
Étude de la biodiversité fongique à l'aide de techniques de pyroséquençage

Biodiversité fongique

- **Biodiversité**: diversité spécifique d'une communauté écologique, correspondant au nombre d'espèces et à leur abondance relative.
 - Qui? Combien?
 - Originellement, l'identification des espèces se faisait sur l'observation morphologique des carpophores
-

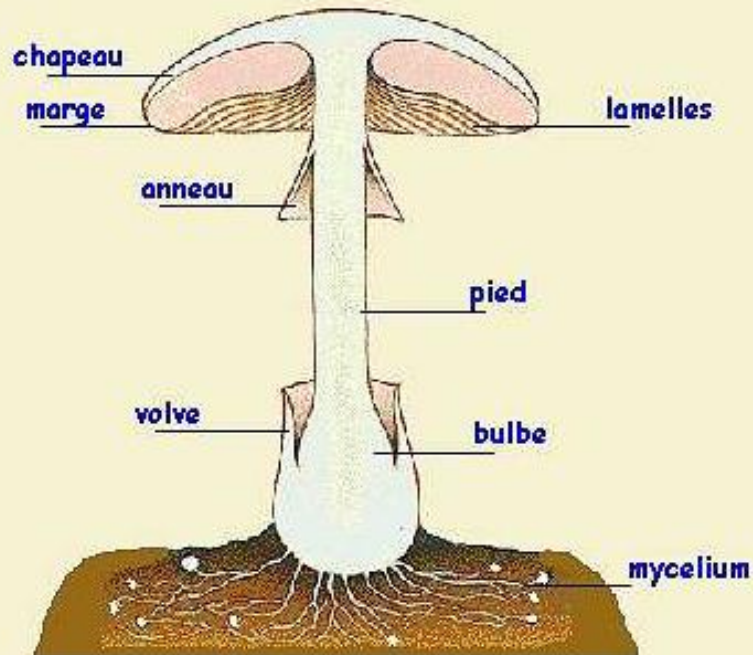


Source: http://www.botanicalgarden.ubc.ca/potd/fungus_diversity.jpg

Carpophores

Anatomie et description du champignon typique

La grande famille des champignons comporte des espèces aux formes les plus diverses parmi lesquelles se distingue la plus commune et la plus connue : un pied surmonté d'un chapeau (carpophore) généralement convexe.



Sous le chapeau, il peut y avoir des lames, des lamelles, des lamellules, des plis, des aiguilles ou des tubes



Les lames peuvent être :

libres, échancrées, décurrentes, adnées, émarginées, anastomosées, fourchues, collariées, espacées, serrées, fines, épaisses, larges Cliquez !

Pourquoi séquencer ?

- Carpophores ne sont disponibles qu'à un moment restreint de l'année
 - Métagénomique
 - Mise en culture en Pétri: mycélium se développe, mais pas de fructification/carpophore
 - "Accessibilité visuelle"
 - Structures microscopiques – Microorganismes
-

