

**Objectif :** Construire un arbre de parentés (phylogénétique) des 8 espèces d'ursidés actuels. Le panda (*Ailuropoda melanoleuca*) servira d'extra-groupe.

## 1 Récupération des séquences des gènes dans GenBank <http://www.ncbi.nlm.nih.gov/genbank/>

Le choix s'est porté sur le gène mitochondrial : Cox1 (cyclo-oxydase) pour les 8 espèces.

The screenshot shows the NCBI Gene database search results for the gene symbol 'cox1'. The search results page includes a search bar with 'cox1' entered, a 'Search' button, and a 'Create alert' link. Below the search bar, there is a message: 'Did you mean cox1 as a gene symbol? Search Gene for cox1 as a symbol.' The search results are displayed in a table with columns for Name/Gene ID, Description, Location, and Aliases. The first result is for COX1 in *Saccharomyces cerevisiae*, and the second is for COX1 in *Drosophila melanogaster*. The page also shows a 'Results by taxon' section with a list of top organisms and a 'Find related data' section.

Name/Gene ID	Description	Location	Aliases
COX1 ID: 854598	cytochrome c oxidase subunit 1 [ <i>Saccharomyces cerevisiae</i> S288c]	Chromosome MT, NC_001224.1 (13818..26701)	Q0045, OX13
COX1 ID: 10803533	cytochrome c oxidase subunit I [ <i>Drosophila melanogaster</i> (fruit fly)]	Chromosome MT, NC_024511.2 (1474..2000)	

- Choisir « Gene » dans la « DataBase » et taper cox → 8072 résultats !!!  
Cliquez sur « Advanced »
- Taper « cox1 » et « Ursus » dans les 2 champs → 14 résultats !!!

The screenshot shows the NCBI Gene database search builder interface. The search query is '(cox1) AND ursus'. The interface includes a search bar with the query, an 'Edit' link, and a 'Builder' section with three rows of search criteria. The first row is 'All Fields' with 'cox1' entered. The second row is 'AND' with 'All Fields' and 'ursus' entered. The third row is 'AND' with 'All Fields' and an empty field. Below the builder section, there is a 'Search' button and a link to 'Add to history'.

- Choisir une espèce en cliquant sur le gène

## COX1 cytochrome c oxidase subunit I [ *Ursus spelaeus* (cave bear) ]

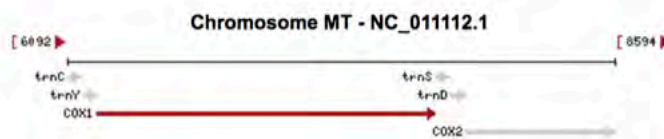
Gene ID: 6740794, updated on 26-Jun-2015

### Summary

<b>Gene symbol</b>	COX1
<b>Gene description</b>	cytochrome c oxidase subunit I
<b>Gene type</b>	protein coding
<b>RefSeq status</b>	REVIEWED
<b>Organism</b>	<a href="#">Ursus spelaeus</a>
<b>Lineage</b>	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Laurasiatheria; Carnivora; Caniformia; Ursidae; Ursus

### Genomic context

Sequence: Chromosome: MT; NC\_011112.1 (6227..7771)



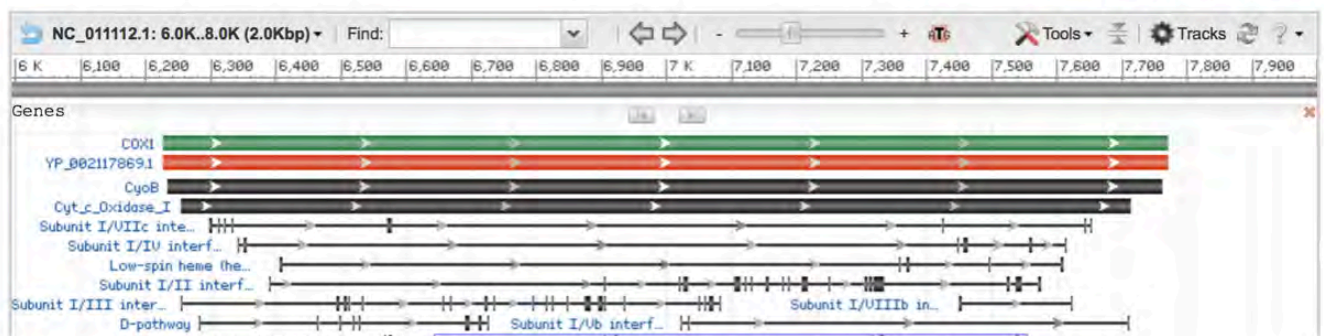
- Cliquer sur « FASTA »

### Genomic regions, transcripts, and products

Genomic Sequence: NC\_011112.1

[Go to reference sequence details](#)

[Go to nucleotide: Graphics FASTA GenBank](#)



2

## Création des fichiers de séquences des gènes pour ANAGENE et PHYLOGENE (cox1\_ana.adn et cox1\_phylo.adn)

- Surligner - copier puis coller la séquence dans un éditeur de texte type TextEdit ou WorPad/notePad en respectant le formatage des séquences.

**Pour Anagène :** Enregistrer avec l'extension .adn

Ouvrir le fichier joint pour le modèle du formatage ou voir le site acces.

