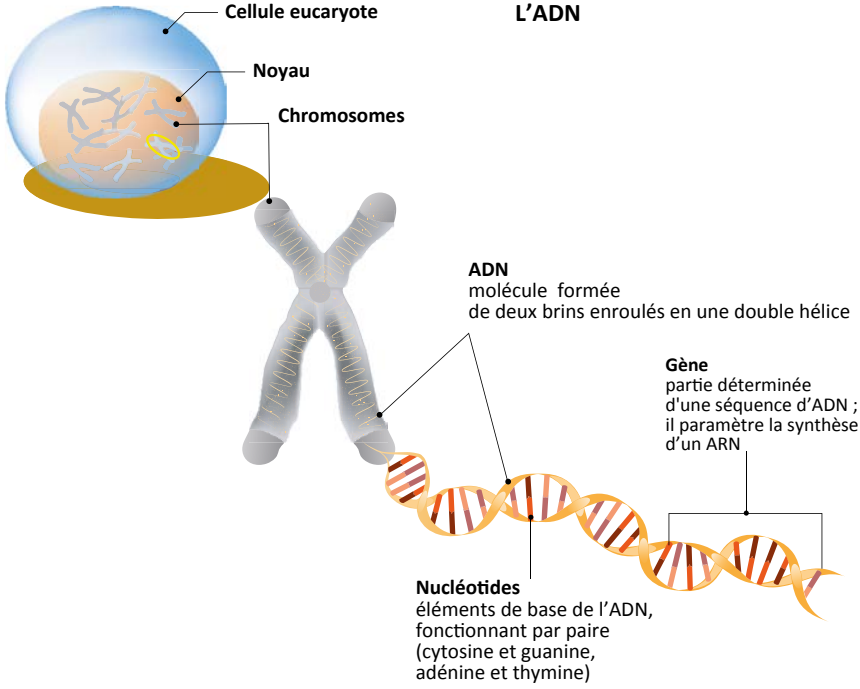
Pourquoi l'ADN ?

Plusieurs caractéristiques de l'ADN sont déterminantes pour comprendre à la fois l'information qu'il délivre aux chercheurs mais aussi les principales méthodes avec lesquelles il est étudié.

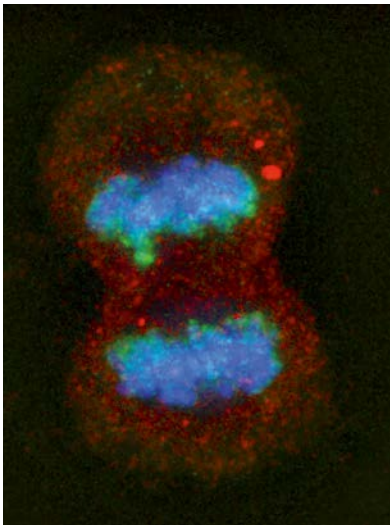
L'ADN est associé à la vie. L'ADN est l'acronyme d'acide désoxyribonucléique. C'est un polymère linéaire qui résulte de l'assemblage de plusieurs milliers voire milliards de briques moléculaires appelées nucléotides. Ces briques moléculaires sont de 4 types chimiques et

sont symbolisées par les lettres A, T, G et C. Dans les cellules, deux polymères d'ADN s'enroulent l'un autour de l'autre pour former une double hélice. La découverte de la structure de l'ADN a été récompensée par un prix Nobel en 1962 attribué à Francis Crick (1916-2004) et James Watson (né en 1928). L'ADN est produit par tous les organismes vivants, microscopiques ou macroscopiques, unicellulaires ou pluricellulaires. L'ADN est partout où s'est développée la vie sur Terre. Une vie sans ADN eu-t-elle été possible ? Peut-être. Celle que nous connaissons y est indissociablement liée. Étudier l'ADN, c'est étudier le vivant dans toutes les composantes de la diversité biologique, tel qu'il a été dans le passé, tel qu'il pourrait être demain en construisant des *scenarii* prédictifs de diversification, ou tel qu'il pourrait être sur une exoplanète.

Déterminer la séquence de l'ADN. Dans chaque organisme vivant, les milliers de briques moléculaires A, T, G et C de l'ADN sont agencées dans un ordre unique. Connaître cet ordre, c'est connaître la séquence ADN (le génome) d'un organisme. Cette séquence unique permet immédiate-



ment d'identifier un organisme, de le nommer et de le comparer à d'autres pour apprécier leurs similitudes et différences. Ainsi, l'ADN informe sur la taxinomie et la diversité du vivant. L'ADN est extrait d'un organisme, puis est séquencé avec des technologies dédiées. Une des premières techniques de séquençage a été inventée par Frederick Sanger (1918-2013) qui a partagé un prix Nobel pour cette découverte en 1980. Au cours des années 2000, d'autres avancées technologiques ont considérablement accru la rapidité du séquençage. Cette révolution du séquençage à haut débit permet d'accéder à moindre coût à la séquence ADN de tout organisme vivant. Associées à l'analyse informatique des séquences, ces nouvelles générations de séquenceurs ont profondément changé les stratégies d'étude de l'ADN, et ont contribué à l'explosion de nouvelles connaissances dont témoigne cet ouvrage.



Cellules en cours de division, observées en microscopie optique à fluorescence. En bleu, l'ADN, en vert, une des protéines autour desquelles l'ADN s'enroule pour former un chromosome. Les chromosomes se sont ici déjà divisés, et la cellule va le faire elle aussi, afin de donner deux cellules au patrimoine génétique identique. Photographie J.-Chr. Ame / BSC / Unistra / CNRS Photothèque.

Amplifier l'ADN. Une caractéristique originale de l'ADN est son mode de réplication dans les cellules. Une molécule d'ADN matrice se réplique en deux molécules d'ADN identiques. Cette réplication est réalisée par des enzymes appelées ADN polymérases. Leur étude est à l'origine de la découverte de la PCR (ou *polymerase chain reaction*) qui a valu à Kary Mullis (né en 1944) de partager un prix Nobel en 1993 avec Michael Smith. Par cette technique, il est possible de répliquer facilement, à moindre coût et en grande quantité, un fragment d'ADN grâce à l'action d'une ADN polymérase. En quelques heures, dans un volume de réactifs chimiques équivalent à une goutte d'eau, 30 cycles d'amplification suffisent pour obtenir 2^{30}, soit environ 1 milliard de fragments d'ADN identiques à partir d'un seul fragment ADN matrice. Ainsi, même si l'ADN est extrait en de faibles quantités à partir de cellules, d'organismes ou d'échantillons environnementaux

(sol, eau, etc.), l'étape d'amplification permet d'en obtenir une quantité suffisante pour réaliser son séquençage et son analyse. Cette révolution a considérablement facilité l'étude de l'ADN. La plupart des méthodes de détection et de caractérisation de l'ADN, que ce soit à des fins de recherche ou de diagnostic (médical, sanitaire, judiciaire…), font intervenir une étape d'amplification, par PCR ou par une méthode analogue. Aujourd'hui, certaines techniques visent à s'affranchir de cette étape d'amplification pour séquencer directement l'ADN issus des échantillons biologiques.

