

République Algérienne Démocratique et Populaire
Ministère de l'Enseignement Supérieur et de la Recherche Scientifique
Université de Blida 1
Faculté de biologie
Département de Biologie et Physiologie Cellulaire



Mémoire de fin d'études
En vue de l'obtention du diplôme de MASTER II
Option : *Génétique*

Thème :

**Position des deux espèces de hérisson d'Algérie dans
l'arbre phylogénétique de la famille des Erinaceidae –
étude in silico.**

Présenté par :

M^{elle} HADJ NACER Lamya & M^{elle} ALLAL Rihab

Soutenu devant le jury composé de :

<i>Présidente :</i>	<i>M^{me} CHAKHMA A.</i>	<i>MAA</i>	<i>USDB</i>
<i>Examinatrice :</i>	<i>M^{me} AMOKRANE A.</i>	<i>MAA</i>	<i>USDB</i>
<i>Promotrice :</i>	<i>M^{me} DEROUICHE L.</i>	<i>MCA</i>	<i>ESSAIA</i>

Remerciements

*Nous remercions ALLAH le tout puissant de nous avoir donné le courage,
la volonté et la patience ainsi que la force afin de réaliser ce travail.*

Nous sommes très heureuses d'exprimer ici Tous nos remerciements et gratitude à notre promotrice M^{me} DEROUICHE maitre de conférences à L'École Supérieure des Sciences de l'Aliment et Industries Agroalimentaires– Alger, pour sa patience, et surtout pour sa confiance, ses remarques et ses conseils, sa disponibilité et sa bienveillance.

Et a

notre Co-promotrice M^{me} EDDAKRA maitre de conférences au département de biologie

on voudraist également remercier les membres du jurys:

La présidente M^{me} CHAKHMA d'avoir bien voulue nous faire l'honneur de présider ce jury

L'examinatrice M^{me} AMOKRANE pour avoir accepté d'examiner notre travail.

Nos remerciements s'adressent également à notre chef d'option Mr. Mohamed Said et à tous nos enseignants, qui nous ont donné les bases de la science .

Merci

Dédicaces

Je dédie ce mémoire à

*Mes très chers parents
qui ont consacré leur vie à mon éducation et ma réussite, qui m'ont encouragé et
soutenu dans les moments les plus difficiles de ma vie.*

Que Dieu les garde et les protège pour nous.

Mon frère:

ABD EL MOUMEN

Mes sœurs:

HAYET. MAROUA .KAOUTER

*Ainsi qu'à mon binôme et sœur : LAMYA à qui je souhaite une vie pleine de
bonheur et de succès. Ainsi que sa famille.*

*Toutes mes amies en particulier : LINA, ZINEB, SAMAH, IBTISEM
CHAIMA.*

Toutes les étudiantes de l'option Génétique promotion 2020.

Tous ceux qui m'aiment et que j'aime.

RIHAB

Dédicaces

Je dédie ce mémoire à

Mes très chers parents qui ont consacré leur vie à mon éducation et ma réussite, qui m'ont soutenu et encouragé dans les moments les plus difficiles de ma vie.

Que Dieu les garde et les protège.

Mes frères :

AHMED, EMBAREK, MALIK, TOUFIK, TARIK et leurs femmes

Mes sœurs :

AMEL qui m'a toujours soutenue ainsi qu'à LYNDA et son mari.

Mes nièces et mes neveux :

*SARAH, SAMY, MELISSA, SERINE, YUCEF, ADEM, AMIRA,
MARIA, MOUAD, TINHINANE ET ELINE.*

Mes chères bonnes tantes

Mon fiancé WALID qui m'a toujours encouragé et à sa maman.

Mon binôme et ma sœur RIHAB à qui je souhaite une vie pleine de bonheur et de succès.

*Toutes mes amies en particulier : ROUMAÏSSA, NADJET, ZINEB, SAMAH
ET IBTISEM*

Tous les étudiants de l'option génétique promotion 2020.

Tous ceux qui m'aiment et que j'aime.

LAMYA

Résumé

Le hérisson, de la famille des *Erinaceidae*, est un petit mammifère insectivore possédant une très vaste aire de répartition, cette dernière regroupe 21 espèces connues.

Les outils de bioinformatique permettent l'analyse phylogénétique tel que la reconstruction de l'arbre phylogénétique, l'identification des haplotypes, et avoir l'évolution des espèces. Dans ce travail nous avons tout d'abord présenté la famille des *Erinaceidae*, leur morphologie, leur distribution dans le monde et en Algérie en particulier ; nous avons exposé les méthodes de reconstruction phylogénétique pour inférer les phylogénies de la famille des *Erinaceidae*.

Les analyses phylogénétiques obtenues dans ce travail ont été élaborées à l'aide de différents logiciels tel que, MEGA5 pour la reconstruction de l'arbre phylogénétique, DnaSP pour l'identification des haplotypes, NETWORK pour avoir l'évolution des espèces, et DAMBE pour faire des calculs statistiques, avec l'utilisation de différentes méthodes comme UPGMA, Kimura- 2 paramètres ; ces analyses appliquées sur 57 séquences de Cytochrome b et 25 séquences de BRCA1 de 21 espèces appartenant à la famille *Erinaceidae* prises dans différentes régions géographiques démontrent la monophylie de la famille *Erinaceidae*.

Mots clés : *Erinaceidae*, Cytochrome b, BRCA1, phylogénétique, bioinformatique.

Summary

The hedgehog, from the Erinaceidae family, is a small insectivorous mammal with a very large distribution area, which includes 21 known species.

Bioinformatics tools allow phylogenetic analysis such as phylogenetic tree reconstruction, haplotype identification, and species evolution. In this memoir, we first presented the family of *Erinaceidae*, their morphology, their distribution in the world and in Algeria in particular; we explained the methods of phylogenetic reconstruction to infer the phylogenies of the family *Erinaceidae*.

The phylogenetic analyses obtained in this work were developed using different software such as, MEGA5 for the reconstruction of the phylogenetic tree, DnaSP for the identification of haplotypes, NETWORK for the evolution of species, and DAMBE to make statistical calculations, using different methods such as UPGMA, Kimura- 2parameters; these analyses applied to 57 Cytochrome b sequences and 25 BRCA1 sequences of 21 species belong to the *Erinaceidae* family taken in different geographical regions demonstrate monophyly of the *Erinaceidae* family.

Keywords: *Erinaceidae*, Cytochrome b, BRCA1, phylogenetic, bioinformatics.

ملخص

القنفذ من عائلة Erinaceidae هو حيوان ثديي صغير أكل للحشرات مع منطقة توزيع كبيرة جداً ، والتي تضم 21 نوعاً معروفاً. تتضمن هذه المذكرة استخدام مختلف وسائل الاعلام الالي الحيوي في التحليل دراسة النشوء لجنس القنفذيات, مثل إعادة بناء شجرة النشوء والتطور وتحديد les haplotypes, و تطور الأنواع قدمنا دراسة جنس القنفذيات, مورفولوجيا و توزيعه على المستوى العالمي و المحلي بصفة خاصة, قدمنا أساليب إعادة الإعمار النشوء لاستنتاج نشوء جنس القنفذيات.

كما اننا تطرقنا لتحليلات دراسة النشوء المحصل عليها في هذا العمل التي تم وضعها في مختلف البرامج مثل MEGA5 لإعادة بناء شجرة النشوء, DnaSP من اجل تحديد les haplotypes, NETWORK يعمل على دراسة تطور انواع DAMBE, يقوم بعمل حسابات احصائية. و استخدام مختلفة الطرق (UPGMA, كيمورا - parameters 2). قمنا بمعالجة 79 سلسلة من مورثة Cytb و 53 سلسلة من مورثة BRCA1 لـ 21 نوع من جنس القنفذيات, لـ مختلف المناطق الجغرافية. وفي النهاية تحصلنا على النتائج التي تدل على monophyliela التتخصص نوع القنفذيات.

المصطلحات: القنفذيات, BRCA1, Cytb, شجرة النشوء, المعلوماتية الحيوية.

ABREVIATIONS

- **ADN:**Acide DésoxyriboNucléique.
- **ADNmt :** ADN mitochondrial.
- **ARN:**Acide RiboNucléique.
- **BLAST:**Basic Local Alignment Search Tool.
- **BRCA1 :** Breast Cancer 1
- **Cyt b :**Cytochrome b.
- **DAMBE:** Data Analysis in Molecular Biology and Evolution.
- **DDBJ:**DNA Data Bank of Japan.
- **DnaSP:**DNASequence Polymorphism.
- **EMBL:** European Molecular Biology Laboratory.
- **HAMAP:**High-quality Automated and Manual Annotation of microbial Proteomes.
- **IUCN:** International Union for Conservation of Nature.
- **MEGA:**Molecular Evolutionary Genetics Analysis using Maximum Likelihood.
- **MPI:**Message Passing Interface.
- **NCBI :** National Center for Biotechnology Information.
- **PAUP*:**Phylogenetic Analysis Using Parsimony *and other methods.
- **PCR :** Polymerase Chain Reaction / Réaction de polymérisation en chaîne.
- **PDB:**Protein Data Bank.
- **PHYLIP:** PHYLogeny Inference Package.
- **PIR:**Protein Information Resource.
- **UE:**Unités Evolutives.
- **UPGMA:** Unweighted Pair-Group Method using Arithmetic Averages.

LISTE DES FIGURES

N° de figure	Titre	Page
01	Répartition de la famille des <i>Erinaceidae</i> dans le monde.	05
02	Répartition géographique de la famille des <i>Erinaceidae</i> en Algérie.	06
03	Exemple d'un arbre phylogénétique.	12
04	Les différents arbres phylogénétiques : a) arbre enraciné, b) arbre non enraciné.	12
05	Diverses catégories de groupes taxonomiques. Les taxa X, Y, Z forment un groupe monophylétique. Les taxa 1, 2 et 3 forment un groupe paraphylétique. Les taxa A, B et C forment un groupe polyphylétique (TOURASSE, 1992).	13
06	Les étapes initiales pour récupérer les séquences de la GenBank.	23
07	Les dix-neuf espèces de la famille <i>Erinaceidae</i> obtenues dans la GenBank ; ces espèces sont indiquées par des flèches en rouge.	23
08	Exemple des séquences nucléotidiques trouvées pour l'espèce <i>Atelerix algirus</i> .	24
09	Exemple d'une séquence nucléotidique au format FASTA.	24
10	La page principale du programme FaBox.	25
11	La conversion des séquences sous format FASTA par l'utilisation de FaBox.	26
12	Alignement trimmer (format FASTA).	27
13	Barre d'outils de MEGA5.	27
14	Le programme ClustalW utilisé pour l'alignement multiple des séquences du marqueur moléculaire choisi.	28
15	La méthode UPGMA utilisée pour la construction de l'arbre.	28

16	Les paramètres utilisés pour la reconstruction phylogénétique.	29
17	Barre d'outils de DnaSP.	30
18	Interface utilisateur de DnaSP.	31
19	Barre d'outil du programme DAMBE.	32
20	Barre d'outil pour tracer les graphes de C/G et A/T.	32
21	La page d'accueil de NETWORK 5.0.	34
22	Ouverture du fichier de données de rdf pour le calcul de MJ NETWORK.	35
23	Un fichier sous forme *.out a été sauvegardé.	35
24	Reconstruction de réseau haplotypique.	36
25	Schéma montrant l'organisation de l'ADN mitochondrial. la position de Cytb.	39
26	La position du gène BRCA1 sur le chromosome 17(17q21.31).	39
27	Les séquences de Cytb ont été convertit sous forme FASTA à partir de bloc note grâce à FaBox.	41
28	Arbre phylogénétique reconstruit par méthodes UPGMA basé sur 57 séquences de nucléotides de cytochrome-b de la famille des <i>Erinaceidae</i> en utilisant MEGA5.	42
29	Arbre phylogénétique des séquences de BRCA1 de la famille des <i>Erinaceidae</i> en utilisant la méthode UPGMA par le logiciel MEGA5.	44
30	(A) <i>Atelerix algirus</i> ; (B) <i>Paraechinus aethiopicus</i> (DEROUICHE et al., 2016).	46
31	Résultat d'alignement des séquences BRCA1 de la famille des <i>Erinaceidae</i> en utilisant le logiciel FaBox.	47
32	Les haplotypes des séquences Cytb de la famille des <i>Erinaceidae</i> en utilisant le logiciel DnaSP.	49
33	Les haplotypes des séquences BRCA1 de la famille des <i>Erinaceidae</i> en utilisant le logiciel DnaSP.	50
34	Variations des fréquences GC (A) et de AT(B) de Cytb produites à partir	52

DAMBE.

- | | | |
|-----------|--|----|
| 35 | Variations des fréquences GC (A) et de AT (B) de BRCA1 produites à partir de DAMBE. | 53 |
| 36 | Comparaison des fréquences de CG et AT entre Cytb et BRCA1 réalisés par le logiciel DAMBE. | 54 |
| 37 | Réseau haplotypique réalisé par le logiciel NETWORK à partir de 57séquences de Cytb de la famille des <i>Erinaceidae</i> . | 55 |
| 38 | La croissance des haplotypes desséquences de BRCA1de la famille des <i>Erinaceidae</i> en utilisant le logiciel NETWORK. | 56 |

LISTE DES TABLEAUX

N° de tableau	Titre	Page
I	Description morphologique des deux espèces du hérisson d'Algérie.	7
II	Les formats de données utilisés par logiciel DAMBE(XIA, 2001).	33
III	Les séquences d'ADN nucléaire et mitochondriale de la famille des <i>Erinaceidae</i> extraites de la GenBank.	37
IV	Comparaison des indices de diversités calculés avec DnaSP sur les échantillons étudiés.	51
V	Les séquences extraites à partir de la GenBank (types des gènes et leur origine géographique).	Annexe

SOMMAIRE

INTRODUCTION..... 01

CHAPITRE I: RAPPELS BIBLIOGRAPHIQUES

I.1. Terminologie et modèle d'étude 03

I.1.1. Présentation de l'espèce 03

I.1.2. Distribution géographique de la famille des Erinaceidae 04

I.1.3. Description morphologique 06

I.2. Etude phylogénétique 09

I.2.1. Définition de la phylogénie 09

I.2.2. Phylogénie moléculaire 10

I.2.3. Représentation phylogénétique 11

I.2.4. Arbre phylogénétique 11

I.2.4.1. Caractéristiques générales des arbres 12

I.2.4.2. Différentes représentations graphiques pour les arbres 13

I.2.5. Reconstruction d'arbres phylogénétiques 14

I.2.6. Critères de choix entre les méthodes 15

I.3. Stockage de données biologiques 16

I.3.1. Banques des données 16

I.3.1.1. Les banques généralistes **Error! Bookmark not defined.**16

I.3.1.2. Les banques spécialistes **Error! Bookmark not defined.**17

I.3.2. Différentes étapes et logiciels d'une étude phylogénétique 17

I.3.2.1. Logiciels de nettoyage 18

I.3.2.2. Logiciels d'alignement 18

I.3.2.3. Calcul des distances 19

I.3.2.4. Logiciels de reconstruction de l'arbre phylogénétique 19

I.3.2.5. Logiciels de visualisation d'arbres phylogénétiques 20

CHAPITRE II : MATERIEL ET METHODES

II.1. Analyses moléculaires et phylogénétiques : 21

II.1.1. Etapes d'Analyses Moléculaires	21
II.1.1.1. Echantillonnage.....	21
II.1.1.2. Extraction et quantification de l'ADN	21
II.1.1.3. Amplification de l'ADN par PCR.....	22
II.1.1.4. Séquençage	22
II.2. Etude de la GenBank :	22
II.2.1. Définition de la GenBank	22
II.2.2. Traitement des séquences de la GenBank.....	25
CHAPITRE III: RESULTATS ET DISCUSSION	
III.1. Résultats de la GenBank	37
III.2. Analyses phylogénétiques.....	39
III.2.1. Construction des arbres phylogénétiques	39
III.2.2. Détermination des haplotypes.....	46
III.2.3. Analyses statistique	51
III.2.4. Résultat de NETWORK	54
CONCLUSION.....	58
RÉFÉRENCES BIBLIOGRAPHIQUES.....	60
ANNEXE	

Introduction

La famille des *Erinaceidae* (**FISCHER, 1817**) est appelée ainsi pour la première fois par **BONAPARTE** en **1838** ; revue plus récemment par **HONACKI et al. (1982)** ; **CORBET, (1988)** ; **DARREL et al. (1991)** (in : **WILSON et REEDER, 1993**). Selon **SCHILLING et al. (1986)**, les *Erinaceidae* sont des mammifères insectivores assez gros, pourvus de piquants dorsaux les protégeant des prédateurs. Cette dernière regroupe 21 espèces. Les traits morphologiques décrits par **Le BERRE (1990)** permettent de distinguer aisément les deux espèces: *Atelerix algirus* et *Paraechinus aethiopicus*.

La bioinformatique est un domaine multidisciplinaire qui associe la biologie, l'informatique et les mathématiques. Le développement récent et rapide de la génomique et de la protéomique suscite une collaboration de plus en plus étroite entre les spécialistes des sciences de la vie et de l'informatique, afin de développer de nouvelles approches et de nouvelles méthodes analytiques permettant d'examiner la quantité massive de données biologiques disponibles. Pour la réalisation de l'arbre phylogénétique (**DARLU et al., 1993** ; **LE GUYADER, 2003**), il existe cependant quelques limites à l'utilisation de l'outil bioinformatique pour la reconstruction de l'arbre phylogénétique. Cette reconstruction se base sur le fait que certains caractères ont été hérités d'une espèce à l'autre, tandis que d'autres se sont transformés au cours du temps, laissant ainsi des traces de l'évolution.

L'évolution des espèces est comprise comme un processus divergent, modélisé le plus souvent sous la forme d'un arbre dont les feuilles représentent les espèces contemporaines et les nœuds, les espèces ancestrales. On parle alors d'arbre d'évolution ou plus communément de phylogénie (**BERRY, 1997**).

Notre travail est basé sur les analyses phylogénétiques. Le but de ce mémoire repose sur l'utilisation de plusieurs logiciels de bioinformatique afin d'étudier une base des données de biologie moléculaire pour trouver l'arbre phylogénétique et la position de deux espèces de hérisson d'Algérie dans cet arbre.

On commence le chapitre I par un rappel bibliographique, il est divisé en trois parties :

❖ Présenter notre modèle d'étude « la famille des *Erinaceidae* » leur classification et leur distribution géographique dans le monde et particulièrement en Algérie ainsi que leur morphologie.

❖ Définir : la phylogénie, l'arbre phylogénétique et les caractéristiques générales de l'arbre ; citer aussi les différents marqueurs phylogénétiques qui aident à la reconstruction phylogénétique des espèces.

❖ Présenter les différentes banques des données et les différentes étapes et logiciels d'une étude phylogénétique.

Le chapitre II, fournit le matériel et les méthodes que nous utiliserons pour réaliser notre projet.

Le chapitre III, rapporte les résultats obtenus ainsi qu'une discussion à la lumière des données bibliographiques.

Enfin, on terminera ce mémoire par une conclusion générale où nous dégagerons les principaux résultats de ce travail.

Rappels Bibliographiques

I.1. Terminologie et modèle d'étude :**I.1.1. Présentation systématique de la famille des *Erinaceidae* :**

La famille des *Erinaceidae* regroupe des espèces qui se représentent systématiquement comme suit :

Classification	
Règne	Animalia
Embranchement	Chordata
Sous embranchement	Vertebrata
Super classe	Tetrapoda
Classe	Mammalia
Super ordre	Lamasiatheria
Ordre	Erinaceimorpha
Famille	<u>Erinaceidae</u>
Espèce	Sous espèce
<i>Atelerix albiventris</i>	
<i>Atelerix algirus</i>	
<i>Atelerix frontalis</i>	
<i>Erinaceus amurensis</i>	
<i>Erinaceus concolor</i>	<i>Erinaceus concolor nesiotes</i>
<i>Erinaceus ecaudatus</i>	
<i>Erinaceus europaeus</i>	
<i>Erinaceus roumanicus</i>	
<i>Hemiechinus auritus</i>	<i>Hemiechinus auritus megalotis</i>
<i>Mesechinus dauuricus</i>	
<i>Mesechinus hughii</i>	<i>Mesechinus hughii miodon</i>
<i>Paraechinus aethiopicus</i>	<i>Paraechinus aethiopicus deserti</i>
<i>Paraechinus hypomelas</i>	
<i>Echinosorex gymnura</i>	
<i>Hylomys megalotis</i>	
<i>Hylomys parvus</i>	
<i>Hylomys suillus</i>	<i>Hylomys suillus dorsalis</i>

	<i>Hylomys suillus maxi</i> <i>Hylomys suillus microtinus</i> <i>Hylomys suillus siamensis</i> <i>Hylomys suillus suillus</i>
<i>Hylomys cf. suillus JJS-2016</i>	
<i>Neohylomys hainanensis</i>	
<i>Neotetracus sinensis</i>	
<i>Podogymnura truei</i>	

I.1.2. Les espèces du hérisson d'Algérie :

Plusieurs chercheurs ont essayé d'étudier le hérisson en Algérie, leurs résultats ont mené à le diviser en deux catégories, la première catégorie regroupe le hérisson du désert qui est une espèce saharienne et la deuxième catégorie regroupe, le hérisson d'Algérie qui est une espèce méditerranéenne, avec pour chacune une répartition géographique précise que ce soit dans le monde ou en Algérie.

a. Distribution dans le monde :

- *Atelerix algirus* : Espèce méditerranéenne, sa répartition est limitée à la côte orientale d'Espagne, en France à la côte du Languedoc et du Var, aux îles Baléares (MATTHEWS, 1972), à l'Afrique du Nord, aux îles Canaries (SAINT-GIRONS, 1973) et enfin du Maroc jusqu'à la Libye (BURTON, 1976).

Paraechinus aethiopicus : Il est signalé de la Mauritanie à l'Éthiopie à travers le Maroc, l'Algérie, la Tunisie, la Libye, le Tchad, l'Égypte, le Soudan, dans la péninsule arabique, de même qu'en Iraq et en Iran (MADKOUR, 1982 ; HALTENORTH et DILLER, 1985 ; AULAGNIER et THEVENOT, 1986; CORBET, 1988 ; DELANY et FAROOK, 1989 ; LEBERRE, 1991 ; KOWALSKI et RZEBIK-KOWALSKA, 1991) (Figure 1).

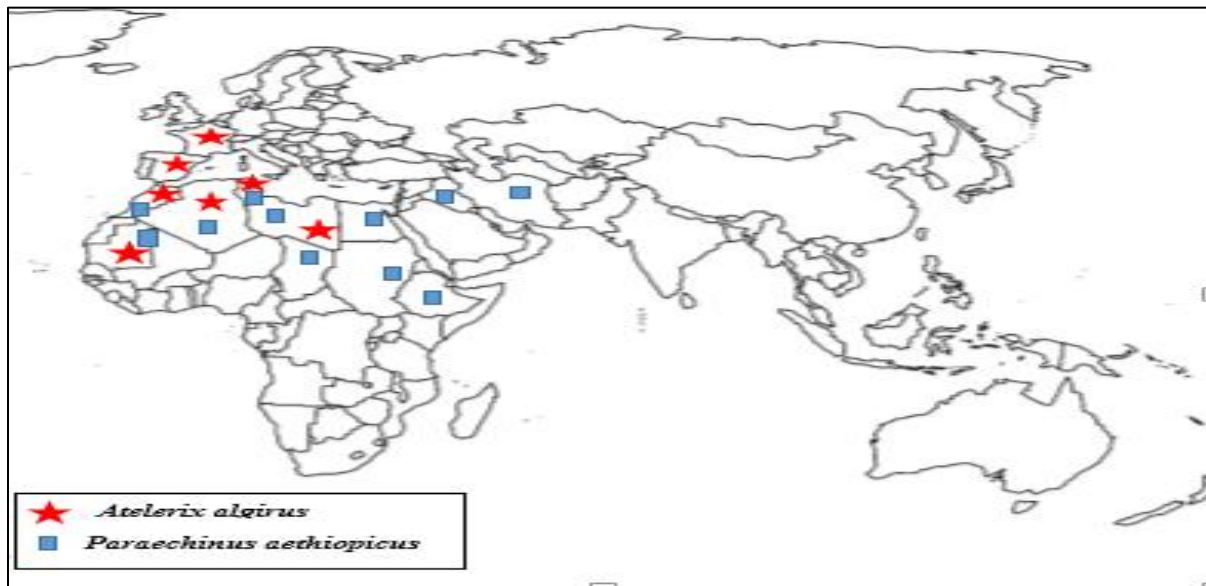


Figure 1 : Répartition de famille des *Erinaceidae* dans le monde.

b. Distribution en Algérie :

- Atelerix algirus* : Il est cité par plusieurs auteurs. Il a été signalé sur le littoral et les plaines intérieures par quelques auteurs, Larbaa (SEURAT, 1924), près d'El Kala (TELAÏLIA, 1990), à Blida, Boumerdès, Jijel, Annaba, El Taraf (HARBI, 1991), dans les environs d'Alger, El Harrach (DOUMANDJI et DOUMANDJI, 1992) et à Cap Djanet (METREF, 1994), en Mitidja et à Oued Smar, à Bab Ezzouar, à Soumaa et à Baraki (DOUMANDJI. Com. pers.). Il a été signalé aussi au niveau de l'Atlas Tellien, dans le parc national de Thniet El-Had (BAICHI, 1987), dans la forêt de l'Akfadou (CHEBINI, 1987), à Tikjda (GUERMAS, 1987 ; SAYAH, 1988), dans le parc national de Chréa (MAZARI, 1988), dans la réserve naturelle du Mont Babor (MORDJI, 1988), à Tala Gilef (HAMDINE, 1991), dans le parc national de Taza (MOSTEFAI, 1990 ; KISSERLI, 1992) et à Souk-Ahras (HARBI, 1991). D'autres auteurs notent sa présence au niveau des Hauts Plateaux. Il existe dans la région de Senabla Chergui (KHIREDDINE, 1977), dans la région sétifienne (GAISLER, 1984) à travers toute la wilaya de Médéa (ANONYME, 1985 ; HARBI, 1991), à Oued El Biod (LAAMARI, 1985), dans la forêt d'Oum-Graf à Saida (TALBI, 1985), au Djebel El Achch (AMMAM, 1987), au niveau du barrage de Boughzoul (BAZIZ, 1991), à Sidi Bel Abbès (HARBI, 1991) et au niveau du lac de Boulhilet à Oum el-Bouaghi (SI BACHIR, 1991). D'après KOWALSKI et RZEBIK-KOWALSKA (1991), *Atelerix algirus* cohabite avec *Paraechinus aethiopicus* au niveau des Hauts

Plateaux. En 1988, **ATHMANI (1988)** note sa présence au niveau de l'Atlas Saharien dans le parc national de Belezma.

