

## LA BIOLOGIE MOLÉCULAIRE AU SERVICE DU DIAGNOSTIC POUR LA SANTÉ DES FORÊTS

Marie-Laure Desprez-Loustau (INRA-Univ Bordeaux UMR BIOGECO) et Renaud Ioos (ANSES, LSV-Mycologie)

La biologie moléculaire est devenue indispensable pour le diagnostic des maladies des arbres forestiers : ADN, PCR, SNP, ITS, BLAST, autant de sigles couramment utilisés par les chercheurs dans leur jargon, devenu incompréhensible pour les hommes et femmes de terrain. Cet article se propose de fournir les principales connaissances de base pour comprendre à quoi sert la biologie moléculaire pour le diagnostic des maladies forestières et "comment ça marche", en illustrant avec quelques exemples d'application.

### Pourquoi avoir recours à la biologie moléculaire?

**Identifier l'agent responsable d'une maladie** est une étape cruciale avant d'envisager de prendre des mesures de prévention ou de lutte. C'est la première étape qui va éventuellement permettre de savoir si l'agent pathogène est indigène ou pas (d'où la décision éventuelle de mesures de quarantaine), de rechercher des connaissances dans la littérature, etc...

Dans le cas des maladies des arbres forestiers, une grande partie des agents pathogènes sont des champignons microscopiques. La reconnaissance et l'identification des champignons consiste à leur attribuer un nom de genre et un nom d'espèce (ex. *Ophiostoma ulmi* correspond à l'espèce « *ulmi* » du genre « *Ophiostoma* ») en fonction de critères de classification. L'identification des champignons a longtemps fait appel à des caractères morphologiques portant sur les organes de reproduction sexuée ou asexuée, tels que l'aspect (forme, taille, couleur, etc.) ou les modes de production et d'éjection des spores. Cette reconnaissance ne pouvait se faire que par quelques experts très spécialisés. La mycologie "classique", basée sur cette description morphologique des espèces, a permis de décrire environ 100 000 espèces de champignons.

L'avènement des techniques de biologie moléculaire, **en permettant d'accéder au génome\* des espèces a permis d'avoir une vue beaucoup plus fine de la diversité des organismes**, et en particulier des champignons. On estime maintenant que le nombre total d'espèces de champignons est compris entre un et cinq millions, nous ne connaissons donc pour l'instant qu'une infime partie de la diversité totale. De plus, comme nous le verrons après, les nouvelles techniques de biologie moléculaire remettent parfois en cause fortement les classifications anciennes : une "ancienne" espèce (= espèce morphologique) peut s'avérer en fait correspondre à plusieurs "vraies" espèces (phylogénétiques\*). La distinction de ces espèces au niveau du génome permet ultérieurement de réaliser que ces espèces peuvent avoir des aires de distribution ou une biologie (gamme d'hôtes par exemple) différentes. A l'inverse, certaines espèces décrites sous des noms différents, par exemple dans des régions géographiques ou associées à des arbres hôtes différents, peuvent en fait correspondre à la même espèce.

Aujourd'hui, tous les laboratoires de mycologie, en particulier le laboratoire de l'ANSES à Nancy, en charge de l'identification des champignons pathogènes forestiers à partir des échantillons soumis par le DSF, ou les laboratoires de pathologie forestière de l'INRA recourent maintenant largement à l'utilisation de techniques de biologie moléculaire pour l'identification des champignons pathogènes.

## Comment ça marche?

**Qui dit biologie moléculaire dit ADN\***. L'ADN est la molécule, présente dans toutes les cellules vivantes, qui porte l'information génétique nécessaire au développement et fonctionnement de l'organisme. La structure particulière de la molécule d'ADN est à l'origine de ses propriétés et des techniques d'analyses qui en découlent. La molécule d'ADN est composée de deux brins complémentaires enroulés en double-hélice.

Chaque brin d'ADN est constitué d'un enchaînement de nucléotides, avec **4 types de nucléotides possibles** : **A, G, C et T**, pour Adénine, Guanine, Cytosine, Thymines. Ces nucléotides sont complémentaires et s'associent par paires, assurant l'appariement des 2 brins de l'ADN : A avec T ; C avec G (Figure 1).