Transmission de l'ADN d'une cellule à l'autre. La transmission de l'ADN s'opère lors de la division des cellules. Une cellule mère duplique son ADN en deux molécules identiques – ou quasi-identiques en fonction des erreurs faites par exemple par les ADN polymérases –, puis la cellule mère se divise en deux cellules filles qui acquièrent chacune une copie de l'ADN. Ce mode de transmission est appelé vertical car il s'opère en une seule fois d'une cellule mère aux cellules filles et il existe chez tous les organismes vivants. Ainsi, la descendance d'une cellule mère (aussi appelée cellule ancestrale) forme une lignée de cellules filles (ou de descendants) ayant un ADN identique, à l'exception près des rares variations qui apparaissent lors de la réplication de l'ADN.

Transmission de l'ADN d'un individu à l'autre. Chez les organismes unicellulaires comme les bactéries, chaque cellule est un individu capable de se diviser par scissiparité : une lignée cellulaire bactérienne correspond à un ensemble d'individus identiques, c'est-à-dire des clones. Chez les organismes pluricellulaires comme les plantes et animaux, chaque individu est constitué de l'assemblage des cellules appartenant à une même lignée de cellules, donc ayant le même ADN. Ces cellules à l'ADN identique constituent les différents tissus des organismes

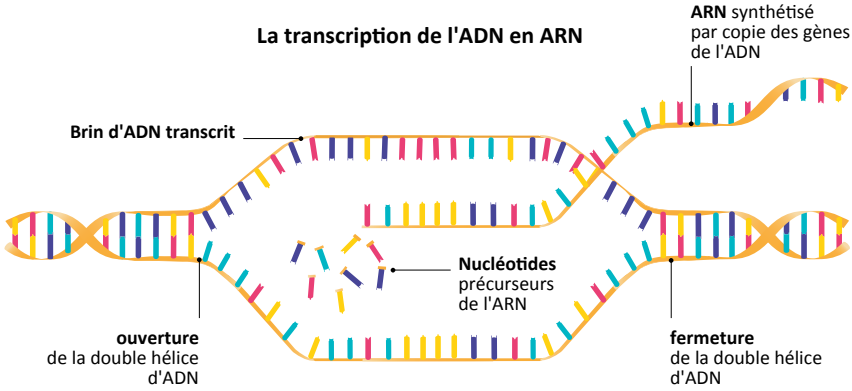


Phase initiale de la fécondation chez l'arabette des dames, observée en microscopie à balayage et colorée artificiellement : grains de pollen déposée sur un pistil. Chr. Mezard / INRA / CNRS Photothèque.

pluricellulaires. Chez eux, la cellule mère fondatrice de la lignée de cellules à l'origine d'un individu est issue soit d'un seul parent (lors d'une reproduction par parthénogénèse, par autofécondation, ou par prolifération végétative), soit de la fusion des cellules sexuelles provenant de deux parents (lors d'une reproduction sexuée). Au contraire des autres modes de reproduction, la reproduction sexuée génère à chaque génération des individus dont la séquence ADN est différente (même si elle présente de fortes similarités) de celle de chacun des parents. Bien que très présente chez certains groupes d'organismes vivants (tels les mammifères), la reproduction sexuée n'est pas le mode de reproduction le plus fréquemment observé dans le monde vivant.

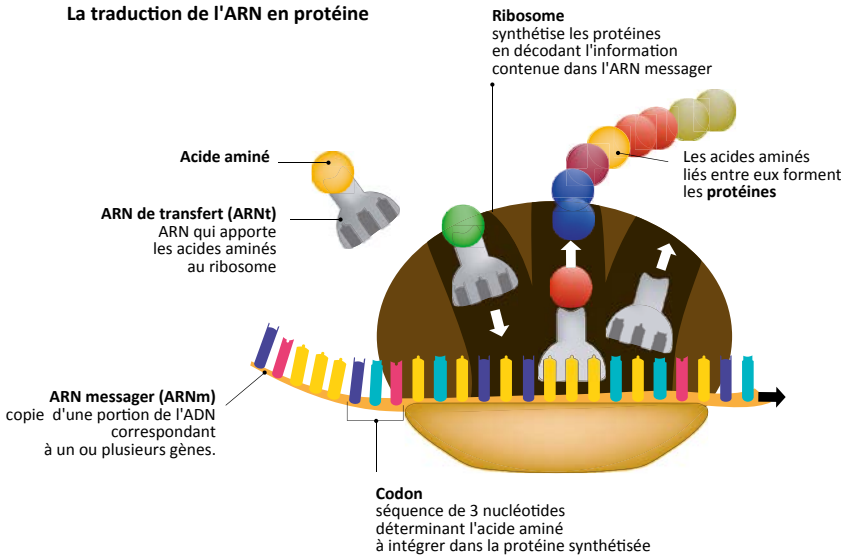
L'ADN révèle la parenté entre organismes. La séquence ADN permet de mesurer la parenté entre organismes : plus le nombre de différences entre les séquences ADN de deux organismes est réduit, plus faible est aussi le nombre de générations qui les séparent de leur ancêtre commun. Ce principe permet la reconstitution de la parenté entre les organismes appartenant à une même espèce, mais aussi à des espèces différentes. Il permet de comprendre l'évolution des organismes au fil des générations. L'arbre du vivant représente les liens de parenté de tous les organismes vivants obtenu par comparaison de leur séquence ADN. Ces méthodes d'analyse des organismes actuels permettent aussi de reconstruire, sur la base d'hypothèses évolutives, la séquence ADN de leur ancêtre disparu et ainsi d'élaborer des hypothèses quant aux conditions d'émergence des premiers organismes vivants. Par ailleurs, l'apparition d'incohérences entre les liens de parenté reconstruits grâce aux séquences ADN a révélé toute l'importance d'un autre mode de transmission : le transfert horizontal. Ce transfert d'ADN s'opère entre deux organismes, indépendamment de leur lien de parenté. Il a lieu également entre des espèces différentes, et agit comme un véritable accélérateur de l'évolution de l'ADN et du vivant.

L'ADN des gènes. La séquence ADN contient des informations essentielles à la forme (morphologie) et au fonctionnement (physiologie) d'un organisme : il s'agit, bien sûr, des gènes. Un gène correspond à une portion (quelques dizaines à milliers de nucléotides) de la séquence ADN qui va servir de matrice pour produire un autre polymère : l'ARN. L'ARN est l'acronyme d'acide ribonucléique. C'est un polymère linéaire qui résulte aussi de l'assemblage de nucléotides, légèrement différents de ceux présents dans l'ADN. Certaines de ces molécules vont à leur tour servir de matrice pour produire des protéines qui sont des constituants essentiels des cellules, dont des enzymes qui agissent comme catalyseurs pour la production d'autres constituants des cellules que sont les



métabolites comme les sucres, les lipides, les acides aminés, etc. Cette chaîne de production de molécules (ADN, ARN, protéines et métabolites) fait partie du métabolisme des cellules. Leur assemblage et leurs interactions définissent la forme et la physiologie des organismes. Et les virus ? Les particules virales ne sont pas des cellules, mais toutes les particules virales contiennent de l'ADN ou de l'ARN qui font l'objet de nombreuses recherches. Les virus influencent la viabilité, le fonctionnement et l'évolution des organismes infectés. L'ADN et l'ARN des particules virales se retrouvent soit dans les cellules infectées, soit à l'extérieur des organismes, dans l'environnement.