Du carpophore au moléculaire

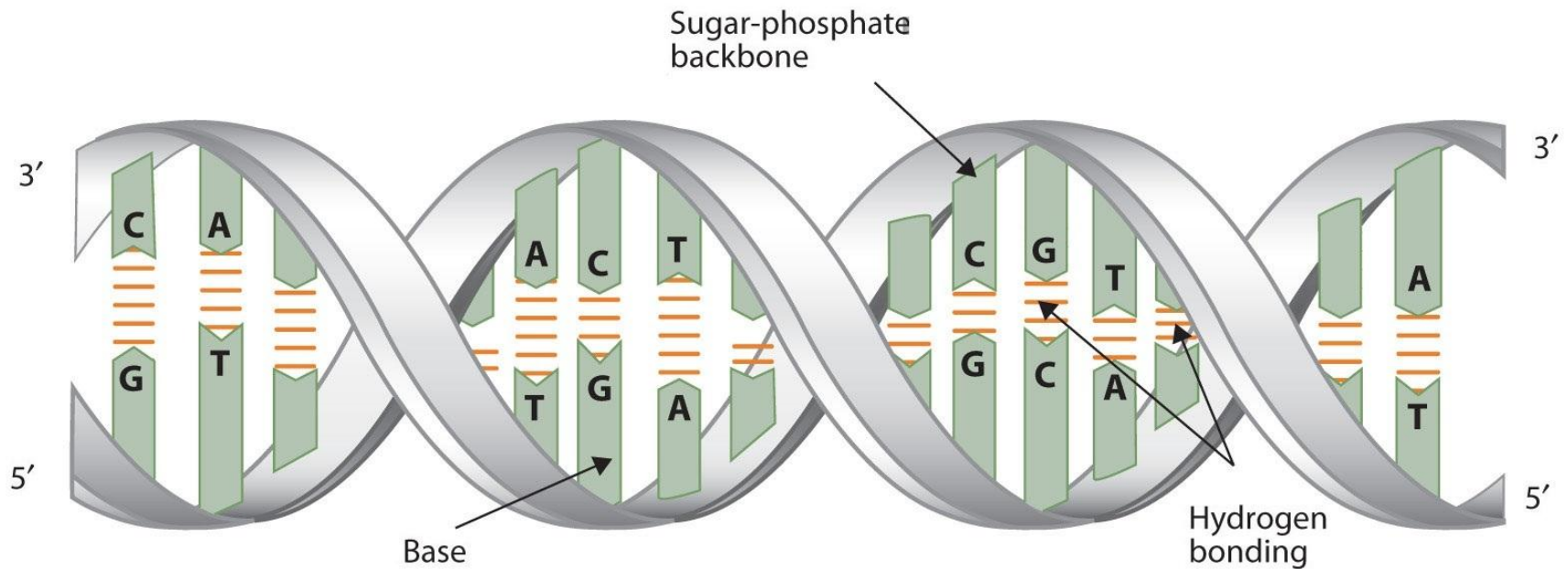


>Morchella elata

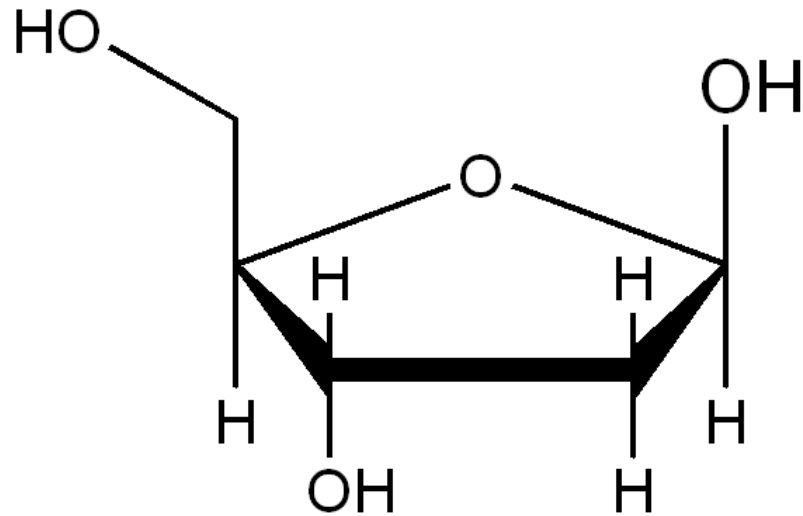
```
CCTGCGGAAGGATCATTACCAAGAACCACACAGAAAAGG
GAGGCAAAGGGGCCGACAGGGCTAGTAGCTTATACGTTG
TTGAACGTCCTGCCTGGACCCGGAGCCGCCCCCATCTAA
ACCCTCTGCGTACCTGTCCCTTCTTGCTTCCCCCGGCATC
TCGTCGGGGGGGAGGTAACAACCAAAACTCTCTGTGAATCA
AACAGCCGTCAGAATTATAAAACAAAACAAAAGTTAAACTT
TCAACAACGGATCTCTTGGTTCCACATCGATGAAGAACG
CAGCGAAATGCGATAAGTAATGTGAATTGCAGAATTCAGT
GAATCATCGAATCTTTGAACGCACATTGCGCCCCCTGGTA
TTCCGGGGGGGCATGCCTGTTGAGCGTCATAAAAACCTC
CTCCCCCTTCGGGTTTTGTTACTATCGTTGGGGGGTTTTG
GCCTAATGGGATAGCGATTGGCAATTCGTTTCCAATGTC
CTAAATAGACGTAGACCCGCCTCCAGATGCGACAGCACC
GAGGCCATCAACCGTGGAGTTATGGGATATAATAGGCTTG
CAGTAAAATGCTCACCTCTCTCCACACGCCGATGGCGCG
ACAGTTGCAGTTGCGGGCGTAAATTGGAGCCCTTTTCAGG
ACCCTTGTGGCCTAGCATCCACCATAACATAATTG
```

Séquencer?

