

TD – Micro-organismes S3

Partie 2 : Classification des micro-organismes par une approche génétique

Méthode UPGMA

Lors de la première partie du TD, vous avez effectué une identification des bactéries par une approche phénotypique : les bactéries sont identifiées selon les propriétés phénotypiques comme la coloration de Gram, la morphologie, le mode respiratoire ou les propriétés métaboliques. Cette méthode traditionnelle d'identification bactérienne a montré des limites, en particulier pour la détection des micro-organismes non ou difficilement cultivables.

Elles ont conduit à la description d'une très faible partie de la diversité bactérienne existante et à la sous-estimation même de la richesse du monde vivant qui nous entoure. Toutefois, les tests phénotypiques fournissent les informations sur la capacité de l'organisme qui peuvent nous aider à comprendre son rôle dans l'environnement où il vit.

Dans cette partie, vous utiliserez une autre approche pour identifier les bactéries. Il s'agit de classification phylogénétique fondée sur l'analyse des séquences d'ARN 16S ou de gènes de protéines. Les méthodes phylogénétiques permettent, aujourd'hui, une détection fiable des bactéries, notamment les bactéries qui sont difficilement cultivées au laboratoire.

Les ARNr (ARN ribosomiaux) ont été choisis en taxonomie pour trois raisons principales :

- . Ils sont présents dans toutes les cellules ce qui permet des comparaisons entre procaryotes et eucaryotes.

- . Ils ont une structure bien conservée car toutes modifications importantes pourraient avoir des conséquences importantes sur les synthèses protéiques. Il existe d'ailleurs des portions d'ARNr dont la séquence est identique chez tous les êtres vivants.

- . Ils sont abondants dans la cellule et faciles à purifier.

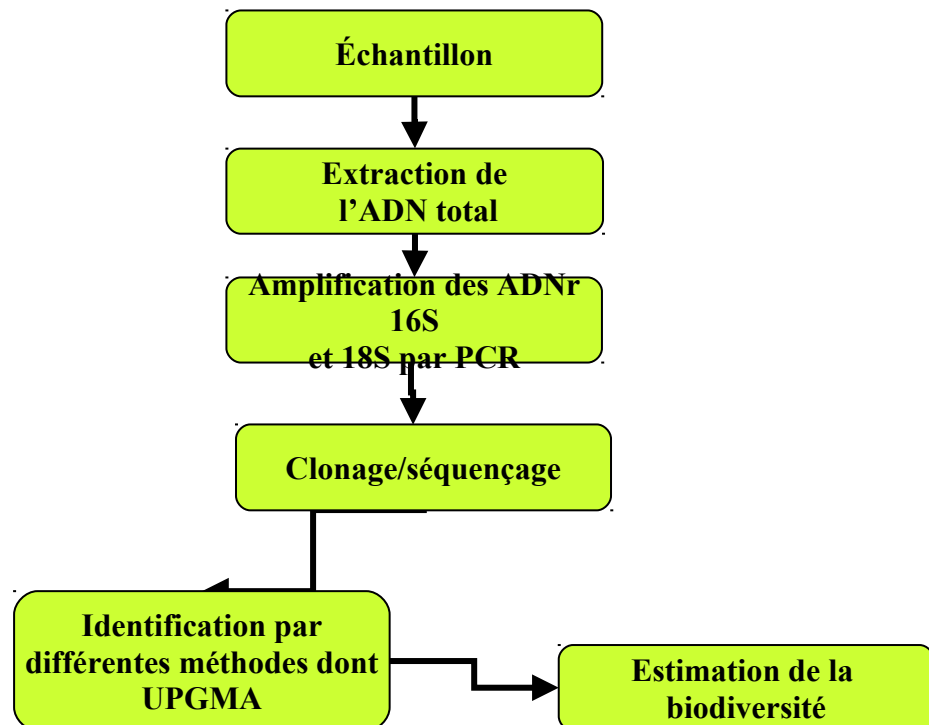
Les ARNr s'associent à des protéines pour former les ribosomes qui, chez les procaryotes, sont constitués d'une sous-unité 30S et d'une sous-unité 50S. La sous-unité 30S d'un ribosome contient de l'ARNr 16S et la sous-unité 50S contient de l'ARNr 5S et de l'ARNr 23S. L'ARNr 5S est formé d'environ 120 nucléotides, l'ARNr 16S d'environ 1 540 nucléotides et l'ARNr 23S comprend environ 2 900 nucléotides. Tous les ARNr sont monocaténaux sauf dans quelques régions où se forment des boucles par appariement de bases complémentaires situées sur la même chaîne.