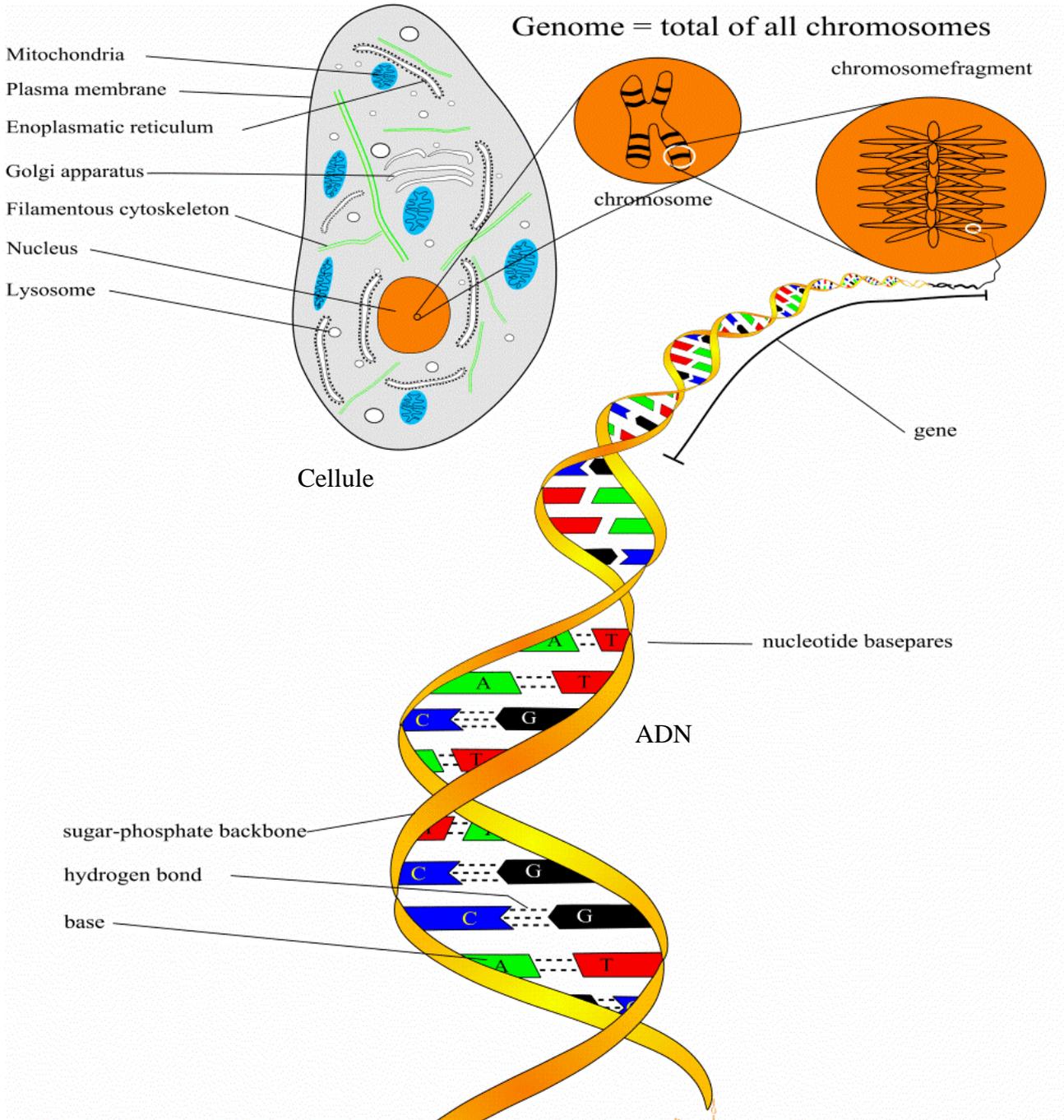


Figure 1 : Localisation de l'ADN dans la cellule et structure de la molécule en double-hélice résultant de l'enchaînement sur chaque brin de 4 nucléotides, appariés par paires entre les 2 brins (figure de Andy Vierstraete, 1999 : <http://users.ugent.be/~avierstr/>)

Le génome\* d'un organisme contient ainsi plusieurs millions de paires de nucléotides : 40 à 200 pour la plupart des champignons pathogènes d'arbres, plus de 3000 pour le génome humain. **L'enchaînement des différents nucléotides** le long d'un brin, par exemple ATGGCT...., **constitue une séquence, qui est caractéristique de chaque individu.** Le long de la séquence, des suites particulières (triplets) de nucléotides servent à coder les acides aminés qui constitueront les protéines après transcription\* et traduction de l'ADN. D'autres triplets correspondent au début et fin de transcription. De façon imagée, on peut dire que **l'information génétique de tous les êtres vivants est écrite avec un alphabet de 4 lettres** (ATCG), avec dans chaque cellule une version du livre complet (l'enchaînement de tous les nucléotides), avec sa copie (les deux brins d'ADN complémentaires).