- ADN: Acide DésoxyriboNucléique



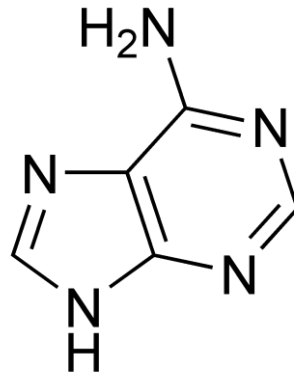
Désoxyribose



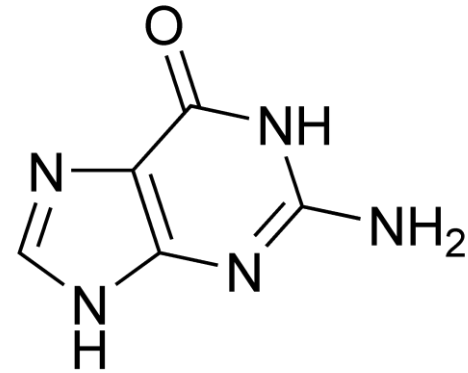
Base azotée

Purines

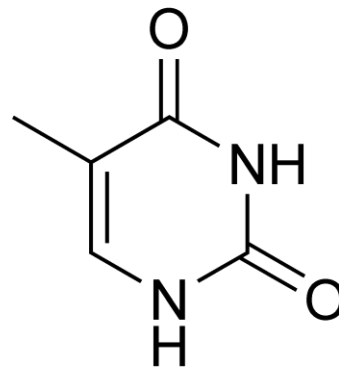
Adenine



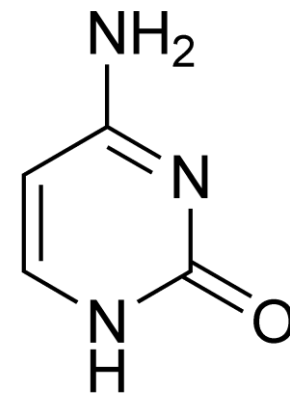
Guanine



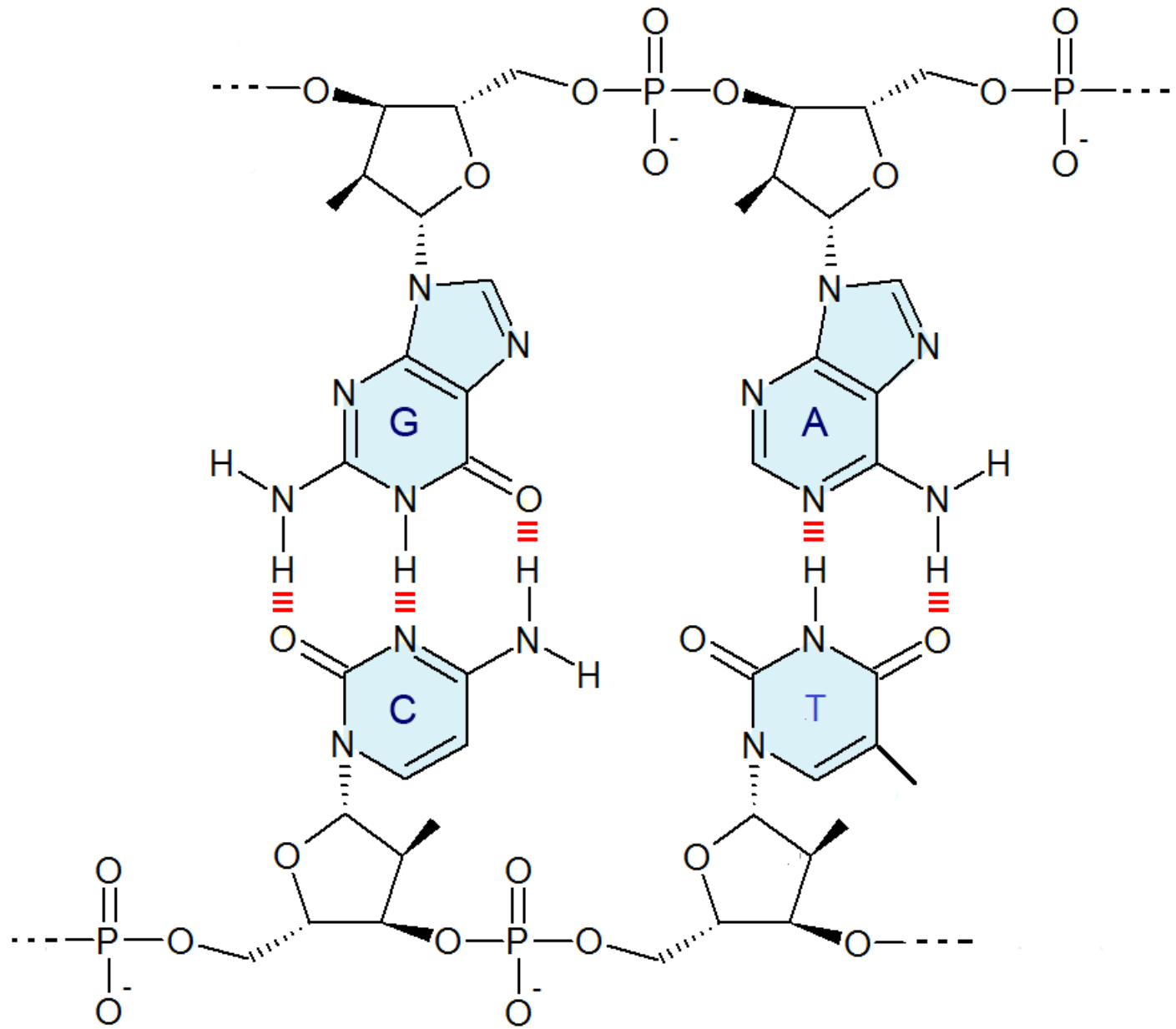
Pyrimidines



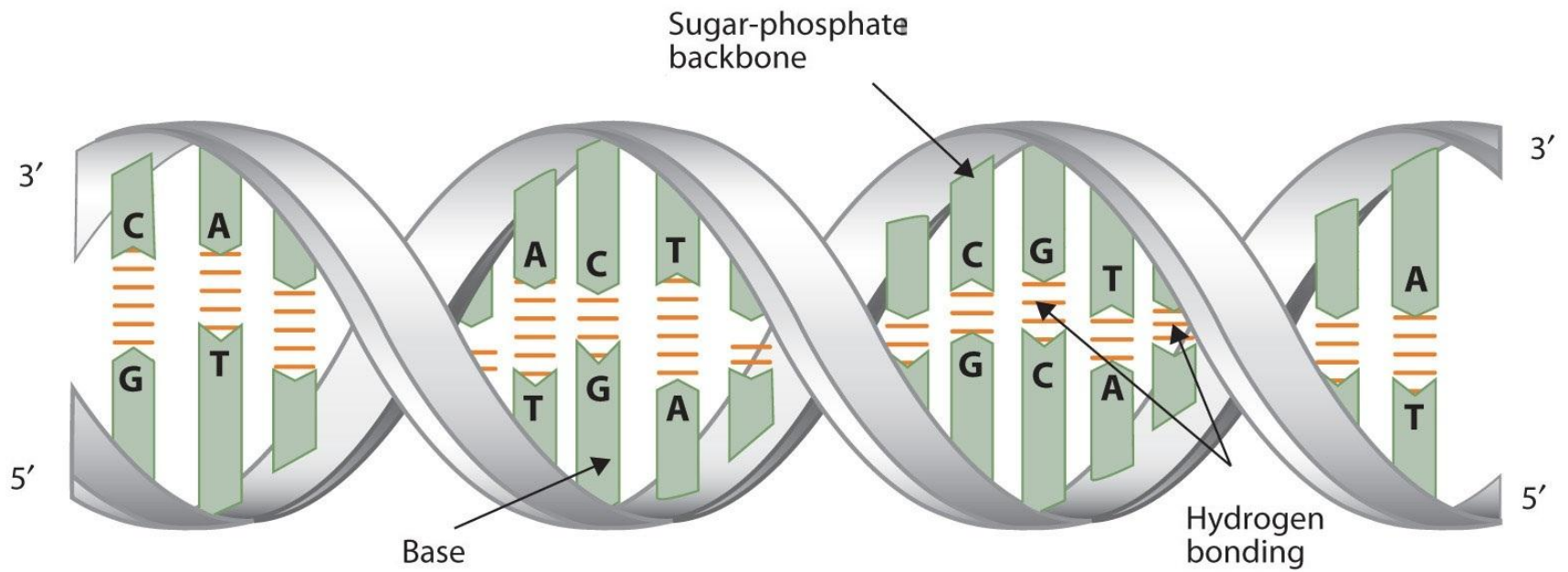
Thymine



Cytosine



ADN



Séquencer

- Séquencer = obtenir l'ordre d'agencement des nucléotides
- >Morchella elata

```
CCTGCGGAAGGATCATTACCAAGAACCACACAGAAAAGGGAGGCAAAGGGGGCCGACAGGG
CTAGTAGCTTATACGTTGTTGAACGTCCTGCCTGGACCCGGAGCCGCCCCCATCTAAACCCT
CTGCGTACCTGTCCCTTCTTGCTTCCCCCGGCATCTCGTCGGGGGGGAGGTAACAACCAAAA
CTCTCTGTGAATCAAACAGCCGTCAGAATTATAAAACAAACAAAAGTTAAACTTTCAACAAC
GGATCTCTTGGTTCCACATCGATGAAGAACGCAGCGAAATGCGATAAGTAATGTGAATTGC
AGAATTCAGTGAATCATCGAATCTTTGAACGCACATTGCGCCCCCTGGTATTCCGGGGGGCA
TGCCTGTTGAGCGTCATAAAAACCTCCTCCCCCTTCGGGTTTTGTTACTATCGTTGGGGGG
TTTTGGCCTAATGGGATAGCGATTGGCAATTCGTTTTCCAATGTCCTAAATAGACGTAGACCC
GCCTCCAGATGCGACAGCACCGAGGCCATCAACCGTGGAGTTATGGGATATAATAGGCTTG
