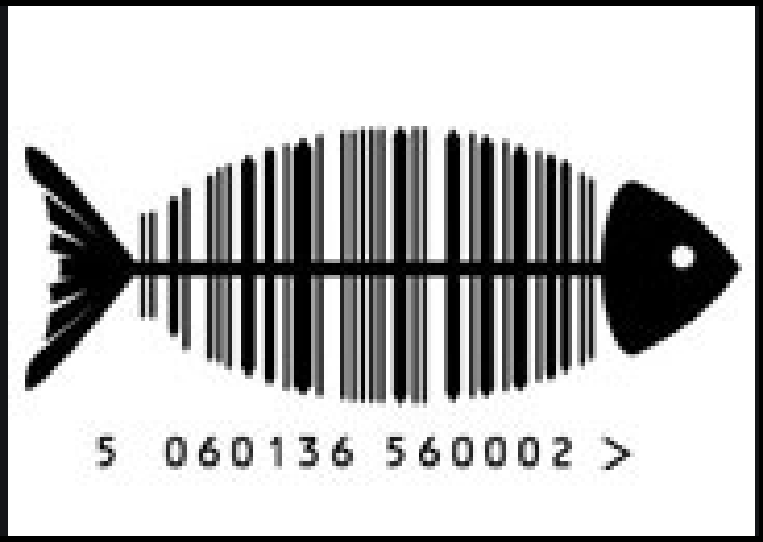
Séquençage génétique des poissons : comment et pourquoi ?