- *Paraechinus aethiopicus* : D'après les recherches de **LE BERRE** il se trouve en Algérie dans les régions suivantes : Hoggar, M'Zab, Ouargla, Touggourt, Laghouat (**FOLEY, 1922**), Idélès (**SEURAT 1934**), Tit (**MEINERTZHAGEN, 1934**), Béni-Abbès (**PETTER, 1955**), Béchar, Béni-Ounif (**JUNQUA, 1966**), et selon **KOWALSKI et RZEBIK-KOWALSKA, (1991)** : la répartition de l'espèce s'étend de Ain Sefra du côté nord de l'Atlas saharien jusqu'à Laghouat et Biskra en passant par Ourka et Brezina du côté Sud et à l'Ouest jusqu'à Béni-Abbès. Au Sud du pays, l'espèce se rencontre au Hoggar, à Adrar et dans les montagnes du Sahara central (Figure 2).

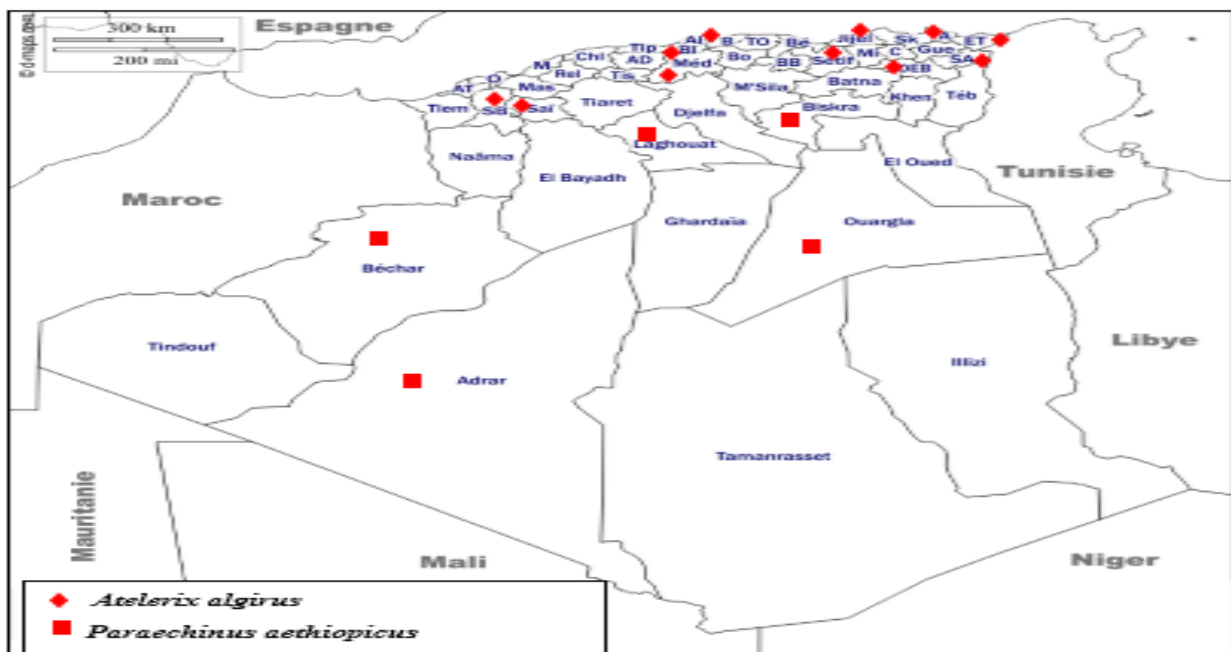


Figure 2 : Répartition géographique de la famille des *Erinaceidae* en Algérie.

I.1.3. Description morphologique :

Chaque espèce a ces propres caractéristiques morphologiques ; mais il y a des points communs entre les différentes espèces (les piquants qui recouvrent le corps de l'animal ; les vibrissiles tactiles ; la formule dentaire ; ressemblances entre mâles et femelles avec un dimorphisme sexuel).

Tableau I : Description morphologique des deux espèces du hérisson d'Algérie.

Espèce	Morphologie
<i>Atelerix algirus</i>	<ul style="list-style-type: none"> • Les oreilles du hérisson d'Algérie sont plus larges et plus longues. • Le dessous du corps est nettement plus clair que chez l'espèce européenne, avec les flancs bruns foncés. • Il présente des piquants plus clairs s'étendant peu en avant sur le front ou une petite échancrure dépourvue de piquants les séparent. • Ongles blancs ou pâles alors que ceux du hérisson d'Europe sont de couleur noirâtre (BRETAGNOLLE et ATTIE, 1989) • La longueur du corps avec la tête est de 200 à 250 mm environ chez le hérisson adulte. • La longueur de la queue 24 à 40 mm. • La mensuration du pied postérieur varie entre 30 et 42 mm. • Le poids moyen est de 800 g (SCHILLING et al., 1986). • La longueur condylobasale varie entre 50 et 60 mm (VESMANIS, 1979). • Le front recouvert par une fourrure blanchâtre. • Les premières incisives sont longues et projetées en avant, les canines sont petites et les molaires développées, adaptées au broyage des insectes (REEVE, 1994). • Ce genre est caractérisé par quatre doigts à chacune des pattes postérieures (BICHE, 2003).
<i>Paraechinus aethiopicus</i>	<ul style="list-style-type: none"> • Une taille petite ou moyenne. • La longueur du corps est d'environ 200-230 mm, la longueur de la queue 20-40 mm, la longueur du pied 32-36 mm (LE

BERRE, 1990).

- Le poids d'un adulte varie entre 400 g et 700 g selon **REEVE (1994)** mais selon **DRAGESCO – JOFFE (1993)** le poids d'un adulte varie entre 165 g et 250 g, il est cinq fois moins lourd ; selon **LE BERRE** le poids d'un adulte est de 250 à 500 g.
- Les piquants recouvrent le corps de l'animal du front jusqu'à la naissance de la queue qui se présente comme un moignon velu, ils sont orientés en tous sens de façon à ce que l'individu ne puisse être saisi sans piquer son agresseur, les piquants sont fortement papilleux (**CORBET, 1988**).
- Chaque piquant mesure 27 mm de longueur environ et compte 20 à 30 compartiments (**CORBET, 1988**) à ce sujet, **GRASSE(1955)** parle de stries longitudinales ou rainures longitudinales, l'extrémité des pics est soit claire soit sombre, déterminant la couleur d'ensemble de l'animal (**LE BERRE, 1990**).
- La couverture épineuse est partagée en deux sur le front par une raie médiane nue de 3 cm ; parfois indistincte.
- Les poils abdominaux sont fins et denses (**LE BERRE, 1990**).
- La tête se termine par un museau long, pointu qui s'avance au-delà de la bouche, très mobile et toujours humide.
- La longueur condylobasale est en moyenne de 48 mm (**VESMANIS, 1979**).
- La tête comporte de grandes oreilles plus longues que les épines voisines ; ils sont entre 48 et 55 mm (**LE BERRE, 1990**).
- Les bulles tympaniques sont hypertrophiées et cela provoque un renflement des régions auditives (**CORBET, 1988**).
- Les deux côtés de l'animal possèdent de nombreux vibrissiles tactiles très développés l'aidant à repérer ses proies (**DRAGESCO-JOFFE, 1993**).
- Ils sont terminés par cinq doigts avec le gros orteil très petit (**HALTENORTH et DILLER, 1985**).
- Pour **HALTENORTH et DILLER (1985)** le hérisson du

désert possède une face de couleur brun foncé avec une rainure noire passant entre les yeux, le front, les oreilles, le menton, la gorge et le haut de la poitrine sont de couleur blanche, les membres, le bas-ventre et la queue sont brun-noir, les mêmes auteurs notent aussi l'existence d'individus complètement blanc, avec seulement les pattes et la queue d'un brun-noir.

- Le nombre de dents est de 36, les premières incisives supérieures et inférieures ordinaires et plus grosses que les autres.
- Les mâles ont tendance à être plus lourds que les femelles.

I.2. Etude phylogénétique :

I.2.1. Définition de la phylogénie :

La phylogénie ou phylogenèse se définit, en général, dans l'encyclopédie comme « l'histoire de la formation et de l'évolution d'une espèce ». Le terme phylogenèse provient du grec phûlon « tribu » et genesis « origine » et il a été présenté par **HAECKEL** dès **1860**, qui l'a définie comme « l'histoire du développement paléontologique des organismes par analogie avec l'ontogénie ou histoire du développement individuel ». La phylogénie constitue un procédé pour construire des classifications d'espèces. La phylogénie moléculaire, dont certaines méthodes de reconstruction seront présentées par la suite, l'étude de l'histoire évolutive des espèces en se basant sur une portion de leur séquence moléculaire.

I.2.2. Phylogénie moléculaire :

La phylogénie moléculaire utilisée pour comparer les séquences des molécules d'ADN ou de protéines des êtres vivants dans le but de déterminer les liens de parenté qui les unissent ainsi que pour appréhender leur histoire évolutive (phylogenèse). Elle étudie l'histoire évolutive des espèces étudiées à la base d'une portion de leur séquence moléculaire (**DIALLO, 2009**).

La phylogénie moléculaire est basée sur les principes suivants :

- Les mutations se produisent au hasard.
- Les mutations s'accumulent au cours de temps.
- Les mutations se produisant chez un individu et se fixent dans sa descendance.

Il existe aujourd'hui une large gamme de marqueurs moléculaires qui peuvent être utilisés pour accomplir ce type de recherche.

Un marqueur génétique est un caractère mesurable qui peut détecter une variation dans la séquence soit protéique, soit nucléique. Selon le niveau taxonomique étudié, il est essentiel de choisir un type de marqueur qui lui est adapté. Par exemple, si le but est de distinguer des groupes d'individus au niveau de l'espèce, les marqueurs utilisés devront être plus conservés que dans le cas des études de populations (**BEAULIEU, 2007**).

Les marqueurs moléculaires concernant la molécule d'ADN elle-même, et, comme tels, sont considérés comme des mesures objectives de la variation (**VICENTE *et al.*, 2003**).

✓ **Marqueurs d'ADN nucléaire :**

Un certain nombre de marqueurs sont à présent disponibles pour détecter les polymorphismes d'ADN nucléaire. Dans les études sur la diversité génétique, les marqueurs les plus fréquemment utilisés sont les microsatellites. Les microsatellites sont à présent les marqueurs les plus utilisés dans les études de caractérisation génétique des animaux d'élevage (**SUNNUCKS, 2001**).

✓ **Marqueurs d'ADN mitochondrial :**

Le génome mitochondrial des animaux est formé par une petite molécule unique d'ADN circulaire dont la taille varie dans une gamme assez étroite chez les vertébrés (16-20kb) (**AVISE *et al.*, 1987 ; BOORE, 1999**). Sa séquence code pour 37 gènes dont 24 contiennent l'information pour une partie de la machinerie de traduction de la molécule elle-même (ARNt et ARNr), et les 13 autres pour les sous unités de la chaîne de transport des électrons. En plus, il est reconnu une région régulatrice (ou de contrôle) des processus de réplication et de transcription de l'ADNmt d'environ 0.8Kb nommée D-Loop chez les vertébrés (**AVISE *et al.*, 1987**). Par ailleurs, il est décrit la présence, dans plusieurs ADNmt, de petits fragments non codants avec des fonctions qui peuvent être aussi régulatrices (**BOORE, 1999**).

I.2.3. Représentation phylogénétique :

L'élaboration de technologies à haut débit serait inutile si l'on ne disposait pas des capacités d'analyse des données biologiques en croissance exponentielle. Ces données doivent

se stocker dans des bases de données électroniques associées à des logiciels spécialement conçus pour la mise à jour, l'interrogation et l'extraction. Les informations doivent être facilement accessibles et flexibles aux interrogations pour faciliter l'extraction des informations pouvant être analysées pour éclaircir les voies métaboliques et le rôle des protéines et des gènes impliqués (**RISCHKOWSKY *et al.*, 2008**).

La bioinformatique est fondamentale pour associer les informations provenant de sources différentes et générer une connaissance nouvelle à partir de données existantes. Elle dispose également des potentialités pour simuler la structure, la fonction et la dynamique des systèmes moléculaires et est donc utile pour la formulation des hypothèses et pour la conduite du travail expérimental (**RISCHKOWSKY *et al.*, 2008**).

I.2.4. Arbre phylogénétique :

Un arbre phylogénétique ou un dendrogramme est une représentation graphique de la phylogénie. Il exprime les liens entre les taxa sous la forme d'une succession de branchements (**RASMONT, 1997 ; GATTOLLIAT, 2002 ; SCHMIDT, 2003**).

Il contient quatre éléments principaux (Figure 3) :

- ✚ **La racine** (au cas où l'arbre serait enraciné) qui indique l'ancêtre commun des espèces représentées dans l'arbre.
- ✚ **Les nœuds externes** (ou feuilles) représentent les espèces contemporaines pour lesquelles les informations ont été disponibles lors de la construction de l'arbre. Ils sont communément appelés taxons.
- ✚ **Les nœuds internes** qui représentent des ancêtres inférés, hypothétiques.
- ✚ **Les branches** (ou arêtes) de l'arbre qui montrent les relations de descendance entre les nœuds (les taxons). Ces arêtes peuvent avoir des longueurs. Ces longueurs peuvent correspondre à plusieurs informations dont, le taux de mutation, la distance génomique, etc.

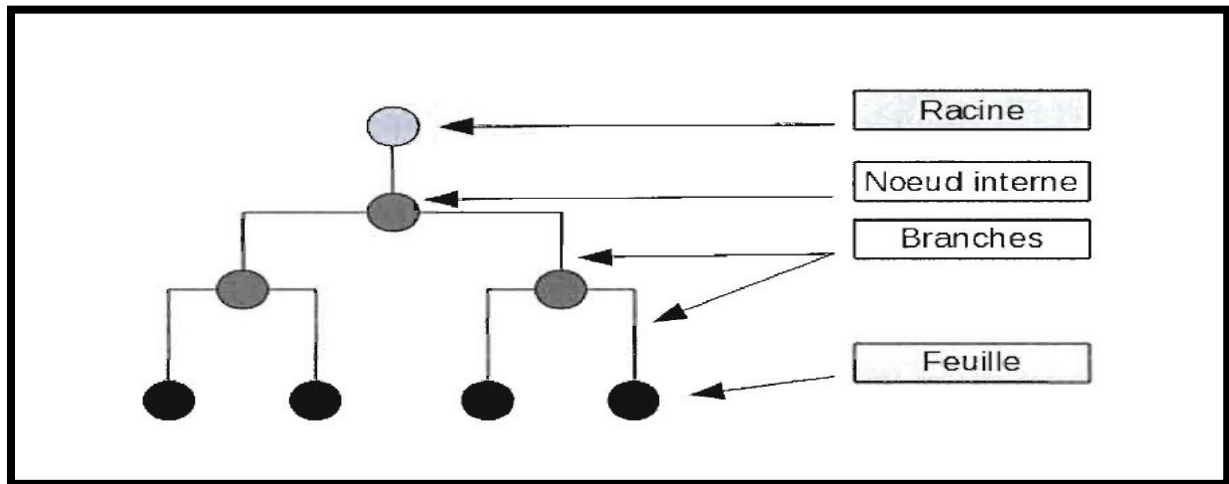


Figure 3 : Exemple d'un arbre phylogénétique.

I.2.4.1. Caractéristiques générales des arbres :

Il est nécessaire de distinguer d'abord entre réseaux et arbres, ces derniers pouvant être non enracinés ou enracinés.

- Un arbre enraciné** est un arbre dans lequel un des nœuds est désigné pour être la racine, et la direction des rapports héréditaires est déterminée (Figure 4).
- Un arbre non enraciné** est une représentation intemporelle des relations phylogénétiques, les liens entre nœuds ne sont pas orientés, et un seul et unique chemin permet de passer d'un sommet à l'autre. Cet arbre n'induit donc aucune hiérarchie (Figure 4. b).

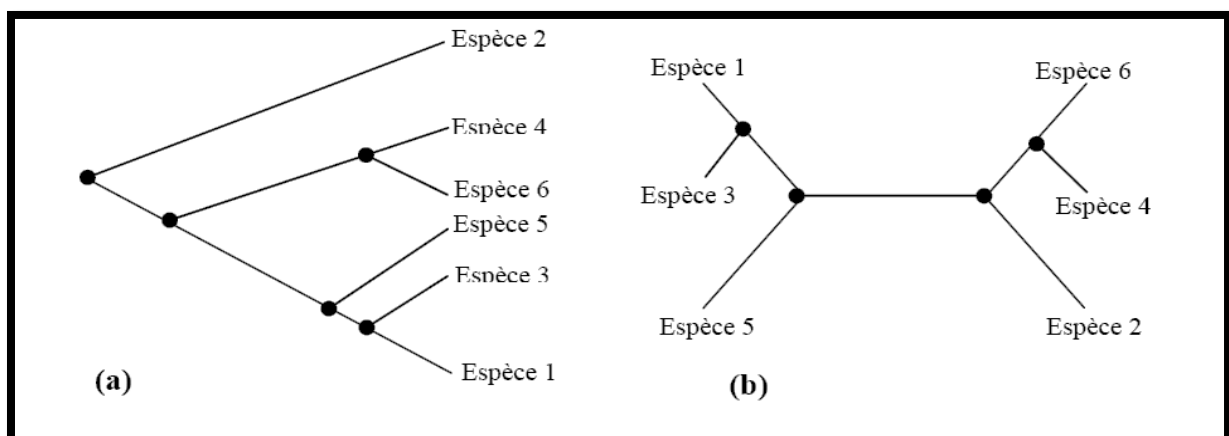


Figure 4 : Les différents arbres phylogénétiques : a) arbre enraciné, b) arbre non enraciné.

- Monophylie, paraphylie et polyphylie** : D'après RASMONT (1997) et TOURASSE (1992), dans le domaine de la reconstruction phylogénétique, on distingue trois

groupes taxonomiques : Le groupe monophylétique, le groupe paraphylétique et le groupe polyphylétique :

- ❖ **Groupe monophylétique** : Est un ensemble d'espèces issues d'un même ancêtre commun, c'est le cas de groupe (X, Y, Z) de la figure 5
- ❖ **Groupe paraphylétique** : C'est lorsqu'une ou plusieurs espèces d'un groupe monophylétique partage(nt) un ancêtre commun avec des espèces appartenant à d'autres lignées, comme par exemple l'ensemble (1, 2, 3) (Figure 5)
- ❖ **Groupe polyphylétique** : Si les différentes espèces d'un groupe dérivent d'ancêtres différents, celui-ci est dit polyphylétique, c'est le cas de groupe d'espèces (A, B, C) (Figure 5).

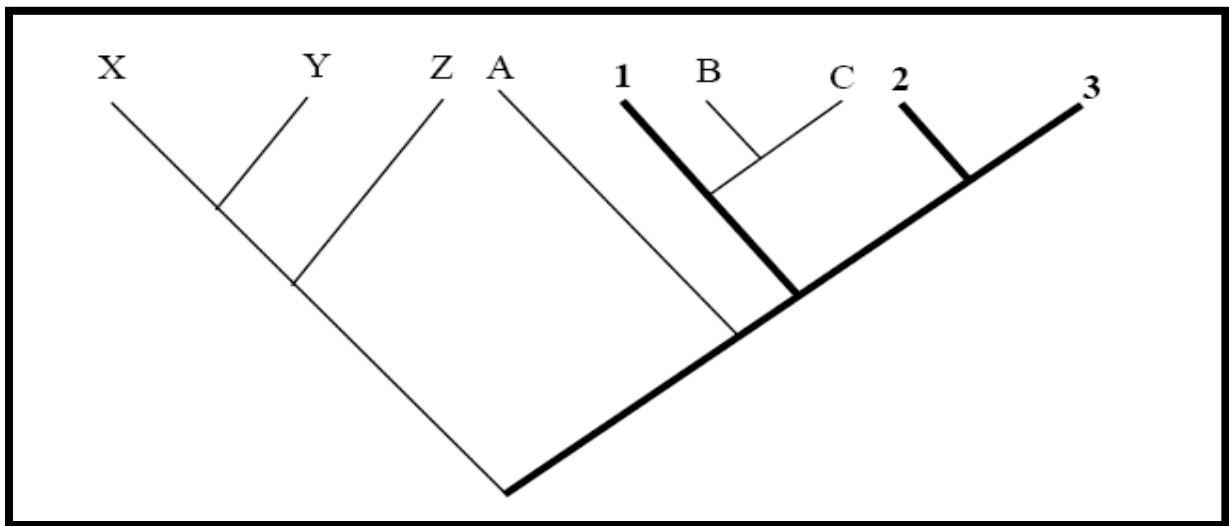


Figure 5 : Diverses catégories de groupes taxonomiques. Les taxa X, Y, Z forment un groupe monophylétique. Les taxons 1, 2 et 3 forment un groupe paraphylétique. Les taxons A, B et C forment un groupe polyphylétique (TOURASSE, 1992).

I.2.4.2. Différentes représentations graphiques pour les arbres :

Les arbres phylogénétiques ont la capacité de véhiculer de nombreuses informations. Toutefois, avoir la capacité ne signifie pas forcément que celles-ci soient toujours exploitées. Selon la représentation graphique adoptée, il est possible de reconnaître le type d'arbre et généralement le type d'information véhiculée.

a) **Le dendrogramme** : Le dendrogramme est un arbre exprimant les liens entre taxons sous la forme d'une succession de branchements. Il ne désigne rien d'autre qu'un arbre dont les

éléments terminaux sont les taxons ou UE observés. Ce terme est assez large pour ne rien exprimer quant à la procédure utilisée pour son obtention.

b) Le cladogramme : Le cladogramme est un dendrogramme exprimant les relations phylogénétiques entre taxons et construit à partir de l'analyse cladistique où les points de branchements (les nœuds) sont définis par des synapomorphies. Ce mot a été créé la même année par **MAYR (1965)** et par **CAMIN (1965)** avec des sens un peu différents.

c) Les phylogrammes : Le phylogramme est un dendrogramme exprimant les branchements cladistiques et le degré de divergence adaptative subséquente aux branchements (**MAYR, 1969**).

d) Les phénogrammes : Le phénogramme est un dendrogramme produit par la taxinomie numérique où les relations entre taxons expriment les degrés de similitude globale, défini simultanément par **MAYR (1965)** et par **CAMIN (1965)**.

I.2.5. Reconstruction d'arbres phylogénétiques

La manière classique d'illustrer les relations phylogénétiques entre les espèces est de les modéliser en utilisant un arbre phylogénétique. La manière moderne consiste en la modélisation d'un réseau phylogénétique comprenant les réticulations nécessaires. A ce niveau nous présenterons seulement quelques méthodes d'inférences d'arbres phylogénétiques. Les méthodes d'inférence phylogénétique présentées sont exposées en détails dans **SWOFFORD *et al.* (1996)**, **LI (1997)** et **FELSENSTEIN (2004)**.

La reconstruction d'un arbre phylogénétique débute par ce que l'on appelle « l'alignement » qui consiste à mettre en correspondance les sites des séquences des espèces de manière à pouvoir les comparer les unes aux autres. Les séquences utilisées pour la reconstruction peuvent être de l'ADN ou de l'ARN.