CAGTAAAATGCTCACCTCTCTCCACACGCCGATGGCGCGACAGTTGCAGTTGCGGGGCGTAA
ATTGGAGCCCTTTTCAGGACCCTTGTGGCCTAGCATCCACCATAACATAATTG
```

Séquencer quoi?

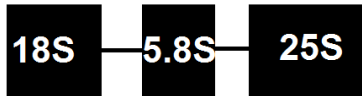
Nuclear ribosomal internal transcribed spacer (ITS) region as a universal DNA barcode marker for *Fungi*

Conrad L. Schoch^{a,1}, Keith A. Seifert^{b,1}, Sabine Huhndorf^c, Vincent Robert^d, John L. Spouge^a, C. André Levesque^b, Wen Chen^b, and Fungal Barcoding Consortium^{a,2}

^aNational Center for Biotechnology Information, National Library of Medicine, National Institutes of Health, Bethesda, MD 20892; ^bBiodiversity (Mycology and Microbiology), Agriculture and Agri-Food Canada, Ottawa, ON, Canada K1A 0C6; ^cDepartment of Botany, The Field Museum, Chicago, IL 60605; and ^dCentraalbureau voor Schimmelcultures Fungal Biodiversity Centre (CBS-KNAW), 3508 AD, Utrecht, The Netherlands

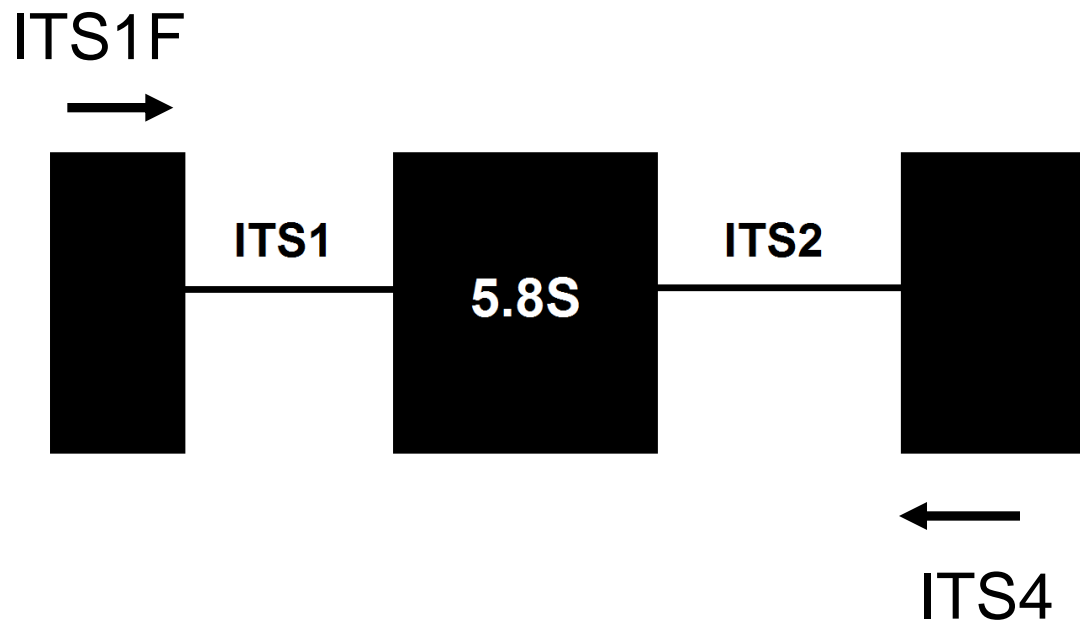
Edited* by Daniel H. Janzen, University of Pennsylvania, Philadelphia, PA, and approved February 24, 2012 (received for review October 18, 2011)