Reconstruire un organisme virtuel à partir de l'ADN. À partir de la séquence ADN d'un organisme, il est possible de prédire, avec certaines approximations, le nombre et la séquence ADN de ses gènes, ainsi que les caractéristiques des ARNs, protéines et métabolites qui peuvent être produits. Cette démarche prédictive qui analyse l'information contenu dans l'ADN, et plus particulièrement les gènes, relève de la génomique. L'intérêt de l'ADN est d'être stable (nonobstant les rares variations qui peuvent apparaître lors de la division cellulaire ou lors de transferts horizontaux) tout au long de la vie d'un individu. Ainsi, à tout moment de la vie, voire même après la vie d'un individu, il est possible par séquençage et analyse de sa séquence ADN de connaître non seulement sa taxinomie (l'espèce à laquelle il appartient), mais aussi de prédire (avec certains degrés d'approximation) sa morphologie et sa physiologie, sans l'avoir vu ni observé. La sophistication de la reconstruction d'un organisme virtuel à partir d'un ADN fait de plus en plus appel à un champ de recherche en émergence, la biologie des systèmes. Cette reconstruction du vivant peut être affinée par l'intégration d'autres sources d'information que l'ADN.



Seulement l'ADN ? Peut-on comprendre avec précision la morphologie et la physiologie des organismes par la seule analyse de leur ADN ? La réponse est non, car la production des ARNs, protéines et métabolites est influencée, voire conditionnée par l'environnement dans lequel vit chaque organisme. Par conséquent, si l'ADN permet de prédire une typologie de morphologie et de physiologie, seule l'analyse d'autres composants du métabolisme (ARNs, protéines, métabolites) permet de tenir compte de l'influence de l'environnement sur le développement et le fonctionnement d'un organisme. La séquence ADN informe sur ce qui est possible, l'analyse des autres composantes du métabolisme (ARNs, protéines, métabolites) informe sur ce que réalise un organisme dans un environnement donné. L'analyse à haut débit des variations de la production des ARNs (aussi appelés les transcrits), des protéines et des métabolites relève, respectivement, de la transcriptomique, protéomique et métabolomique. Les approches de génomique, transcriptomique, protéomique et métabolomique sont collectivement appelées les omiques. Outre les omiques, de nombreux paramètres sont mesurés chez les organismes vivants et dans leur environnement, comme certains paramètres physico-chimiques et rayonnements (son, lumière, UV, rayon X, etc.), grâce à une instrumentation très diversifiée capable de mesures de l'échelle moléculaire à l'échelle planétaire. Toutes ces approches permettent de décrire la morphologie, la physiologie

et le comportement d'un individu ou de groupes d'individus, et ainsi de comprendre comment ils vivent et s'adaptent à leur environnement.

L'ADN environnemental. L'ADN présent dans les écosystèmes comprend celui des organismes vivants et des virus, mais aussi celui que ces organismes libèrent tout au long de leur vie (notamment les fèces, poils, plumes) et qui constituent autant de traces signalant leur présence et leurs déplacements. L'ADN des écosystèmes comprend aussi celui des organismes morts, qui peut être préservé au cours du temps, et refléter la biodiversité du passé. Cet ADN prélevé dans les écosystèmes est communément appelé ADN environnemental. Plusieurs difficultés contraignent son étude. La première est l'accès aux échantillons : réaliser des prélèvements dans des sols, des forêts tropicales ou des océans n'est pas une opération triviale. Une deuxième est sa purification : l'ADN environnemental est parfois prisonnier de matériaux (présence d'acides humiques dans le sol, exosquelette minéral de certains organismes…) qui contraignent son extraction et sa purification. L'ADN qui passe cette étape d'extraction est parfois altéré et fragmenté, obtenu en de faibles quantités, et pas très pur, ce qui peut affecter l'étape de séquençage. Enfin, l'étape d'analyse informatique des séquences obtenues est d'autant plus complexe que la richesse en espèces est élevée dans l'échantillon étudié. En effet, l'ADN extrait des échantillons environnementaux est un mélange d'ADN de différentes espèces : savoir à quel organisme appartient une séquence devient une étape complexe quand il s'agit d'en analyser des milliards, provenant d'organismes appartenant à des milliers d'espèces. Les difficultés de ces quatre étapes d'analyse de l'ADN environnemental (échantillonnage, extraction et séquençage, puis analyse informatique des séquences) sont décuplées lorsqu'il s'agit d'ADN ancien provenant d'organismes morts il y a des milliers ou des centaines de milliers d'années. Qu'il s'agisse d'ADN contemporain ou ancien, les chercheurs doivent prendre des précautions – parfois drastiques – pour réduire les contaminations



Prélèvement et étiquetage de micro et macro-organismes à la station des Nouragues, Guyane. L'identification des bactéries et champignons utilisera l'ADN environnemental. Photographie Cl. Delhaye / CNRS Photothèque.

des échantillons étudiés par de l'ADN exogènes introduits accidentellement, et en premier chef celui du manipulateur.

Au cours de la dernière décennie des efforts considérables ont été réalisés par la communauté scientifique pour étudier l'ADN environnemental, et ainsi comprendre les écosystèmes présents et passés, leur complexité, leur dynamique et leur fragilité face aux changements environnementaux et aux pressions anthropiques.

Interroger l'ADN

nterroger l'ADN concerne plusieurs champs de recherche qui explorent la vie des organismes contemporains et du passé, y compris l'espèce humaine, en interaction avec leur environnement.

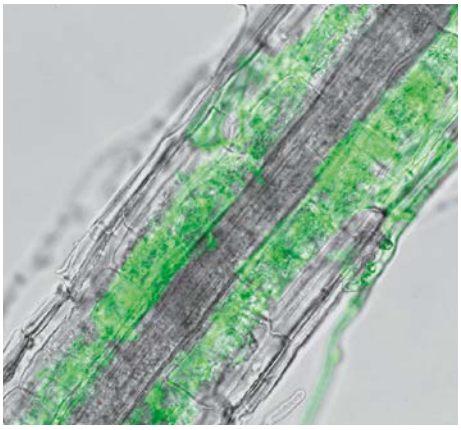
Taxinomie et biodiversité. Ce champ de recherche décrit et inventorie le vivant, il quantifie la biodiversité et ses variations dans les écosystèmes. Il a des applications majeures dans l'usage, le commerce et la protection du vivant. La classification taxinomique partagée par la communauté mondiale des scientifiques nomme chaque espèce par un nom et une épithète : le panda géant (nom commun) est appelé *Ailuropoda melanoleuca* (nom scientifique). Ces espèces sont regroupées en genres, eux-mêmes en familles, etc. selon une classification hiérarchique des taxons dont la nomenclature a été établie par Carl von Linné (1707-1778). L'ADN est utilisé pour définir avec précision les limites entre les



Un scientifique relâche un requin citron, *Negaprion acutidens*, après l'avoir examiné, et avoir prélevé un petit morceau de peau pour des analyses génétiques. Atoll de Tetiaroa, archipel de la Société, Polynésie française. Photographie Th. Vignaud / CNRS Photothèque.

espèces, donc définir une espèce et pour assigner un individu à une espèce connue. Pour ce faire, les chercheurs utilisent soit toute la séquence ADN des organismes (leur génome), soit des portions de cette séquence ADN. Les séquences ADN qui distinguent les taxons (espèce, genre, famille, etc.) sont appelées des marqueurs génétiques ou code-barres ADN. Le séquençage ADN à haut débit a changé certaines procédures de dénomination du vivant et de mesure de la diversité biologique. Par exemple, l'utilisation des codes-barres ADN permet d'identifier et de compter les espèces connues, mais aussi celles qui n'ont pas fait l'objet d'une description taxinomique. Le séquençage ADN des codes-barres permet ce qui était impossible hier : inventorier le vivant à grande échelle et mesurer les variations de la biodiversité dans les écosystèmes, notamment en réponse aux changements environnementaux et aux pressions anthropiques. Ces recherches alimentent les travaux de la plateforme intergouvernementale scientifique et politique sur la biodiversité et les services écosystémiques (IPBES) chargée de fournir les meilleures données disponibles pour éclairer les décisions affectant la nature.