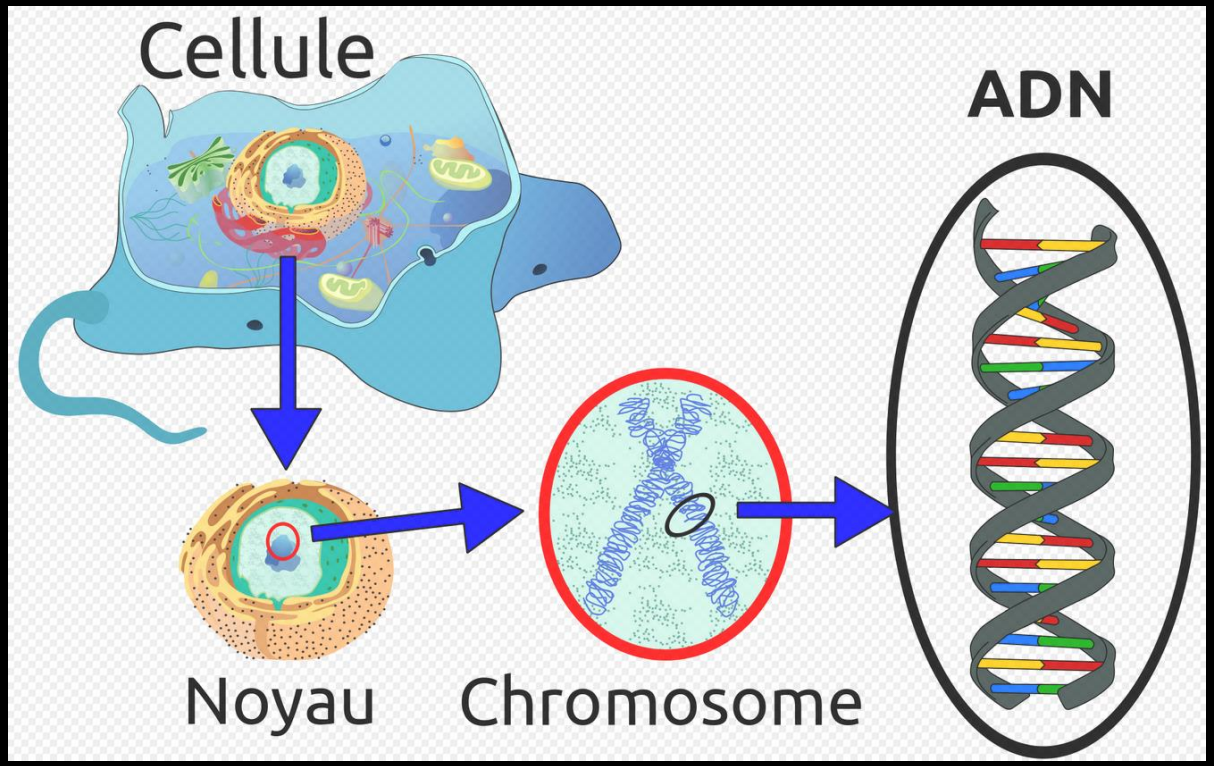
AQUARIUM TROPICAL

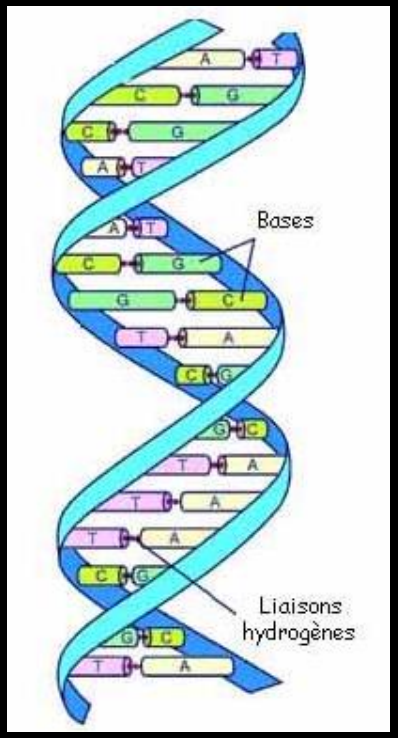
#FDS2019 fetedelascience.fr

arte 20 minutes SCIENCE&VIE TV JUNIOR SCIENCE&VIE JUNIOR SCIENCE&VIE DÉCOUVERTES THE CONVERSATION

MINISTÈRE DU L'ENSEIGNEMENT SUPÉRIEUR, DE LA RECHERCHE ET DE L'INNOVATION

Qu'est-ce qui identifie les êtres vivants ?





L'ADN est une molécule présente dans le noyau de toutes les cellules. C'est le support de l'information génétique qui varie selon l'enchainement des 4 nucléotides (A,T,C et G) = séquence caractéristique d'une espèce.

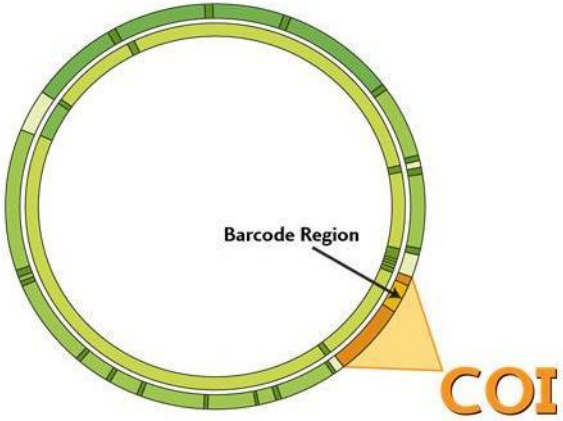
Quel est le principe du barcoding ?

- Détermination séquence d'ADN d'un individu
- Identification par comparaison de la séquence dans une bibliothèque (BOLD)

Si les individus sont de même espèce, on s'attend à trouver des codes-barres (séquence d'ADN) très proches.
Si les individus ne sont pas de la même espèce, on obtiendra des séquences d'ADN très différentes.

Quel gène est utilisé dans le barcoding ?

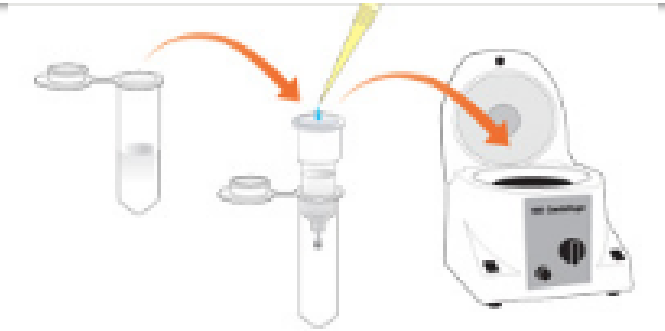
Il faut cibler un gène présent chez tous les poissons : celui de la cytochrome c oxydase qui permet la respiration cellulaire