Il existe trois grands types de méthodes de reconstruction d'arbres phylogénétiques : les méthodes de distances, les méthodes du maximum de parcimonie et les méthodes du maximum de vraisemblance, ces dernières sont appelées également les méthodes probabilistes (**TOURASSE, 1992 ; NEI, 1996 ; ROBINSON, 1997 ; COMET, 1998 ; HAUBOLD, 2000**).

○ **L'approche phénétiq**ue : Ne tient pas compte du processus de l'évolution. Elle se contente de mesurer les distances entre les espèces et de reconstruire le meilleur arbre possible à l'aide d'une stratégie de regroupement hiérarchique.

- **L'approche cladistique** : Cherche à établir des relations de parenté en s'intéressant aux caractères (bases ou acides aminés) dérivés, partagés par les taxons. On considère ainsi tous les scénarios d'évolution en inférant les caractères des ancêtres potentiels à chaque nœud, et on choisit l'arbre qui correspond au meilleur scénario d'évolution selon un critère préalablement choisi. Les méthodes utilisées sont essentiellement basées sur le maximum de parcimonie.
- **L'approche probabiliste** : Ou maximum de vraisemblance, quant à elle, évalue en termes de probabilités l'ordre des branchements et la longueur des arêtes d'un arbre sous un - modèle évolutif donné. Les méthodes bayésiennes font aussi partie de cette approche.

La première approche étudie la parenté entre les taxons en s'intéressant à leur degré de similarité alors que la deuxième est basée sur la généalogie.

I.2.6. Critères de choix entre les méthodes :

Les méthodes de reconstruction d'arbres phylogénétiques sont évaluées et comparées en fonction de cinq critères (**HUELSENBECK, 1995 ; MORET *et al.*, 2002**) :

- **L'efficacité** : Rapidité de la méthode.
 - **La puissance** : Plus une méthode est puissante, moins elle nécessite de données pour obtenir un résultat fiable.
 - **La consistance** : Caractérise une méthode qui converge vers le bon arbre si on lui fournit suffisamment de données.
 - **La robustesse** : Une méthode est robuste si de petites violations des hypothèses de départ ne résultent pas dans une mauvaise estimation de la phylogénie.
 - **La fiabilité** : Une méthode est fiable si elle construit des arbres phylogénétiques « exacts ».
- De tous ces critères, la robustesse peut être la plus importante, parce que, avec des vraies données, c'est-à-dire, données n_{0+n} issues des simulations, il y a un grand risque que les hypothèses posées ne soient pas respectées.

D'après **KUHNER *et al.* (1994)**, et suite à leurs études de comparaison par simulation sur ordinateur de plusieurs algorithmes de reconstruction phylogénétique, l'algorithme de Neighbor-Joining, l'algorithme de Fitch-Margoliash, Parcimonie et Maximum de vraisemblance, ils fournissent les conclusions suivantes :

- ☞ Pour l'estimation des longueurs des branches, les deux algorithmes de distances (Neighbor-Joining et Fitch-Margoliash) ont été les plus performants.

- ↪ L'algorithme de Parcimonie et l'algorithme de compatibilité fournissent des résultats similaires, et ils sont moins performants que les algorithmes précédents.
- ↪ L'algorithme de Maximum de Vraisemblance est le plus robuste de tous les algorithmes étudiés.

I.3. Stockage de données biologiques :

I.3.1. Banques des données :

Aujourd'hui les méthodes rapides de séquençages sont utilisées fréquemment et le nombre de nouvelles séquences augmente rapidement. Toutes les données issues du séquençage doivent être traitées et analysées afin d'obtenir le plus grand nombre d'informations. Il faut ainsi stocker ces séquences et toutes les informations obtenues. Pour cela, de grandes bases de données de séquences ont été mises en œuvre pour permettre un accès facile aux données. Les premières banques de données en biologie moléculaire ont traité des informations structurales sur les protéines, puis très rapidement, des séquences protéiques et nucléotidiques. Il existe différents types de bases de données biologiques : celles qui sont dites généralistes et qui stockent des séquences provenant de tous les organismes et celles dites spécialisées qui se consacrent plus particulièrement à un organisme ou à une thématique donnée.

I.3.1.1. Les banques généralistes :

Il existe plusieurs banques généralistes publiquement accessibles. La principale banque généraliste de séquences nucléotidiques est produite par trois partenaires : EMBL data library (COCHRANE *et al.*, 2006) en Europe, GenBank (BENSON *et al.*, 2006) aux Etats-Unis et DDBJ (OKUBO *et al.*, 2006) au Japon. La plupart des données de ces banques proviennent de soumissions effectuées par les auteurs. D'autres regroupent des séquences protéiques telles que UNIPROT (WU *et al.*, 2006), GenPept, HAMAP (GATTIKER *et al.*, 2003), etc. De la même manière que pour les banques de séquences nucléotidiques, leur organisation se base autour des annotations biologiques et biochimiques d'une part, et des séquences d'autre part. GenPept correspond à la traduction de l'ensemble des parties codantes de GenBank. La principale banque de protéines est UNIPROT. En effet, elle possède de nombreux atouts : redondance minimale, références croisées, qualité d'annotation, etc. Elle correspond à la fusion de SWISS-PROT (WU *et al.*, 2006), TrEMBL et PIR (WU *et al.*, 2003).

Les séquences contenues dans SWISS-PROT sont issues de la traduction des gènes annotés dans EMBL, d'autres banques protéiques, de publication scientifiques et de quelques soumissions d'auteurs. TrEMBL est la version protéique de la banque nucléotidique EMBL. Elle contient la traduction de toutes les parties codantes annotées d'EMBL en excluant les protéines présentes dans SWISS-PROT. PIR, qui maintenant n'existe plus, fournissait des informations organisées selon des critères taxonomiques et de similarité. Enfin, HAMAP est un projet développé par le groupe SWISS-PROT. Son but est d'annoter automatiquement les protéines provenant des projets de séquençage des génomes microbiens.

La banque contient également des collections de familles de protéines microbiennes générées par des experts et utilisées pour l'annotation automatique.

Ces banques généralistes permettent donc de centraliser toutes les séquences connues. Cependant, il existe tout de même un grand nombre d'erreurs, notamment au niveau des annotations des séquences ainsi qu'une redondance des informations dans certaines banques.

I.3.1.2. Les banques spécialistes :

Pour pallier ces inconvénients, l'augmentation exponentielle du volume, de la diversité des séquences et la diversité des études, des banques spécialisées ont été développées. Ces développements ont permis l'introduction d'informations spécifiques à chacune permettant ainsi d'avoir des banques adaptées aux besoins des utilisateurs. Elles répondent pour la plupart soit à des besoins ponctuels, soit à des besoins liés à des secteurs d'activité bien précis.

Parmi celles-ci, ont été développées des banques thématiques se consacrant à un domaine bien précis. Ainsi, certaines regroupent des données sur les structures moléculaires tridimensionnelles telles que la PDB (**BERMAN *et al.*, 2000**). D'autres s'intéressent à la structure en domaine des séquences protéiques comme la banque ProDom (**SERVANT *et al.*, 2002**). Il y en a également qui centralisent des données sur les signatures, caractéristiques de certaines protéines telle que PROSITE (**HULO *et al.*, 2006**). D'autres encore traitent des séquences et des structures d'ARN.

I.3.2. Différentes étapes et logiciels d'une étude phylogénétique :

Il existe une myriade de logiciels, implantés sur différentes machines depuis le microordinateur jusqu'à des ordinateurs les plus puissants, écrits dans tel ou tel langage, traitant tel ou tel problème. Dans le domaine de la reconstruction phylogénétique,

(FELSENSTEIN, 2004). Dans cette section, nous allons décrire quelques grandes catégories de logiciels utilisés pour les analyses phylogénétiques.

I.3.2.1. Logiciels de nettoyage :

Tout d'abord toutes les séquences ont été expurgées, à l'aide de logiciels de nettoyage comme :

❖ **Sequencher :**

Est un logiciel de bio-informatique produit par la société Gene Codes Corporation, ce logiciel assemble et aligne plusieurs séquences d'ADN contigües relativement courtes afin de créer des séquences plus longues.

❖ **BioEdit :**

Beaucoup de chercheurs dans le domaine de la biologie moléculaire ont utilisé des modèles de BioEdit pendant leur recherche. Il a été employé pour des études moléculaires de différents organismes tels que des génomes de virus (RON *et al.*, 2005 ; CHEN *et al.*, 2006).

I.3.2.2. Logiciels d'alignement :

D'une manière informelle, l'alignement de deux séquences consiste à mettre en évidence les similitudes et les différences entre les deux séquences. L'alignement de séquences a pour objectif de mettre en correspondance les portions homologues des molécules, afin de retrouver de la façon la plus cohérente possible le signal phylogénétique.

Evidemment, on s'offre aussi la possibilité de sauter quelques lettres (COMET, 1998). Donc l'alignement peut être vu comme une série de transformations permettant de passer d'une séquence à l'autre (BERARD, 2003 ; DE CRAVALHO JUNIOR, 2003). Parmi les logiciels qui font l'alignement on a :

❖ **Clustal W :**

Est un outil pour aligner la protéine multiple ou séquences de nucléotides. L'alignement est réalisé par l'intermédiaire de trois étapes : par paires alignement, génération de guide-arbre et alignement progressif. Clustal W-MPI est distribué et exécution parallèle de Clustal W. Chacune des trois étapes a été parallélisé pour réduire le temps d'exécution (LI, 2003).

❖ **Blast :**

Basic Local Alignment Search Tool ; permis d'identifier des régions de similarité locale entre les séquences. Le programme compare des séquences nucléotidiques ou protéiques et calcule la significativité des résultats.

❖ **Exonerate :**

Outil d'alignement de séquences deux à deux. Il vous permet d'aligner à l'aide d'un alignement des séquences de nombreux modèles, soit la programmation dynamique exhaustive ou une variété d'heuristiques.

I.3.2.3. Calcul des distances :

Actuellement, il y a beaucoup de programmes spécialisés pour estimer des distances évolutives entre le nucléotide ou les séquences des acides aminés et la reconstruction d'arbres phylogénétiques :

❖ DNAdist:

Calcule une matrice de distances à partir de l'alignement multiple de ClustalW.

❖ PAUP :

PAUP (Phylogenetic Analysis Using Parsimony) est un logiciel de reconstruction phylogénétique mis au point par **Swofford** entre **1989** et **1998**. C'est le logiciel le plus cité dans la littérature scientifique moderne. Il a été conçu spécialement pour faire des analyses phylogénétiques selon la méthode de parcimonie, puis il a été élargie pour inclure d'autres types d'analyses phylogénétiques telles que les méthodes de distances et les méthodes de maximum de vraisemblance, et pour réaliser des tests statistiques tels que : le bootstrap et le jackknife (**SWOFFORD, 1998**).

I.3.2.4. Logiciels de reconstruction de l'arbre phylogénétique :

❖ PHYML :

PHYML est un logiciel qui implémente une nouvelle méthode de reconstruction de phylogénie à partir de séquences en utilisant le principe de maximum de vraisemblance. Cette méthode démarre avec un arbre initial fourni par l'utilisateur ou construit à partir d'un algorithme rapide basé sur les distances, puis elle améliore cet arbre à travers des réarrangements topologiques (**GUINDON, 2003**).

❖ DnaMLK :

Est un programme d'inférence d'arbres phylogénétique, pour les séquences d'ADN, basé sur le maximum de vraisemblance avec la contrainte que les arbres inférés soient en conformité avec l'horloge moléculaire. L'horloge moléculaire stipule que toutes les feuilles de l'arbre sont équidistantes par rapport à la racine (au niveau de la longueur des branches). C'est un programme connexe à DnaMLK. Les mêmes suppositions au niveau du modèle sont valides (plus l'hypothèse de l'horloge moléculaire). Le programme utilise également un modèle de Markov caché pour inférer différents taux d'évolution pour des sites différents. Au niveau de la parallélisation du programme DNAMLK, un parallélisme de données a été

réalisé en fonction du nombre d'ensembles de données reçues en entrée et en fonction du nombre d'arbres à inférer (**FELSENSTEIN, 1993**).

I.3.2.5. Logiciels de visualisation d'arbres phylogénétiques :

Actuellement, Il existe plusieurs programmes qui permettent la visualisation des arbres, les arbres finalisés ont alors été visualisés et annotés à l'aide des programmes : PhyloDraw, NJplot, GeneTree, Phylip (DRAWTREE et DRAWGRAM), Genedoc, Dambe, Treecon, TreeView, et Spectrum (**CHOI *et al.*, 2000**).

❖ TreeView :

TreeView est l'un des programmes les plus utilisés pour visualiser des arbres phylogénétiques sous forme graphique. Ce programme est comme le programme PhyloDraw, il utilise les résultats des autres programmes (Phylip, PAUP, ClustalW...etc.), c'est à dire les matrices de distances des longueurs des branches, pour dessiner des arbres phylogénétiques sous différentes formes : forme radial, cladogramme, phylogramme...etc (**PAGE, 1996**).

❖ Dendroscope :

Un éditeur et un visualisateur d'arbre.

❖ PhyloDraw :

PhyloDraw est un logiciel de dessin d'arbres phylogénétiques. Il utilise les résultats des autres programmes de construction (Phylip, PUAP, ClustalW...etc) sous forme d'une matrice de distances, pour visualiser divers types de dendrogrammes, par exemple, les cladogrammes rectangulaires, les cladogrammes inclinés, les phylogrammes et les arbres phylogénétiques radiaux. Avec PhyloDraw, les utilisateurs peuvent ajuster la forme d'un arbre phylogénétique facilement et interactivement en employant plusieurs paramètres de commande. Ce programme peut exporter la disposition finale d'arbre vers le format d'image (**CHOI *et al.*, 2000**).

Matériels et Méthodes

Cette étude est réalisée au niveau du laboratoire de génétique des populations et biologie de conservation des populations animales de l'Université des Sciences et de la Technologie Houari Boumediene (USTHB) d'Alger.

Ce mémoire reposant en grande partie sur l'utilisation des outils de bioinformatique dans les analyses phylogénétiques de la famille des *Erinaceidae*, tel que la reconstruction de l'arbre phylogénétique, identifier des haplotypes, et voir l'évolution des espèces.

Dans ce chapitre, nous présenterons quelques fonctions des logiciels utilisés dans notre travail.

II.1. Analyses moléculaires et phylogénétiques :

L'objectif de cette partie repose essentiellement sur l'utilisation de plusieurs logiciels de bioinformatique afin d'étudier une base de données de biologie moléculaire pour trouver l'arbre phylogénétique et la position des différentes espèces de la famille *Erinaceidae*. Pour cela nous décrirons les étapes d'analyse moléculaire et la méthodologie utilisée depuis l'échantillonnage, l'extraction d'ADN, la PCR, le séquençage, jusqu'au traitement des séquences qui se trouvent au niveau de la GenBank en vue de leur utilisation dans l'étude phylogénétique.

II.1.1. Etapes d'analyses moléculaires :

II.1.1.1. Echantillonnage :

Les échantillons employés pour extraire l'ADN sont composés de : os, tissu, poils, sang ou matériel fécal.

II.1.1.2. Extraction et quantification de l'ADN :

L'extraction d'acides nucléiques d'un matériau biologique requiert la lyse cellulaire, l'inactivation des nucléases cellulaires et la séparation de l'acide nucléique souhaité de débris cellulaires. Il existe beaucoup de kits d'extraction et de purification de l'ADN qui sont commercialisés. Cependant il a été montré que la sensibilité de détection par PCR de l'ADN extrait varie selon le kit utilisé (YOSHIKAWA *et al.*, 2011).

II.1.1.3. Amplification de l'ADN par PCR :

La Polymérase Chain Réaction (PCR) (KLEPPE *et al.*, 1971) est une méthode in vitro d'amplification de séquences spécifiques d'ADN en un très grand nombre de copies. Elle permet d'amplifier une séquence spécifique d'ADN dans un mélange réactionnel approprié.

II.1.1.4. Séquençage :

Le séquençage de l'ADN consiste à déterminer l'ordre des nucléotides sur la molécule d'ADN précédemment amplifié par PCR. La méthode utilisée aujourd'hui (établie par SANGER en 1977) repose sur l'utilisation de didésoxynucléotides, qui bloquent la synthèse de l'ADN après leur incorporation.

II.2. Etude de la GenBank :

II.2.1. Définition de la GenBank :

La GenBank est une collection annotée de toutes les séquences d'ADN publiquement disponibles. Cette banque est mise à jour régulièrement grâce à des échanges quotidiens de séquences avec la banque européenne EMBL et la banque japonaise DDBJ.

Toutes les séquences étudiées de la famille des *Erinaceidae* dans ce travail ont été téléchargées à partir de la GenBank, nous passons par plusieurs étapes distinctes :

1. Construire l'URL suivante : <https://www.ncbi.nlm.nih.gov/genbank/>
2. Choisir recherche par taxonomie.
3. Ecrire le nom de notre genre « *Erinaceidae* » (Figure 6).

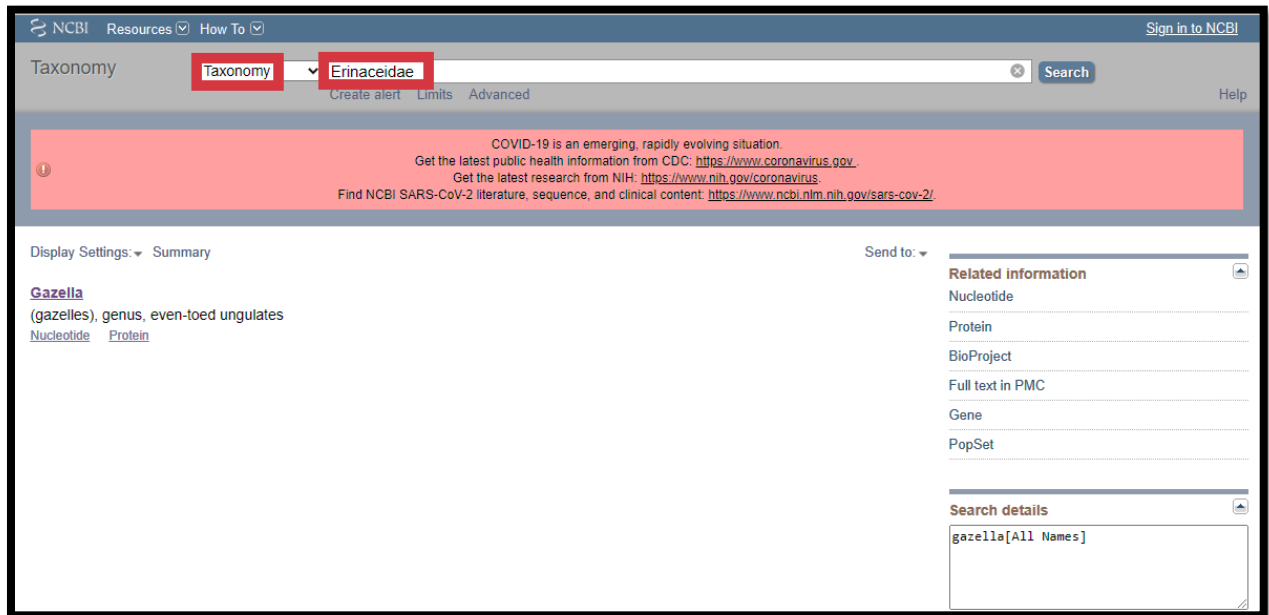


Figure 6 : Les étapes initiales pour récupérer les séquences de la GenBank.

4. En entrant sur Erinaceidae, nous obtenons toutes les espèces qui appartiennent à ce genre (Figure 7).

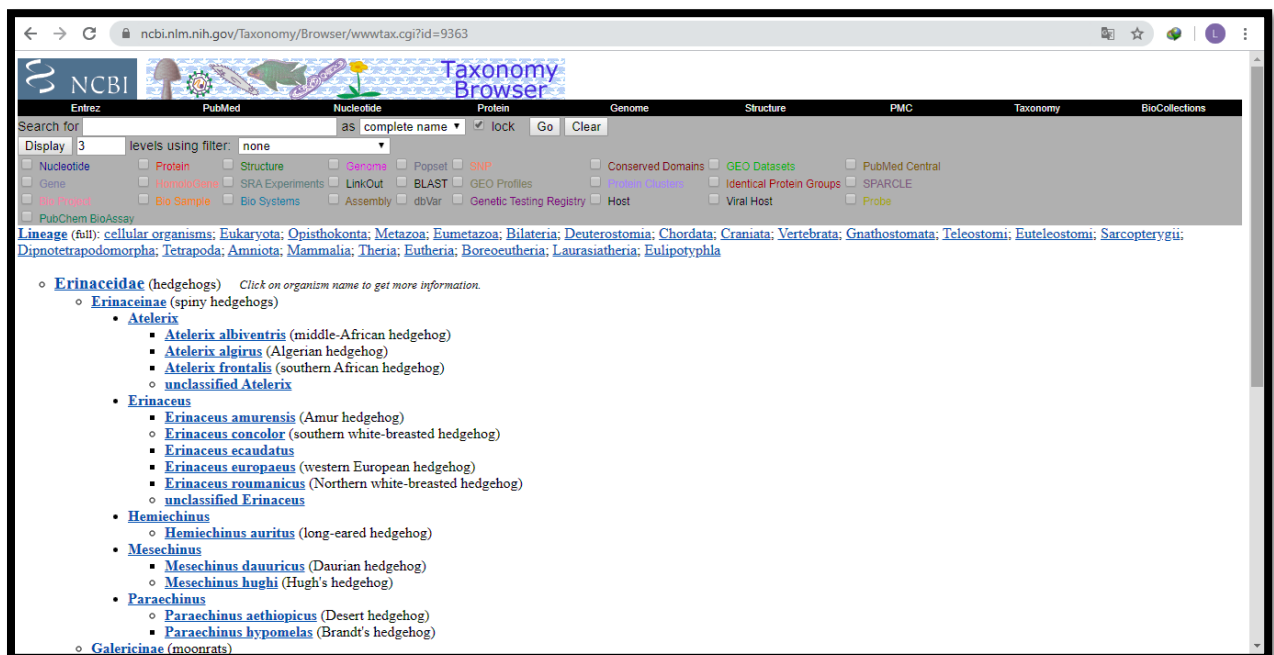


Figure 7 : Les 21 espèces de la famille des *Erinaceidae* obtenues dans la GenBank.

5. Nous avons analysé toutes les espèces une par une pour identifier le nombre des séquences nucléotidiques trouvées dans cette banque (Figure 8).

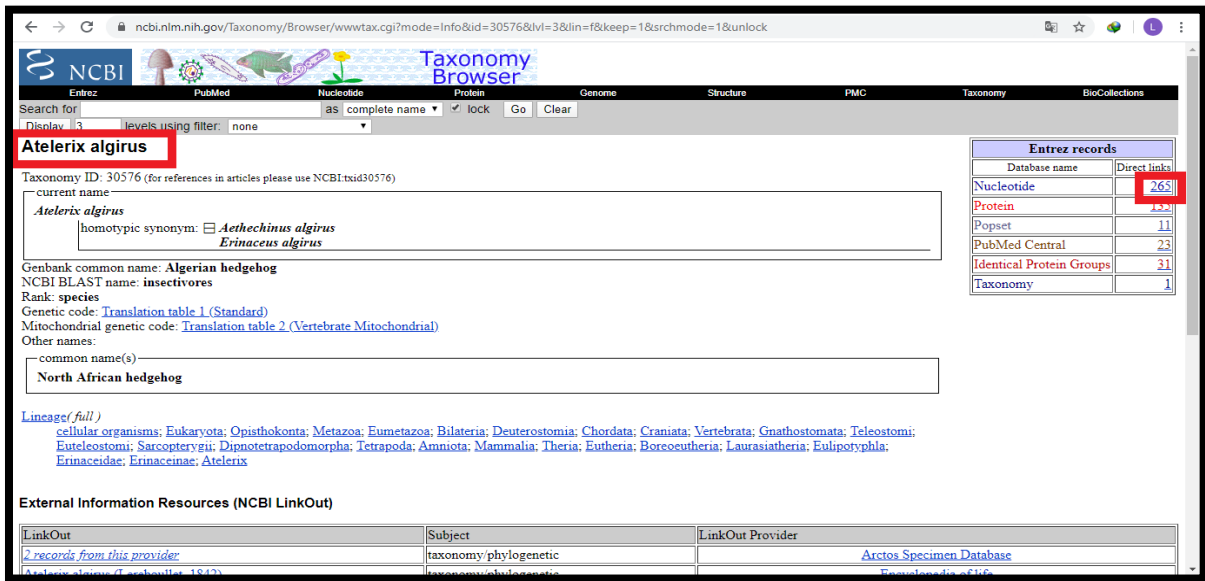


Figure 8 : Exemple des séquences nucléotidiques trouvées pour l'espèce *Atelerix algirus*.