Fonctionnement du vivant : de la cellule aux écosystèmes. Ce champ de recherche analyse comment se développe et s'organise le vivant, que ce soit à l'échelle d'une cellule, d'un organisme ou d'un assemblage d'organismes (populations et communautés) au sein d'un écosystème. Il étudie aussi les interactions entre les organismes vivants, et entre ceux-ci et leur environnement, ainsi que les innovations du vivant qui permettent aux organismes de se développer et se multiplier dans un environnement dont ils exploitent les ressources et subissent les contraintes. Il prend aussi en compte les effets des activités humaines sur leur environnement. Ainsi, à l'interface de la physique et de la biologie, se développe l'étude du cycle



Racine d'une plante terrestre dans laquelle est présent un champignon symbiotique mycorhizien, marqué en vert et observé en microscopie confocale grâce à un colorant spécifique. Photographie A. Le Ru / FRAIB / LRSV université Toulouse 3 / CNRS Photothèque.

des éléments fondamentaux du vivant (carbone, azote, phosphore, oxygène, hydrogène, fer…) : de leur incorporation dans les organismes vivants à leur retour dans l'environnement après la mort des organismes, en passant par les transferts de matière entre les organismes de la chaîne trophique. L'acquisition et l'analyse des données omiques (génomique, transcriptomique, protéomique, métabolomique) permettent d'identifier les organismes, les gènes et les activités biologiques (comme celles catalysées par les enzymes) qui jouent un rôle clé dans les réseaux trophiques. Les omiques sont aussi mobilisées pour comprendre la complexité des interactions entre organismes (de la symbiose à la pathogénie), révélant l'interdépendance des organismes qui se côtoient dans les écosystèmes : un exemple est le lien qui unit un microbiote et son hôte végétal ou animal. Cette interdépendance des organismes au sein des écosystèmes est à l'origine de la définition des services écosystémiques (par exemples la pollinisation ou le recyclage de la matière organique) qui sont essentiels au fonctionnement des écosystèmes et dont bénéficient les sociétés humaines.

Évolution du vivant : la vie au fil des générations. Ce champ de recherche propose une lecture temporelle du vivant en étudiant l'ADN des organismes contemporains mais aussi l'ADN ancien des organismes morts. En combinant différentes approches omiques, ce champ de recherche explore aussi l'hérédité sans modification de la séquence ADN : l'épigénétique. Même si la massification des données ADN peut être source d'erreurs, le séquençage ADN à haut débit a considérablement enrichi les bases de données des séquences qui permettent, grâce à des logiciels appropriés, des reconstructions solides des liens de parenté entre organismes. La comparaison de l'ADN des génomes des organismes actuels permet aussi la reconstruction des génomes de leurs ancêtres disparus : l'obtention de ces génomes ancestraux permet de prédire la physiologie de ces organismes disparus, voire les caractéristiques de



Crépis de Nîmes, qui pousse en abondance en ville. Elle produit deux types de graines : des plumeuses qui peuvent être dispersées par le vent et des plus lourdes qui tombent à son pied. Chaque année, les crépis des villes produisent de plus en plus de graines lourdes, plus aptes à germer dans un milieu urbain : c'est un exemple d'évolution rapide d'une espèce végétale. Photographie G. Przetak/ CEFE/CNRS Photothèque.

l'environnement dans lequel ils évoluaient. L'ensemble de ces recherches soutiennent de nouveaux *scenarii* quant à l'origine et aux conditions d'émergence du vivant. Par ailleurs, la possibilité de séquencer et de comparer l'ADN de dizaines ou centaines d'organismes révèle les variations de la séquence, que l'ADN soit transmis verticalement (de l'ancêtre vers la descendance) ou horizontalement (entre individus sans lien direct de parenté). L'analyse de ces variations révèle comment les espèces s'adaptent aux contraintes environnementales et anthropiques, et comment émergent de nouvelles espèces. La révolution du séquençage ADN à haut débit contribue à l'explosion des découvertes en paléogénomique, qui étudie l'ADN ancien. L'amélioration des techniques d'extraction, de séquençage et d'analyse permettent de d'obtenir la séquence ADN d'organismes morts il y a des milliers voire des centaines de milliers d'années. Ces nouvelles approches enrichissent considérablement les connaissances des écosystèmes passés, leur biodiversité, l'histoire évolutive des espèces, y compris *Homo sapiens*. Néanmoins, la paléobiologie ne se résume pas à l'étude de l'ADN ancien, d'autres vestiges du vivant sont analysés comme les os par exemple.

Observer, expérimenter, modéliser. Les recherches sur l'ADN s'articulent autour de trois activités : observer les écosystèmes pour décrire leur biodiversité et leur fonctionnement ; construire et réaliser des expérimentations pour tester les propriétés du vivant ; proposer des modèles conceptuels ou mathématiques pour représenter, simuler ou prédire l'évolution et le fonctionnement du vivant. Il apparaît immédiatement que ces recherches ne se limitent pas à la manipulation et l'analyse de l'ADN. Outre la biologie et l'écologie, elles mobilisent d'autres disciplines comme la chimie, la physique, les mathématiques, l'informatique et les sciences humaines. La révolution du séquençage ADN à haut débit, c'est, bien sûr, une baisse des coûts qui en facilite l'accès. C'est aussi une augmentation de plusieurs



Pots contenant chacun un écosystème différent, dans simulateur environnemental (écotron) dédié à l'expérimentation en écologie. Photographie C. Frésillon / CEREEP / CNRS Photothèque.

ordres de grandeur du volume de données mises à la disposition de tous (pas uniquement des chercheurs). C'est encore une amélioration de la qualité des données avec l'explosion des génomes et codes-barres ADN disponibles, et c'est surtout l'accès aux séquences ADN d'organismes de plus en plus différents tant du point de vue de leur taxinomie, de leur histoire évolutive, de leur mode de vie et de leur écosystème. L'étude de l'ADN n'est plus l'apanage de quelques organismes choisis pour leur intérêt économique (santé humaine, productions animales et végétales) et/ou méthodologique (facilité d'étude) comme la souris *Mus musculus*, la plante *Arabidopsis thaliana*, la bactérie *Escherichia coli*. Grâce au séquençage ADN à haut débit, le savoir s'ouvre à toute la diversité du vivant.

Pourquoi 101 secrets de l'ADN ?

Nous avons souhaité aborder dans cet ouvrage un ensemble de 101 notions biologiques qui représentent les principaux champs d'investigation actuellement en cours dans les laboratoires de recherche en France, en collaborations avec les partenaires étrangers.