Gène COI (cytochrome C oxydase)

Etape 1 : Extraction de l'ADN des échantillons prélevés

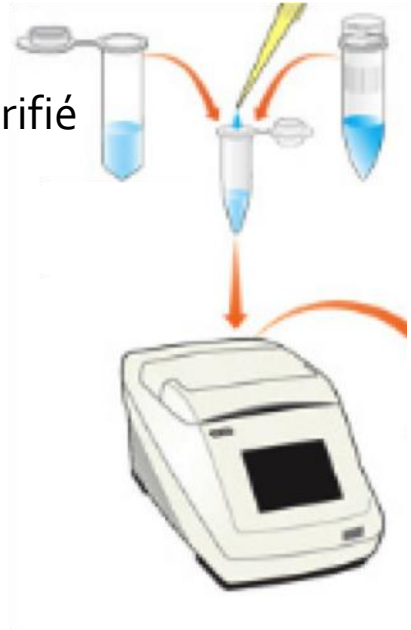
Prélèvement d'un fragment de tissu



Préparation des échantillons par l'utilisation de produits d'extraction puis purification par centrifugation

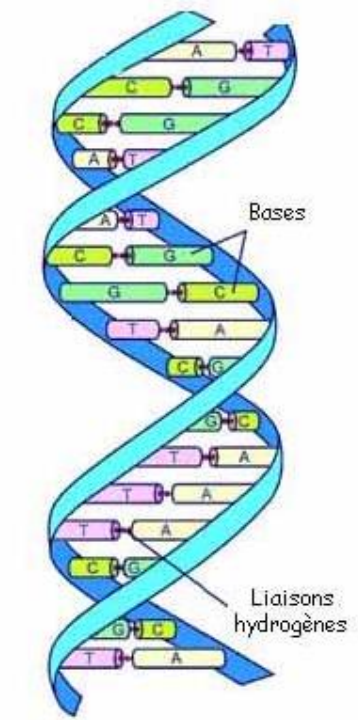
Etape 2 : Amplification des fragments d'ADN extraits par PCR