<http://acces.ens-lyon.fr/evolution/logiciels/anagene/produire-un-fichier.edi/fichiers-pour-anagene/>

**Pour Phylogène :** Enregistrer avec l'extension .adn.

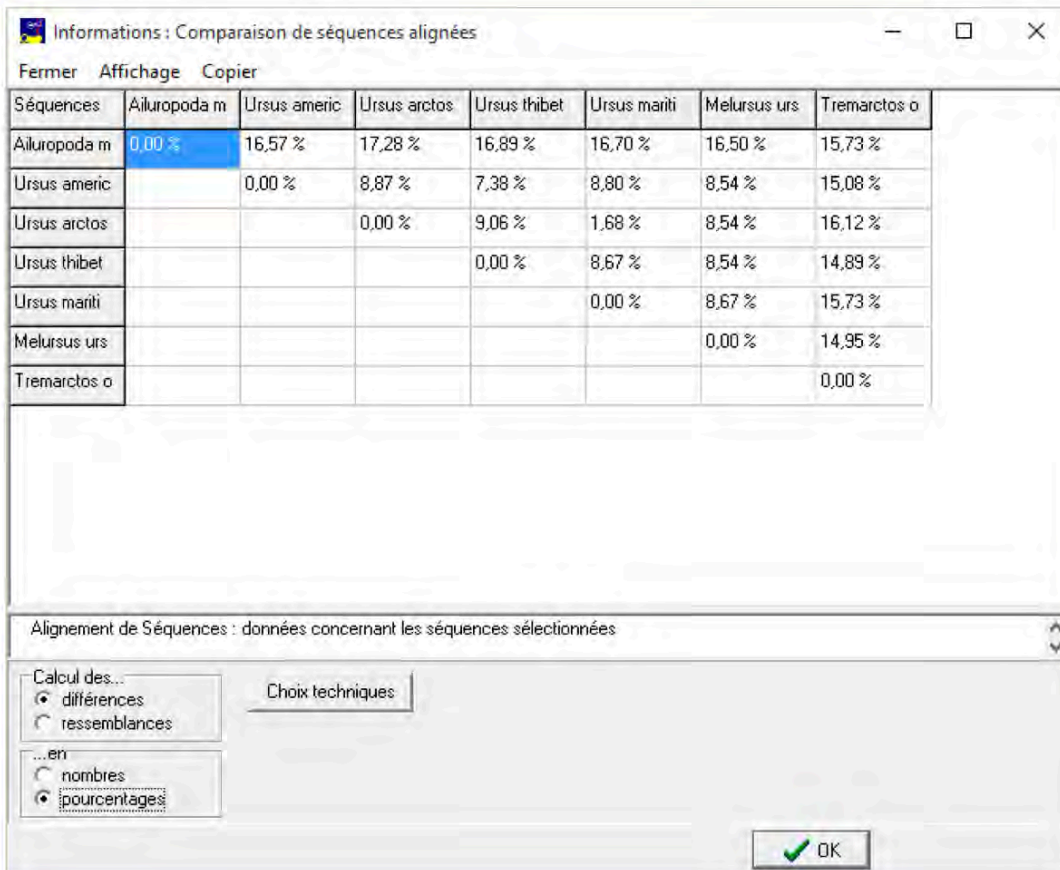
Ouvrir le fichier joint pour le modèle du formatage ou voir le site : [http://svt.tice.ac-orleans-tours.fr/php5/publis/phylogene/nouv\\_donn\\_phylo/html/ajouter.htm](http://svt.tice.ac-orleans-tours.fr/php5/publis/phylogene/nouv_donn_phylo/html/ajouter.htm)

**Vous avez maintenant 2 fichiers .adn différents pour anagène et pour phylogène**

### 3 Construction d'une matrice des distances

Comparer les séquences et construire une matrice des distances : matrice du nombre de différences ou de % de différences. (Comparaison simple)

- Ouvrir le fichier `cox1_ana.adn` avec Genieen
- Sélectionner toutes les séquences et remonter *Ailuropoda melanoleuca* (panda) en 1ère position (Elle sera la séquence de référence).
- Aligned les séquences : Menu Action > Alignement
- Ouvrir le menu « Informations » (Choix : Calcul des différences - en pourcentages)



Séquences	Ailuropoda m	Ursus americ	Ursus arctos	Ursus thibet	Ursus mariti	Melursus urs	Tremarctos o
Ailuropoda m	0,00 %	16,57 %	17,28 %	16,89 %	16,70 %	16,50 %	15,73 %
Ursus americ		0,00 %	8,87 %	7,38 %	8,80 %	8,54 %	15,08 %
Ursus arctos			0,00 %	9,06 %	1,68 %	8,54 %	16,12 %
Ursus thibet				0,00 %	8,67 %	8,54 %	14,89 %
Ursus mariti					0,00 %	8,67 %	15,73 %
Melursus urs						0,00 %	14,95 %
Tremarctos o							0,00 %