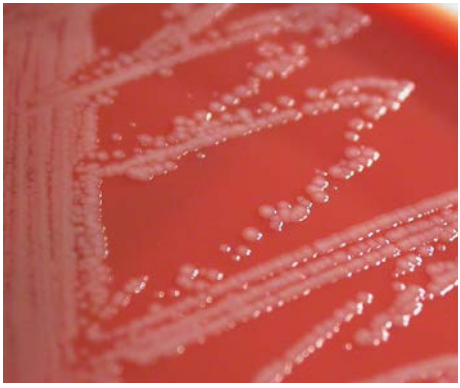
Les différentes escales. Cet ensemble est regroupé en huit chapitres distincts. Chacun d'eux aborde des connaissances fondamentales en biologie, notamment celles ayant trait à l'information génétique, à son organisation et à son étude. Des recherches parmi les plus récentes, comme celles qui régissent les interactions entre les organismes et leur environnement (biotique et abiotique) aux différentes échelles spatiales et temporelles sont aussi décrites. Sont également disséminées au cours des pages des considérations méthodologiques sur le développement des nouvelles technologies de séquençage haut débit de l'ADN et de ses dérivés, ou de biologie moléculaire, qui permettent de manipuler de façon ciblée des gènes d'intérêt. Sont également présentés des travaux au front des recherches les plus débattues à l'heure actuelle dans la communauté scientifique, comme celle ayant trait à la notion d'holobionte ou de système complexe.

Chaque article se lit de façon indépendante mais peut parfois faire écho à d'autres qui détaillent de façon plus approfondie certaines notions, ou qui les illustrent dans un contexte différent. Ainsi par exemple les notions de transferts de gènes abordées dans l'article 14 pour comprendre l'évolution des espèces à la racine de l'arbre du vivant sont également abordées

dans l'article 56 au sein de réseau d'interactions interspécifiques entre un hôte et son parasite et également dans l'article 92 entre un champignon et son fromage. Cet exemple illustre la diversité des échelles temporelles et biologiques qui sont abordées par l'étude d'un mécanisme de transfert d'information génétique tant vertical par des relations de parenté, qu'horizontal par des échanges de gènes entre espèces qui interagissent au sein d'un même environnement.

Il n'en demeure pas moins qu'au sein de chaque chapitre, la succession des articles répond à une logique d'acquisition graduelle de connaissances et également d'appréhension de la complexité des systèmes étudiés, que ce soit au sein même d'un organisme, comme par exemple pour la sépiole et ses symbiontes luminescents de l'article 54 ou de la boîte à outils génétiques utilisée dans le mimétisme chez les papillons de l'article 60.

Le chapitre 1 a ainsi pour ambition de montrer comment l'information génétique, par l'étude de l'ADN et de ses dérivés, permet de compléter notre connaissance, de façon quasi exhaustive, de la biodiversité actuelle et passée, visible ou invisible, mais également de reconstituer l'histoire évolutive des espèces, tant sur des échelles temporelles millénaires, depuis l'apparition de la vie jusqu'à nos jours, que sur les relations de parenté entre groupes taxinomiques plus ou moins éloignés. La puissance des technologies de séquençage en quantité et qualité permet en effet aujourd'hui d'analyser tous les niveaux de complexité et de tester des hypothèses restées jusque-là au stade de la réflexion par manque de preuve expérimentale. Ainsi nous y découvrirons que l'étude des archées, ces micro-organismes unicellulaires procaryotes formant une des trois grandes branches du vivant largement méconnue jusqu'à présent, permet de revisiter l'origine des eucaryotes.



Bactéries (ici *Acinetobacter baumannii*) en culture. Cette bactérie a acquis des gènes de résistance aux antibiotiques grâce au transfert horizontal. Photographie P.-E.Fournier / CNRS Photothèque.

Le chapitre 2 aborde la question de la biodiversité, de sa quantification, de son explora-

tion et de sa dynamique. Avec la génomique, la notion d'espèce, classiquement définie comme étant l'ensemble des individus inter-fertiles au sein d'une population ou d'un groupe reconstitué, s'affranchit d'observations dans la nature ou de tests expérimentaux au laboratoire. Dorénavant ce sont les relations de proximité phylogénétique, établis à partir de la reconstitution des relations de parenté de l'ensemble des séquences génomiques analysées, qui sont largement utilisées pour définir des entités qui partagent le même ancêtre commun. Plus la portion des génomes analysés est grande et plus le degré de confiance dans les entités ainsi distinguées sera important. La conséquence directe de cette approche est de permettre l'accès aux organismes non cultivables ainsi qu'aux organismes jusque-là inconnus, en les replaçant dans l'arbre du vivant. Ces méthodes permettent également une approche globale, par exemple en analysant l'ensemble de l'ADN d'un sol ou d'un océan, sans aucune connaissance *a priori*. Cette quantification de la biodiversité a évidemment des conséquences importantes en terme de protection et de conservation, et ce d'autant plus lorsque des spécificités fonctionnelles intéressantes sont identifiées.

À une échelle biologique limitée à l'individu ou à la population, le chapitre 3 traite de l'évolution des génomes, tant dans la reconstitution de la genèse des espèces et de leur délimitation qu'au niveau des réarrangements génomiques ou des modifications de séquences qui en sont à l'origine. On y aborde la question des éléments génétiques, mobiles ou non, qui viennent grossir la taille de certains génomes, notamment pour les espèces cavernicoles, ou des éléments répétés, parfois à l'origine d'instabilité génomique, voire d'incompatibilité et qui conduisent à la séparation d'espèces. On y apprend ainsi comment l'environnement façonne le génome et permet aux organismes de s'adapter.

Le chapitre 4 développe la question de l'adaptation, particulièrement d'actualité dans le contexte des changements climatiques et des pressions anthropiques. Certains articles sont dédiés aux environnements extrêmes, particulièrement sensibles à des changements mineurs, comme en antarctique ou dans les profondeurs des océans et des abysses, d'autres ciblent des milieux moins contraints mais tout aussi redoutables, comme ces fruits toxiques sur lesquels seules certaines espèces de drosophiles viennent s'alimenter et se reproduire. Enfin, le milieu urbain est aussi un vecteur d'adaptation pour la plupart des organismes, qu'ils soient domestiques ou issus de milieux moins anthropisés. Les questions de pollution, et de cycles biogéochimiques dans les systèmes aquifères fédèrent également tout un pan de recherche sur les plantes et leur potentiel de tolérance aux métaux, voire même de remédiation ainsi que sur les micro-organismes, et leur

potentiel pathogène. Enfin, quelques articles développent la question des mécanismes épigénétiques par lesquels les organismes sont capables de répondre rapidement aux changements environnementaux grâce aux capacités de variation d'expression des génomes.



Stromatolite fossile (roches feuilletées coniques ou en dômes formées par le dépôt calcaire d'un tapis de micro-organismes) datant de 3,45 milliards d'années, découvert dans la formation de North Pole, dans le Pilbara, en Australie. Photographie P. Lopez Garcia / CNRS Photothèque.

Les interactions entre espèces sont centrales dans la compréhension de l'écosystème, comme cela est développé dans le chapitre 5. Aucune espèce n'est indépendante des autres, mais certaines ont des relations plus qu'intimes, avec des interactions parfois obligatoires entre une espèce et son hôte, comme pour la sépiole ou les micro-organismes de la rhizosphère. D'autres enfin ont besoin d'intermédiaires comme ces guêpes qui parasitent des papillons en utilisant les capacités de transfert de gènes de certains virus. Les études fonctionnelles de transcriptomique ou de méta-transcriptomique permettent également d'accéder à l'ensemble des fonctions de l'holobionte et de déterminer qui fait quoi de façon très précise.