Pour une même zone du génome, par exemple un gène\* donné, une différence de séquence entre individus sera qualifiée de polymorphisme. Un SNP\* est une variation pour un seul nucléotide, généralement produite par une mutation\*. En dehors des jumeaux (et des clones !), deux individus auront donc un livre un peu différent.

L'utilisation des données moléculaires pour **l'identification des espèces repose sur le principe d'utiliser des zones du génome qui soient à la fois identiques pour tous les individus d'une même espèce et différentes (= porteuses de polymorphismes) entre individus appartenant à des espèces différentes.** Le gène ou la portion de génome qui correspond le mieux à ces critères est appelé 'code barre génétique' ou **en anglais "barcode"**. **Le barcode des champignons**, couramment utilisé maintenant en phylogénie et identification des espèces, **est l'ITS\*** (Figure 2).

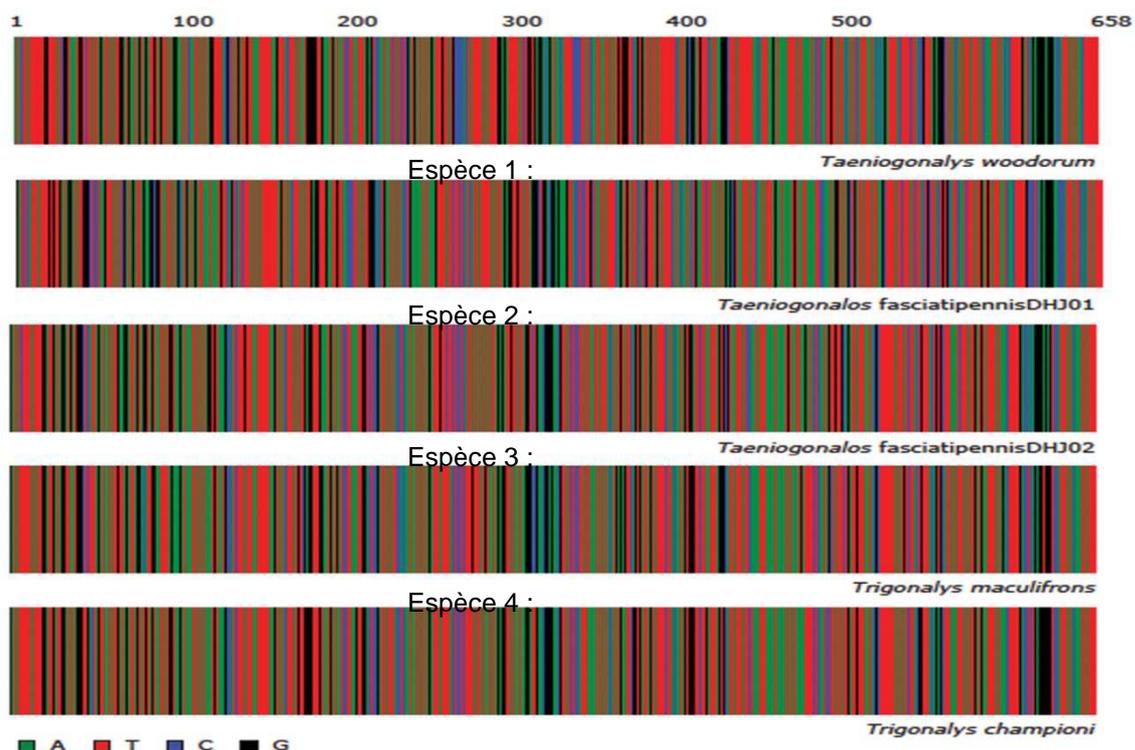


Figure 2 : Illustration du "barcode" des champignons.