ADN purifié



ADN polymérase thermorésistante

Thermocycleurs

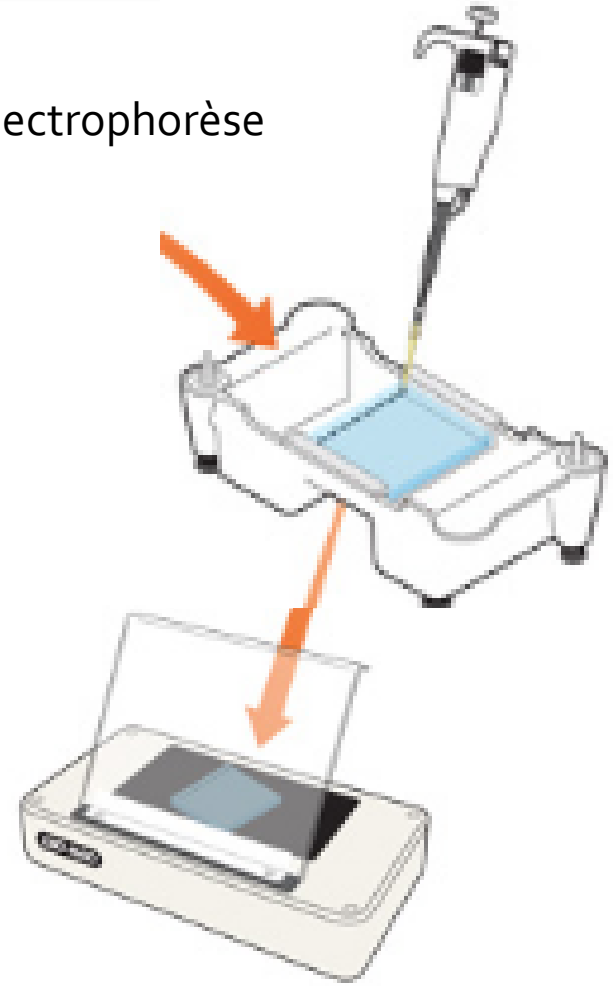


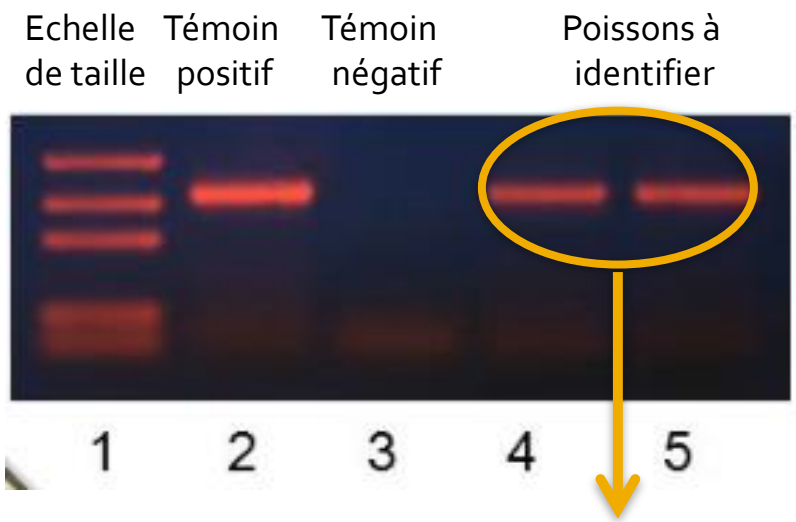
Objectif : Copier l'ADN pour avoir suffisamment de fragments d'ADN pour pouvoir les séquencer.

2 x plus d'ADN à chaque cycle, 30 à 50 cycles => 1100 milliards x nombre de copies initiales