Alignement de Séquences : données concernant les séquences sélectionnées

Calcul des...  
 différences  
 ressemblances

Choix techniques

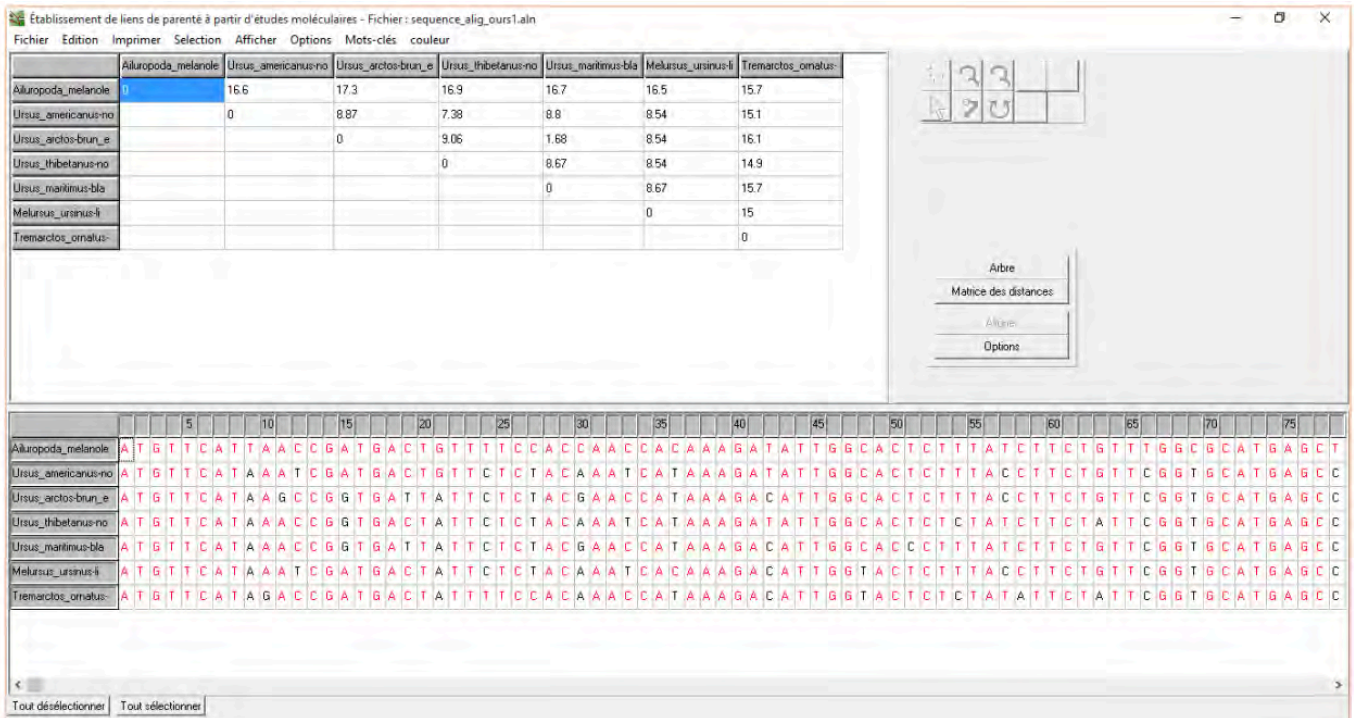
...en  
 nombres  
 pourcentages

OK

- Dessinez une ébauche d'arbre de parentés à partir des résultats en vous posant la question : Qui est le prés de qui ? (Moins de % de différences).  
Exemples de questions à vous poser :  
Quels sont les 2 espèces les plus proches ? *U.maritimus* et *U.arctos*  
L'espèce la plus éloignée des autres ? *Ailuropoda*. Réfléchir à la notion d'espèce extra-groupe  
L'espèce la plus proche de *maritimus* et *arctos* ? *U. melursus*  
2 autres espèces très proches ? *U. americanus* et *U. thibetanus*
- Enregistrez les séquences alignées pour les ouvrir plus tard dans phylogène :  
Sélectionnez la fenêtre « Alignement de séquences ».  
Menu Fichier > Enregistrez sous ...Type « phylogène » (`aln`)

## Construction d'un arbre des parentés (phylogénétique)

- Ouvrir le fichier des séquences alignées (.aln) avec phylogène :  
Menu Fichier > Ouvrir > Fichier de molécules
- Afficher la matrice des distances :
  - Bouton Options - onglet Distances : Pourcentage)
  - Sélectionnez les séquences
  - Menu Afficher > Matrice des distances



- Comparez-la avec votre ébauche à partir de la matrice de Geniegen.
- Afficher arbre :  
Bouton Options – Onglet Arbres – Choisir UPGMA  
Menu Afficher > Arbre

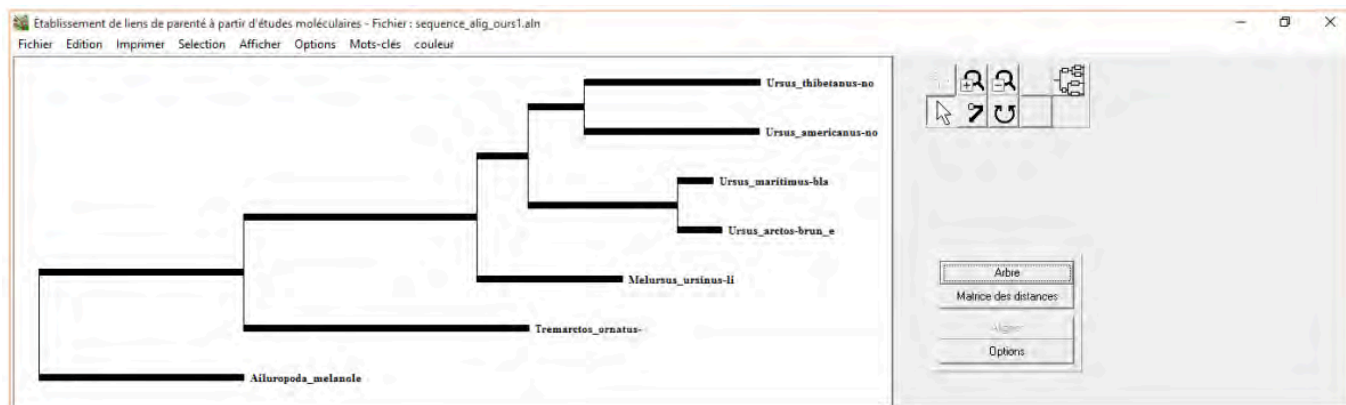


- Comparez cet arbre avec celui que vous avez construit précédemment.