La séquence de l'ITS, constituée d'environ 650 nucléotides, présente des polymorphismes entre espèces. La séquence particulière de chaque espèce, résultant de l'enchaînement des nucléotides, peut être matérialisée par un ensemble de bandes colorées (chaque couleur correspondant à une base différente: A, T, C, G) donnant une signature unique (code-barre) pour chaque espèce (illustration extraite de Hebert et al 2003- Proceedings of the Royal Society - B Biological sciences: doi : 10.1098/rspb.2002.2218)



Figure 2B : Exemple d'alignements de séquences ITS de diverses espèces de champignons (Source : Nilsson et al 2012 : MycoKeys 4: 37-63 : doi: 10.3897/mycokeys.4.3606cart)

Il s'agit d'une partie du génome impliquée dans le codage des ARN constitutifs des ribosomes. Les ribosomes, présents en grand nombre dans les cellules, sont les organites cellulaires qui assurent la synthèse des protéines. Ainsi, les gènes codant pour les ARN ribosomiques sont présents en de nombreuses copies dans le génome (environ 100 chez les champignons). Une autre caractéristique de ces gènes ribosomiques est que seule une partie de l'ARN issu de la transcription va produire les molécules constitutives des ribosomes (zones dites "codantes"), d'autres zones étant non fonctionnelles (non codantes = ITS\*) et éliminées lors du processus de maturation de l'ARN (Figure 3).

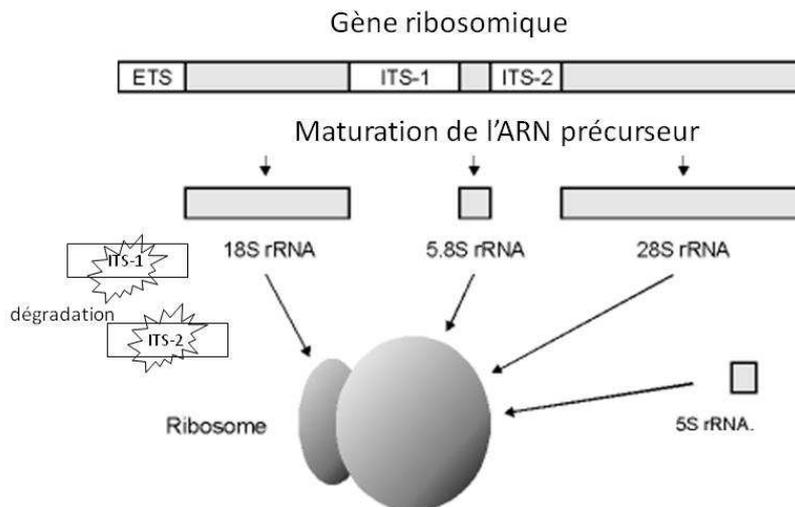


Figure 3 : Structure de l'ADN ribosomique. Une seule unité est représentée: en gris, zones codantes, en blanc, zones non codantes, en particulier l'ITS, composé de 2 parties

Comme toutes les zones du génome, l'ADN ribosomique peut subir des mutations. Toutefois, les conséquences vont être très différentes dans les zones codantes et non codantes : pour les premières, une mutation aura beaucoup de chances d'altérer la fonction de la molécule et donc, compte tenu de l'importance des ribosomes pour le fonctionnement cellulaire, d'être éliminée par sélection naturelle; par contre pour les zones non codantes, les mutations peuvent s'accumuler sans problème. C'est en effet ce que l'on observe : **les zones codantes sont extrêmement "conservées" entre organismes alors que les zones non codantes sont beaucoup plus diverses.**

Dans le cas des champignons par exemple, la divergence entre deux espèces d'un même genre est généralement inférieure à 1 % pour les séquences des zones codantes mais de 3 à 10 % pour les zones non codantes. **Ces caractéristiques ont été mises à profit pour identifier les champignons à partir de l'ADN ribosomique, en utilisant la technique de PCR\*** que nous allons maintenant expliquer.