6. Les séquences représentatives de chacune des 21 espèces ont été téléchargées sous format FASTA à partir de la "GenBank" (Figure 9).

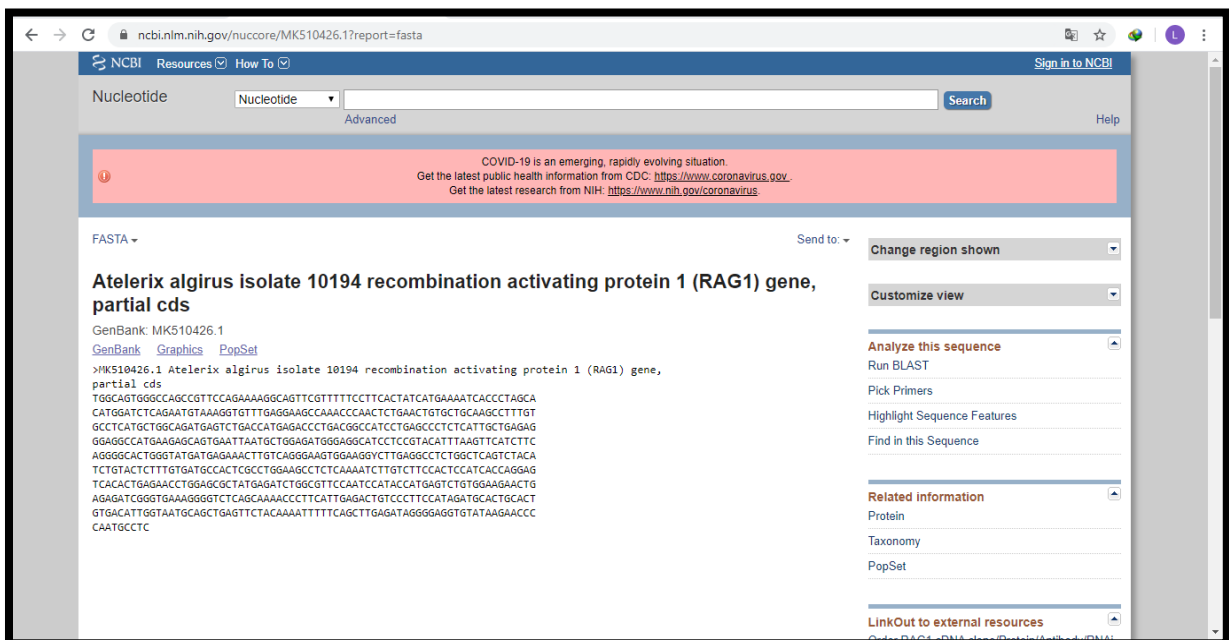


Figure 9 : Exemple d'une séquence nucléotidique au format FASTA.

La base de données a été construite pour sélectionner les isolats à étudier et générer les fichiers de sortie correspondant au format informatique « FASTA ». Ces fichiers peuvent alors être directement utilisés dans plusieurs logiciels de traitement et d'analyses des séquences.

II.2.2. Traitement des séquences de la GenBank:

Pour réaliser une phylogénie, un bon nombre de programmes informatiques sont disponibles, pour traiter les séquences environ 92 paquets de phylogénie et 54 web server gratuits parmi eux : PHYLIP, PAUP*, MEGA, Phylo_win, ARB, DAMBE, PAL, Bionumerics, Mesquite, PaupUp, BIRCH, Bosque, EMBOSS, phangorn, Bio++, ETE, DendroPy, SeaView, Crux.

Nous avons choisi le logiciel MEGA version 5 (TAMURA *et al.*, 2011) ; MEGA a été développé pour faciliter des analyses statistiques d'évolution moléculaire à l'aide des PCs. Le MEGA est conçu pour réaliser diverses analyses statistiques dans un programme et pour produire des résultats dans des sorties de publication-qualité (KUMARA *et al.*, 1994).

Nous pouvons alors procéder à l'analyse phylogénétique à partir des séquences nucléotidiques de chaque espèce de la famille des *Erinaceidae* par différentes étapes :

Le programme FaBox a été employé pour convertir les fichiers, de la page principale (Figure 10) l'utilisateur peut choisir une liste de services, chacune va effectuer une tâche simple. En outre, il y a des liens aux services de conversion où l'utilisateur peut convertir n'importe quel format de données en format FASTA.

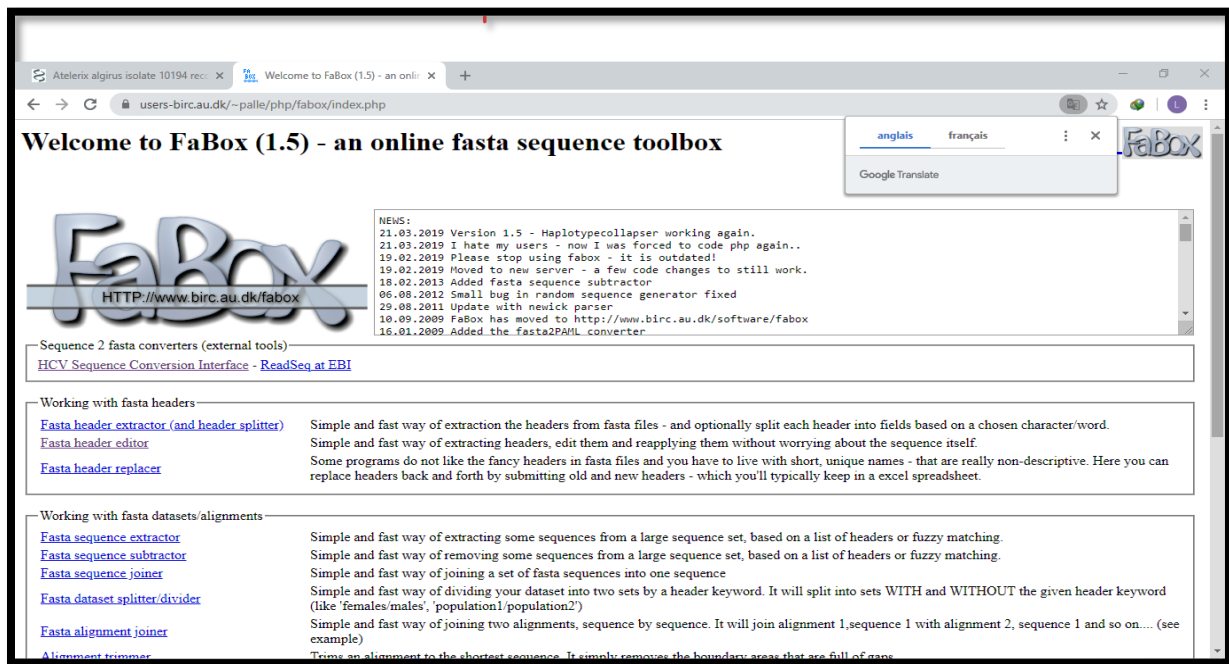


Figure 10 : La page principale du programme FaBox.

L'utilisateur peut faire entrer des données en téléchargeant un fichier de données ou simplement en collant des données directement dans le web browser.

On s'attend à ce que la boîte à outils se développe sur la base des demandes des services particuliers et des convertisseurs à l'avenir (Figure 11).

Housekeeping Retrieving data Analysis Tools Links Search

To our users Please note that the HCV database site is no longer funded. We try to keep the database updated and the tools running, but unfortunately, won't be manually curated either.

Format Converter

Purpose: Convert an input sequence or alignment to a user-specified format. See [Explanation](#).

Input

Paste your sequences here [Sample Input]

Or upload your file: Choisir un fichier | Aucun fichier choisi

Options

Select input format Automatic

Select output format Fasta

Output line width 50 characters as wide as possible

Enforce sequence name uniqueness

Convert Genbank to GFF3

Replace IUPAC codes with Ns

Submit Reset

Figure 11 : La conversion des séquences sous forme FASTA par l'utilisation de FaBox.

Alignement cropper va réduire l'alignement (en format FASTA) à la longueur de la séquence la plus courte (suppression des gaps dans le début et la fin de l'alignement). Appuyez sur "show example" pour l'essayer sans soumettre notre propre séquence (Figure 12).

Alignment trimmer (fasta files)

Upload or paste your fasta formatted sequences

Choose File No file chosen

Action

Trim alignment Show example

Figure 12 :Alignement trimmer (format FASTA).

Les alignements multiples des séquences nucléotidiques et protéiques ont été réalisés au moyen du logiciel MEGA version 5 (TAMURA *et al.*, 2011)(Figure 13) par la méthode d'alignement progressif ClustalW (THOMPSON *et al.*, 1994)(Figure 14).

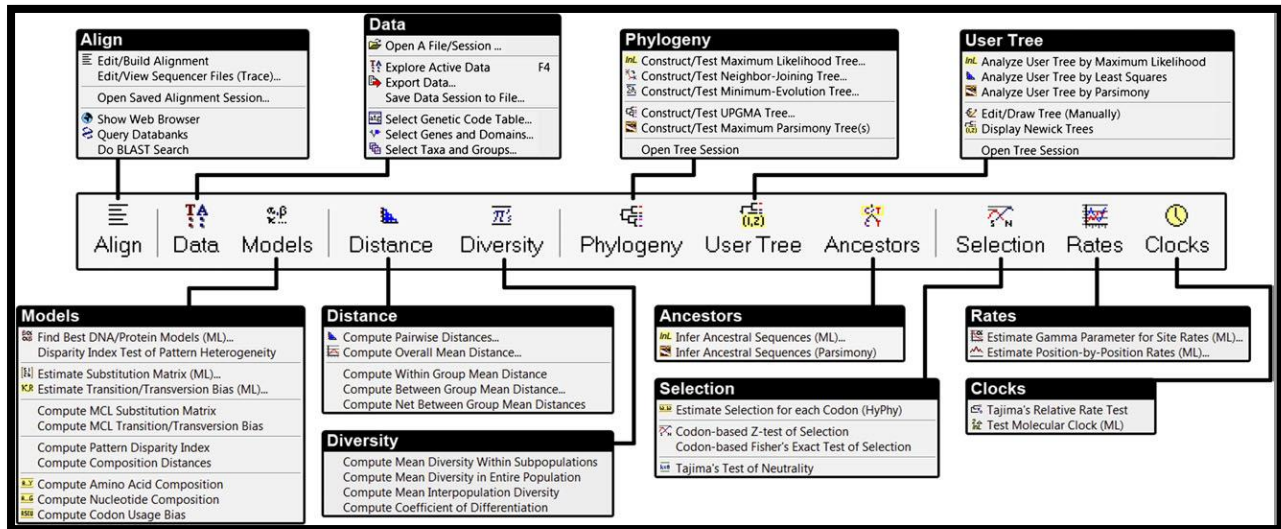


Figure 13 :Barre d'outils de MEGA5.

Cette méthode dite heuristique parce qu'elle propose un alignement réalisable mais pas nécessairement optimal, nous avons choisi ce type d'alignement par ce qu'il est très utile et permet d'étudier les différences et les similitudes entre les séquences d'un même marqueur moléculaire.

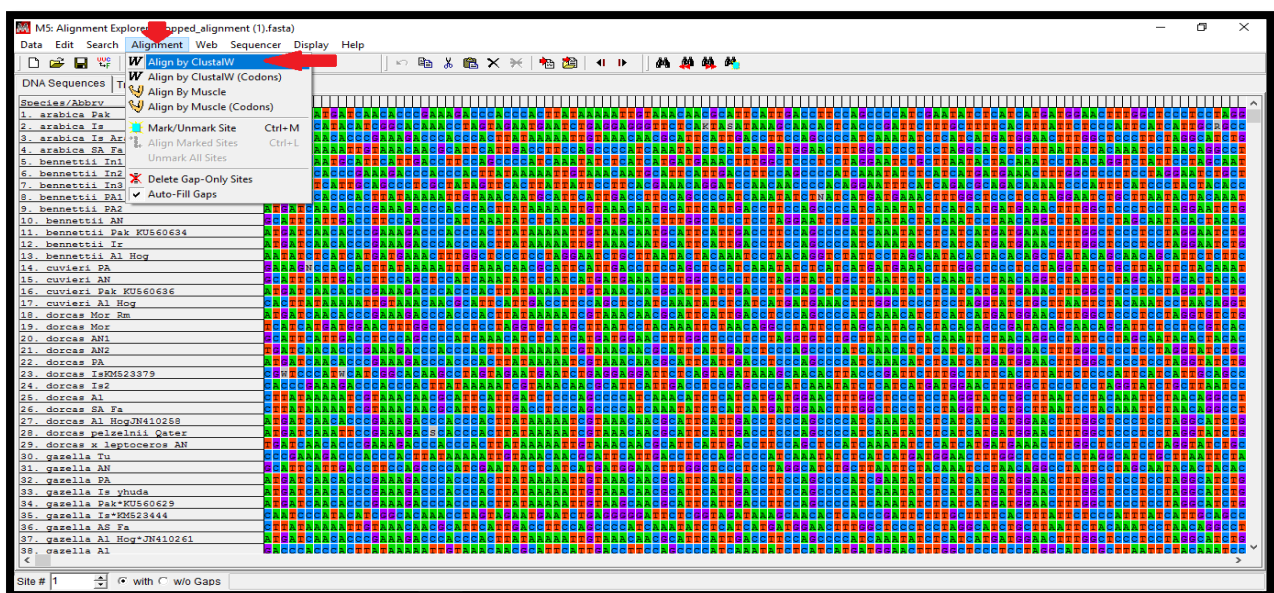


Figure 14 : Le programme ClustalW utilisé pour l'alignement multiple des séquences du marqueur moléculaire choisi.

Une fois les séquences alignées elles sont ensuite comparées et servent à la construction d'arbres phylogénétiques en utilisant l'une des méthodes fondées sur les distances, la méthode UPGMA (Figure 15), la plus simple, basée sur l'hypothèse que les taux de mutation et donc les vitesses d'évolution sont identiques sur les différentes branches de l'arbre, en utilisant le modèle de « Kimura 2-paramètres ».

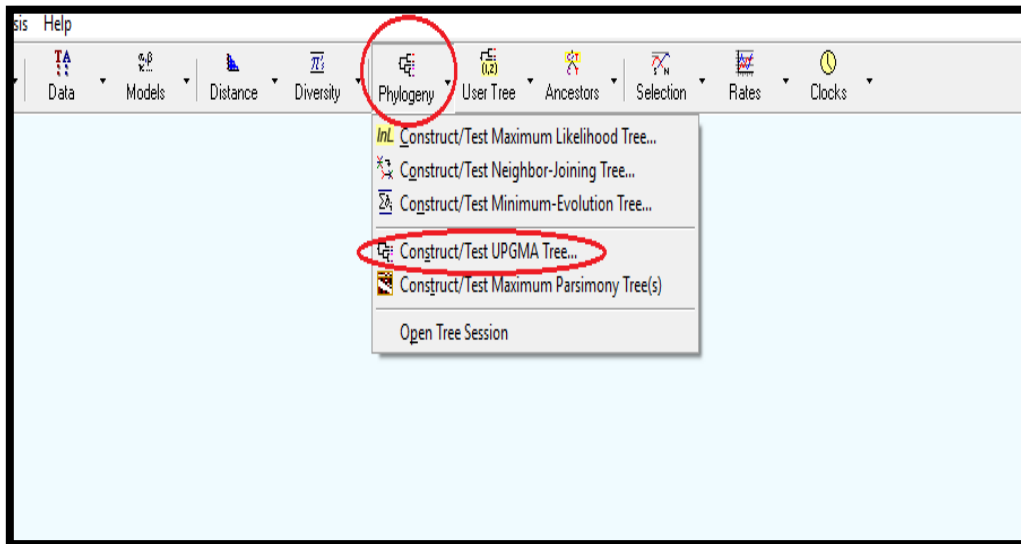


Figure 15 : La méthode UPGMA utilisée pour la construction de l'arbre.

Dans ce modèle, les transitions et les transversions ne sont pas équivalentes en termes de proportions et leur fiabilité a été évaluée par le programme "Bootstrap" qui a été appliqué avec un nombre de répétitions égal à 1000 (Figure 16). Les répliquations sont ramenées en pourcentages et ils sont indiqués au niveau des nœuds.

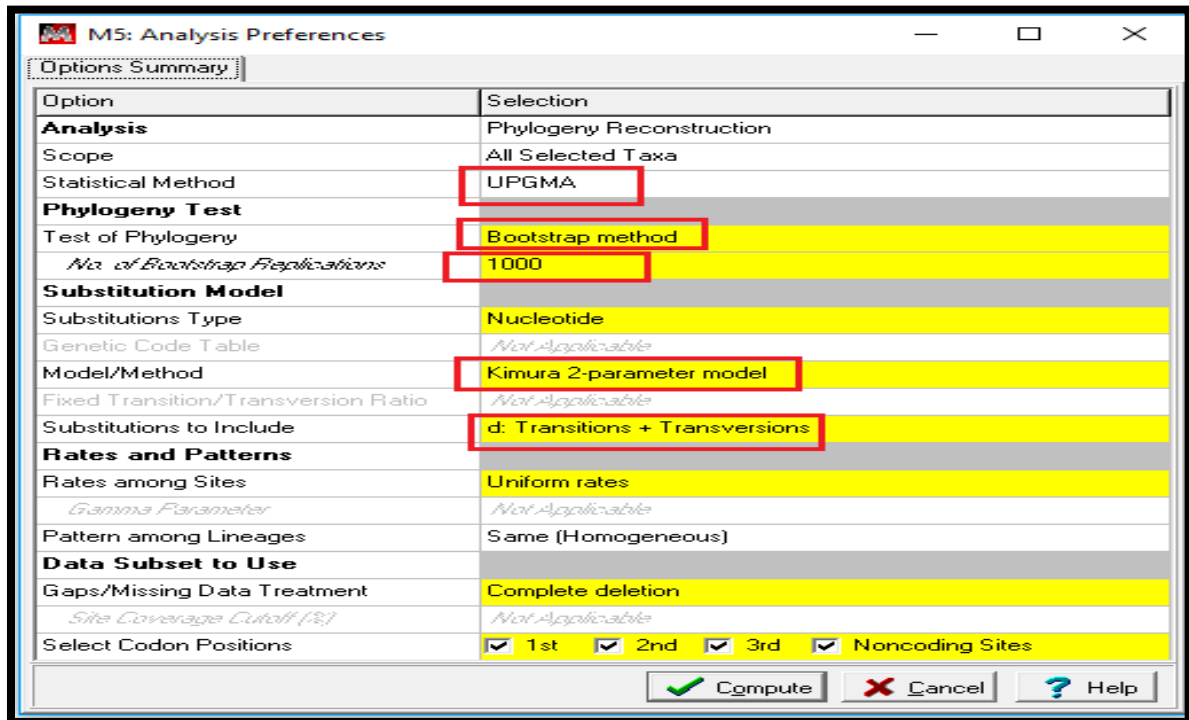


Figure 16 : Les paramètres utilisés pour la reconstruction phylogénétique.

Après avoir expurgé les alignements des séquences erronées et des recombinants, la composition nucléotidique des alignements a été déterminée à l'aide du logiciel DnaSP version 5 (LIBRADO *et al.*, 2009). Ce logiciel permet d'analyser le polymorphisme des séquences d'un alignement. Le nombre de sites ségrégatifs (sites pour lesquels au moins une substitution est observée au sein de l'alignement) a été calculé, ainsi que le nombre de transversions et de transitions.

Diversité mitochondrique pour l'ensemble des séquences et à chaque population séparément ont été évalués pour plusieurs paramètres comprenant le nombre de haplotypes (h), et les diversités des haplotypes (H_d) et des nucléotides (p) (LIBRADO *et al.*, 2009).

DnaSP peut également effectuer plusieurs tests de neutralité. En outre, il peut estimer les intervalles de confiance de certaines statistiques d'essai par le coalescent. DnaSP accepte cinq formats de fichier de données d'entrée : FASTA, MEGA, NBRF/PIR, CONNEXION, et PHYLIP. Les résultats des analyses sont présentés sous forme de tableau et de graphes (Figure 17).

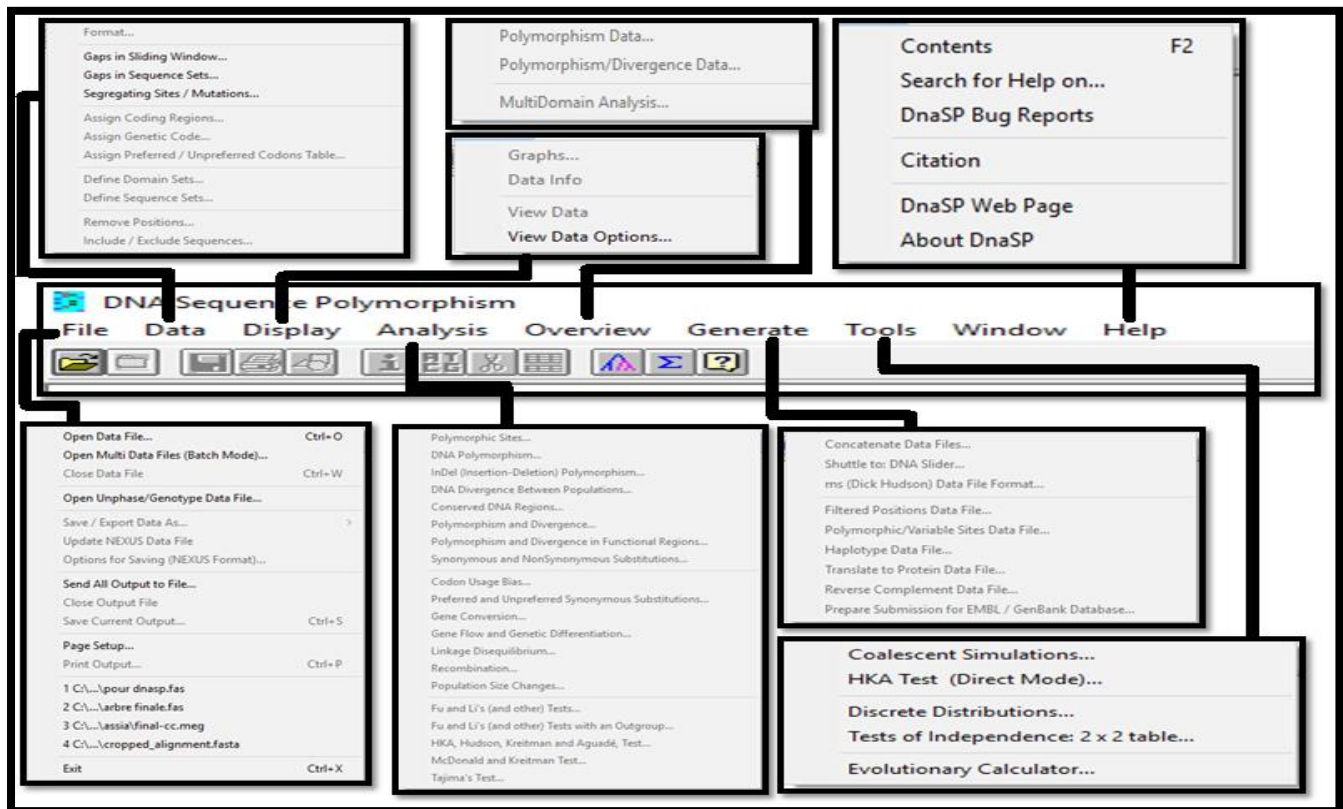


Figure 17 : Barre d'outils de DnaSP.

Le DnaSP produit des fichiers de données sur les haplotypes. Les résultats peuvent être enregistrés sur des fichiers NEXUS ou de données de Roehl(Figure 18).

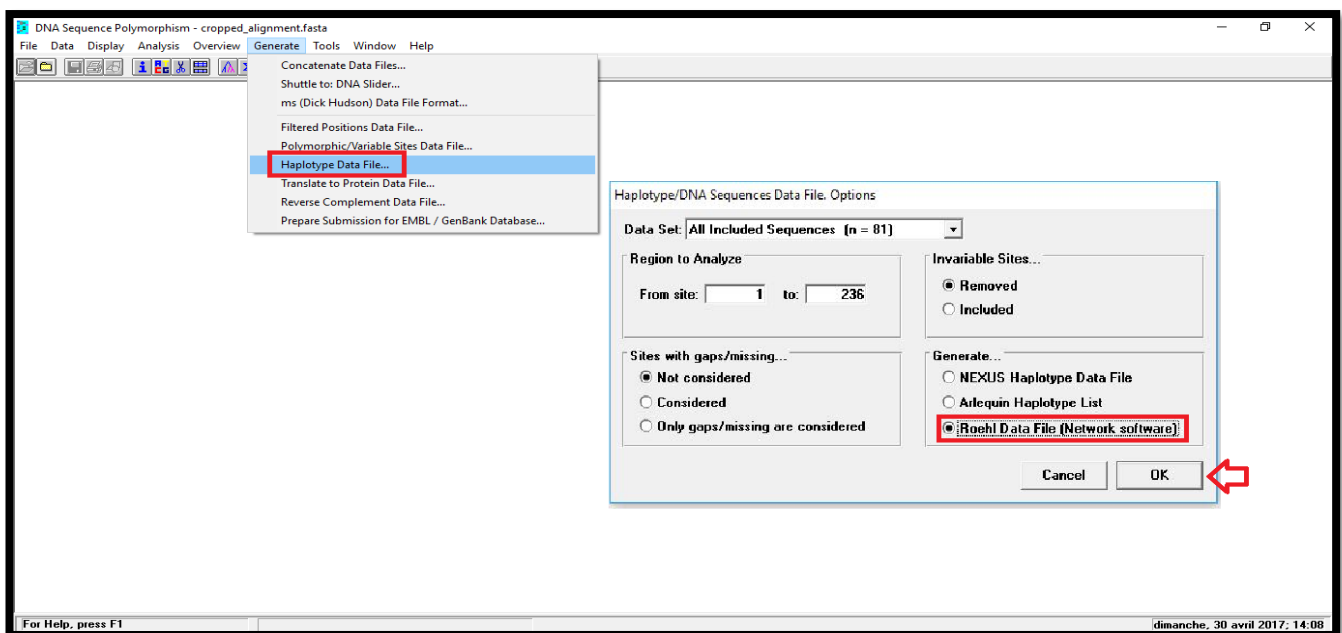


Figure 18 :Interface utilisateur de DnaSP.