**La méthode UPGMA** (*Unweighted Pair Group Method with Arithmetic Mean*) est une méthode dite de distance, c'est-à-dire une méthode basée sur les similarités entre paires de séquences. Elle est délaissée au profit de sa cousine (NJ) qui est plus adaptée aux études phylogéniques moléculaires.

- Modifier le type d'arbre : Menu Options > Onglet Arbres > Choisir NJ



**La méthode du Neighbour Joining (*Neighbor Joining*- NJ)** est aussi une méthode de distance.

Elle a l'avantage d'être rapide. En général, elle est utilisée pour faire des arbres de plusieurs milliers de séquences.

#### **D'autres méthodes :**

la méthode du maximum de vraisemblance (Maximum Likelihood- ML)

la méthode du maximum de parcimonie (Maximum Parsimony)

Méthodes basées sur l'apprentissage artificiel (via un réseau de neurones ou encore de l'inférence bayésienne).

5

## **Construction d'un arbre phylogénique en ligne**

Pour construire un arbre phylogénétique, copier les différentes séquences récupérées sur le site de ncbi (<http://www.ncbi.nlm.nih.gov>) et les coller dans le programme de création d'un arbre phylogénétique du site Web MABL ([www.phylogeny.fr](http://www.phylogeny.fr)).

Les séquences sont regroupées dans le fichier texte : phylogeny\_sequences.txt mais on peut demander aux élèves de créer un fichier identique par copier-coller dans un traitement de texte des différentes séquences des différentes espèces puis de copier-coller le résultat dans le site phylogeny.fr

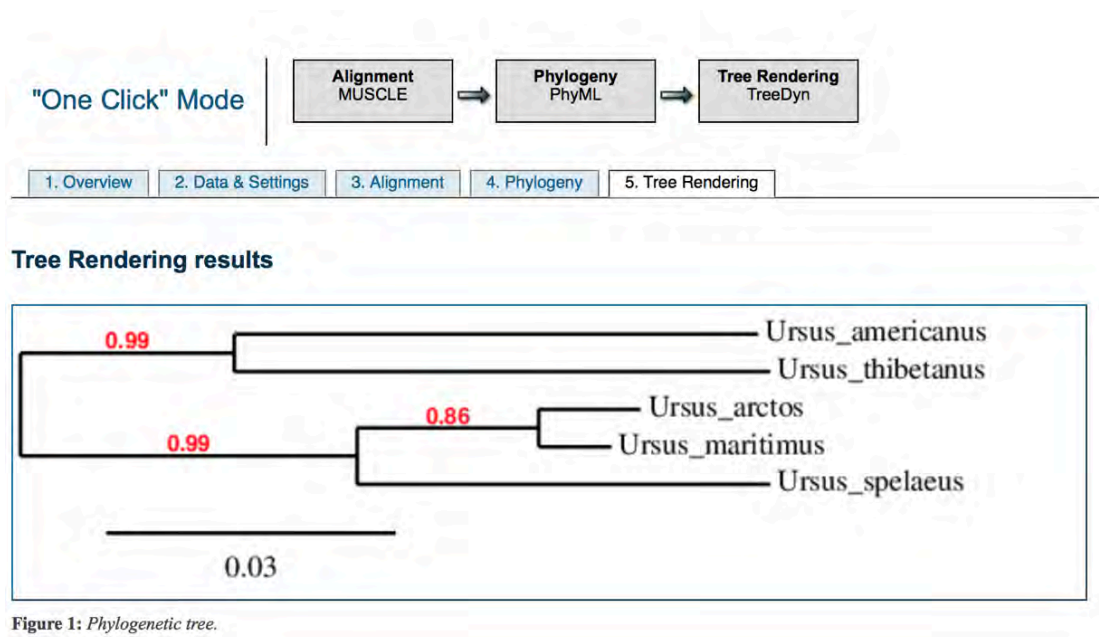
Ce site permet de suivre les différentes étapes de la construction :

**Alignement des séquences - Comparaison (Module Muscle) → Création de l'arbre (Module PhyML) → Tracé de l'arbre (Module TreeDyn)**

- Connectez vous sur : [www.phylogeny.fr](http://www.phylogeny.fr)

Décocher « Blocks Program »