### Polymerase chain reaction - PCR

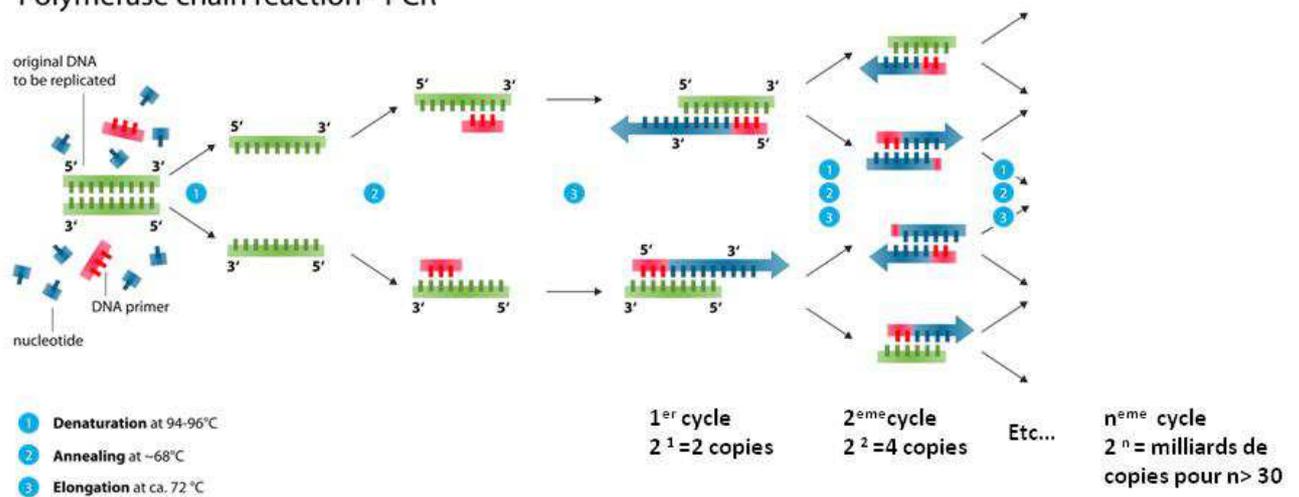


Figure 4 : Principe de la PCR = amplification exponentielle de l'ADN (Source = Wikipedia)

**La PCR est une sorte de "photocopieuse" à ADN**, c'est-à-dire qu'à partir d'un fragment d'ADN donné, on va pouvoir « l'amplifier », soit obtenir de nombreuses copies, ce qui est nécessaire pour les analyses ultérieures : on ne peut pas "voir" une seule molécule (par exemple déterminer sa séquence). La PCR est basée sur le principe d'une succession de réactions de réplifications\* au cours desquelles le fragment à amplifier va être "accroché" par des "amorces\*", les deux brins d'ADN étant séparés et chacun servant alors de matrice à la formation d'un fragment d'ADN complémentaire, l'amplification obtenue est ainsi exponentielle (Figure 4).



## Exemples d'application

Plusieurs exemples illustrent les bénéfices apportés par l'utilisation de l'ITS dans l'identification d'agents pathogènes forestiers.

**La maladie à Phytophthora des aulnes**, caractérisée par des chancres au collet aboutissant à la mort des arbres, s'est manifestée pour la première fois dans les années 1990 en Europe. L'agent causal a rapidement été identifié comme un *Phytophthora*, proche de *P. cambivora* mais avec des caractéristiques particulières. Grâce à des études morphologiques et génétiques, des chercheurs de la Forestry Commission, ont mis en évidence qu'il s'agissait d'un hybride. L'équipe INRA de Nancy a montré par la suite, grâce à une étude moléculaire de l'ADN nucléaire et mitochondrial, que les deux espèces parentes étaient *P. alni ssp. uniformis* et *P. alni ssp. multififormis*, inconnues auparavant et très peu virulentes sur aulne. Le scénario le plus probable de l'apparition de l'espèce virulente implique une mise en contact récente des deux espèces parentes, l'une, *P. alni ssp. multififormis*, d'origine européenne et l'autre, *P. alni ssp. uniformis*, d'Amérique du Nord Ouest (Alaska, Oregon). Les deux espèces n'ayant pas évolué ensemble n'ont pas de "barrière au croisement" et se seraient donc hybridées. Les nombreux échanges de matériel végétal dans et entre pays, en particulier le commerce de plants, sont très probablement impliqués sinon dans l'événement initial de mise en contact ayant produit les hybrides, du moins dans leur diffusion dans toute l'Europe et au delà.