Etape 3 : Vérification de l'amplification du gène COI par électrophorèse

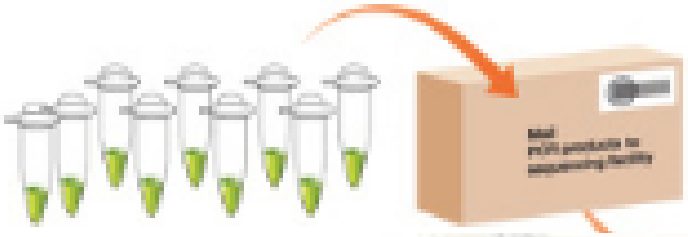
Electrophorèse



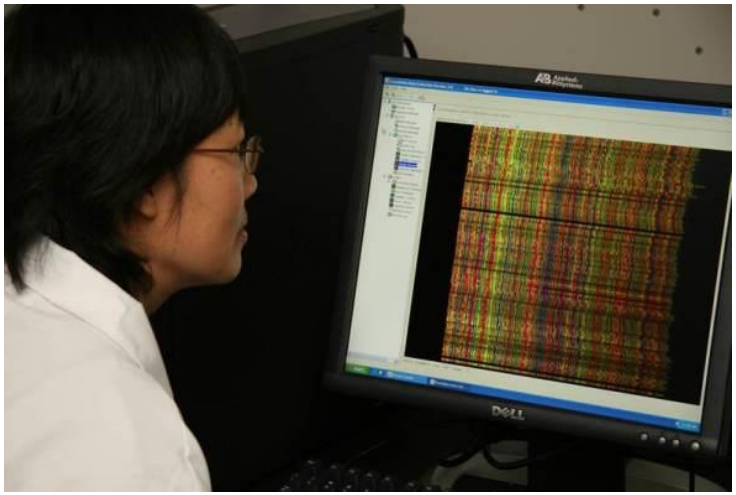


Gène COI localisé
= amplification réussie

Etape 3 : Séquençage du fragment d'ADN correspondant au code-barres



Fragments D'ADN d'intérêt

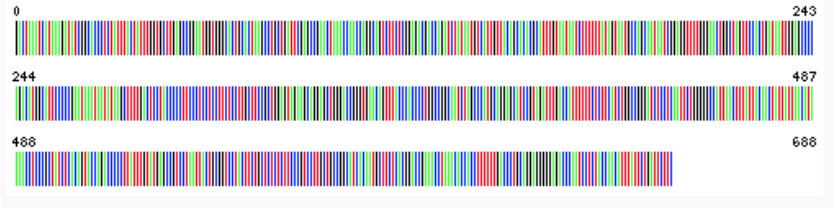


Séquençage des fragments d'ADN afin de déterminer leur séquence en nucléotides

Etape 4 : Identification du poisson inconnu

Espèce inconnue

GCACCCTCTATTTACTATTTGGTGCTTGAGCCGGAATGGTGGGCACTGCA
CTAAGCCTGCTGATTCGAGCCGAACTAAACCAACCAGGCAGTCTTCTCGG
AGACGACCAGATCTATAATGTCATTGTTACAGCACACGCATTTGTAATAA
TCTTTTTTATAGTAATGCCAGCAATAATCGGAGGTTTTGGGAACTGGCTA
GTTCCACTGATAATTGGAGCCCCAGACATGGCATTCCCCCGAATAAATAA
TATAAGCTTTTGACTCCTACCCCCTTCCCTCCTCCTTCTCCTTGCTTCCGCT
GGAGTAGAAGCAGGAGCTGGGACCGGGTGAACAGTTTACCCACCCCTGG
CCGGCAATCTAGCCCATGCAGGAGCATCCGTAGATTTGACCATTTTTTCT
CTTCATCTGGCCGGGATCTCCTCAATTCTGGGGGCCATTAACTTTATTAC
AACTATTATTAACATAAAACCTCCCGCTATCTCACAGTACCAGACCCCCTT
ATTTGTATGAGCCGTCCTAATTACTGCGGTCCTACTTCTGCTCTCTCTTCC
CGTCCTTGCTGCCGGCATCACAATGCTTCTCACGGACCGAAACCTAAACA
CCACCTTTTTTGACCCTGCAGGAGGGGGAGACCCAATCCTCTACCAACAT
TTATTC