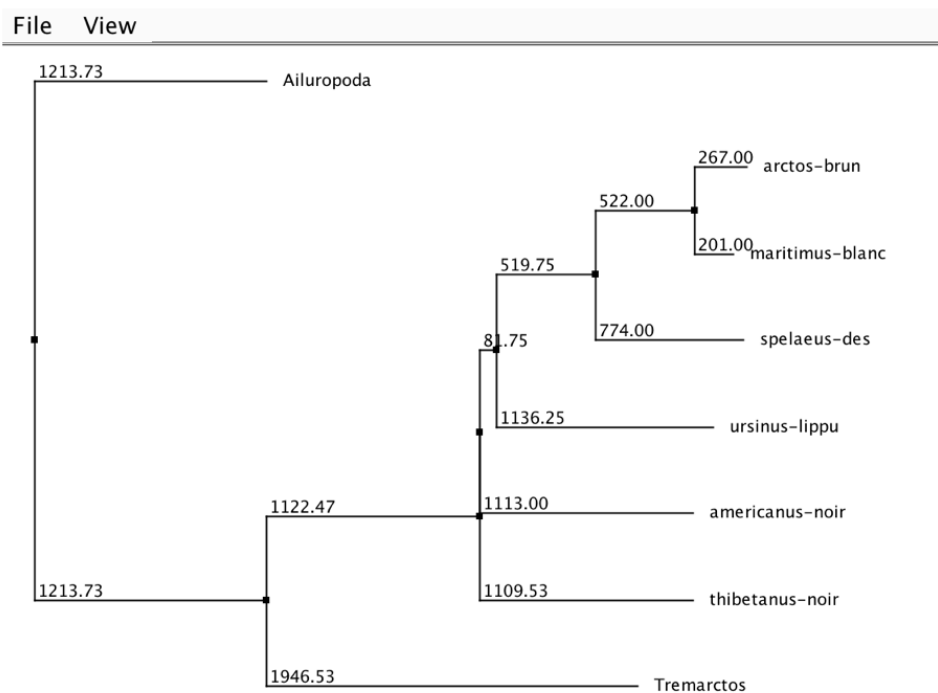
Cliquer sur bouton « Submit » et suivre les différentes étapes



## 6 D'autres outils

### ● Construction des arbres avec Jalview

#### Arbre NJ





4. Enregistrer : Data > Save session (format de fichier : .mas)

5. Alignement des séquences : soit avec ClustalW soit MUSCLE (Sélectionner les séquences : Menu Alignment > Align by Muscle)

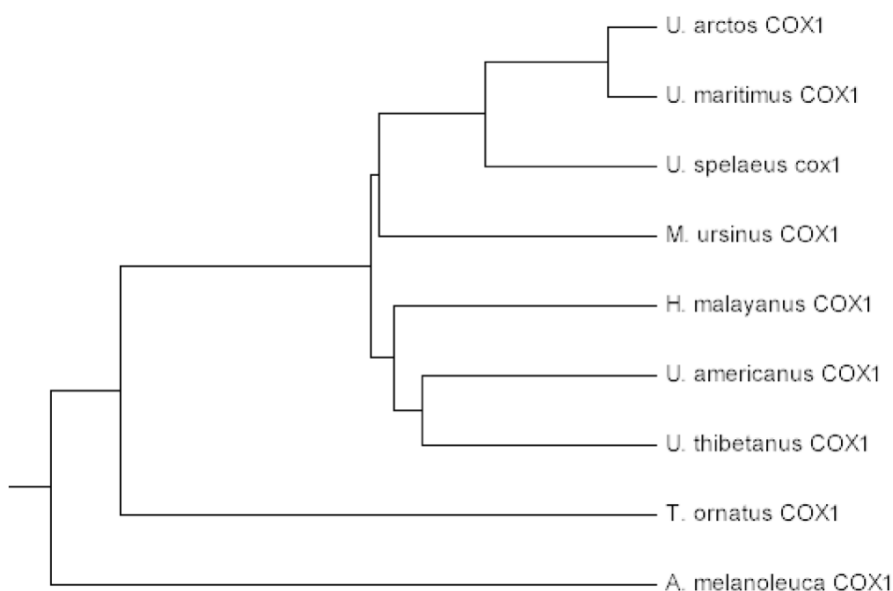
6. Exporter les séquences alignées : Menu Data > Export Alignment > MEGA Format (Remarque : possibilité d'exporter en Format Fasta réutilisable avec Phylogène)

7. Tracer les arbres phylogénétiques

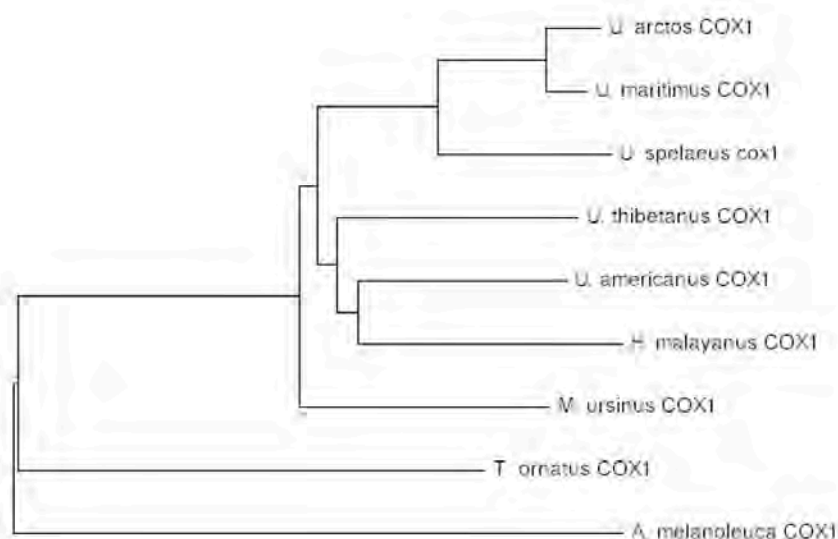
Dans la fenêtre principale de MEGA : Menu Phylogeny > Construct/Test ... Choisir une des méthodes (UPGMA tree - Neighbor Joining tree - Maximum Parsimony ....

Avantage de MEGA : Permet de tracer des arbres par les méthodes des distances (UPGMA et NJ), par la méthode du maximum de vraisemblance (*Maximum Likelihood*- ML) et la méthode du maximum de parcimonie (*Maximum Parsimony*).