ITS

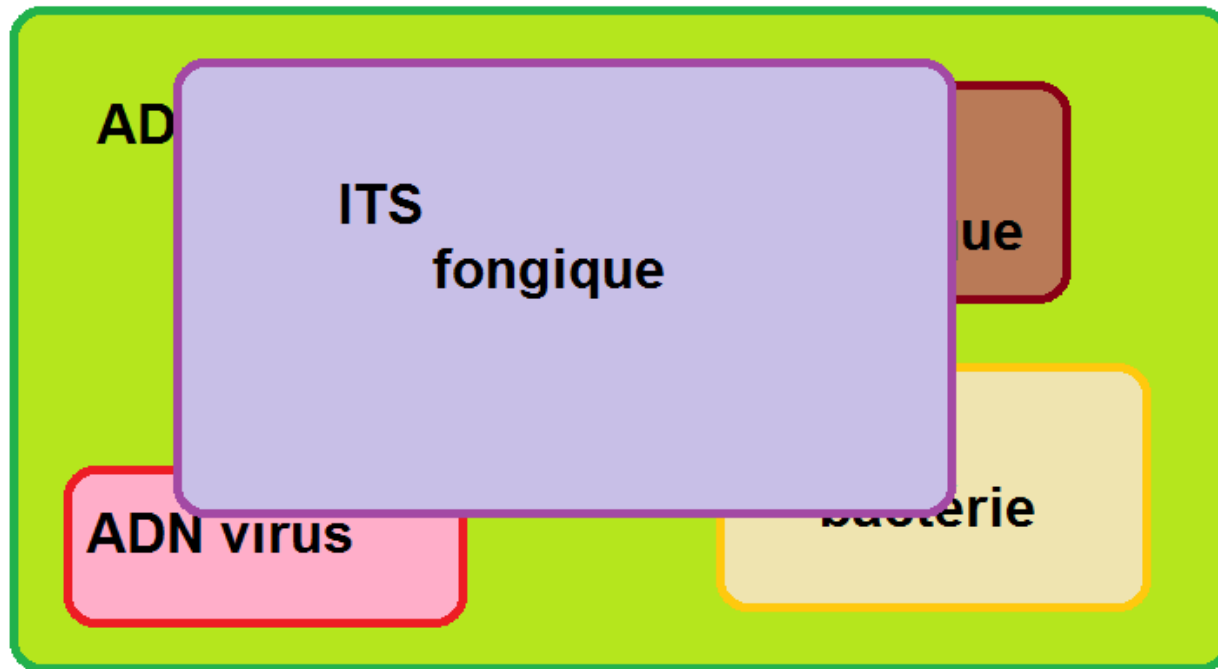


60 à 220 copies dépendamment de l'espèce

ITS1-5.8S-ITS2



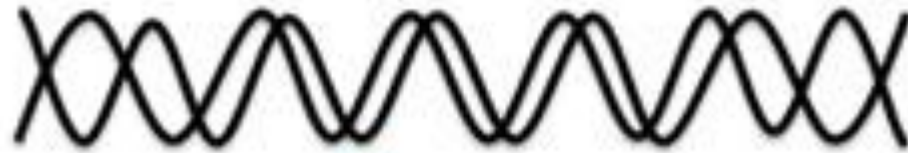
PCR



PCR

- Amorces:
 - ITS1F: CTTGGTTCATTTAGAGGAAGTAA
 - ITS4: TCCTCCGCTTATTGATATGC
-

PCR



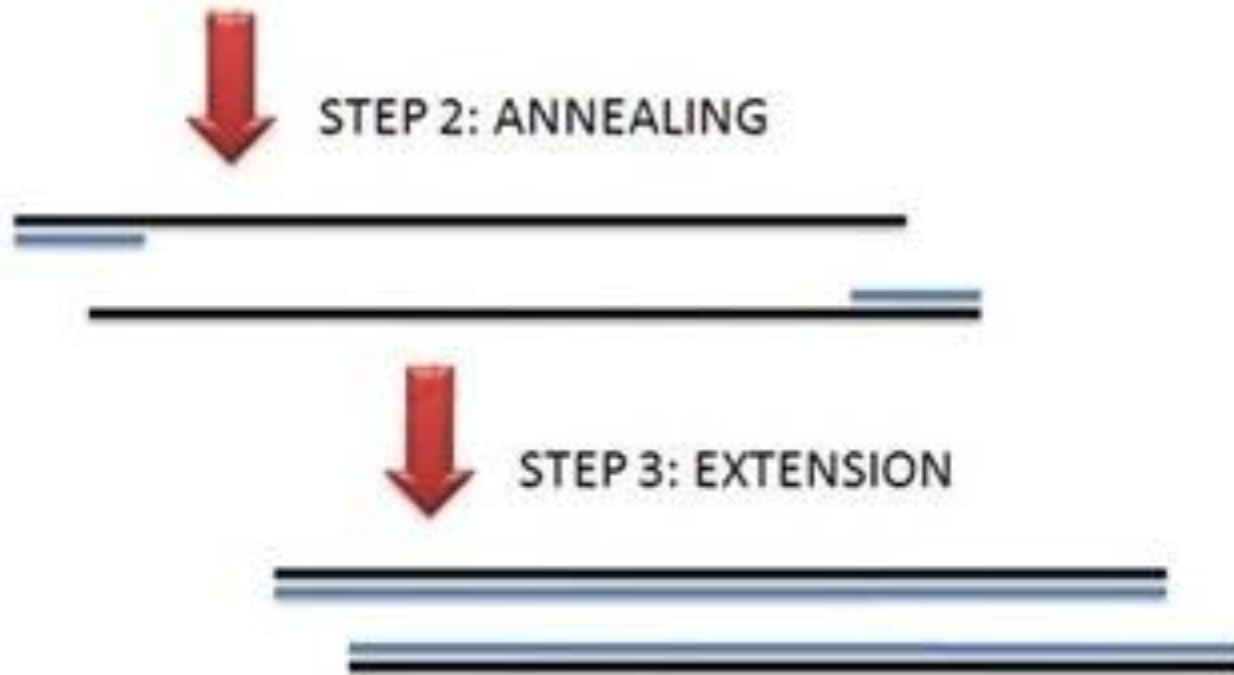
STEP 1: DENATURATION



STEP 2: ANNEALING



PCR





STEP 1: DENATURATION



STEP 2: ANNEALING



STEP 3: EXTENSION



PCR

- Amorces:
 - ITS1F: CTTGGTTCATTTAGAGGAAGTAA
 - CCATCTCATCCCTGCGTGTCTCCGACTCAG
 - ACGAGTGCGT
 - CTTGGTTCATTTAGAGGAAGTAA
-

PCR

- Amorces:

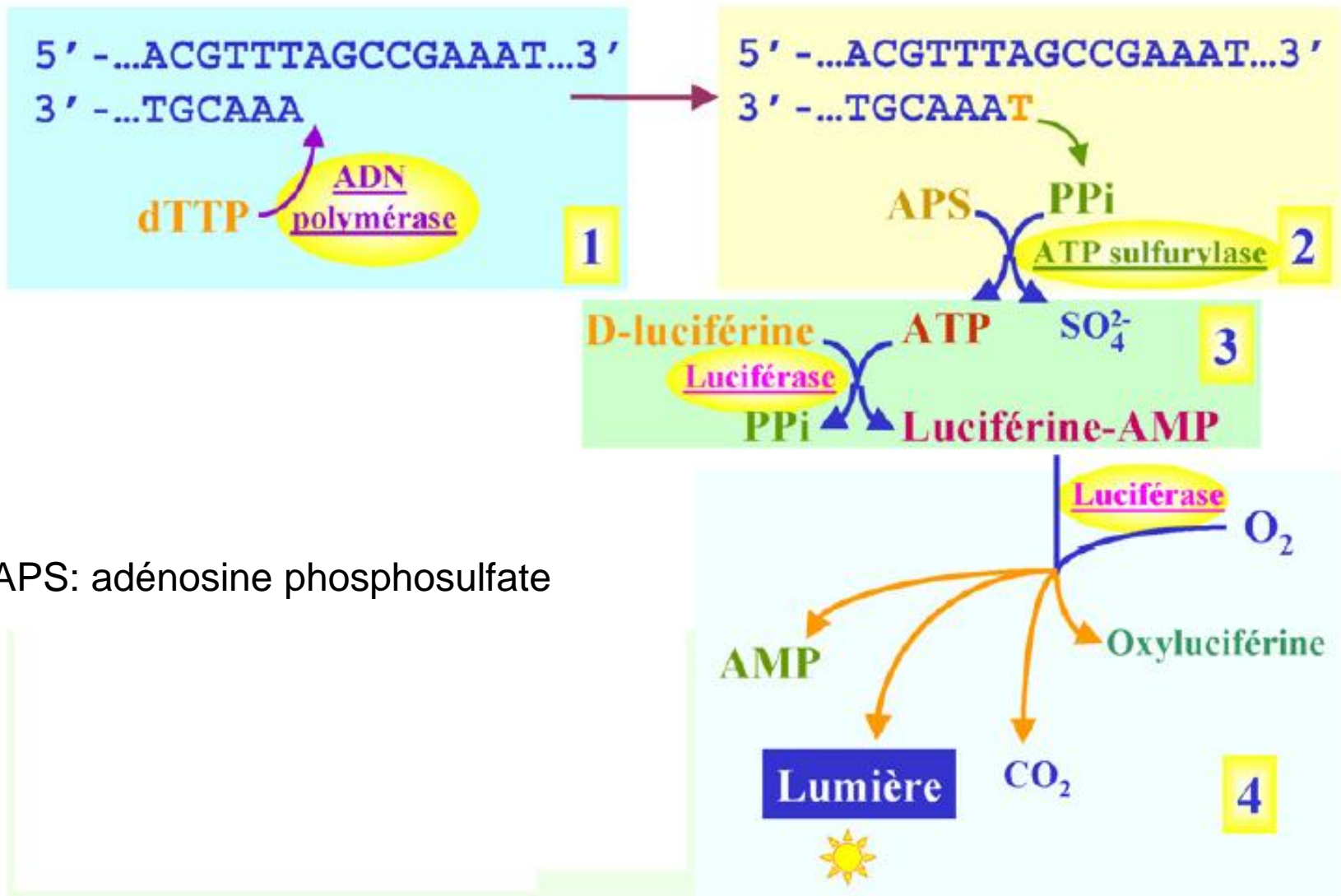
- ITS4: TCCTCCGCTTATTGATATGC

- CCTATCCCCTGTGTGCCTTGGCAGTCTCAG

Pyroséquençage

- Synthèse d'ADN cible en temps réel
 - Chaque base marquée par un fluorophore
 - Signal par bioluminescence si incorporée à base complémentaire
-

Principe



APS: adénosine phosphosulfate

Illumina vs 454 Roche

- Illumina/Solexa: fixation des amplicons sur plaque
 - Pour chaque position, les 4 nucléotides en compétition
 - Ajout d'un seul nucléotide à la fois
 - Moins d'incertitudes
 - Pb: longueur = 150bp
-

Illumina vs 454 Roche

- 454 Roche: fixation sur des billes (emPCR)
 - Ajout séquentiel des nucléotides (A, C, G, T)
 - Pour une position: pas de compétition entre les différents nucléotides
 - Possibilité d'ajouter plusieurs nucléotides à la fois (homopolymères)
 - Longueur: 500bp
-

Fichiers sortie 454 Roche

- .sff: Standard Flowgram Format
 - Contient l'ensemble des données en format codé
 - .fna: fichier fasta: succession de nucléotides
 - .qual: fichier qualité succession de Phred score
-

Sortie 454 .fna

- >FM02SVAb\$BEWIY0O8

```
AAGTCGTAACAAGGTTTCCGTAGGTGAACCTGCGGAAGGATCA
TTAGTTGACAGGCCGTTTCGCGCGGTACCTGAAACCCTCCTTAC
AGCGGCCGGAGTTCGGCGCAAGGGCGAAGCCCTCGCGGGGC
GGAGTCCGCCGGAATCTCGCAAACCATGCCTTGAAACTGCGGT
CTGAGAGAATGATGAATTAGCGAAAACTTTCAACAACGGATCTC
TTGGTTCTGGCAACGATGAAGAACGCAGCGAAATGCGATAAGT
AATGTGAATTGCAGAATTCAGTGAATCATCGAATCTTTGAACGC
ACATTGCGCCCCTTGGTATTCCGGGGGGGCACGCCTGTTCGAG
CGCCATTTCAACCCTCGAGCCCCGCTCGGTGATGGGCCCGTC
```

Sortie 454 .qual

■ >FM02SVAb\$BEWIY008

32 32 40 40 40 40 40 40 40 40 40 40 40 40 40 39 39 39 40 40 40 40 40 40
40
40
40 40 40 40 39 39 39 38 38 38 40 40 40 40 40 40 40 40 40 40 40 40 40
40 40 40 40 40 40 40 40 40 40 40 40 40 40 40 39 39 39 40 40 40 40 39
21 21 21 26 40 40 40 34 34 34 34 40 40 40 40 40 40 40 40 40 40 40 40
40 40 40 40 40 40 40 40 40 39 39 39 40 40 40 40 40 40 40 40 40 40 39
39 39 40
40 40 40 40 40 40 40 40 31 31 31 31 40 39 39 39 40 40 40 40 40 40
40
40 40 40 40 40 40 40 40 40 40 40 40 40 40 40 34 34 34 40 40 40 40
40
40 40 40 40 40 40 40 40 40 40 40 40 40 40 40 40 39 39 39 40 40 40
40 40 40 40 40 40 40 40 40 40 34 34 34 34 40 40 40 34 35 35 40 40
40 19 17 17 17 17 17 40 40 40 40 40 40 40 40 40 40 40 40 40 40 40
40 40 40 39 39 39 40 39 39 34 34 34 40 32 40 40 36 19 19 19 19 32 26
32 22 20 20 28 26 30 30 15 15 15 15 15 15 12 25 12