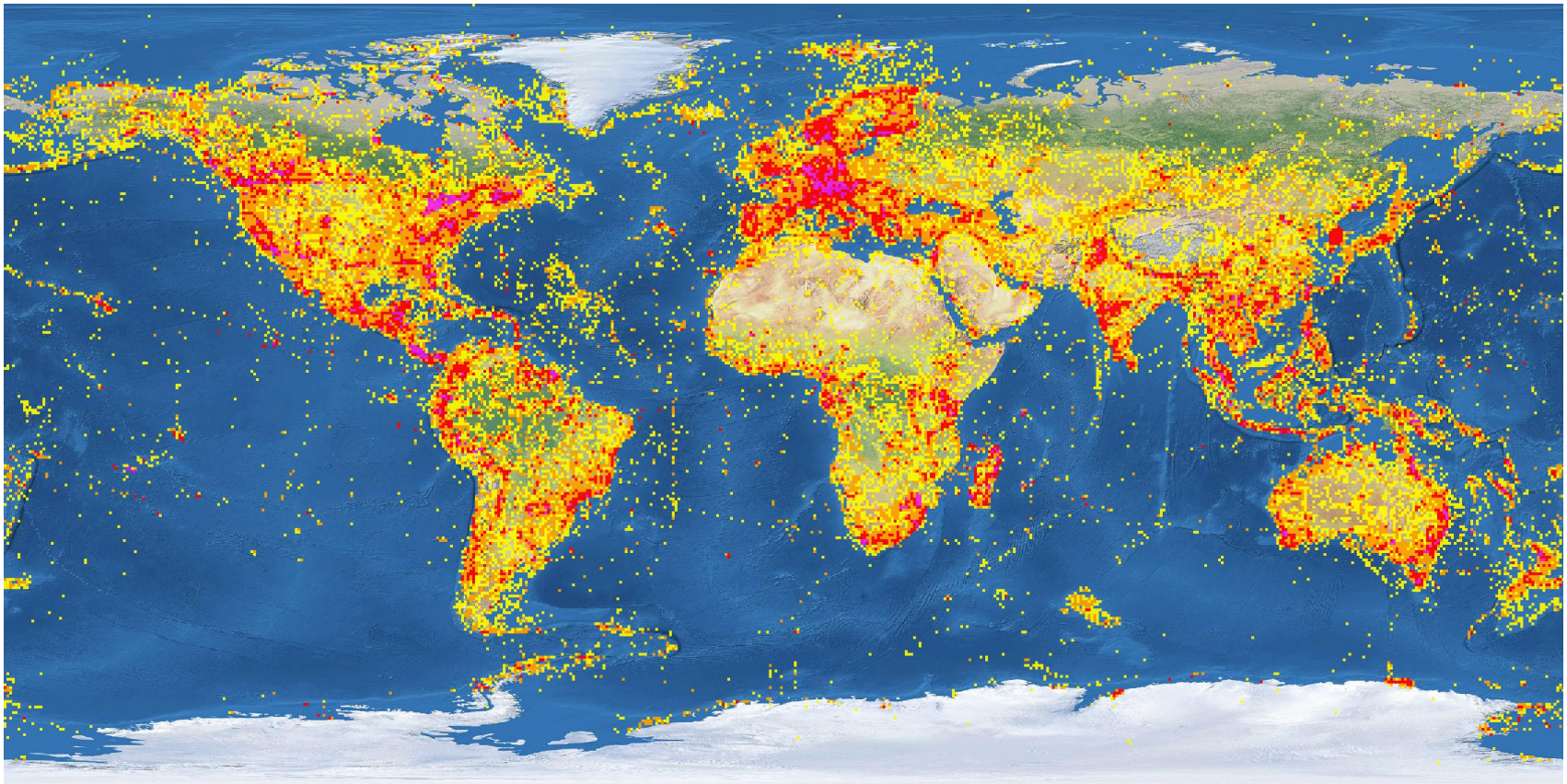


Code-barres du fragment d'ADN de l'espèce inconnue

Pour identifier le poisson, on compare la séquence obtenue avec l'ensemble des données de séquences stockées dans la base BOLD

BARCODE OF LIFE DATA SYSTEM v4

Advancing biodiversity science through DNA-based species identification.



On colle la séquence obtenue et on lance la recherche dans la base BOLD

BOLD SYSTEMS

DATABASES IDENTIFICATION TAXONOMY WORKBENCH RESOURCES LOGIN

IDENTIFICATION ENGINE

ANIMAL IDENTIFICATION [COI] | FUNGAL IDENTIFICATION [ITS] | PLANT IDENTIFICATION [RBCL & MATK]

The BOLD Identification System (IDS) for COI accepts sequences from the 5' region of the mitochondrial Cytochrome c oxidase subunit I gene and returns a species-level identification when one is possible. Further validation with independent genetic markers will be desirable in some forensic applications.

Historical Databases: Current Jul-2019 Jul-2018 Jul-2017 Jul-2016 Jul-2015 Jul-2014 Jul-2013 Jul-2012 Jul-2011 Jul-2010 Jul-2009

Search Databases:

- ○ **All Barcode Records on BOLD (6,676,722 Sequences)**
  Every COI barcode record on BOLD with a minimum sequence length of 500bp (warning: unvalidated library and includes records without species level identification). This includes many species represented by only one or two specimens as well as all species with interim taxonomy. This search only returns a list of the nearest matches and does not provide a probability of placement to a taxon.

- ◉ **Species Level Barcode Records (3,535,614 Sequences/209,838 Species/90,891 Interim Species)**
  Every COI barcode record with a species level identification and a minimum sequence length of 500bp. This includes many species represented by only one or two specimens as well as all species with interim taxonomy.

- ○ **Public Record Barcode Database (1,424,064 Sequences/118,485 Species/37,444 Interim Species)**
  All published COI records from BOLD and GenBank with a minimum sequence length of 500bp. This library is a collection of records from the published projects section of BOLD.

- ○ **Full Length Record Barcode Database (2,259,620 Sequences/188,132 Species/74,010 Interim Species)**
  Subset of the Species library with a minimum sequence length of 640bp and containing both public and private records. This library is intended for short sequence identification as it provides maximum overlap with short reads from the barcode region of COI.