Le chapitre 6 élargit encore le champ des investigations, en posant les questions de dynamique et d'évolution des écosystèmes, ainsi que des services rendus pour le bien-être des humains et leur soutenabilité. Des approches méthodologiques d'analyses globales sont présentées, y compris aux échelles rétrospectives qui permettent un ancrage dans le temps. L'écologie des systèmes, grâce à l'expertise des collègues mathématiciens, permet d'appréhender les réseaux d'interactions fonctionnelles à tous les niveaux d'information biologique (du gène au cycle biogéochimique) dans une approche de modélisation. C'est évidemment un pas essentiel pour développer des approches prédictives de la réponse des écosystèmes aux changements environnementaux.

L'avant dernier chapitre est consacré à la paléogénomique et à la reconstitution d'environnements anciens ou d'histoires évolutives. L'étude des sédiments, qu'ils soient marins ou côtiers, est à ce titre particulièrement prometteuse, puisqu'elle permet de remonter dans le temps au fur et à mesure

que l'on descend en profondeur. Ainsi, le lecteur est transporté aux origines de la vie 1 900 mètres sous le plancher océanique et au développement de l'élevage en Europe et des activités agropastorales sur les trois derniers millénaires. L'étude de spécimens anciens, fossiles ou conservés dans le permafrost ainsi que ceux préservés dans les musées permettent de reconstituer les histoires évolutives des espèces, comme pour le bison, le chien ou les mollusques. L'homme aussi est scruté à la loupe, tant pour reconstituer son histoire biogéographique durant le dernier maximum glaciaire, que pour étudier l'évolution de son microbiome, notamment d'un point de vue pathogénique. L'intégration de ces échelles temporelles est très importante pour mieux comprendre le présent et anticiper le futur.

Finalement l'ouvrage s'achève par un ensemble de travaux dans lesquels les humains sont au centre des pressions environnementales qu'ils font subir aux écosystèmes et qu'ils subissent en retour. Il sera question de microbiome, avec toutes les retombées et développement possibles en termes de médecine personnalisée, et d'antibiorésistances dont le développement gagne quasiment tous les milieux. Seront également évoqués l'évolution des génomes en réponse aux pratiques humaines, concernant notamment les nanoparticules, les toxines environnementales ou encore les pollutions minières, et celles à grande échelle qui font du 7e continent un enjeu majeur pour le traitement des plastiques.

Embarquement immédiat !

101 escales, c'est à la fois peu et beaucoup.

C'est peu, car la révolution des technologies de séquençage ADN à haut débit ouvre des pans entiers de recherche jusque-là inexplorée pour étudier le vivant, son fonctionnement, son évolution. Il a fallu faire des choix, en équilibrant les questions scientifiques posées, les modèles biologiques, les écosystèmes et les approches méthodologiques. Ces 101 articles n'ont pas vocation d'exhaustivité, mais sont autant de portes ouvertes vers d'autres en génomique environnementale.

C'est beaucoup, car le vivant nécessite du temps pour en expliquer la diversité et la complexité. Tous les exemples ont été choisis parmi une littérature scientifique publiée depuis moins de 5 ans. Pour chaque texte, un article scientifique original en anglais est proposée au lecteur, à la fois pour qu'il puisse comprendre comment sont publiés les travaux scientifiques

actuels, mais aussi qu'il puisse poursuivre sa lecture de manière approfondie, voire découvrir d'autres articles scientifiques par une recherche bibliographique par sujet ou par auteur.

Au terme de cette aventure collective de plusieurs mois, riche d'échanges et d'interactions avec les chercheurs de tous horizons, nous espérons susciter aux lecteurs autant de plaisirs, de surprises et de réflexions entremêlés que nous en avons ressentis lors de l'élaboration de cet ouvrage. Nous espérons qu'il éveillera la curiosité de chacun, et peut-être quelques vocations.



1 Exploration haut débit du vivant

1 Explorer et représenter la totalité de l'arbre du vivant

L'arbre de la vie représente l'ensemble des liens de parenté, ou liens évolutifs, qui unissent les espèces de notre planète. Sa construction et sa représentation constituent un défi scientifique, soutenu par l'essor des nouvelles techniques de séquençage de l'ADN à haut débit.

Depuis plus de 3,5 milliards d'années, l'évolution des espèces a conduit à la biodiversité actuelle. Les liens évolutifs qui existent entre toutes ces espèces forment un arbre, nommé arbre de la vie ou arbre du vivant. Chaque feuille représente une espèce, et chaque intersection (appelée nœud) représente un ancêtre (connu ou inconnu), à la manière des arbres généalogiques. La reconstitution de cet arbre est un enjeu pour comprendre les relations entre espèces, ainsi que leurs liens avec les espèces du passé, autrement dit leur histoire évolutive.

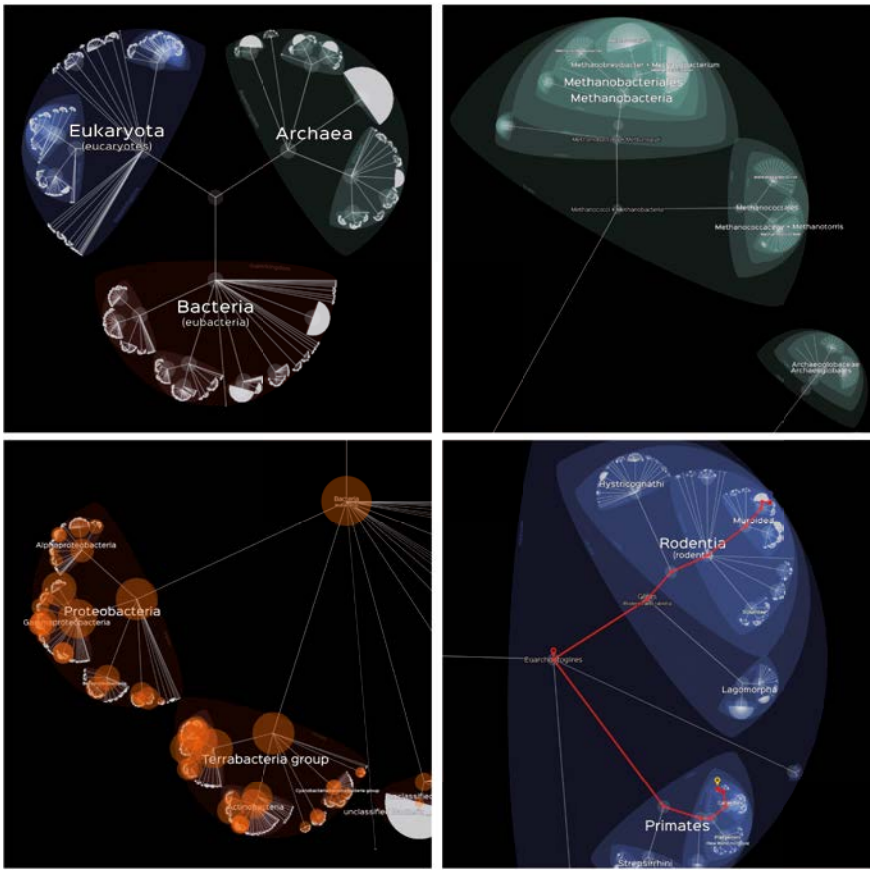
La connaissance de l'arbre de la vie est encore très partielle. À ce jour, le nombre d'espèces connues est faible : le plus grand arbre construit contient 2,3 millions d'espèces, alors que le nombre d'espèces vivant sur Terre est estimé à quelques dizaines de millions, et si l'on ajoute les espèces qui ont existé dans le passé et sont aujourd'hui éteintes, l'on peut atteindre quelques dizaines de milliards.