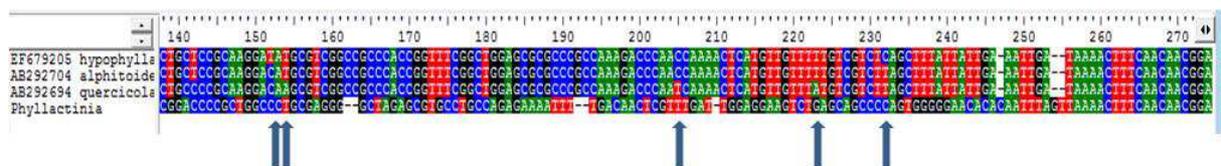


**Phytophthora alni sur aulne, Brasier & S.A. Kirk, forestryimages**

**L'oïdium du chêne** est une des maladies forestières les plus courantes en France actuellement, bien qu'elle ne soit apparue que relativement récemment, il y a un siècle. L'INRA (Bordeaux) a initié un travail de recherche sur cette maladie au début des années 2000. Dans le cadre de la thèse d'Amira Mougou, nous avons analysé un certain nombre d'échantillons de diverses localités françaises pour caractériser l'ITS de l'oïdium du chêne. A notre surprise, nous avons obtenu 4 séquences d'ITS différentes (Figure 6).



**Oïdium sur jeune chêne, DSF**



**Figure 6 : Alignement d'une partie de l'ITS des espèces associées à l'oïdium du chêne : les 3 espèces d'*Erysiphe* (*E. hypophylla*, *E. alphitoides* et *E. quercicola*) se distinguent très nettement de l'espèce de *Phyllactinia* (sauf dans la zone conservée, en bout de séquence); elles ne se distinguent entre elles que par quelques SNP.**

A la même époque Susumu Takamatsu, mycologue japonais spécialiste des oïdiums, a produit une phylogénie de ces espèces basées sur l'ITS, au niveau mondial. Nous avons ainsi pu mettre en évidence que la maladie appelée "oïdium du chêne" en France peut être causée par au moins 4 espèces différentes: *Phyllactinia guttata*, *Erysiphe alphitoides*, *E. hypophylla*, et *E. quercicola*. *E. quercicola* n'avait jamais été mentionné jusqu'alors en Europe. A l'exception sans doute de la première, les 3 autres espèces semblent avoir été introduites en Europe mais leur aire d'origine reste inconnue. L'identification moléculaire aboutit ainsi à mettre en évidence qu'une même maladie est causée par des espèces pathogènes différentes, mais à l'inverse que des oïdiums sur plantes différentes pourraient être causés par la même espèce, contredisant le "dogme" de la spécialisation parasitaire de ce type de maladie. Les analyses moléculaires suggèrent en effet très fortement que *E. alphitoides* et *E. quercicola* seraient également associées à l'oïdium du manguier et de l'hévéa, plantes très éloignées des chênes!

**La Chalarose des frênes** est un autre exemple très récent d'identification d'une nouvelle espèce liée à une maladie émergente grâce aux outils et méthodologies de biologie moléculaire. Des dépérissements de frênes ont été observés et rapportés depuis les années 1990 en Pologne. L'identification d'un agent pathogène associé aux symptômes a pris plusieurs années, la nouvelle espèce *Chalara fraxinea* étant décrite en 2006. Il ne s'agissait alors que de la forme dite imparfaite, asexuée, du champignon. La mise en relation avec la forme sexuée n'est réalisée que trois ans plus tard; *Chalara fraxinea* est alors décrit comme la forme asexuée du champignon *Hymenoscyphus albidus*. Ce champignon, déjà décrit en Europe, était jusqu'alors considéré comme un saprophyte banal dont les apothécies (fructifications sexuées se présentant comme des petites cupules) se développent sur les pétioles des feuilles mortes de frêne. Restait à expliquer le brusque changement de virulence du champignon. Des études approfondies portant sur l'ITS et quelques autres gènes ont alors montré que les isolats pathogènes (*Chalara fraxinea*) et saprophytes (*Hymenoscyphus albidus*) appartenaient à des lignées génétiques différentes: *C. fraxinea* ne pouvait donc pas être considéré comme la même espèce que *H. albidus*. La nouvelle espèce *Hymenoscyphus pseudoalbidus* a ainsi été décrite (récemment renommée *H. fraxineus*), la distinction avec *H. albidus* étant fondée non pas sur des caractéristiques morphologiques mais des caractéristiques moléculaires. Récemment, l'origine asiatique de *H. fraxineus* a été fortement suggérée par le fait qu'il est identique à une espèce indigène du Japon associée aux frênes locaux sans causer de dommages (à l'image de *H. albidus* en Europe).