Après vérification des haplotypes, la pertinence de l'information génétique contenue dans le jeu de données a été contrôlée à l'aide du logiciel DAMBE, c'est un logiciel qui intègre des paquets pour manipulation et analyse des séquences comportant une interface conviviale et un grand choix de fonctions analytiques de bioinformatique, phylogénétiques, et génomique descriptive et comparative (SALEMI *et al.*, 2003; FELSENSTEIN, 2004; LEMEY *et al.*, 2009). Permet de convertir, manœuvrant, statistiquement et graphiquement décrivant (Figure 19).

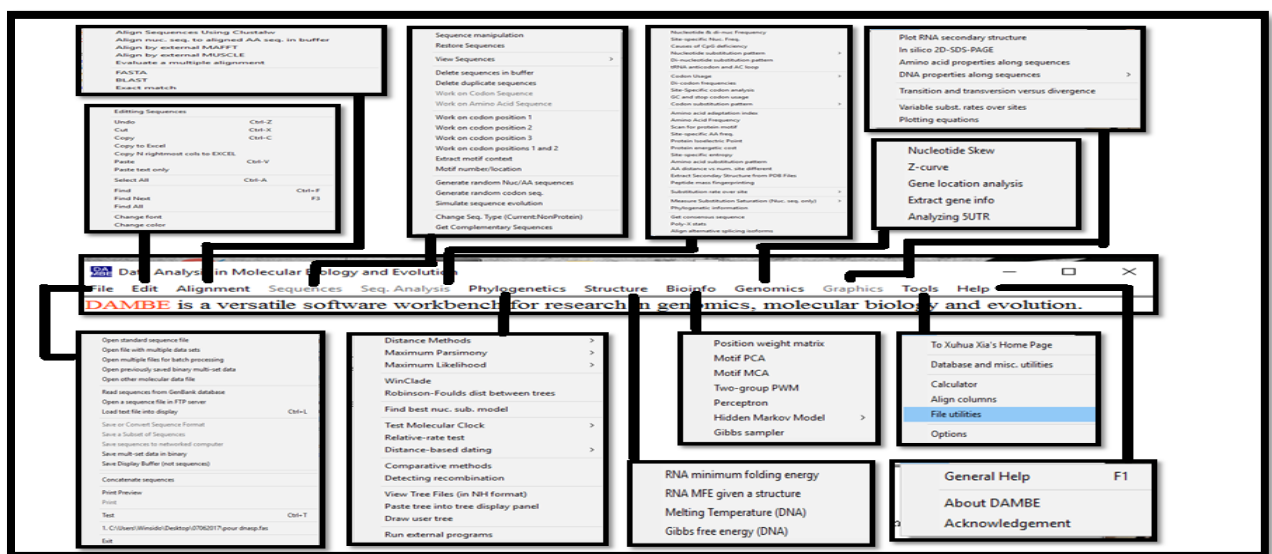


Figure 19 :Barre d'outils du programme DAMBE

DAMBE peut représenter graphiquement le taux de C/G et A/T (Figure 20).

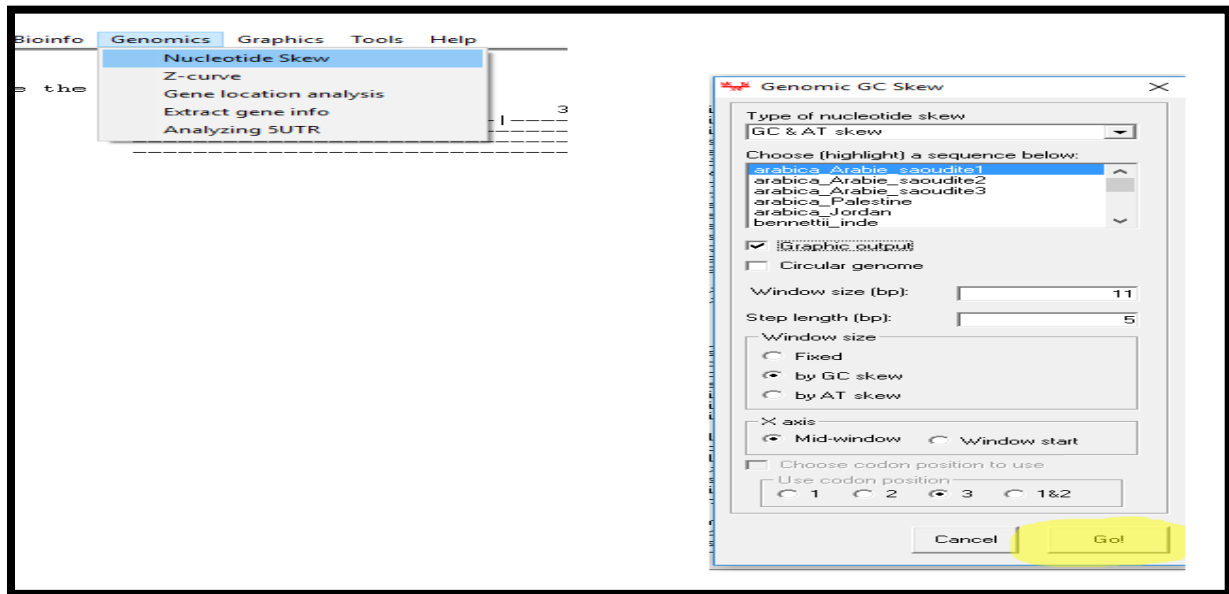


Figure 20 :Barre d'outils pour tracer les graphes de C/G et A/T.

DAMBE lit et écrit presque tous formats de données moléculaires utilisés (TableauII).

Tableau II : Les formats de données utilisés par logiciel DAMBE (XIA, 2001).

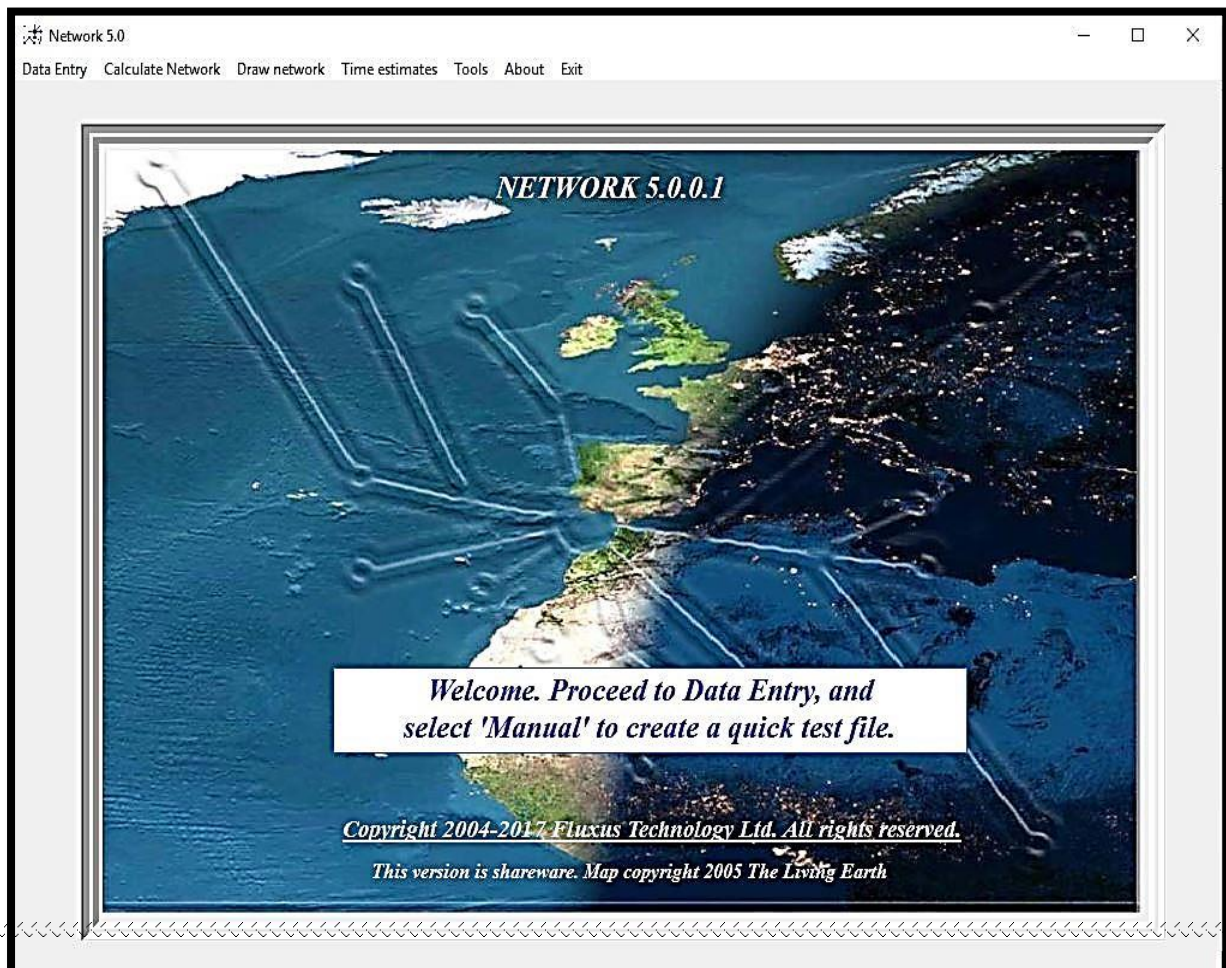
Format des séquences	Lire par DAMBE	Convertir par DAMBE
PHYLIP	+	+
PAUP	+	+
MEGA	+	+
CLUSTAL	+	-
FASTA	+	+
GenBank	+	+
GCG	+	+
MSF	+	+
DNA strider	+	+
PAML	+	-
RST,MPa	-	+
PHYLTEST	+	+

IG/Stanford	+	+
NBRF	+	+
EMBL	+	+
FITCH	+	+
PIR/CODATA	+	+
Plain text b	+	+
Allele frequency	+	-
Distance matrix	+	+

- ❖ *a* : Format des séquence pour stocker des séquences générales et des séquences héréditaires reconstruire des séquences générales.
- ❖ *b* : La mise en forme de texte d'un ordre par dossier des programmes comme le navigateur des séquences et l'étoile d'ADN.

DAMBE est un programme convivial pour la plate-forme de Windows qui comporte une suite des caractéristiques uniques aussi bien que la capacité d'exécuter la plupart des routines (analyses de données en biologie moléculaire, écologie, et évolution).

Pour compléter les analyses de cette étude on a terminé avec les réseaux de Median-



Joining (MJ) (BANDELT *et al.*, 1999), ont été construits avec le programme NETWORK 5.0 (Figure 21). Les méthodes de network sont maintenant disponibles et sont très utilisées par les chercheurs dans des domaines aussi différents que la phylogéographie (SCHAAL *et al.*, 2000).

Figure 21 : La page d'accueil de NETWORK 5.0.

Pour le calcul du réseau initial, nous passons par différentes étapes :

1. Calculate NETWORK menu.
2. NETWORK calculations.
3. Median Joining.
4. File.
5. Open.
6. DNA data file (rdf).
7. Sélectionnez le dossier de rdf dont nous avons manuellement créé dans le rédacteur de données de NETWORK ou exporté du programme DnaSP(Figure 22).

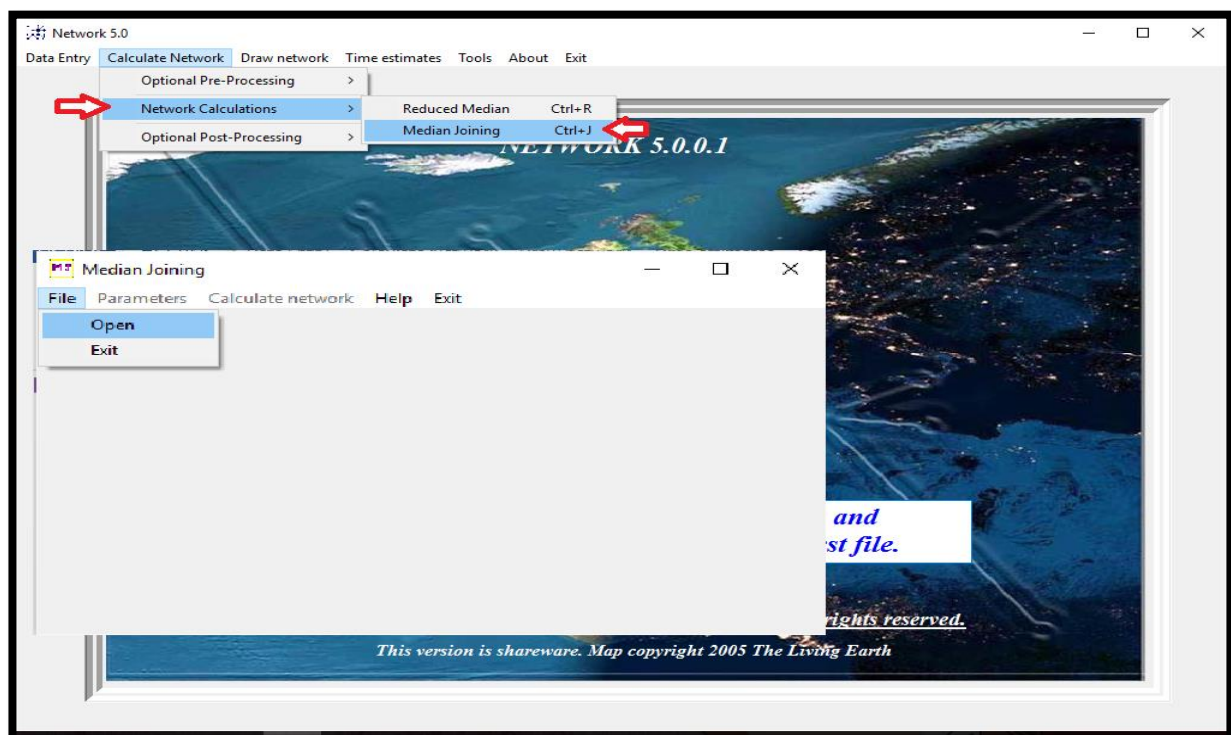
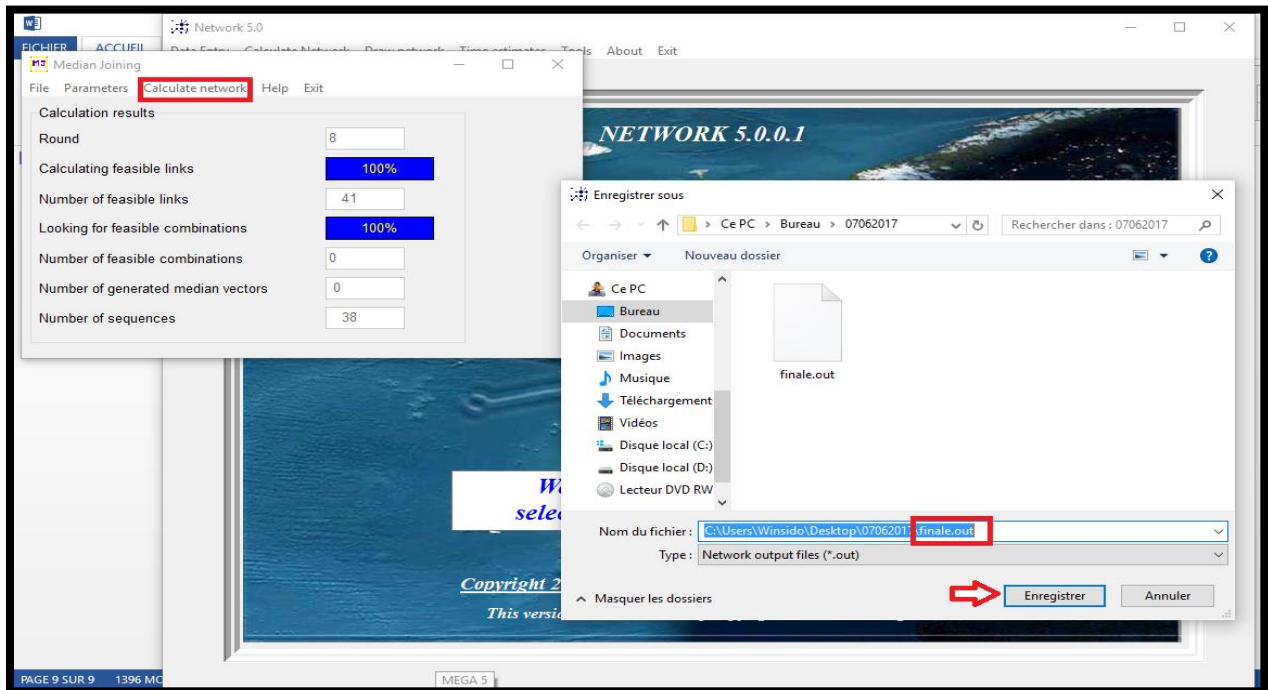


Figure 22 : Ouverture du fichier de données de rdf pour le calcul de MJ NETWORK.

Après l'ouverture du dossier, le NETWORK peut produire des fichiers sous forme *.out afin



que le calcul NETWORK peut reconstruire le réseau (Figure 23).

Figure 23 : Un fichier sous forme *.out a été sauvegardé.

Le NETWORK a présenté la phylogénétique des espèces avec les cycles graphiques sous forme d'un réseau par la fonction Draw NETWORK (Figure 24).

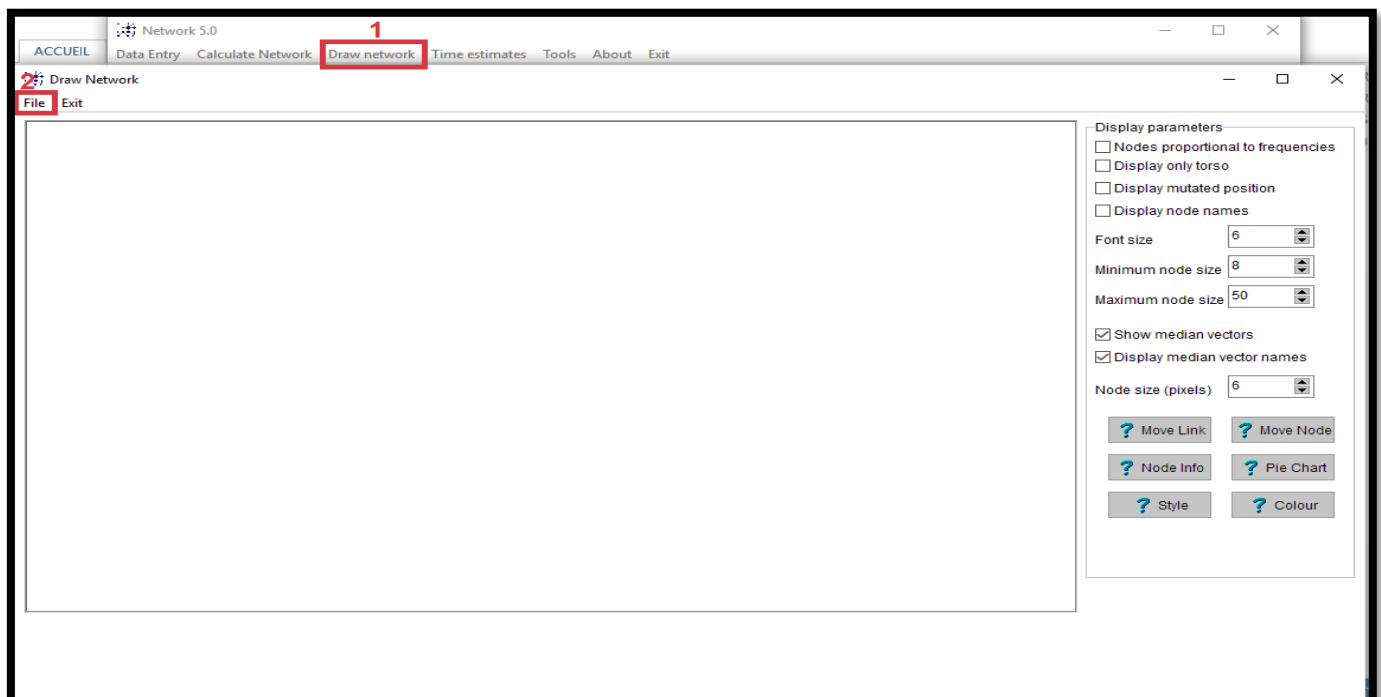


Figure 24 :Reconstruction de réseau haplotypique.

Résultats et Discussion

Cette partie est basée sur la comparaison entre les espèces de la famille des *Erinaceidae* sur le plan moléculaire à partir des données de la GenBank, afin de déterminer la position de différentes espèces dans l'arbre phylogénétique et voir leur évolution.

III.1. Résultats de la GenBank :

21 espèces appartenant à la famille des *Erinaceidae*, ont été récupérées auprès de la GenBank (Tableau III). Les espèces qui ont été utilisées dans ces analyses phylogénétiques sont : *Atelerix albiventris*, *Atelerix algirus*, *Atelerix frontalis*, *Erinaceus amurensis*, *Erinaceus concolor*, *Erinaceus ecaudatus*, *Erinaceus europaeus*, *Erinaceus roumanicus*, *Hemiechinus auritus*, *Mesechinus dauuricus*, *Mesechinus hughi*, *Paraechinus aethiopicus*, *Paraechinus hypomelas*, *Echinosorex gymnura*, *Hylomys megalotis*, *Hylomys parvus*, *Hylomys suillus*, *Hylomys cf. suillus JJS-2016*, *Neohylomys hainanensis*, *Neotetracus sinensis* et *Podogymnura truei* localisées dans différentes régions géographiques.

Les séquences de chaque espèce extraite à partir de la GenBank, leur type de gène et leurs origines géographiques sont mentionnées dans le tableau en annexe (Tableau V).

Il y a plusieurs séquences identiques à 100%, pour cela nous ne prendrons en considération qu'une seule séquence représentative.

Tableau III : Les séquences d'ADN nucléaires et mitochondriales de la famille des *Erinaceidae* extraites de la GenBank.

Espèce	Nombre total des séquences	Séquences d'AND nucléaire	Séquences d'ADN mitochondrial	Nombre total de Cytb	Nombre total de BRCA1	Les séquences non identifiées
<i>Atelerix albiventris</i>	663	599	22	5	2	42
<i>Atelerix algirus</i>	265	127	138	61	4	/
<i>Atelerix frontalis</i>	8	6	2	1	2	/
<i>Erinaceus amurensis</i>	41	17	24	3	4	/
<i>Erinaceus concolor</i>	132	50	82	17	2	/
<i>Erinaceus ecaudatus</i>	3	3	0	0	0	/
<i>Erinaceus europaeus</i>	227868	191857	285	17	9	35726
<i>Erinaceus roumanicus</i>	293	211	82	3	2	/

<i>Hemiechinus auritus</i>	43	15	27	18	5	1
<i>Mesechinus dauuricus</i>	39	21	18	7	2	/
<i>Mesechinus hughi</i>	9	2	7	3	0	/
<i>Paraechinus aethiopicus</i>	77	37	40	6	6	/
<i>Paraechinus hypomelas</i>	18	11	7	4	3	/
<i>Echinosorex gymnura</i>	15	4	10	6	0	1
<i>Hylomys megalotis</i>	10	5	5	2	1	/
<i>Hylomys parvus</i>	12	4	8	4	1	/
<i>Hylomys suillus</i>	31	10	20	4	3	1
<i>Hylomys cf. suillus</i> JJS-2016	8	8	0	0	2	/
<i>Neohylomys hainanensis</i>	13	0	13	4	0	/
<i>Neotetracus sinensis</i>	23	7	15	3	2	1
<i>Podogymnura truei</i>	32	28	4	2	1	
Total : 21 espèces	229.603	193022	809	170	58	35775

D'après le tableau ci-dessus, nous remarquons qu'il y a :

- ✓ 193022 séquences d'ADN génomique : BRCA1, GHR...ect
- ✓ 809 séquences d'ADN mitochondriale : Cytb, D-loop ...ect

Afin d'établir des liens de parenté entre les espèces de la famille des Erinaceidae, notre étude est basée sur deux gènes ;

1. un gène de l'ADN mitochondrial : Le Cytochrome b qui fonctionne dans le cadre de la chaîne de transport d'électrons et est la sous-unité principale des complexes transmembranaires (Figure 25).
2. un gène de l'ADN génomique : Breast Cancer 1 (BRCA1) est un gène suppresseur de tumeur impliquée dans la réparation des dommages de l'ADN, l'ubiquitination, la régulation transcriptionnelle, ainsi que dans d'autres fonctions (Figure 26).