Phred scores

- Indice de confiance dans l'attribution de la base (inversement probabilité d'erreur)

Phred quality scores are logarithmically linked to error probabilities

Phred Quality Score	Probability of incorrect base call	Base call accuracy
10	1 in 10	90 %
20	1 in 100	99 %
30	1 in 1000	99.9 %
40	1 in 10000	99.99 %
50	1 in 100000	99.999 %

Qualité 454

- Développé pour séquençage génome: qualité moyenne mais compensé par grand nombre de séquences pour le même fragment (consensus)
 - Pb métagénomique: grand nombre de séquences du même fragment mais pas du même individu
-

Trim.seqs

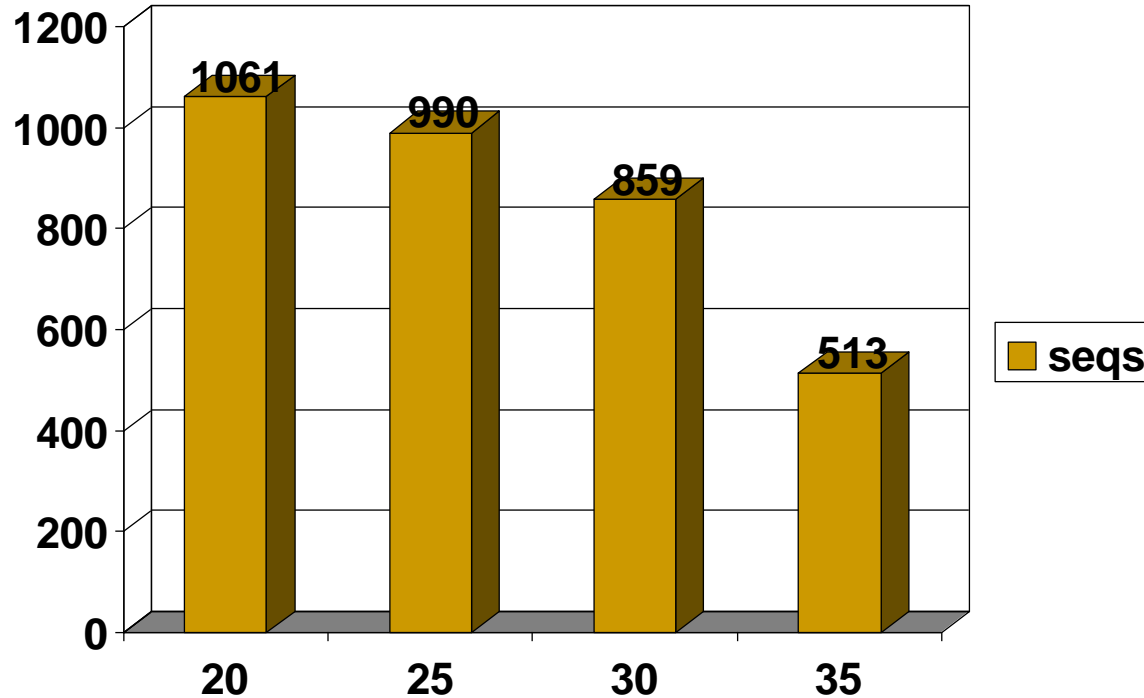
- 1/8 de ronde de 454 : 120 000 seqs
 - Seqs interrompues aléatoirement
 - Étiquette (éventuellement intégrer infos dans noms des séquences)
 - Séquence ITS1F
-

Trim.seqs

- Moyenne Phred score: 30
 - Fenêtre chevauchante de 50bp
 - Pas d'ambiguïté
 - Homopolymères < 8bp
 - Minlength = 150bp
-

Variation Phred scores

- Fichier brut de 7250 seqs (minlength = 300)



Recouvrement

- But recouvrement: définir un fragment pour l'ensemble des séquences qui contient la même quantité d'information
 - Éviter surestimation de la biodiversité
 - Extrémité 3' dans 5.8S (conservés)
 - Longueur moy du recouvrement: 316bp
-

Alignement

- Mesure de similarité entre les séquences
 - \neq homologie
 - Global alignement: force l'alignement sur l'ensemble des séquences
 - Local alignement: cherche les régions conservées contenues dans les séquences
-

Matrice alignement

- Indice similarité: nombre de bp à une position identique ou score

	A	C	G	T
A	1	0	0	0
C	0	1	0	0
G	0	0	1	0
T	0	0	0	1

	A	C	G	T
A	3	0	1	0
C	0	3	0	1
G	1	0	3	0
T	0	1	0	3

Gap

- 1 gap = 1 événement biologique
 - Une insertion ou une deletion
- Une valeur négative au placement d'un gap
 - Amplifiée par la longueur du gap

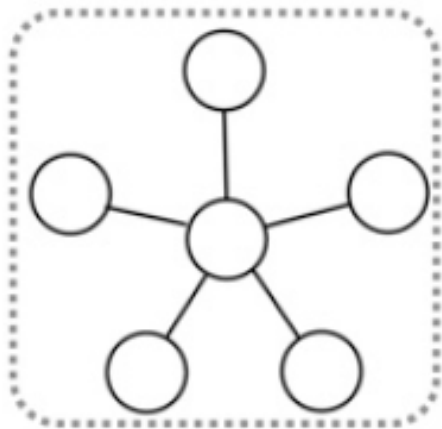
- AAGCT AAG_CT
- AGGCT _AGGCT

Alignement

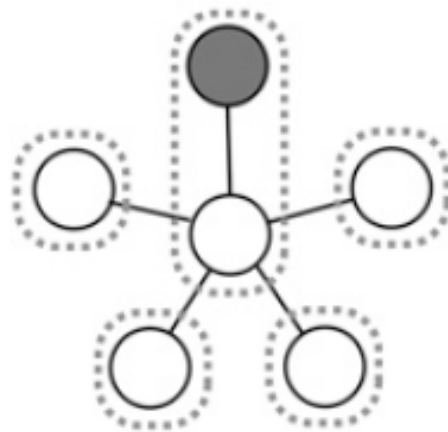
- Aligner = disposer les nucléotides de deux séquences différentes afin de maximiser le nombre de nucléotides positionnés au même endroit.
 - Mettre en évidence les similarités entre les séquences
-