Chez une bactérie, les gènes qui codent pour les ARNr sont organisés en opérons dont la structure est semblable : trois gènes, séparés par un espace, codent pour les ARNr 16S (gène *rrs*), les ARNr 23S (gène *rrl*) et pour les ARNr 5S (gène *rrf*). L'ARNr 16S est le plus utilisé car il est plus facile à analyser que l'ARN 23S et plus riche en information que l'ARNr 5S.

L'inconvénient est que les résultats obtenus peuvent être variables selon les séquences génétiques étudiées.

Parmi plusieurs méthodes basées sur cette approche, la UPGMA est la méthode la plus simple, la plus rapide et la plus utilisée

Schéma de la méthodologie basée sur l'ARNr 16S



● **UPGMA (Unweight Pair Group Method with Arithmetic mean)**

C'est une méthode agglomérative (cluster analysis) qui regroupe ensemble les séquences les plus proches. Les démarches sont :

1. Etablir la matrice à partir de l'alignement des séquences
2. Calculer les pourcentages de divergence nucléotidique (toutes différences incluses) entre les différentes séquences prises deux à deux.
3. Regrouper les micro-organismes selon la similitude entre les micro-organismes.
4. Reconstruire la nouvelle matrice et continuer de regrouper les micro-organismes

Exemple :

(A) Méthode UPGMA

Ce tableau décrit une région de 9 bases dans les ARNr 16S de quatre souches.

Organisme (souche)	Nombre de sites								
	1	2	3	4	5	6	7	8	9
a	G	C	G	G	A	C	A	A	A
b	G	A	C	G	C	C	A	A	G
c	G	A	A	A	U	C	U	A	A
d	G	A	A	A	G	C	U	A	G

1 Construction d'une matrice de distance montrant la parenté (distance d) entre les souches en vue de déterminer les plus proches, dans ce cas, c et d.

Première matrice

	a	b	c
b	$d_{ab} = 4$	—	—
c	$d_{ac} = 5$	$d_{bc} = 5$	—
d	$d_{ad} = 6$	$d_{bd} = 4$	$d_{cd} = 2$

Premier arbre

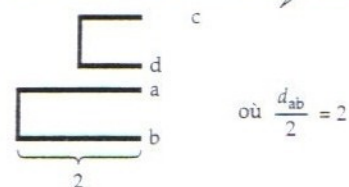


3 Construction d'une deuxième matrice pour tester la distance entre les deux groupes associés (cd) et les deux souches (a et b).

Deuxième matrice

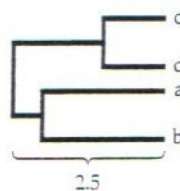
	(cd)	a
a	$d_{(cd)a} = 11/2$	—
b	$d_{(cd)b} = 9/2$	$d_{ab} = 4$

Deuxième arbre



5 Détermination de la parenté entre ac et (a et b), et tracé d'un diagramme.

Arbre final



$$\text{où } d_{(cd)(ab)} = \frac{(11/2 + 9/2)/2}{2} = 2.5$$

Exercices de la méthode UPGMA

Vous avez une bactérie inconnue à identifier par la méthode UPGMA. Vous allez comparer la séquence de cette bactérie par rapport les trois autres espèces de la famille de Streptococcus : *Streptococcus.canis*, *Streptococcus.equis* et *Streptococcus hyointestinalis*

Après l'extraction de l'ADN, des échantillons ont été amplifiés par PCR. Des produits de PCR ont été séquencés. Les séquences de 150 bases nucléotidiques correspondant à une même région des produits amplifiés provenant de quatre micro-organismes différents ont ensuite été alignées en utilisant un logiciel d'alignement multiple de séquences. Le pourcentage de divergence nucléotidique (la différence) est donné dans le tableau 1. C'est la première matrice

1 ^{ère} Matrice	bactérie inconnue	<i>Streptococcus.equis</i>	<i>Streptococcus.canis</i>	<i>Streptococcus hyointestinalis</i>
bactérie inconnue	0			
<i>Streptococcus.equis</i>	22,87	0		
<i>Streptococcus.canis</i>	<u>14,89</u>	18,08	0	
<i>Streptococcus hyointestinalis</i>	50,53	52,13	45,21	0

Note : Les bactéries ayant de plus de 98,5 % de similitude appartient à la même espèce.
Continuez de regrouper les bactéries par la méthode UPGMA