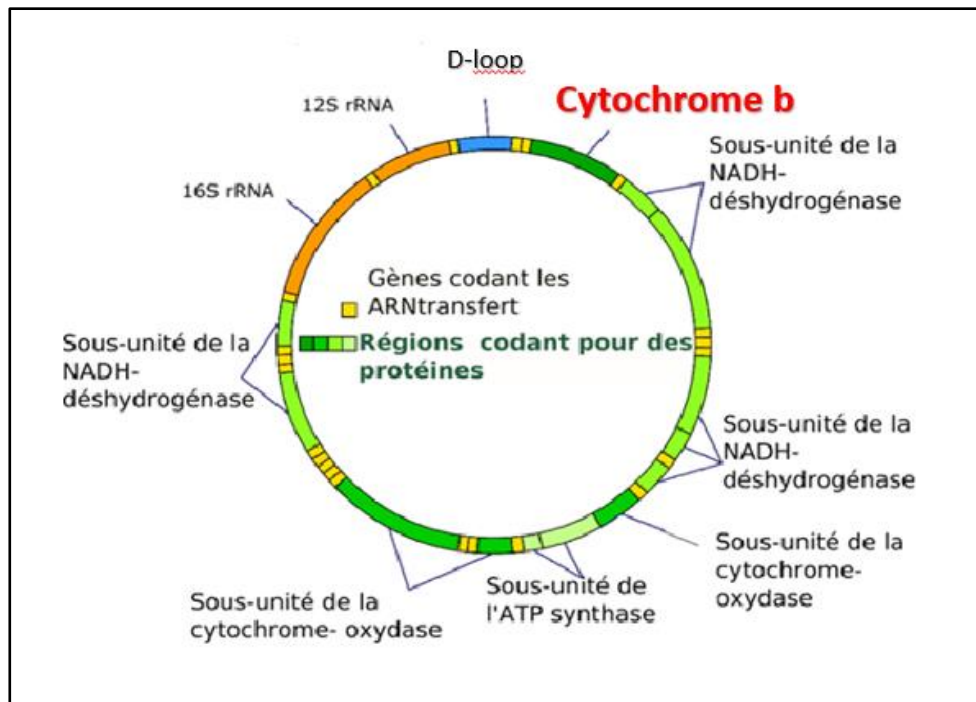


Figure 25: Schéma montrant l'organisation de l'ADN mitochondrial et la position de Cytb.

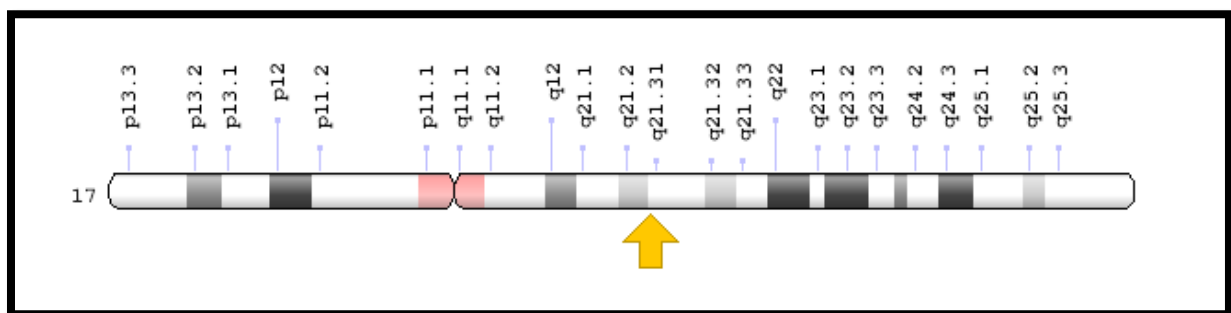


Figure 26 : La position du gène BRCA1 sur le chromosome 17(17q21.31) ; localisation cytogénétique sur le bras long (q) du chromosome 17 à la position 21.31.

III.2. Analyses phylogénétiques :

III.2.1 Construction des arbres phylogénétiques :

La construction de l'arbre phylogénétique a été réalisée à l'aide du programme MEGA5. Premièrement, nous avons choisi les séquences représentatives de Cytb et BRCA1 en se basant sur leurs origines géographiques :

➤ **Cytb:**

- *Atelerix algirus*: Au total, il y a 61 séquences, nous avons choisi 3 séquences de l'Algérie de la même taille (743pb) et une séquence de l'Espagne (1132pb).
- *Atelerix albiventris* : Au total, il y a 5 séquences, nous avons choisi une seule séquence du Sénégal (1140pb).
- *Atelerix frontalis* : Il y a une seule séquence du Sud d'Afrique (1140pb).
- *Erinaceus amurensis* : Nous avons au total 4 séquences, on a choisi deux séquences de la Chine (1140pb) et une séquence de la Russie (1134pb).
- *Erinaceus concolor* : Au total il y a 17 séquences, nous avons choisi deux séquences de la même taille (1140pb) une d'Israël et l'autre séquence d'Abkhazie.
- *Erinaceus europaeus* : Il y a 20 séquences, on a choisi six, deux séquences du Portugal (743pb), deux séquences de la Russie (1140pb), une séquence d'Italie (1140pb) et une séquence de la Grande-Bretagne (1140pb).
- *Erinaceus roumanicus* : Au total il y a 3 séquences, on a choisi deux séquences de la Russie de différentes tailles (1127pb ; 1128pb).
- *Hemiechinus auritus* : Il y a 20 séquences, on a choisi 15 séquences, trois séquences du Koweït (743pb), quatre séquences de Mongolie (1140pb), deux séquences de Turkménistan (1140pb), deux séquences de Ouzbékistan (1140pb), une séquence de Kazakhstan (1140pb), une séquence de la Russie (1140pb), une séquence d'Égypte (1140pb) et une séquence de Jordanie (1140pb).
- *Mesechinus dauuricus* : Il y a au total 7 séquences, on a choisi cinq séquences de la Chine (1140pb) et deux séquences de la Russie (1140pb).
- *Mesechinus hughi* : Il y a 3 séquences, on a choisi deux séquences de la Chine (1140pb).
- *Paraechinus aethiopicus* : On a trouvé au total 6 séquences de la même taille (1140pb) et on a choisi, deux séquences de Qatar, deux séquences d'Algérie, une séquence de l'Arabie Saoudite et une séquence des Émirats Arabes Unis.
- *Paraechinus hypomelas* : On a trouvé au total 4 séquences, on a choisi deux séquences des Émirats Arabes Unis (743pb), une séquence du Yémen (743pb) et une séquence de Turkménistan (1140pb).
- *Hylomys megalotis* : On a choisi une séquence du Laos (885pb).
- *Hylomys parvus* : On a trouvé 6 séquences et on a choisi une séquence (1123pb).
- *Hylomys suillus* : Il y a 4 séquences, on a pris une séquence de la Chine (1140pb)

Neotetracus sinensis : Il y a 5 séquences, on a choisi une seule du Vietnam (1140pb).

Nous avons étudié 57 séquences sélectionnées de Cytb pour faire l'analyse phylogénétique.

Nous avons utilisé le FaBox pour convertir notre fichier de séquences sous le format FASTA afin d'organiser les bases azotiques qui forment une séquence d'une seule ligne (Figure 27).

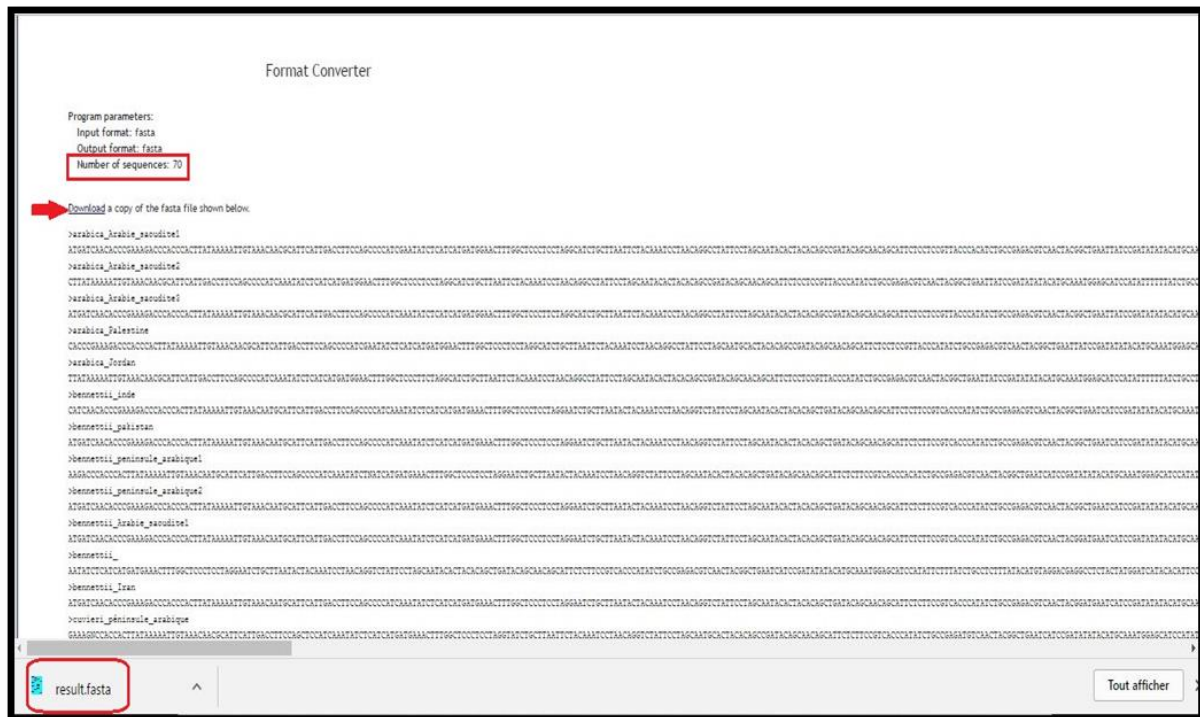


Figure 27 : Les séquences de Cytb ont été convertit sous format FASTA à partir de bloc note grâce à FaBox.

L'arbre phylogénétique a été construit avec la méthode UPGMA. La distance utilisée est celle du Kimura-2p (= 2 paramètres) qui intègre le fait que les transitions et les transversions ne sont pas équiprobables lors de l'évolution des séquences, notamment pour l'ADN mitochondrial où les transitions peuvent représenter 90% des mutations (**KIMURA, 1980**).

Dans la réalité biologique, les transitions représentent la majorité des substitutions nucléotidiques. Ainsi, lorsque la distance génétique entre deux séquences augmente, le nombre de transitions et de transversions augmente proportionnellement, le nombre de transitions étant toujours supérieur. Cependant, dans le cas de séquences de plus en plus éloignées, la saturation des substitutions peut être atteinte et les transversions devenir plus nombreuses que les transitions.

Il faut 1000 répliquions de "Bootstrap" pour que cette méthode soit statistiquement valable.

Les relations entre les séquences de Cytb de la famille des *Erinaceidae* ont été représentées à l'aide d'un arbre réalisé grâce au programme MEGA5 (Figure 28).

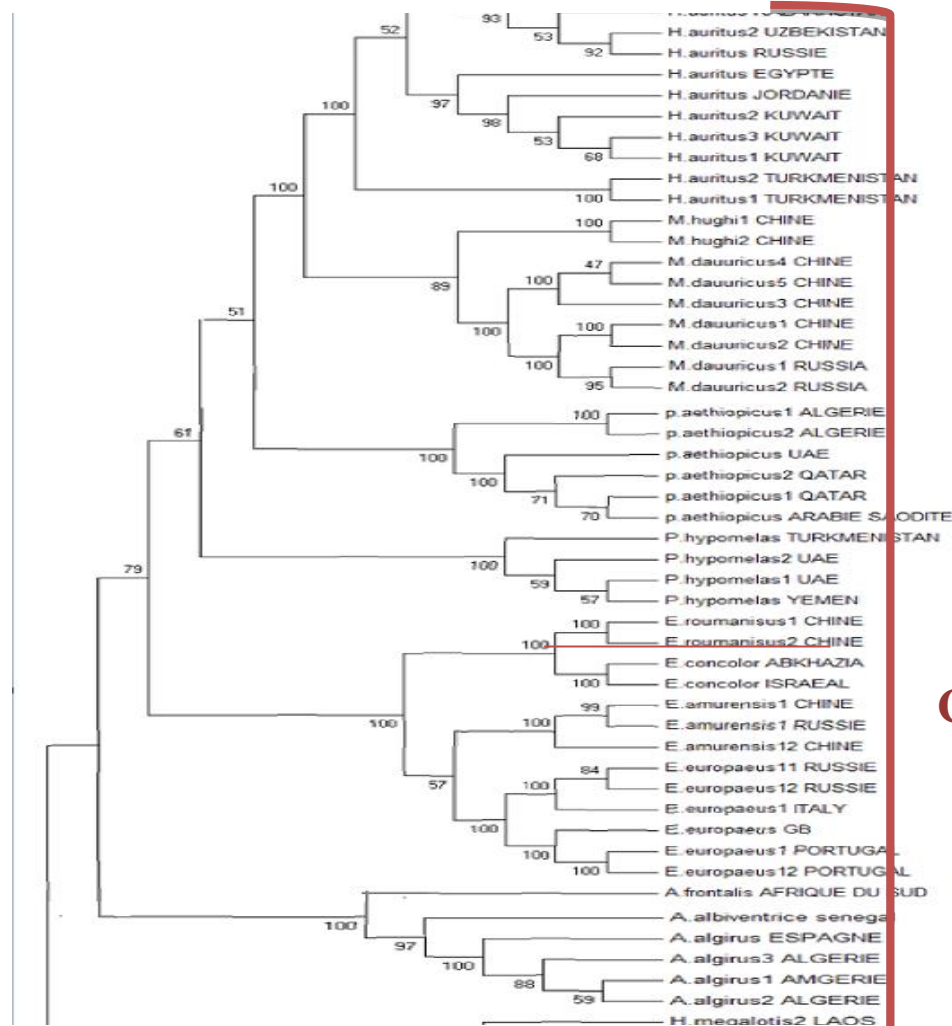


Figure 28 : Arbre phylogénétique reconstruit par méthodes UPGMA basées sur 57 séquences de nucléotides de Cytochrome b de la famille des *Erinaceidae* en utilisant MEGA5.

Une reconstruction phylogénétique est alors faite à partir de chaque alignement généré, et la valeur associée à chaque nœud dans l'arbre correspond au nombre de fois sur 1000 répliquas où cette même topologie a été trouvée.

D'après l'arbre (Figure 28) les séquences de Cytb se répartissent en 2 clades principaux :

Clade 1 : Il comprend différentes espèces dont : *Hemiechinus auritus*, *Mesechinus hughi*, *Mesechinus dauuricus*, *Paraechinus aethiopicus*, *Erinaceus roumanicus*, *Erinaceus concolor*, *Erinaceus amurensis*, *Erinaceus europaeus*, *Paraechinus hypomelas*, *Atelerix frontalis*, *Atelerix albiventris* et *Algirus atelerix*.

- **Sous clades A :** Il comprend 14 specimens de *H. auritus*, deux specimens de *M. hughi*, sept specimens de *M. dauuricus*, six specimens de *P. aethiopicus*, deux

specimens de *E. roumanicus*, deux specimens de *E. concolor*, trois specimens de *E. amurensis*, six specimens de *E. europaeus*, et quatre specimens de *P. hypomelas*.

- **Sous clade B :** Un specimen de *A. frontalis*, un specimen de *A. albiventris*, et quatre specimens de *A. atelerix*.

Clade 2 : Il comprend : *Hemichinus megalotis*, *Hemichinus suillus*, *Neotetracus sinensis* et *Hylomys parvus*.

- **Sous clade A :** Il est représenté par *H. megalotis*.
- **Sous clade B :** *N. sinensis*, *H. suillus* et *H. parvus*.

➤ **BRCA1:**

- *Atelerix algirus* : Au total il y a 4 séquences, on a choisi deux séquences d'Algérie (683pb) et une séquence d'Espagne (823pb).
- *Atelerix albiventris* : On a trouvé deux séquences, on a choisi une seule séquence du Sénégal (823pb).
- *Atelerix frontalis* : On a choisi une séquence du Sud d'Afrique (823pb) parmi deux séquences.
- *Erinaceus amurensis* : On a au total 4 séquences, on a choisi une séquence de la Russie (823pb).
- *Erinaceus concolor* : Il y a deux séquences, on a pris une séquence d'Israël.
- *Erinaceus europaeus* : On a trouvé 9 séquences au total, on a choisi une séquence du Portugal (683pb), une séquence de Grande Bretagne (856pb) et une séquence de la Russie (823pb).
- *Erinaceus roumanicus* : Au total il y a 2 séquences, on a choisi une séquence de la Russie (823pb).
- *Hemiechinus auritus* : On a trouvé 5 séquences, on a choisi trois séquences une de Turkménistan (823pb), une de l'Egypte (823pb) et une de la Jordanie (683pb).
- *Mesechinus dauuricus* : On a choisi deux séquences de la Russie (837pb ; 823pb).
- *Paraechinus aethiopicus* : Au total il y a 6 séquences, on a choisi une d'Arabie Saoudite (820pb), une de l'Algérie (683pb) et une de UAE (683pb).
- *Paraechinus hypomelas* : On a choisi deux séquences, une de Turkménistan (823pb) et une de Yémen (683pb).
- *Hylomys megalotis* : Une seule séquence du Laos (839pb).
- *Hylomys parvus* : Une seule séquence (801pb).

- *Hylomys cf. suillus* JJS-2016 : On a choisi une séquence du Vietnam (607pb) parmi deux séquences.
- *Neotetracus sinensis* : Deux séquences du Vietnam (849pb ; 810pb).

L'arbre phylogénétique de la famille des *Erinaceidae* a été réalisé à partir des séquences de gène de BRCA1 (Figure 29).

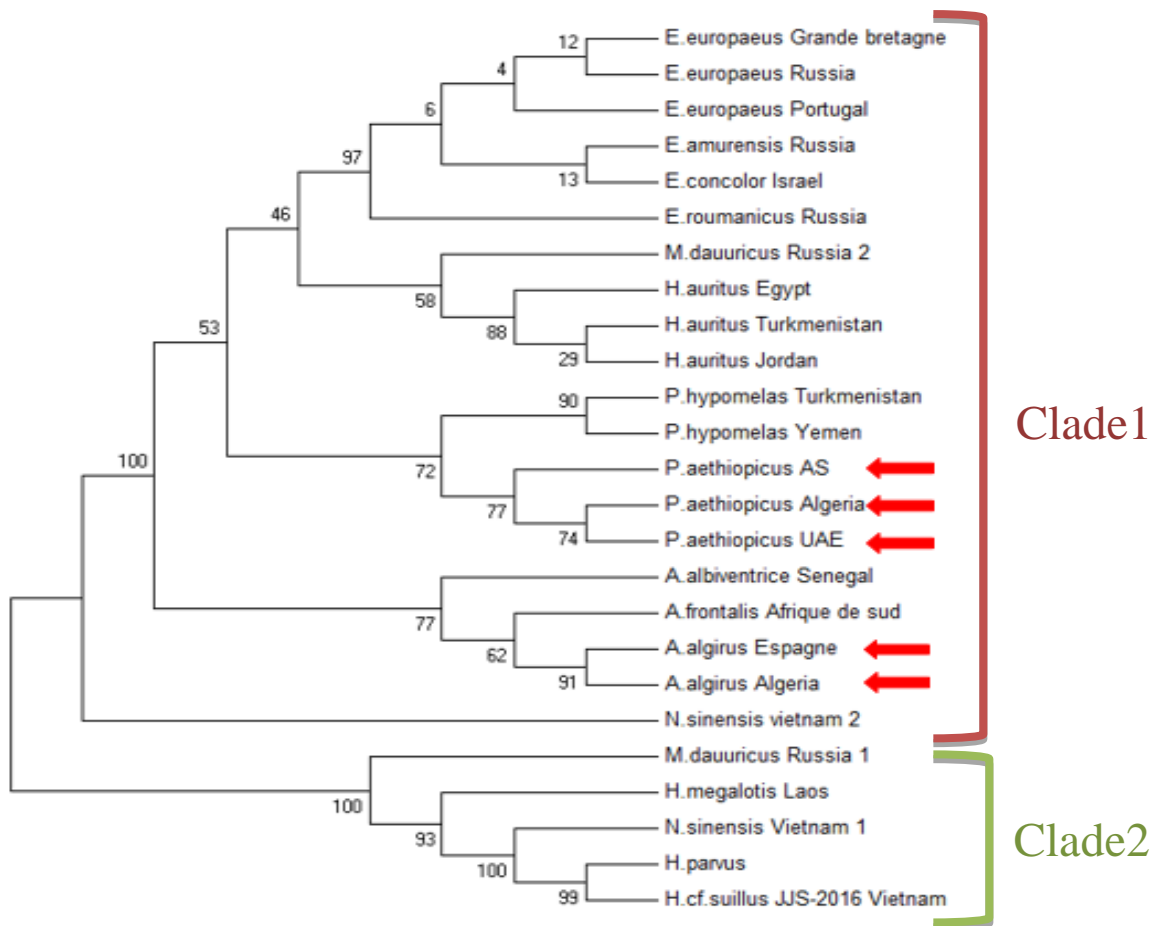


Figure 29 : Arbre phylogénétique des séquences de BRCA1 de la famille des *Erinaceidae* en utilisant la méthode UPGMA par le logiciel MEGA5.

D'après l'arbre de la figure 29, les séquences de BRCA1 se répartissent en 2 clades principaux :

Clade 1 : Il comprend la majorité des espèces dont : *Erinaceus concolor*, *Erinaceus europaeus*, *Erinaceus roumanicus*, *Erinaceus amurensis*, *Mesechinus dauuricus*, *Hemiechinus auritus*, *Paraechinus hypomelas*, *Atelerix albiventris*, *Atelerix frontalis*, *Algirus atelerix*, *Neotetracus sinensis* et *Paraechinus aethiopicus*.

Ce clade est subdivisé en 2 sous clades :

- ✓ **Sous clade A :** Il comprend un spécimen de *E. concolor*, trois spécimens de *E. europaeus*, un spécimen de *E. roumanicus*, un spécimen *E. amurensis*, un spécimen de *M. dauuricus*, trois spécimens de *H. auritus*, deux spécimens de *P. hypomelas*, un spécimen de *A. albiventris*, un spécimen de *A. frontalis* et deux espèces *A. atelerix*.
- ✓ **Sous clade B :** Il comprend *N. sinensis*.

Clade 2 : *Mesechinus dauuricus*, *Hylomys megalotis*, *Neotetracus sinensis*, *Hylomys parvus* et *Hylomys cf. suillus* JJS-2016.

- ✓ **Sous clade A :** Il est représenté par *M. dauuricus*.
- ✓ **Sous clade B :** Il est représenté par *H. megalotis*, *N. sinensis*, *H. parvus* et *H. suillus cf. JJS-2016*.

*D'après les arbres globaux (de Cytb et BRCA1) de la famille des *Erinaceidae*, nous avons deux clades important :

- 1) Clade Asiatique-Africain –Européen : *E. europaeus*, *E. roumanicus*, *E. amurensis*, *E. concolor*, *M. dauuricus*, *H. auritus*, *H. hypomelas*, *P. aethiopicus*, *A. albiventris*, *A. frontalis*, *A. atelerix*, *N. sinensis* et *M. hughi*.
- 2) Clade Asiatique : *H. suillus*, *N. sinensis*, *H. megalotis*, *H. parvus*, *M. dauuricus* et *H. cf. suillus* JJS-2016.

La famille des *Erinaceidae* possède une très vaste aire de répartition à travers le monde. En Algérie, deux espèces nominales ont été décrites, appartenant à deux genres distincts sur la base des caractéristiques morphologiques, l'espèce dite méditerranéenne *Atelerix algirus* et l'espèce désertique *Paraechinus aethiopicus* (DEROUICHE *et al.*, 2016 et 2017)(Figure 30).