Molecular Operational Taxonomic Unit

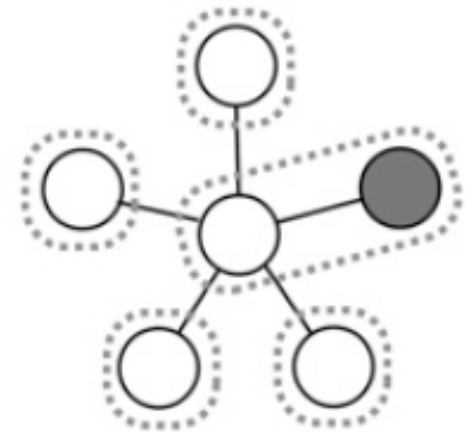
E. Nearest Neighbor = 1 OTU



F. Furthest Neighbor = 5 OTUs



OR



etc.



mOTUs

- Two-step clustering
 - 1er clustering à basse similarité (90%)
 - 2nd clustering par pré-mOTU à plus haute similarité (95-97%)
 - Une séquence référence par mOTU
 - La similarité se base sur la référence
-

Identification



Home R

Description

Legend for links

Sequences

Access

- JF908364
- JQ691485
- JN085121
- JN085120
- GU08323
- GU08322
- EF434146
- JN085111
- GQ24905
- EF081000
- EF017946
- EU292650
- AJ698477
- AJ544199
- GQ22846
- GQ22847
- GQ22847
- JN085119

> [gb|JF908364.1](#) *Morchella elata* voucher 13895 18S ribosomal RNA gene, internal transcribed spacer 1, 5.8S ribosomal RNA gene, internal transcribed spacer 2, and 28S ribosomal RNA gene, region
Length=669

Score = 1207 bits (1338), Expect = 0.0
Identities = 669/669 (100%), Gaps = 0/669 (0%)
Strand=Plus/Plus

```
Query 1 CCTGCGGAAGGATCATTACCAAGAACCACACAGAAAAGGGAGGCAAAGGGGCCGACAGGG 60
      |||
Sbjct 1 CCTGCGGAAGGATCATTACCAAGAACCACACAGAAAAGGGAGGCAAAGGGGCCGACAGGG 60

Query 61 CTAGTAGCTTATACGTTGTTGAACGTCCTGCCTGGACCCGGAGCCGCCCCCATCTAAACC 120
      |||
Sbjct 61 CTAGTAGCTTATACGTTGTTGAACGTCCTGCCTGGACCCGGAGCCGCCCCCATCTAAACC 120

Query 121 CTCTGCGTACCTGTCCCTTCTTGCTTCCCCGGCATCTCGTCGGGGGGAGGTAACAACCA 180
      |||
Sbjct 121 CTCTGCGTACCTGTCCCTTCTTGCTTCCCCGGCATCTCGTCGGGGGGAGGTAACAACCA 180

Query 181 AAACCTCTCTGTGAATCAAACAGCCGTCAGAATTATAAAACAAACAAAAGTTAAAACCTTC 240
      |||
Sbjct 181 AAACCTCTCTGTGAATCAAACAGCCGTCAGAATTATAAAACAAACAAAAGTTAAAACCTTC 240

Query 241 AACCAACGGATCTCTTGGTCCCACATCGATGAAGAACGCAGCGAAATGCGATAAGTAATG 300
      |||
Sbjct 241 AACCAACGGATCTCTTGGTCCCACATCGATGAAGAACGCAGCGAAATGCGATAAGTAATG 300

Query 301 TGAATTGCAGAATTCAGTGAATCATCGAATCTTTGAACGCACATTGCGCCCCCTGGTATT 360
      |||
Sbjct 301 TGAATTGCAGAATTCAGTGAATCATCGAATCTTTGAACGCACATTGCGCCCCCTGGTATT 360

Query 361 CCGGGGGGCATGCCTGTTGAGCGTCATAAAACCTCCTCCCCCTTCGGGTTTTGTTACT 420
      |||
Sbjct 361 CCGGGGGGCATGCCTGTTGAGCGTCATAAAACCTCCTCCCCCTTCGGGTTTTGTTACT 420

Query 421 ATCGTTGGGGGTTTTGGCCAAATGGGATAGCGATTGGCAATTCGTTCCCAATGTCCTA 480
      |||
```

Identification

- Pb: mauvaise identification ~20% séquences
 - Paradoxalement moléculaire prend une ampleur mais besoin de taxonomiste pour créer base de données de référence fiable
-

Identification

- Descriptions carpophores = lien entre spécimen et mOTU
 - Entre 2000 et 2008: 11009 nouvelles espèces à l'index of fungi
 - Seulement 2511 (23%) référencées par un quelconque locus dans GenBank
 - Sur ~100 000 espèces morphologiques décrites, seulement 19848 (20%) dans GenBank
-

Abondance relative

- Se fait par l'utilisation d'indices de diversité et richesse spécifique (Chaos – Shannon)
- Basée sur la fréquence en nombre de séquences
 - Sur catégorie de séquence: singleton - duet



Conclusions

- ITS : pas le fragment parfait mais le meilleur disponible
 - 454 Roche: bcp de données mais pas toutes utilisables
-

Conclusions

- Taxonomie = 100 000esp
 - Via moléculaire: estimation : 1.5millions
 - Via NGS estimation 5millions
 - Moléculaire et taxonomique associés
-