**Chancres au collet dû à *Chalara fraxinea*, L.-M. Nageleisen**

Enfin, en diagnostic courant, l'analyse de code barre génétique est désormais fréquemment utilisée pour identifier des champignons qui ont pu être isolés des plantes symptomatiques et mis en culture au laboratoire, mais qui ne produisent pas suffisamment voire pas de structures caractéristiques. Malgré par exemple l'absence de spores, ou dans le cas d'un champignon à croissance extrêmement lente, il est possible d'identifier le microorganisme avec un très bon degré de certitude. Pour de nombreux genres de champignons, seules certaines espèces bien précises sont effectivement des agents parasites sur certaines plantes, d'autres n'ont aucun pouvoir pathogène. Identifier l'espèce avec précision permet donc en retour de savoir si cette dernière peut être à l'origine des symptômes observés sur la plante analysée, et d'aider à réaliser le diagnostic.

## Conclusion

Les analyses permises par le développement de la biologie moléculaire ont révolutionné la mycologie en apportant de nouveaux caractères diagnostiques permettant de différencier très simplement les espèces quand les différences morphologiques ne le permettaient pas ou n'étaient utilisables que par un très petit nombre d'experts. Cette reconnaissance des espèces basée sur la génétique permet de différencier des espèces très proches mais ayant évolué sur des continents différents par exemple, avec pour conséquence une biologie très différente notamment en terme de pouvoir pathogène. Le cas des *Phytophthora* associés à l'aulne (ou plutôt aux espèces d'aulnes, en Europe et en Amérique du Nord) et des *Hymenoscyphus*, associés aux frênes, en Europe et au Japon, le montre bien. La reconnaissance des espèces phylogénétiques a également permis de montrer que de nombreuses maladies sont causées non pas par un seul agent pathogène mais par des complexes d'espèces dites cryptiques (= non différenciables morphologiquement). Il est alors important d'étudier les implications épidémiologiques de l'existence de ces complexes d'espèces, les espèces fongiques impliquées pouvant varier pour leur répartition géographique, leur sensibilité aux facteurs climatiques, leur pouvoir pathogène vis à vis des espèces forestières... Les recherches en cours sur l'oïdium du chêne s'inscrivent dans cet objectif.

Nous avons assisté au cours des dernières décennies à l'apparition de plus en plus fréquente de "nouvelles maladies". Ces maladies émergentes sont le plus souvent causées par des organismes introduits, inconnus avant les dégâts causés par leur introduction. En effet, dans leur zone d'origine, ces champignons cohabitent avec leurs arbres hôtes sans causer de dommages importants, et la diversité fongique dans de nombreuses régions du monde reste à explorer. Dans ce contexte, il est essentiel de pouvoir caractériser très rapidement les agents pathogènes en cas d'apparition de symptômes anormaux sur une essence donnée. Les outils moléculaires permettent de développer des tests de détection très rapides, spécifiques et sensibles, comme cela a été réalisé par l'ANSES pour caractériser *C. fraxinea* dans les tissus de frêne. Ces tests peuvent alors être mobilisés pour éviter la dissémination de l'agent pathogène, en particulier lorsque sa nature non indigène est fortement suspectée. La mise en évidence tardive de l'origine exotique de *C. fraxinea* a ainsi retardé la mise en place de mesures de quarantaine. Retracer l'origine des pathogènes pourrait ultérieurement permettre de trouver des sources de résistance dans les espèces avec lesquelles ils ont co-évolué. L'accès à l'analyse de la diversité de l'ADN permet en effet d'étudier la structuration des populations et de proposer ainsi des scénarios de dispersion. Enfin, l'essor de la génomique ouvre des perspectives inédites sur la description de la diversité fongique: l'amplification de l'ITS, avec des amorces fongiques universelles, associée à des techniques de séquençage haut-débit\*, a permis de mettre en évidence la richesse en termes d'espèces de champignons associées à la rhizosphère et la phyllosphère des arbres. A l'image des recherches en cours sur le rôle du microbiote intestinal sur la santé humaine, cette diversité cachée est aujourd'hui étudiée en lien avec la résistance des arbres aux stress biotiques et abiotiques.

## Glossaire

(la plupart des définitions sont extraites et adaptées de Wikipedia: <http://fr.wikipedia.org>, qui peut être consultée pour chaque mot pour plus d'informations)

**ADN** = acide désoxy-ribo-nucléique : macro-molécule constituée d'un enchaînement de nucléotides, eux-mêmes composés de bases azotées, d'oses (désoxyribose) et de groupes phosphate. L'ADN détermine la synthèse des protéines, par l'intermédiaire de l'acide ribonucléique (ARN). Dans les cellules eucaryotes (plantes, animaux, champignons), l'ADN est contenu dans le noyau, avec également un peu d'ADN dans les mitochondries et les chloroplastes.