La position des deux espèces du hérisson d'Algérie *Paraechinus aethiopicus* et *Atelerix algirus* sur deux clades différents reflète bien leur distribution spatiale, qui révèle que le hérisson dit du désert habite une bande assez étroite entre l'Atlas saharien et le désert. Plus au nord, son aire de répartition s'étend d'Ain Sefra, situé sur le versant nord de l'Atlas saharien, à Ain Ouarka, Brezina, Laghouat, Biskra au sud, Béni Abbès situés à l'ouest et El Goléa dans la partie centrale de l'Algérie. Il est également observé dans les montagnes du Sahara central, dans l'Ahnet Adrar et dans le Hoggar. Par contre, la deuxième espèce a une aire de distribution couvrant la partie méditerranéenne du nord Algérien. Elle coexiste dans les Hauts Plateaux avec *Paraechinus aethiopicus*. Les plus anciennes archives signalent sa présence au sud de Ain Sefra, Aïn El Orak, Laghouat et Biskra (KOWALSKI *et al.*, 1991). Selon

MÜLLER (1974) le modèle de distribution serait allopatrique, cependant la coexistence des deux espèces notamment sur les hauts plateaux suggère une distribution parapatricque (DEROUICHE *et al.*, 2016 ; BULL, 1991 ; DENNIS et HELLBERG, 2010) avec la possibilité de présence d'hybrides comme résultat du croisement entre les deux espèces (BOLFIKOVA et HULVA, 2012).

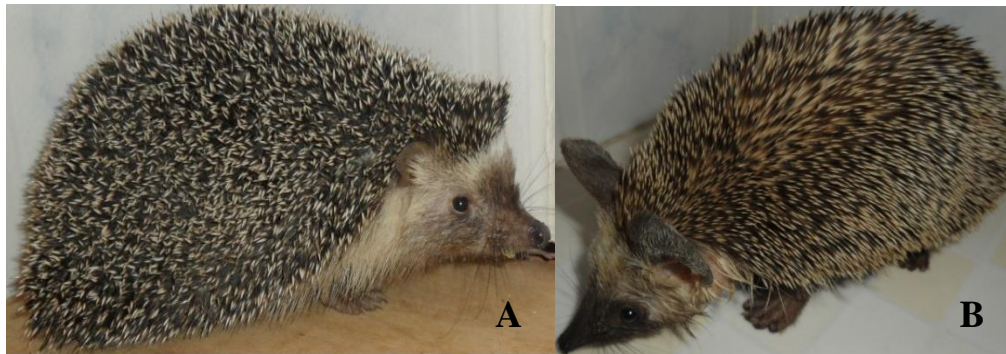


Figure 30 : (A) *Atelerix algirus* ; (B) *Paraechinus aethiopicus* (DEROUICHE *et al.*, 2016).

III.2.2. Détermination des haplotypes :

Nous avons utilisé en ligne le logiciel FaBox pour faire l'alignement de la séquence la plus courte. Elle supprime simplement les zones frontalières qui sont pleines de gaps pour utiliser le résultat d'alignement dans le logiciel DnaSP.

Nous avons utilisé le DnaSP pour la détermination des haplotypes ; le terme haplotype correspond à une séquence nucléotidique, qui peut être commune à plusieurs individus, mais diffère des autres haplotypes par une ou plusieurs substitutions de nucléotides.



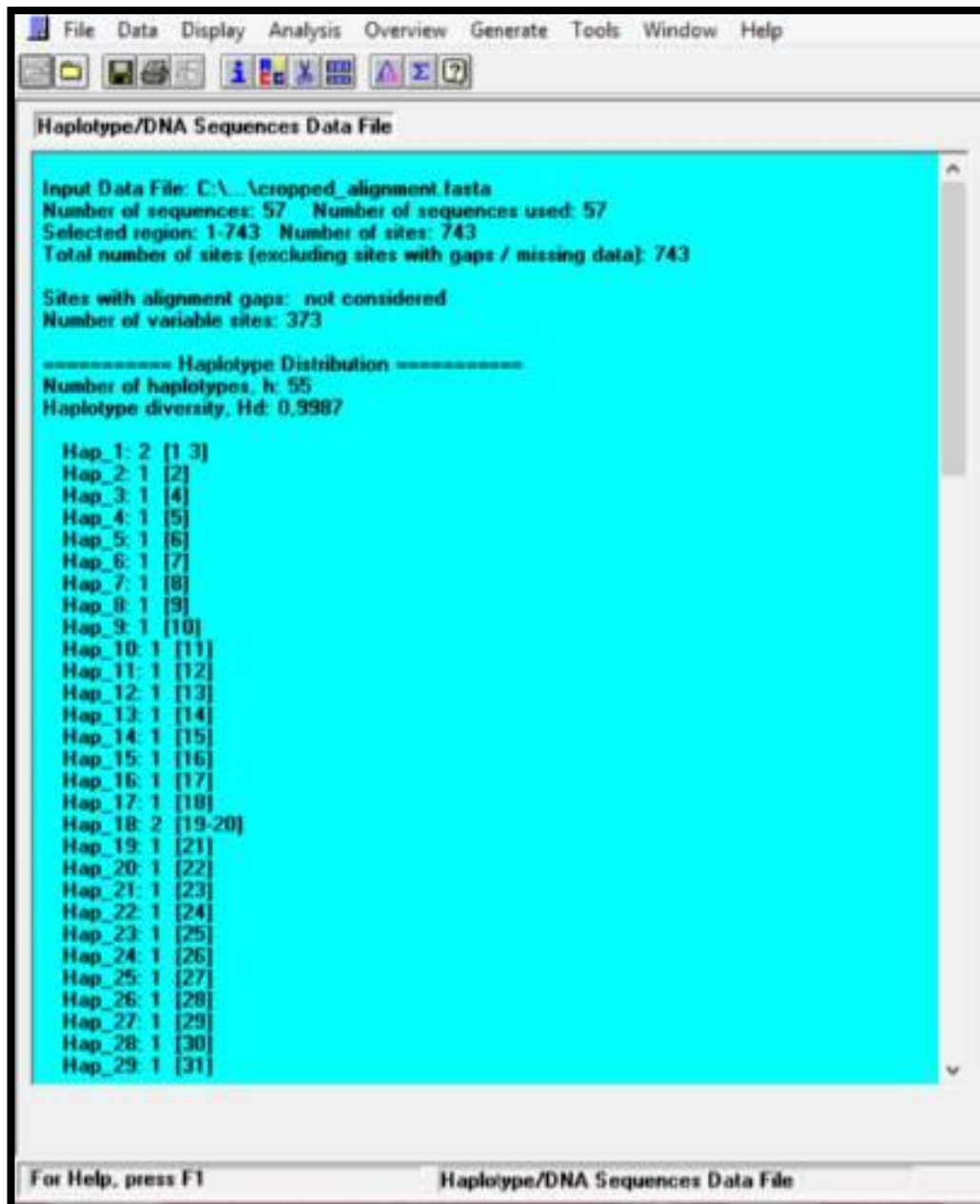
Figure 31 : Résultat d'alignement des séquences BRCA1 de la famille des *Erinaceidae* en utilisant le logiciel FaBox.

❖ **Blocs d'alignement :**

- ✓ Rouge = gaps : La présence des gaps peut être expliquée par 2 possibilités : soit la délétion des bases azotées de séquences qui contiennent les gaps, ou bien l'insertion des bases dans les séquences qui ne contiennent pas de gaps.
- ✓ Vert = conservés : Les zones de similarité entre les séquences.

A. Cytb :

L'analyse de Cytb de l'ADN mitochondrial révèle une diversité génétique très importante ; les 57 individus étudiés ont permis de caractériser 55 haplotypes différents. 53 haplotypes privés ont été détectés ; un haplotype privé est un haplotype qui ne se retrouve que dans une seule population. Les diversités haplotypique et nucléotidique sont donc très élevées au sein de notre échantillonnage pourtant restreint ($Hd = 0.9987$) (Figure 32, 33).



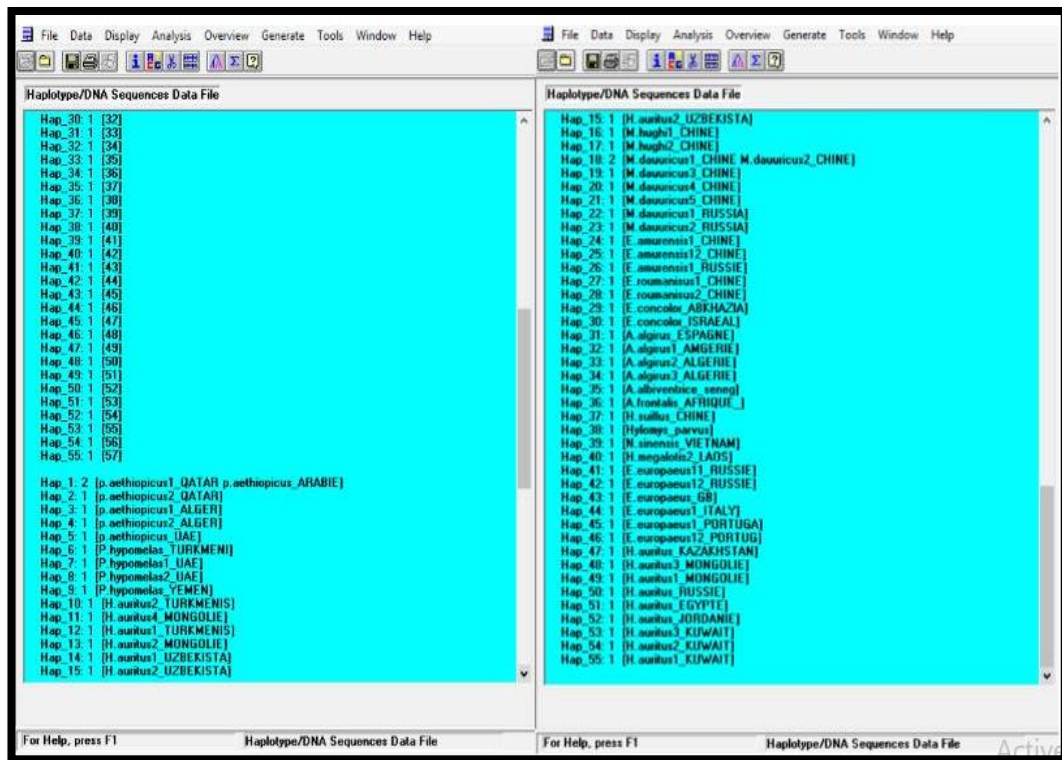


Figure 32 : Les haplotypes des séquences Cytb de la famille des *Erinaceidae* en utilisant le logiciel DnaSP.

- **53 haplotypes privés :**

P. aethiopicus Qatar (01), *P. aethiopicus* Algerie (02), *P. aethiopicus* UAE (01), *P. hypomelas* Turkménistan (01), *P. hypomelas* UAE (02), *P. hypomelas* Yémen (01), *H. auritus* Turkménistan (02), *H. auritus* Mongolie (04), *H. auritus* Uzbekistan (02), *H. auritus* Kazakhstan (01), *H. auritus* Russie (01), *H. auritus* Egypte (01), *H. auritus* Jordanie (01), *H. auritus* Kowait (03), *M. hughi* Chine (02), *M. dauuricus* chine (03), *M. dauuricus* Russie (02), *E. amurensis* Chine (02), *E. amurensis* Russie (01), *E. roumanicus* Chine (02), *E. concolor* Abkhazia (01), *E. concolor* Israel (01), *A. atelerix* Algerie (03), *A. atelerix* Espagne (01), *A. albivetrus* Sénégal (01), *A. frontalis* Sud d'Arique (01), *H. suillu* Chine (01), *H. parvus* (01), *N. sinensis* Vietnam (01), *H. megalotis* Laos (01), *E. europaeus* Russie (02), *E. europaeus* GB (01), *E. europaeus* Italie (01), *E. europaeus* Portugal (02).

- **02 haplotypes comportant 2 séquences :**

Hap 1 : *P. aethiopicus* Qatar et *P. aethiopicus* Arabie Saoudite.

Hap 18 : *M. dauuricus* Chine (02).

B. BRCA1 :

Sur les 25 séquences de BRCA1 on a obtenu 15 haplotypes dont 10 haplotypes privés qui ont pu être identifiés (hd : 0.9300).

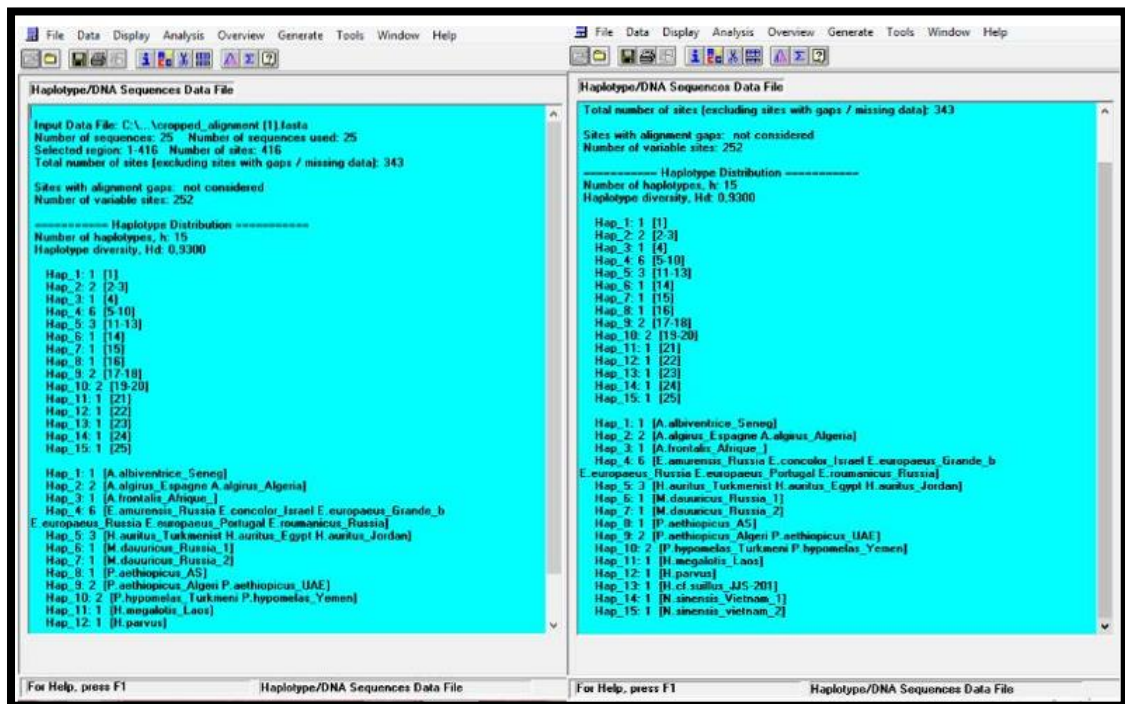


Figure 33 : Les haplotypes des séquences BRCA1 de la famille des *Erinaceidae* en utilisant le logiciel DnaSP.

• 10 haplotypes privés :

A. albiventris Sénégal (01), *A. frontalis* Sud d’Afrique (01), *M. dauuricus* Russie (02), *P. aethiopicus* Arabie Saudite (01), *H. megalotis* Laos (01), *H. parvus* (01), *Hsuillus cf. JJS-2016* Vietnam (01) et *N. sinensis* Vietnam (02).

• 03 haplotypes comportant 2 séquences :

- ✓ Hap 2 : *A. atelerix* Espagne et *A. atelerix* Algerie.
- ✓ Hap 9 : *P. aethiopicus* Algérie et *P. aethiopicus* UAE.
- ✓ Hap 10 : *P. hypomelas* Turkménistan et *P. hypomelas* Yémen.

• 01 haplotypes comportant 3 séquences :

- ✓ Hap 5 : *H. auritus* Turkménistan, *H. auritus* Egypt et *H. auritus* Jordanie.

• 01 haplotypes comportant 6 séquences :

- ✓ **Hap 4 :** *E. amurensis* Russie, *E. concolor* Israel, *E. europaeus* GB, *E. europaeus* Russie, *E. europaeus* Portugal et *E. roumanicus* Russie.

Tableau IV : Comparaison des indices de diversités calculés avec DnaSP sur les échantillons étudiés.

Population	Clade	Taille de l'échantillon	Nombre d'haplotypes h	Diversité haplotypique Hd	Nombre de Sites polymorphes S	Le nombre de site variable
Cytb	1	57	55	0,9300	743	373
	2		2			
BRCA1	1	25	15	0.9987	868	252
	2		5			

*Nous avons remarqué que les résultats obtenus par le DnaSP confirment les résultats de MEGA5, les analyses du polymorphisme mitochondrial fournissent des pistes intéressantes sur l'histoire de l'expansion géographique de chaque espèce. Les relations entre individus, haplotypes ou espèces sont visualisées précédemment via la construction des arbres, par agglomération des plus proches voisins (SAITOU *et al.*, 1987).

Plus la diversité haplotypique est élevée au sein de l'échantillonnage, plus il y a de chances d'observer des haplotypes différents lorsque l'on sélectionne 2 individus au hasard (NEI, 1987).

La diversité nucléotidique mesure le nombre moyen de différences entre 2 séquences choisies aléatoirement dans l'échantillonnage (NEI *et al.*, 1979).

III.2.3. Analyses statistique :

La pertinence de l'information génétique contenue dans le jeu de données a été contrôlée à l'aide du logiciel DAMBE (XIA, 2001). DAMBE est énuméré en tant qu'un des logiciels les plus très utilisés dans la phylogénétique moléculaire (SALEMIET *et al.*, 2003 ; FELSENSTEIN, 2004 ; LEMEY *et al.*, 2009).

Différents descripteurs statistiques sont calculés afin de mieux cerner la diversité génétique qui caractérise les populations de chaque espèce.

La méthode utilisée par le logiciel DAMBE est basée sur une analyse de la fréquence de GC et AT. Le logiciel GenSkew calcule la normale et le skew (biais) cumulatif de deux

nucléotides sélectionnables pour une séquence donnée. Le résultat est montré dans deux graphiques différents (Figures 34 et 35). Le degré d'asymétrie compositionnelle, exprimé en termes de biais de GC et d'AT, peut être calculé à l'aide des formules suivantes (PERNA *et al.*, 1995) :

$$\text{Skew} = (\text{nucleotide 1} - \text{nucleotide 2}) / (\text{nucleotide 1} + \text{nucleotide 2})$$

- GC skew = $(G - C) / (G + C)$
- AT skew = $(A - T) / (A + T)$

Les méthodes comme GC skew, CGC skew, et Z-curve sont des outils pour mieux étudier le mécanisme de la réplication de l'ADN dans différents organismes.

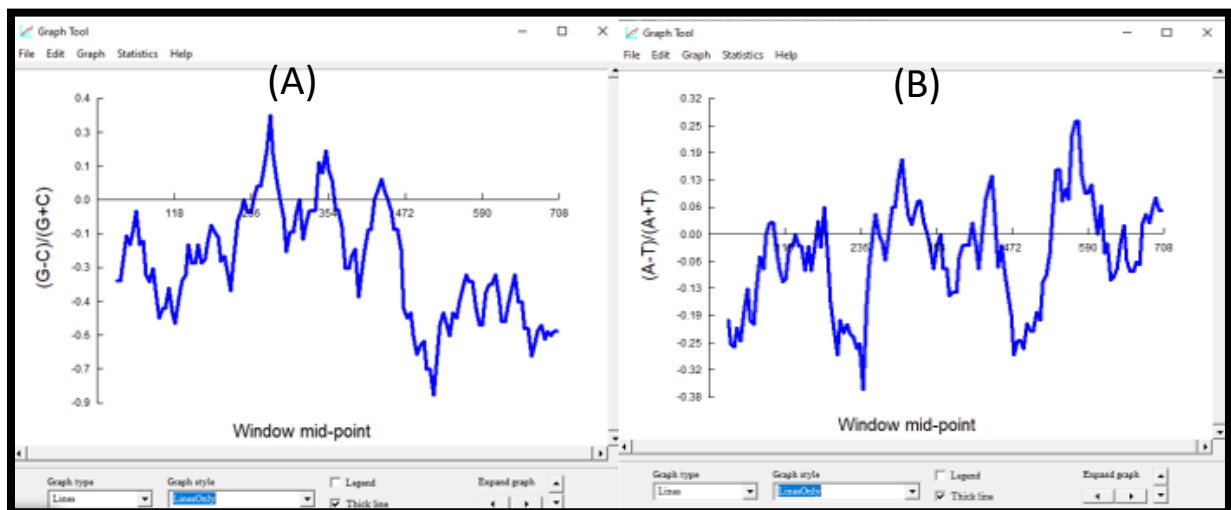


Figure 34 : Variations des fréquences GC (A) et AT (B) de *Cytb* produites à partir de DAMBE.

Nous avons remarqué pour le gène de *Cytb* que le GC skew est négatif (Figure 34 A) et l'AT skew est aussi négatif (Figure 34 B) ; sachant que, GC skew positif représente la richesse de G sur C et le GC skew négatif représente la richesse de C sur G (TILLIER *et al.*, 2000).

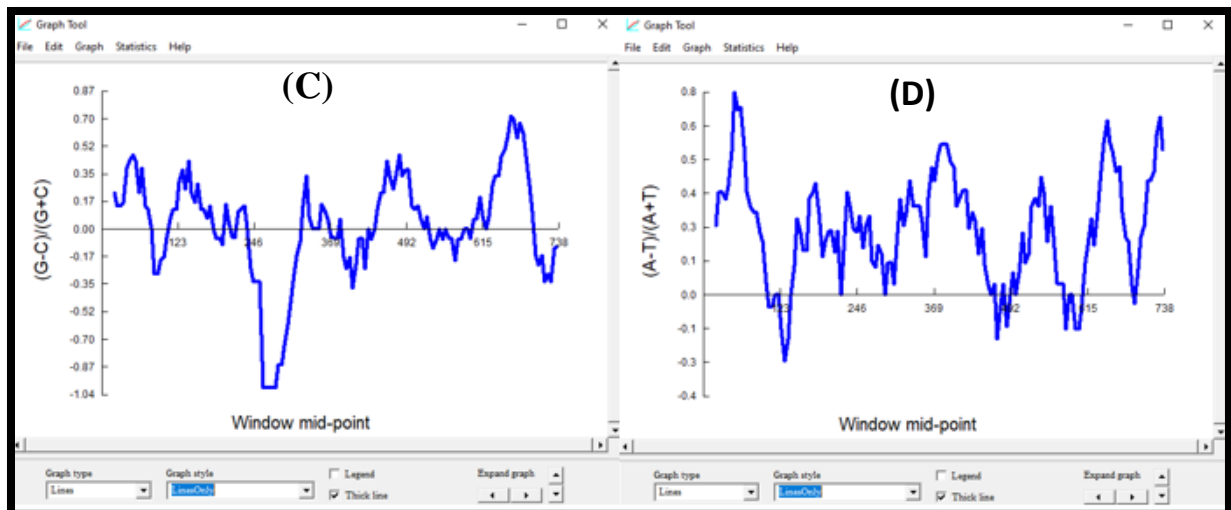


Figure 35: Variations des fréquences GC (A) et de AT (B) de BRCA1 produites à partir de DAMBE.

Nous avons remarqué pour le gène BRCA1 que la fréquence GC Skew est généralement positive (Figure 35 C) cependant l'AT Skew est généralement positif (Figure 35 D).

Plus l'ADN est riche en paires G/C plus l'ADN résiste à la dénaturation par l'augmentation de la température. Le taux de G+C d'une molécule d'ADN est la fréquence relative, exprimé généralement en pourcentage (**GENETIK *et al.*, 2006**). Les segments d'ADN composés de plusieurs bases G-C sont plus stables que les séquences composées de plusieurs bases A-T.

A, T, G et C représentent la fréquence d'occurrence de l'équivalent de la base dans une séquence particulière d'une longueur définie (**LOBRY *et al.*, 1996**). Nous avons comparé les fréquences de CG et AT des gènes Cytb et BRCA1 (Figure 36).