Résultat de la recherche dans la base de données BOLD

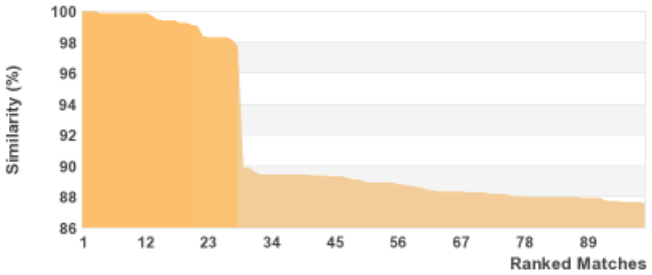
Results Summary Download

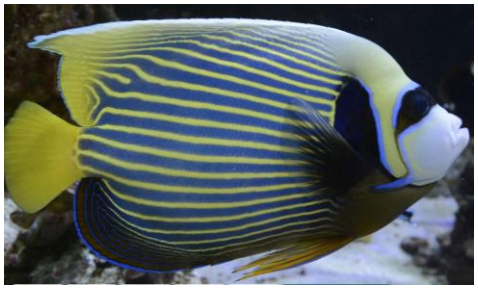
Query ID	Best ID	Search DB	Tree	Top %	Graph	Low %
unlabeled_sequence	*Pomacanthus imperator*	COI SPECIES DATABASE		100.00		87.54

Identification Summary

Taxonomic Level	Taxon Assignment	Probability of Placement (%)
Phylum	Chordata	100
Class	Actinopterygii	100
Order	Perciformes	100
Family	Pomacanthidae	100
Genus	*Pomacanthus*	100
Species	*Pomacanthus imperator*	100

Similarity Scores of Top 99 Matches





=

?

Résultat de la recherche dans la base de données BOLD

Pomacanthus imperator de 97,73 à 100 % d'homologie

Pomacanthus annularis à 89,88 % d'homologie au maximum

Intérêt du barcoding pour les poissons

- Participer au Projet international de code-barres du vivant (International Barcode of Life [iBOL]) = étoffer la base de données (plus de 7 millions de séquences pour ~300 000 espèces)
- Suivre l'évolution de la biodiversité
- Contrôles de fraudes:

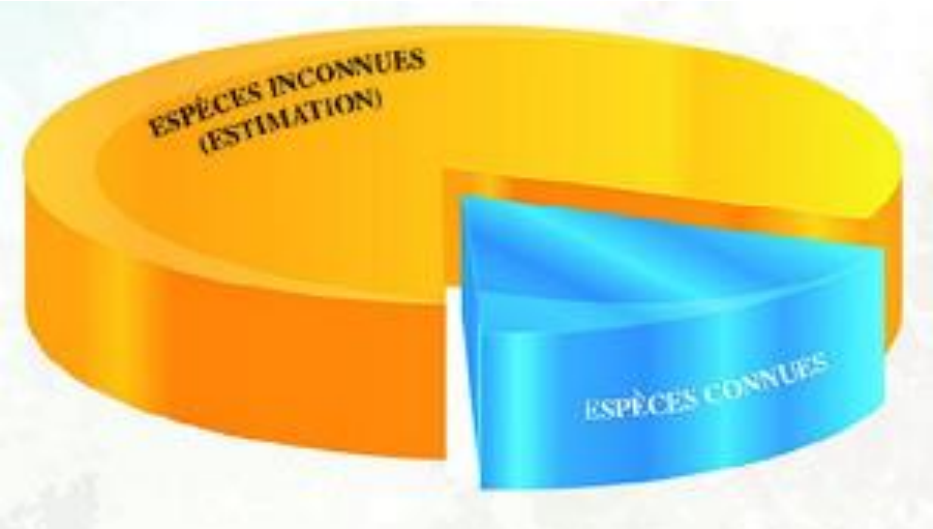
Thon obèse vendu comme de l'albacore

Thon albacore vendu comme du rouge

Et pourtant, cette biodiversité est encore sous estimée

- Les méthodes classiques ont permis de décrire 1,9 million d'espèces

- Selon les estimations, le nombre total d'espèces varie entre 10 et 100 millions



Ce diagramme à secteurs représente la proportion entre les espèces connues et une estimation des espèces inconnues. Ici, l'infographie représente les groupes d'êtres vivants connus.