Par ailleurs, les espèces positionnées actuellement dans l'arbre de la vie ne sont pas représentatives de la diversité réelle. L'arbre est riche en eucaryotes (animaux, champignons, plantes) et en certaines bactéries et archées, qui ont été facilement échantillonnées et/ou sont d'intérêt médical ou agro-industriel. Mais les espèces non cultivées à ce jour et celles vivant dans des environnements extrêmes ou peu accessibles sont peu ou pas représentées.

Enfin, l'arbre de la vie est difficile à reconstruire. Plus un nœud de l'arbre est profond (ancien), plus il est difficile d'en déterminer la position à partir du signal évolutif détectable dans les séquences d'ADN des espèces actuelles. De plus, des transferts d'ADN entre espèces distinctes, appelés transferts horizontaux par opposition aux transferts verticaux entre parents et descendants d'une même espèce, se sont produits et continuent à se produire. Ces échanges d'ADN horizontaux peuvent brouiller le signal évolutif (c'est-à-dire la relation de parenté entre espèces) et donc la topologie de l'arbre

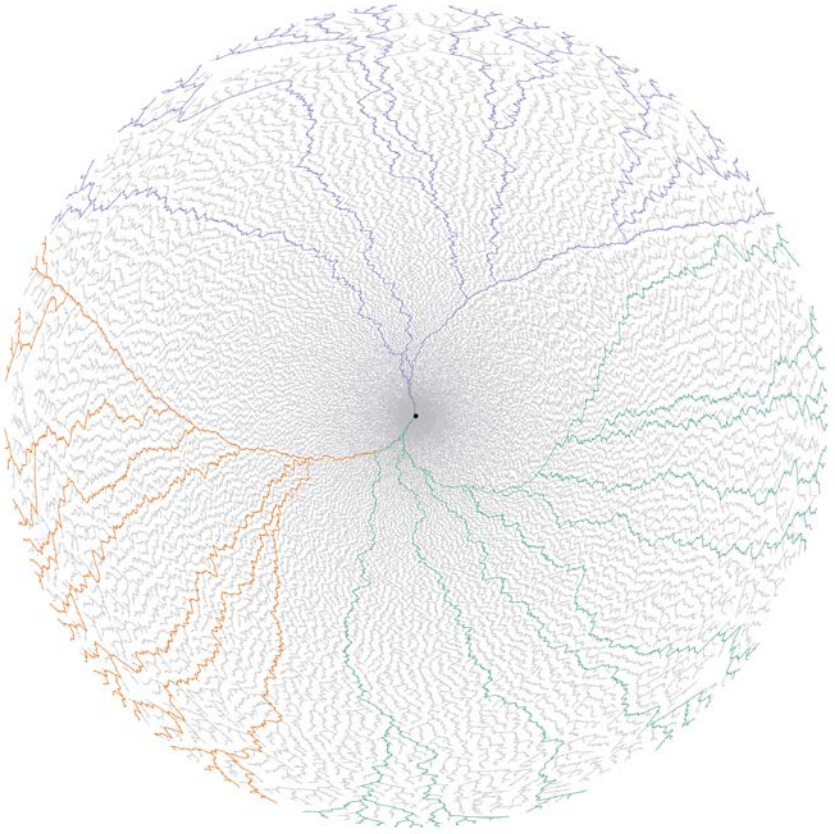
obtenu. Mais on commence à réaliser qu'ils peuvent aussi aider à dater les nœuds de l'arbre de la vie.

Au-delà de sa construction, la représentation de l'arbre est aussi un défi méthodologique en raison de la quantité considérable d'information à y intégrer. La nature hiérarchique de l'organisation des espèces a permis de proposer une solution inspirée des outils de cartographie. De même qu'une rue fait partie d'une ville, qui fait partie d'un département, d'une région, d'un pays, d'un continent, une espèce fait partie d'un genre, d'une famille, d'un ordre, d'une classe, d'un règne et enfin d'un des 3 domaines du vivant



Quelques vues de l'outil *Lifemap*, l'explorateur interactif de l'arbre de la vie (accessible à http://lifemap.univ-lyon1.fr). Image D.M. de Vienne, 2016-2017. Licence CC-BY-NC.

(Eucaryotes, Bactéries, Archées). En s'appuyant sur des outils développés pour la cartographie, l'outil internet *Lifemap* propose une nouvelle façon de représenter les données hiérarchiques. *Lifemap* permet de se déplacer dans l'arbre de la vie comme sur une carte avec *Google Maps* et ainsi d'explorer, de façon interactive et intuitive, l'intégralité de l'arbre de la vie tel qu'il est connu aujourd'hui, facilitant sa compréhension, son utilisation et sa diffusion.



Simulation d'un arbre de la vie représentant les lignages menant à la diversité actuelle (traits colorés) et les lignages menant à des espèces aujourd'hui éteintes (traits gris). Le centre du cercle représente l'apparition de la vie, la périphérie représente le temps présent, et les trois couleurs représentent les trois domaines du vivant. Image D.M. de Vienne et W. Duchemin, 2017. Licence CC-BY-NC.

2 À la recherche de LUCA, notre ancêtre commun

Il y a plus de 20 ans, l'acronyme LUCA (*Last Universal Common Ancestor*) a été proposé pour désigner l'ancêtre commun à tous les êtres vivants actuels. De nombreux chercheurs ont essayé de dresser un portrait de cet organisme ancestral et des conditions dans lesquelles il avait pu apparaitre.

Au XXe siècle, la biologie moléculaire a permis de démontrer l'unicité du monde vivant. En effet, tous les organismes actuels partagent le même support pour l'information génétique ainsi que les mêmes machineries (ARN et protéines ribosomaux) permettant l'expression des gènes. Cette unicité implique une origine commune, c'est-à-dire l'existence d'un ancêtre possédant déjà toutes ces caractéristiques. Au cours d'un colloque organisé en 1996 par la Fondation des Treilles, cet ancêtre commun a été nommé LUCA, en hommage à Lucy, l'Australopithèque trouvée en Éthiopie en 1974.

Si tous les organismes actuels dérivent d'un même ancêtre, les travaux de Carl Woese dans les années 1970 ont permis de montrer qu'il existe trois domaines dans le monde vivant : les bactéries, les archées et les eucaryotes. Une hypothèse largement acceptée est que LUCA a donné naissance à ces trois domaines *via* deux lignées distinctes, d'un côté les bactéries et de l'autre les archées et eucaryotes.