#### Arbre UPGMA



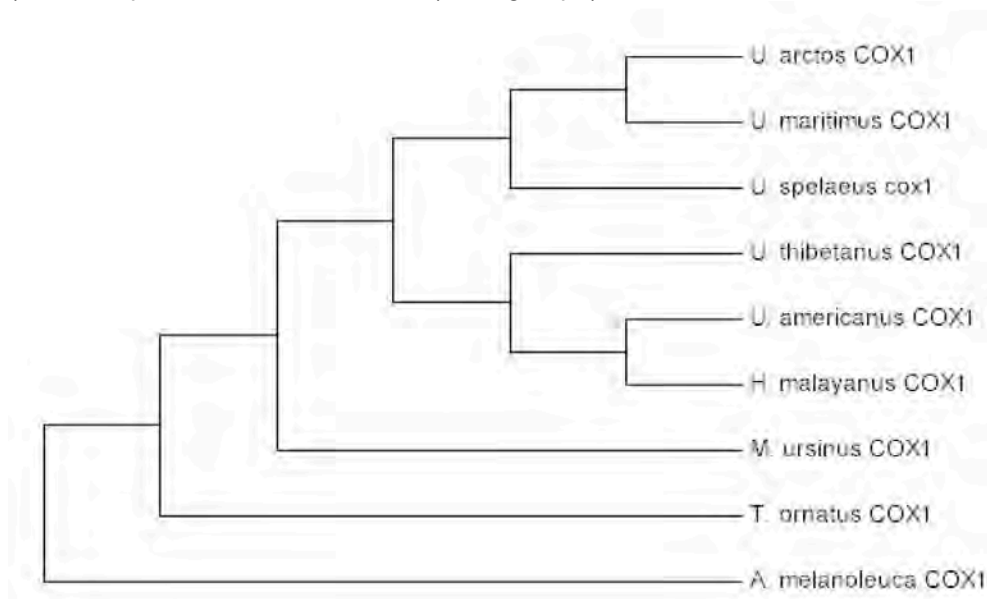
#### Arbre Neighbor Joining





### Arbre du maximum de parcimonie

(Choix du panda comme « racine » (extra groupe))



### ● LIENS

<http://www2.unil.ch/phylo/course/phylogenie-des-plantes>

<http://bioinfo-fr.net/les-arbres-phylogoniques-construction-et-interpretation?s=arbre>

<http://evolution.genetics.washington.edu/phylip/software.html> <http://bioinfo-fr.net/alignements-multiples-quels-logiciels-choisir> <http://doua.prabi.fr/software/seaview>

<https://tel.archives-ouvertes.fr/tel-00711689/document>

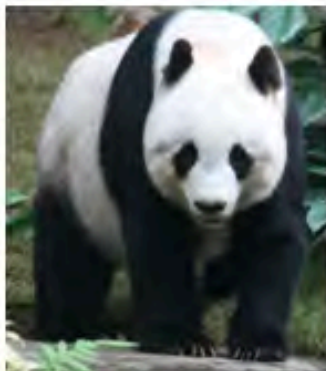
[http://www.ncbi.nlm.nih.gov/data\\_specs/NCBI\\_data\\_conversion.html](http://www.ncbi.nlm.nih.gov/data_specs/NCBI_data_conversion.html)

[http://sequenceconversion.bugaco.com/converter/biology/sequences/ fasta\\_to\\_nexus.php](http://sequenceconversion.bugaco.com/converter/biology/sequences/ fasta_to_nexus.php)

[http://desdevises.free.fr/Adaphyl/Adaptation\\_et\\_Phylogenie/Logiciels.html](http://desdevises.free.fr/Adaphyl/Adaptation_et_Phylogenie/Logiciels.html)

## ANNEXE 1 : LES URSIDES

*Ailuropoda melanoleuca*  
(Panda géant)



*Ursus arctos* (ours brun)  
(ours brun)



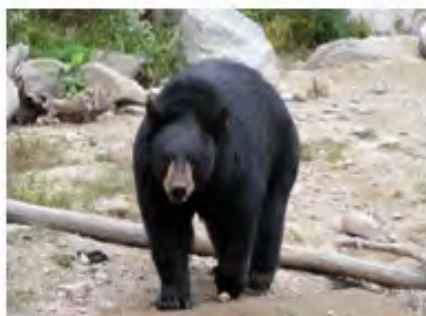
*Ursus maritimus*  
(Ours blanc)



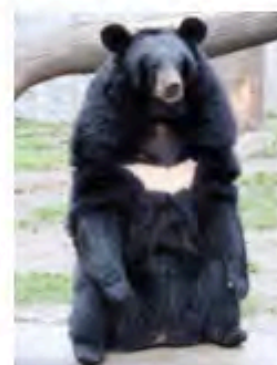
*Melursus ursinus*  
(Ours lippu)



*Ursus americanus*  
(Ours noir)



*Ursus thibetanus*  
(Ours noir d'Asie)



*Tremarctos ornatus*  
(Ours à lunettes)