**Amorce** : courte séquence d'ADN, complémentaire du début de la séquence cible, servant de point de départ à la synthèse du brin complémentaire de cette séquence par une ADN polymérase (voir réplication).

**BLAST** = basic local alignment search tool : méthode de recherche utilisée en bio-informatique permettant de trouver les régions similaires entre deux ou plusieurs séquences de nucléotides ou d'acides aminés et de réaliser un alignement de ces régions homologues.

**Gène** : séquence d'ADN qui code pour la synthèse d'une molécule fonctionnelle pour la cellule, généralement protéine ou acide ribonucléique (ARN) ; le gène peut être défini comme l'unité d'information génétique

**Génome** : ensemble du matériel génétique d'un individu ou d'une espèce codé dans son ADN

**ITS** = Internal Transcribed Spacer (Espaceur Interne Transcrit): fragments non fonctionnels d'ARN situés entre les fragments d'ARN ribosomiques au sein d'un même précurseur; au cours de la maturation de l'ARN les ITS sont excisés et dégradés alors que les fragments fonctionnels vont constituer les ribosomes

**Mutation**: changement permanent dans la séquence nucléotidique du génome d'un organisme, pouvant être causé par différents facteurs ou processus; une mutation n'aboutit pas forcément à une modification visible de l'organisme

**PCR** = Polymerase Chain Reaction (réaction en chaîne par polymérase) : méthode d'amplification génique *in vitro*, qui permet de dupliquer une séquence d'ADN connue, disponible en très faible quantité, avec un facteur de multiplication de l'ordre du milliard. La technique se fonde sur les propriétés d'hybridation et de déshybridation des brins complémentaires d'ADN en fonction de la température et l'utilisation d'ADN polymérases (enzymes) thermostables, grâce à des transitions de température (assurées par un thermocycleur) répétées de manière cyclique (voir animation à : <http://www.snv.jussieu.fr/bmedia/PCR>)

**Phylogénie** : étude des relations de parenté entre êtres vivants. La phylogénie permet de reconstituer l'évolution des organismes vivants, souvent représentée par un arbre phylogénétique. Les espèces phylogénétiques sont ainsi définies comme des lignées dans l'arbre phylogénétique.

**Séquençage**: processus permettant de déterminer l'ordre d'enchaînement des nucléotides d'un fragment d'ADN donné. La méthode classique dite de "Sanger" a valu le prix Nobel à son auteur en 1980. Depuis le milieu des années 2000, de nouvelles méthodes dites à haut débit (HTS pour high-throughput sequencing) aussi appelées NGS (next-generation sequencing) permettent de produire des millions de séquences rapidement et à faible coût (mais plus courtes que par la méthode Sanger). Le pyroséquençage appartient à ces nouvelles techniques.

**SNP**= Single Nucleotide Polymorphism (polymorphisme d'un seul nucléotide): variation (polymorphisme) d'un seul nucléotide (par exemple un C à la place d'un T dans une séquence donnée), entre individus d'une même espèce. Ces variations représentent la forme la plus fréquente de variation génétique entre individus: en moyenne un SNP tous les 300 nucléotides, soit environ 10 million de SNP dans le génome humain.

**Réplication**: processus au cours duquel l'ADN est synthétisé grâce à une enzyme appelée ADN polymérase. A partir d'une molécule d'ADN initiale, les 2 brins sont séparés et deux molécules identiques à la molécule initiale (chacune complémentaire d'un des 2 brins) sont formées .

[http://www.ac-rennes.fr/pedagogie/svt/cartelec/cartelec\\_lyc/premiere\\_s/vegetal/adn/replic1.htm](http://www.ac-rennes.fr/pedagogie/svt/cartelec/cartelec_lyc/premiere_s/vegetal/adn/replic1.htm)

**Transcription**: processus par lequel les régions dites codantes de l'ADN sont transcrites en molécules d'ARN. Ces molécules d'ARN serviront ensuite à la traduction en séquences protéiques. voir animation: [http://www.snv.jussieu.fr/vie/dossiers/ADN\\_Prot/ADN\\_ARN/ADN\\_ARN2.html](http://www.snv.jussieu.fr/vie/dossiers/ADN_Prot/ADN_ARN/ADN_ARN2.html)