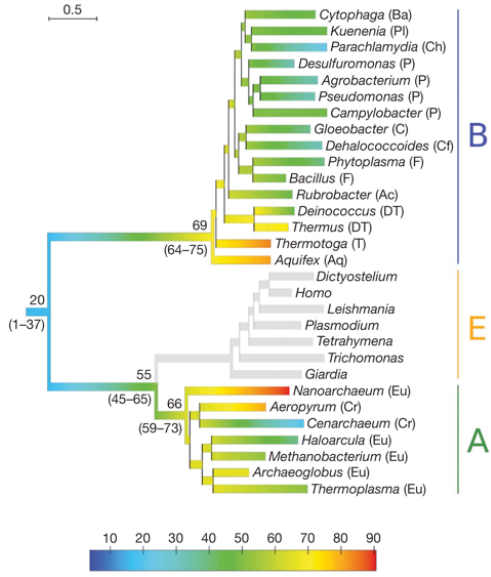
En 1996, au moment où l'acronyme LUCA a été introduit, seul un tout petit nombre de génomes complets étaient disponibles : les bactéries *Haemophilus influenzae* et *Mycoplasma genitalium*, l'archée *Methanococcus jannaschii* et l'eucaryote *Saccharomyces cerevisiae*. Aujourd'hui, ce sont plus de 100 000 génomes (complets ou partiels) qui sont disponibles dans les collections publiques de séquences. Cet accroissement considérable de la quantité de données a permis le développement des approches dites de phylogénomique (contraction de phylogénie et de génomique), qui ont permis d'améliorer notre compréhension de la nature de LUCA et de l'environnement dans lequel il vivait.

La phylogénomique a établi qu'il existe trois ARN ainsi que 34 protéines du ribosome communes à l'ensemble des êtres vivants et donc à LUCA. Si l'on prend en compte le fait que les êtres vivants actuels contiennent environ deux fois plus de protéines ribosomiques, cela signifie que la machinerie de synthèse des gènes présente dans LUCA était beaucoup moins complexe. LUCA était également probablement dépourvu des systèmes complexes

de production d'énergie tels qu'ils sont observés dans les organismes modernes.

Les arbres du vivant publiés à la fin des années 1990 laissaient penser que LUCA vivait à haute température (plus de 80 °C), des résultats récents sont venus invalider cette hypothèse. À l'heure actuelle on estime plutôt que LUCA vivait à une température proche de 20 °C. En revanche, il se serait produit une adaptation progressive aux hautes températures indépendamment dans les deux lignées en réponse aux changements dans le climat de la Terre durant l'éon Archéen (de -3,8 à -2,5 milliards d'années).

Par la suite, dans les différentes branches de l'arbre de la vie, des adaptations secondaires aux conditions de températures existantes dans les différentes niches écologiques de la planète se sont produites, conduisant à la diversité qui est observée aujourd'hui.



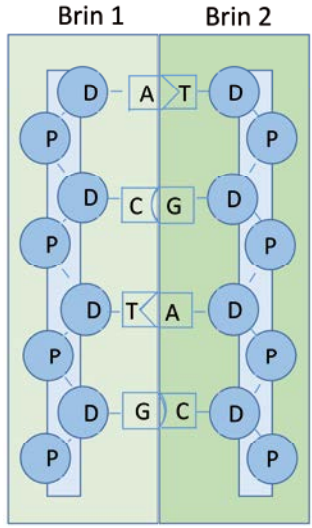
Arbre du vivant intégrant les trois domaines (A : archées, B : bactéries, E : eucaryotes) dans lequel les températures optimales de croissance des organismes ont été indiquées par un code couleur sur les branches. LUCA se situe à la racine de cet arbre et sa température de croissance se situait donc aux alentours de 20 °C. Image B. Boussau / Laboratoire Biométrie et Biologie évolutive, UMR CNRS 5558.

3 L'ADN, support universel de l'information génétique

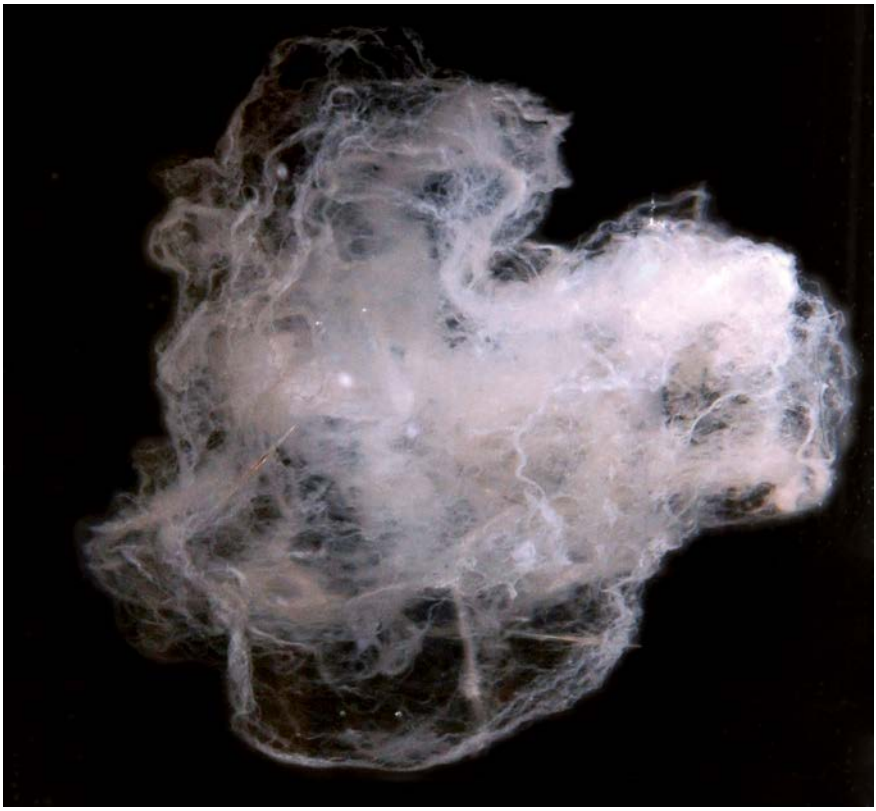
Tous les êtres vivants sur Terre partagent la façon dont l'information génétique est codée dans les cellules. Il y a bien quelques variantes, le vivant étant source inépuisable de diversité, mais le codage de cette information sous la forme d'un alphabet de quatre lettres, l'ADN, est universel.

L'ADN ou acide désoxyribonucléique est un polymère naturel composé de trois types de molécules : un phosphate sur lequel vient s'ancrer un sucre (le désoxyribose) qui porte lui-même une base azotée parmi les 4 possibles que sont l'adénine (A), la thymine (T), la guanine (G) et la cytosine (C).

Les molécules de phosphate et de sucre se lient l'une à l'autre telles les perles alternées d'un collier. En revanche, la base azotée (A, T, G ou C) liée aux sucres peut varier d'un sucre à l'autre, créant une succession (on dit une « séquence ») de bases azotées, par exemple ATTACCGAGT. Une séquence ADN se caractérise donc par deux paramètres, sa longueur, dont l'unité de



La molécule d'ADN est composée de deux brins complémentaires. Image S. Salamitou.



En présence d'alcool et de sels minéraux, de grandes quantités d'ADN précipitent sous la forme de filaments blanchâtres qui sont visibles à l'œil nu. Photographie F. Viala / CNRS Photothèque.

mesure est le nombre de bases « b » (égal à 10 dans l'exemple ci-dessus) et l'ordre de celle-ci (la séquence est énoncée par la lecture lettre à lettre, de gauche à droite).

Dans les cellules, l'ADN est constituée de 2 polymères, soit 2 brins d'ADN qui s'associent par l'interaction moléculaire des bases azotées. Les liaisons entre bases assurent le maintien de la structure en double brin ADN, qui s'enroule sous la forme de double hélice. Lors de cette interaction, A et T s'associent toujours, de même que C et G. Chaque brin correspond à la séquence complémentaire et inversée de l'autre. Ainsi est-il possible de séquencer l'ADN par n'importe lequel de ces deux brins, ce qui permet de connaître l'autre. Pour un ADN double brin, la longueur de la séquence