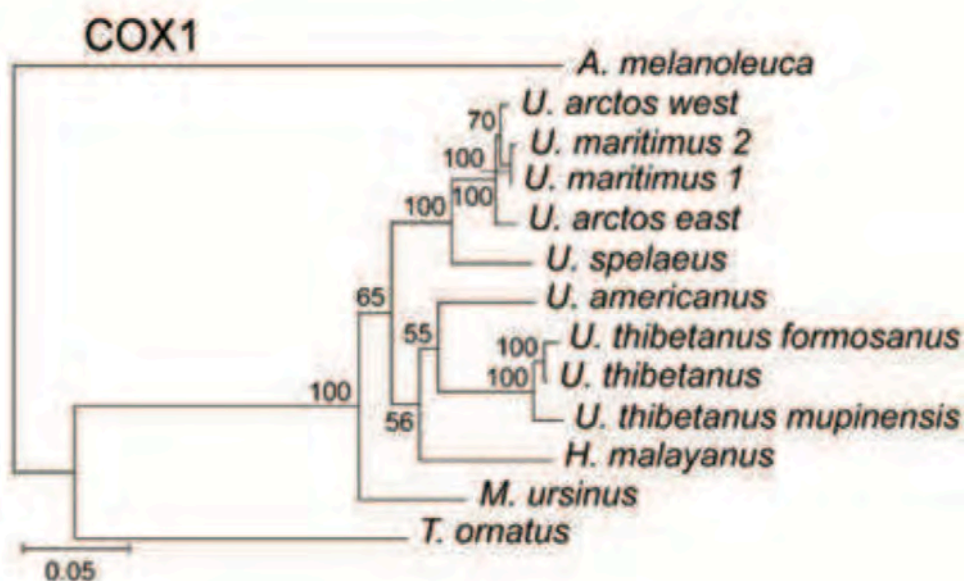
*Ursus malayanus*  
(Ours malais)



*Ursus spelaeus*  
(Ours des cavernes)

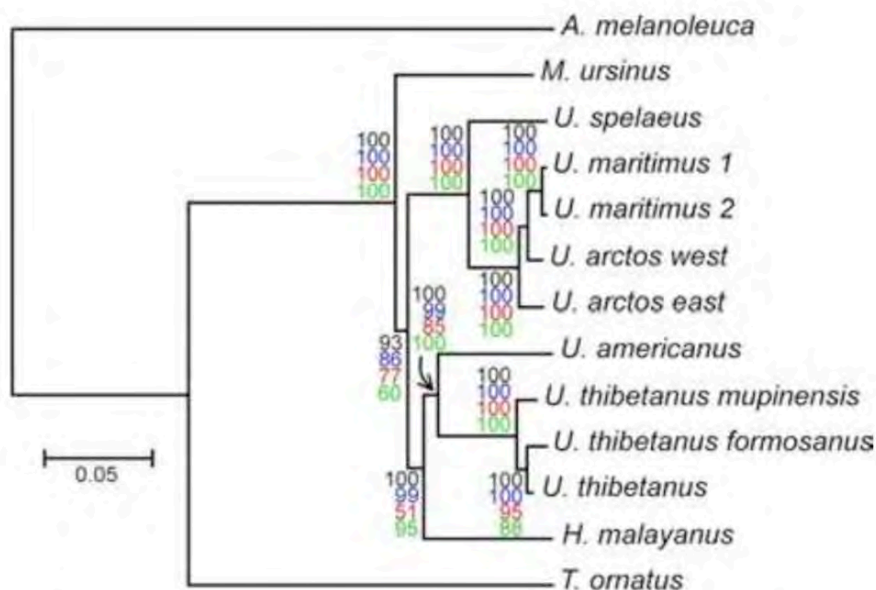


<https://tel.archives-ouvertes.fr/tel-00711689/document>



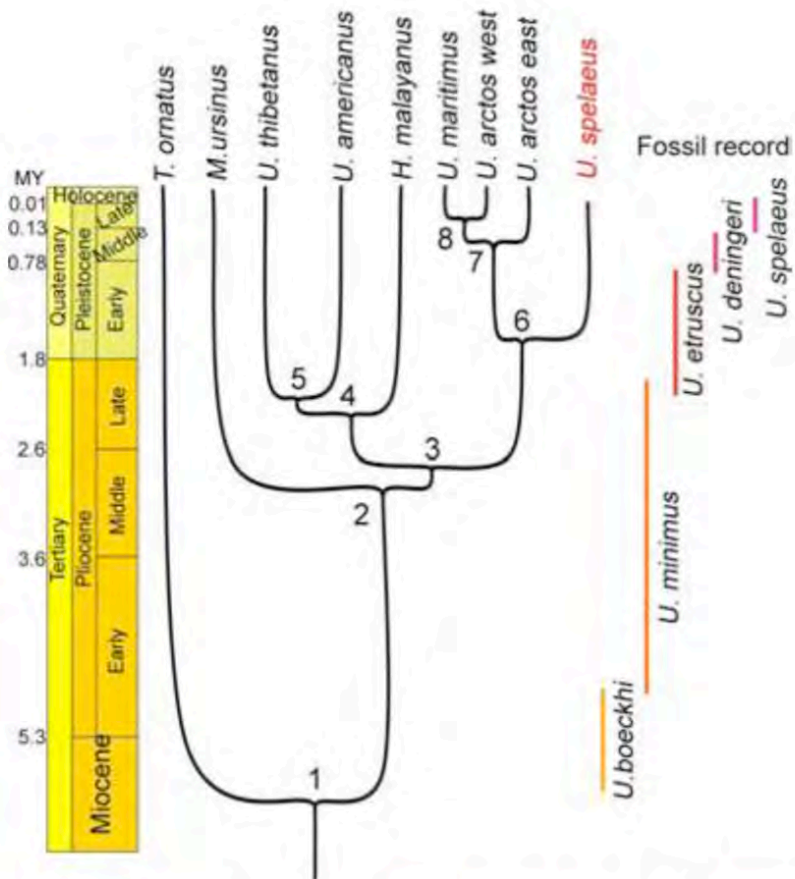
Ailuropoda Melanoleuca : extra-groupe

Logiciel PhyML : méthode du maximum de vraisemblance (ML)

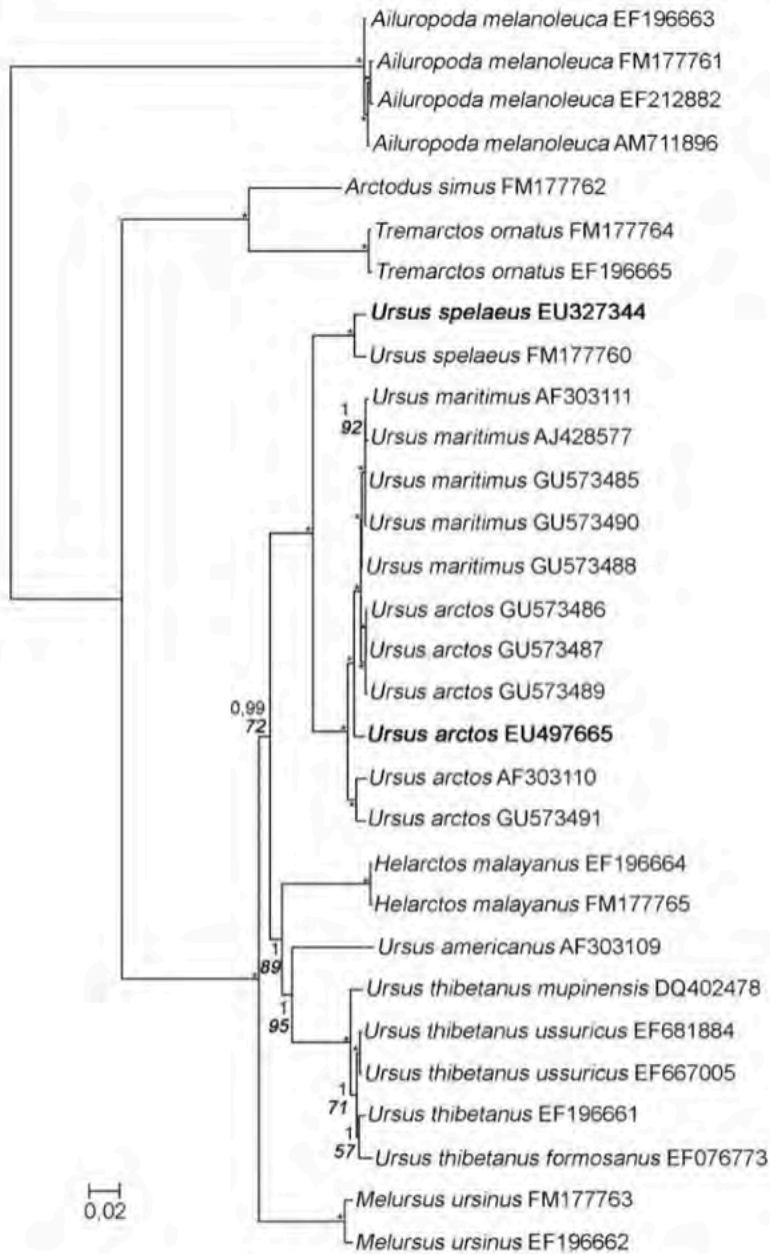


**Fig. 4.** Molecular phylogeny inferred from complete mitochondrial genome sequences. Tree construction was performed by MrBayes analysis, using the giant panda (*Ailuropoda melanoleuca*) as an outgroup. The posterior probability value ( $\times 100$ ) of each node is indicated in black, and the scale for genetic distance is shown at the bottom of the figure. The same tree topology was obtained using three other methods, and bootstrap values are indicated with colored characters for PhyML (blue), maximum parsimony (red), and neighbor joining (green) analysis. The *Ursus spealeus* and *Ursus arctos* western lineage (west) sequences are from this study. GenBank accession numbers (from top to bottom) for the other sequences are as follows: EF196663, EF196662, AJ428577, AF303111, AF303110, AF303109, DQ402478, EF076773, EF196661, EF196664, and EF196665.





**Fig. 5.** Phylogeny and divergence times determined using the mitochondrial genome sequence of the cave bear and of eight extant bears. Divergence times were calculated using BEAST software with the splits between the giant panda and *Ursidae* and between *Ursinae* and *Tremarctinidae* set to 12 and 6 million years (MY), respectively. Age for each node and 95% credibility intervals are as follows: 1, 6.3 MY (5.4–7.2); 2, 3.0 MY (2.2–3.8); 3, 2.8 MY (2.1–3.5); 4, 2.4 MY (1.7–3); 5, 2.1 MY (1.4–2.7); 6, 1.6 MY (1–2.1); 7, 0.6 MY (0.3–0.8); and 8, 0.4 MY (0.2–0.5).



**Figure 26 :** Arbre phylogénétique utilisant les génomes mitochondriaux complets d'Ursidés actuels et éteints, enraciné avec les séquences de génomes mitochondriaux de panda. Les valeurs de probabilité a posteriori et de bootstrap maximales sont indiquées par une astérisque. Les caractères en gras indiquent les séquences publiées dans Bon et collaborateurs. (Bon *et al.*, 2008)