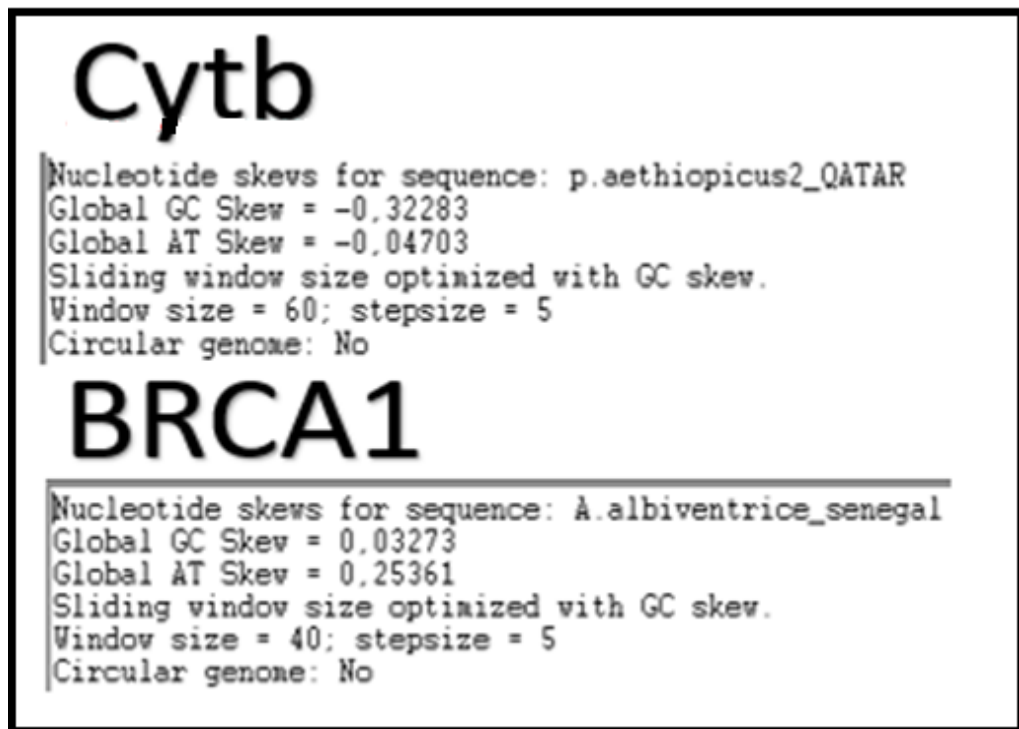


Figure 36 : Comparaison des fréquences de CG et AT entre Cytb et BRCA1 réalisés par le logiciel DAMBE.

Nous avons remarqué qu'il y a une richesse de la cytosine sur la guanine et de la thymine sur l'adénine pour BRCA1, et il y'a une richesse de la guanine sur la cytosine et de l'adénine sur la thymine pour le Cytb, la fréquence de BRCA1 est plus élevée par rapport au Cytb (Figure 36) c.-à-d. le nombre de CG présente dans BRCA1 est plus élevé par rapport au Cytb, parce que cette dernière est plus conservé que le Cytb, c'est due à leurs rôles dans l'organisme. Le gène du Cytb code pour une protéine membranaire intégrale cependant BRCA1 code pour la protéine BRCA1 impliquée dans la réparation des dommages de l'ADN pour cela il est plus conservé.

III.2.4. Résultat de NETWORK :

Les haplotypes, utilisable pour des analyses de traçabilité, mais aussi pour étudier l'évolution de l'espèce. Les relations entre les haplotypes peuvent également être appréhendées par la construction de réseaux d'haplotypes. Ces réseaux d'haplotypes sont construits grâce au logiciel NETWORK 5.0 (Figure 37 et 38).

La construction de ces réseaux est basée sur la théorie de la coalescence (KINGMAN, 2000). Cette théorie consiste en une approche rétrospective qui décrit mathématiquement le processus de fusion binaire de tous les lignages généalogiques d'un échantillon de gènes jusqu'à leur plus proche ancêtre commun.

Dans le réseau d'haplotypes ; chaque cercle correspond à un haplotype ; la taille du cercle est proportionnelle à la fréquence de l'haplotype dans le jeu de données. La longueur des segments entre chaque haplotype est proportionnelle au nombre de mutations qui les sépare.

- **Cytb :**

Sur les 57 séquences de Cytb de la famille *Erinaceidae* obtenues, 55 haplotypes ont pu être identifiés. La structure du réseau apparaît plus complexe et avec de nombreux haplotypes divergents.

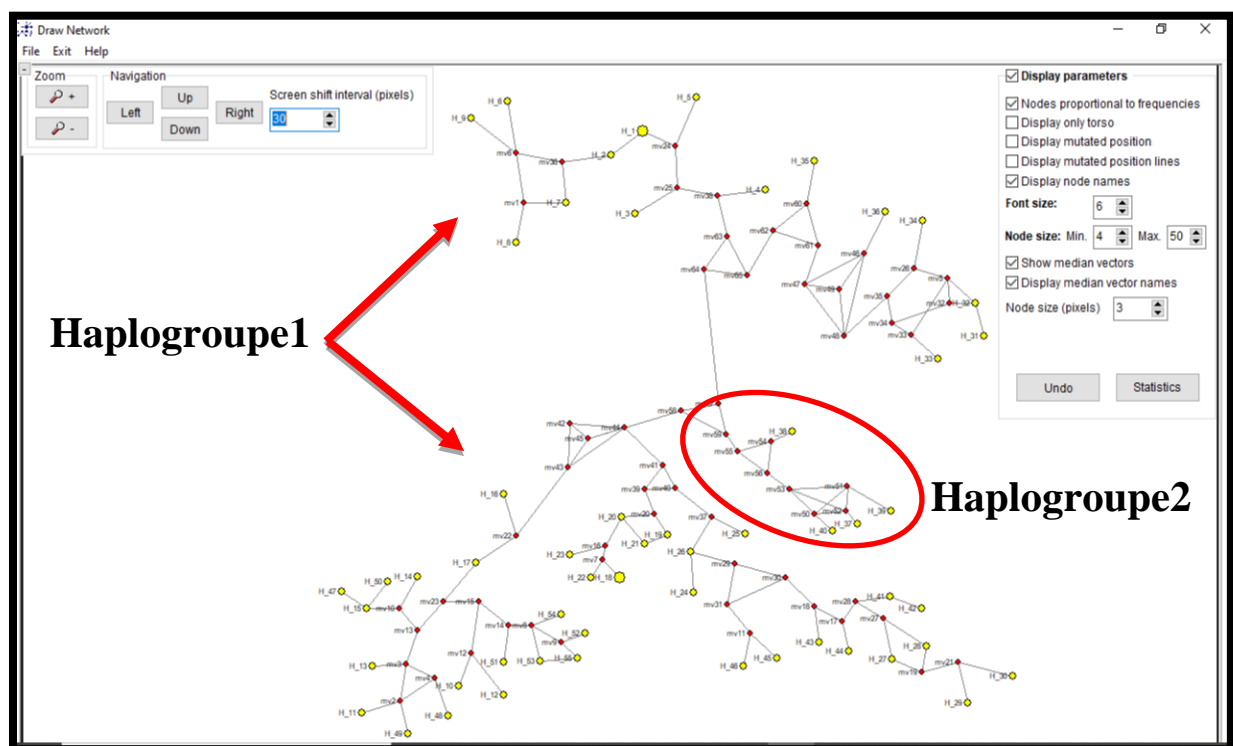


Figure 37 : Réseau haplotypique réalisé par le logiciel NETWORK à partir de 57 séquences de Cytb de la famille des *Erinaceidae*.

Haplogroupe 1 : Correspond au groupe Asiatique-Africain-Européen : *Hemiechinus auritus*, *Mesechinus hughi*, *Mesechinus dauuricus*, *Paraechinus aethiopicus*, *Erinaceus roumanicus*,

Erinaceus concolor, *Erinaceus amurensis*, *Erinaceus europaeus*, *Paraechinus hypomelas*, *Atelerix frontalis*, *Atelerix albiventris* et *Atelerix algirus* pour les haplotypes suivants : H1, H2, H3, H4, H5, H6, H7, H8, H9, H10, H11, H12, H13, H14, H15, H16, H17, H18, H19, H20, H21, H22, H23, H24, H25, H26, H27, H28, H29, H30, H31, H32, H33, H34, H35, H36, H41, H42, H43, H44, H45, H46, H47, H48, H49, H50, H51, H52, H53, H54 et H55.

Haplogroupe 2 : Correspond au groupe Asiatique: *Hemichinus megalotis*, *Hemichinus suillus*, *Neotetracus sinensis* et *Hylomys parvus*. Pour les haplotypes suivants : H37, H38, H39 et H40.

- **BRCA1 :**

D’après la structure du réseau (Figure 38), les 25 séquences de BRCA1 se répartissent en 2 haplogroupes principaux :

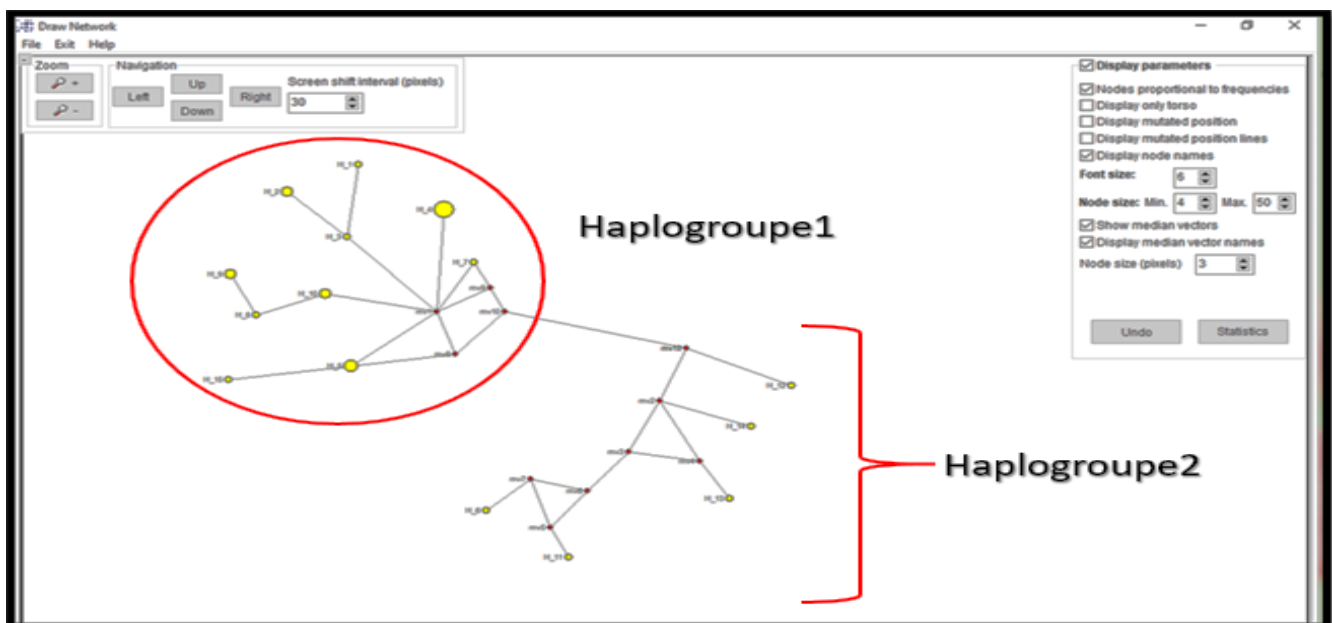


Figure 38 : La croissance des haplotypes des séquences de BRCA1 de la famille des *Erinaceidae* en utilisant le logiciel NETWORK.

Haplogroupe 1: Correspond au groupe Asiatique-Africain-Européen : *Erinaceus concolor*, *Erinaceus europaeus*, *Erinaceus roumanicus*, *Erinaceus amurensis*, *Mesechinus dauuricus*, *Hemiechinus auritus*, *Paraechinus hypomelas*, *Atelerix albiventris*, *Atelerix frontalis*, *Atelerix algirus*, *Neotetracus sinensis* et *Paraechinus aethiopicus*. Pour les haplotypes suivants : H1, H2, H3, H4, H5, H7, H8, H9, H15 et H10.

Haplogroupe 2 : Correspond au groupe Asiatique : *Mesechinus dauuricus*, *Hylomys megalotis*, *Neotetracus sinensis*, *Hylomys parvus* et *Hylomys cf. suillus JJS-2016* ; pour les haplotypes suivants : H6, H11, H13, H14 et H12.

Les résultats obtenus des réseaux haplotypiques des deux gènes que ce soit le Cytb ou BRCA1 ont montré l'existence de deux haplogroupes major qui représentent le clade Asiatique-Africain-Européen et le clade d'Asie.

D'après les travaux de **SAITOU *et al.* (1987)**, plus deux séquences homologues sont différentes, plus la distance évolutive qui les sépare est grande, et donc les séquences ont divergé il y a plus longtemps ; à l'inverse, plus deux séquences homologues sont proches, plus la distance évolutive qui les sépare est faible. Nous avons observé une distance génétique élevée entre les espèces de la famille des *Erinaceidae*, cette croissance à lieu sur une période relativement restreinte.

Conclusion

Le hérisson, petit mammifère insectivore de la famille des *Erinaceidae* possède une très vaste aire de répartition à travers le monde, comporte aujourd'hui 21 espèces : *Atelerix albiventris*, *Atelerix algirus*, *Atelerix frontalis*, *Erinaceus amurensis*, *Erinaceus concolor*, *Erinaceus ecaudatus*, *Erinaceus europaeus*, *Erinaceus roumanicus*, *Hemiechinus auritus*, *Mesechinus dauuricus*, *Mesechinus hughi*, *Paraechinus aethiopicus*, *Paraechinus hypomelas*, *Echinosorex gymnura*, *Hylomys megalotis*, *Hylomys parvus*, *Hylomys suillus*, *Hylomys cf. suillus* JJS-2016, *Neohylomys hainanensis*, *Neotetracus sinensis* et *Podogymnura truei*. En Algérie, deux espèces nominales ont été décrites, appartenant à deux genres distincts sur la base des caractéristiques morphologiques, l'espèce dite méditerranéenne *Atelerix algirus* et l'espèce désertique *Hemiechinus aethiopicus*.

La reconstruction phylogénétique est un outil utilisé dans les domaines aussi divers que la taxonomie, l'épidémiologie et la génétique des populations. C'est une procédure incontournable pour retracer l'histoire évolutive des organismes de tous niveaux taxonomiques. Cette analyse est également d'une grande importance pour clarifier les modèles évolutifs des espèces et pour comprendre l'évolution adaptative au niveau morphologique et moléculaire.

Les analyses phylogénétiques obtenues dans ce travail ont été élaborées à l'aide des différents logiciels tel que, MEGA5 pour la reconstruction de l'arbre phylogénétique, DnaSP pour la détermination des haplotypes (une séquence nucléotidique, qui peut être commune à plusieurs individus, mais diffère des autres haplotypes par une ou plusieurs substitutions de nucléotides), NETWORK pour avoir l'évolution des espèces et DAMBE pour faire des calculs statistiques.

Nous avons remarqué que les résultats obtenus par le DnaSp confirment les résultats de MEGA5, les analyses du polymorphisme mitochondrial fournissent des pistes intéressantes sur l'histoire de l'expansion géographique de chaque espèce. Les relations entre individus, haplotypes ou espèces sont visualisées précédemment via la construction des arbres, par agglomération des plus proches voisins.

Nous avons remarqué aussi que les résultats obtenus par le NETWORK confirment les résultats de MEGA5, ils ont montré l'existence de deux haplogroupes pour les gènes Cytb et BRCA1 qui représentent le clade Asiatique-Africain-Européen et le clade d'Asie.

Les méthodes d'analyses phylogénétiques, que ce soit en bifurcation, les haplotypes ou en réseau par les logiciels MEGA, DnaSP, NETWORK ont montré pour la famille des *Erinaceidae* l'existence de deux clades distincts qui sont le clade Asiatique-Africain-Européen (*E. europaeus*, *E. roumanicus*, *E. amurensis*, *E. concolor*, *M. dauuricus*, *H. auritus*, *H. hypomelas*, *P. aethiopicus*, *A. albiventris*, *A. frontalis*, *A. atelerix*, *N. sinensis* et *M. hughi*) et le clade Asiatique (*H. suillus*, *N. sinensis*, *H. megalotis*, *H. parvus*, *M. dauuricus* et *H. cf. suillus* JJS-2016).

Références

- **AMMAM M., 1987** – Inventaire de la faune du Djebel El Achch (Saïda) en vue d'un aménagement cynégétique. Thèse Ingénieur. Agro., Inst. Nat. Agro., El Harrach (Alger), 100 p.
- **ATHMANI L., 1988** – Comparaison faunistique entre trois stations dans le parc national de Belezma (Batna). Thèse Ingénieur. Agro., Inst. Nat. Agro., El Harrach, (Alger), 97 p.
- **AULAGNIER S. et THEVENOT M., 1986** – Catalogue des mammifères sauvages du Maroc. Trav. Inst. Sci., Série Zool., 41, Rabat, 164 p.
- **AVISE J. C. S., ARNOLD J., BALL MR., BERMINGHAM E., LAMB T., NEIGEL JE., REEB CA., ANDSAUNDRES NC., 1987.** “Intraspecific phylogeography: The mitochondrial DNA Bridge Between population genetics and systematics.” Annual review of ecology and systematics 18: 489-522.
- **BAICHI A., 1987** – Etude faunistique dans le parc National de Thniet El – Had notamment en cédraie. Thèse Ingénieur. Agro., Ins. Nat. Agro., El Harrach (Alger), 101 p.
- **BALLARD J. W. O AND MC WHITLOCK., 2004.** “The incomplete natural history of mitochondria.” Molecular Ecology 13:729-744.
- **BANDELT H. J., FORSTER P., RÖHL A., 1999.** Median-joining networks for inferring intraspecific phylogenies. Molecular Biological Evolution 16:37-48.
- **BAZZIZ B., 1991** – Approche biogéographique de la faune de Boughzoul, régime alimentaire de quelques vertébrés supérieurs. Mémoire Ingénieur. Agr., Inst. Nat. Agro., El Harrach, (Alger) 124 p.
- **BEAULIEU M. E., 2007.** Caractérisation moléculaire des champignons ophiostomatoïdes associés à quatre espèces de scolytes de l'écorce colonisant l'épinette blanche au Québec et phylogénie multigénique d'une nouvelle espèce de leptographium. Thèse doctorat, Université Laval, Québec.
- **BENSON D. A., KARSCH-MIZRACHI I., LIPMAN D. J., OSTELL J., et WHEELER D. L., 2006.** GenBank. NucleicAcidsRes, 34(Database issue), D16–20.
- **BERARD S., 2003.** Comparaison de séquences répétées en tandem et application à la génétique. Thèse de doctorat en Informatique. Dir. Thèse : Gascuel O. Univ. Montpellier II. 33-55. 242p.
- **BERMAN H. M., WESTBROOK J., FENG Z., GILLILAND G., BHAT T. N., WEISSIG H., SHINDYALOV I. N. & BOURNE P. E., 2000-**The ProteinData Bank. Nucleic Acids Res, 28(1), 235–42.

- **BOLFIKOVA B., HULVA P., 2012-** Microevolution of sympatry: landscape genetics of hedgehogs *Erinaceus europaeus* and *E. roumanicus* in Central Europe. *Heredity* (Edinb), 108, 248 - 55.
- **BOORE J. L., 1999-** “Animal mitochondrial genome”. *Nucleic acids Res* 27(8): 1767-1780.
- **BURTON M., 1976** – Tous les mammifères d’Europe en couleurs. Elsevier – Sequoia , Paris, 256 p.
- **CAMIN J. H., ANDSOKAL R. R., 1965-**A method for deducing branching sequences in phylogeny. *Evolution*, 19: 311-326.
- **CHEBINI F., 1987** – Inventaire ornithologique et recherches sur la reproduction des mésanges du genre *Parus* dans trois stations de la forêt de l’Akdadou. Thèse Magister, Inst. Nat. Agro. El Harrach (Alger), 70 p.
- **CHEN H., SMITH G. J. D., LI K. S., WANG J., FAN X. H., RAYNER J. M ET AL., 2006.** Establishment of multiple sublineages of H5N1 influenza virus in Asia: Implications for pandemic control. *Proc. Natl. Acad. Sci. USA.* 103(8):2845-50.
- **CHOI J. H., JUNG H. Y., KIM H. S., AND CHO H.G., 2000.** PHYLODRAW: A Phylogenetic Tree Drawing System. Department of computer science, Pusan National University, Pusan, Korea.
- **COCHRANE G., ALDEBERT P., ALTHORPE N., ANDERSSON M., BAKER W., BALDWIN A., BATES K., BHATTACHARYYA S., BROWNE P., VAN DEN BROEK A., CASTRO M., DUGGAN K., EBERHARDT R., FARUQUE N., GAMBLE J., KANZ C., KULIKOVA T., LEE C., LEINONEN R., LIN Q., LOMBARD V., LOPEZ R., MCHALE M., MCWILLIAM H., MUKHERJEE G., NARDONE F., PASTOR M. P., SOBHANY S., STOEHR P., TZOUVARA K., VAUGHAN R., WU D., ZHU W. & APWEILER R., 2006.**EMBL Nucleotide Sequence Database: developments in 2005. *Nucleic Acids Res*, 34(Database issue), D10–5.
- **COMET J. P., 1998.** Programmation Dynamique et Alignements de Séquences Biologiques. Thèse de doctorat en Informatique. Dir. Thèse : Henry J. Univ de Technologie de Compiègne : 5-38. 216p.
- **CORBET G.B., 1988** – The family of the Erinaceidae: A synthesis of its taxonomy, phylogeny, ecology and zoogeography. *Mammal Rev.*, vol. 18, n° 3 p. p : 117 – 172.
- **DARREL R, FROST W, CHRIS W et ROBERT S. HOFFMAN., 1991** Phylogénétique Relationships of Hedgehogs and Gymnures (Mammalia: Insectivora: Erinaceidae). Smithsonian Institution Press. Washington, D, C. 69 p

- **DE CARVALHOJONIOR S. A., 2003.** Sequence Alignment Algorithms. Thesis of Master of Sciences in Advanced Computing. King's College London. Univ. London. 4-18.
- **DELANY M. J. et FAROOK S. M. S., 1989** – The small mammals of a coastal gravel plain in the Sultanate of Oman. *J. Zool. London*, 218: 319 – 321.
- **DENNIS A. B., HELLBERG M. E., 2010-** Ecological partitioning among parapatric cryptic species. *Molecular Ecology* 19: 3206 - 3225.
- **DEROUICHE L., BOUHADAD R., FERNANDES C., 2016-** Mitochondrial DNA and morphological analysis of hedgehogs (Eulipotyphla: Erinaceidae) in Algeria. *Biochem. Syst. ecol.* 64 (2016) 57-64.
- **DEROUICHE L., VERCAMMEN P., BOUHADAD R., FERNANDES C., 2017-** Genetic evidence supporting the taxonomic separation of the Arabian and Northwest African subspecies of the desert hedgehog (*Paraechinus aethiopicus*). *Gene* 620 (2017) 54-65.
- **DIALLO A.B., 2009.** Bio-informatique avancée. *Université du Québec à Montréal. Canada.*
- **DOUMANDJI S. et DOUMANDJI A., 1992 a** – Note sur le régime alimentaire du hérisson d'Algérie, *Erinaceus algirus*, dans la banlieue d'Alger. *Mammalia*, T. 56, n° 2 p. p : 318 – 321.
- **DRAGESCO – JOFFE A., 1993** – La vie sauvage au Sahara. Ed. Delachaux et Nestlé, Neuchâtel, 286 p.
- **FELSENSTEIN J., 1993.** PHYLIP (PHYLogeny Inference Package) version 3.6a2, Distributed by the author, Department of Genetics, University of Washington, Seattle, WA.
- **FELSENSTEIN J., 2004-** Inferring phylogenies. Sunderland (MA): Sinauer Associates: 664.
- **FOLEY H., 1922** – Contribution de la faune saharienne. *Bull. Soc. Hist. Nat., Afr. N.*, 13p:
- **GAISLER J., 1984** – Mammifères de la région sétifienne. *Bull. Zool. Agri., Inst. Nat. Agro. El Harrach, (Alger) n° 8*, p. p. 1173 – 2300.
- **GATTIKER A., MICHOD K., RIVOIRE C., AUCHINCLOSS A. H., COUDERT E., LIMA T., KERSEY P., PAGNI M., SIGRIST C. J., LACHAIZE C., VEUTHEY A. L., GASTEIGER E., et BAIROCH A., 2003.** Automated annotation of microbial proteomes in SWISS-PROT. *Comput Biol Chem*, 27(1), 49–58.

- **GATTOLIAT J. L., 2002.** Etude systématique, cladistique et biogéographique des Baetidae (Ephemeroptera) de Madagascar, Thèse de doctorat en Zoologie et Ecologie Animale. Dir. Thèse : J-M Elouard. Univ. Lausanne. 111-145. 279p.
- **GENETIK., JOCHENGRAW.,2006.** 4. Auflage, Springer Verlag Berlin Heidelberg.
- **GRASSE P.P., 1955** – Traité de zoologie, Mammifères, Anatomie, Ethologie, Systématique. Ed. Masson et Cie, Paris, T. XVII ; Fasc. II, p. p. 1173 – 2300.
- **GUERMAS F., 1987** – Contribution à l'étude de la faune (Oiseaux – Mammifères) de la région du Djurdjura, station de Tikjda. Diversité et mesures de conservation. Thèse Ingénieur. Agro., Institut. Nat. Agro., El Harrach (Alger), 105 p.
- **GUINDON S., 2003-** Méthodes et algorithmes pour l'approche statistique en phylogénie, Thèse de Doctorat en Biologie. Dir. Thèse : O Gascuel. Univ. Montpellier II, 9-50. 155p.
- **HAECKEL E., 1860.** "Überneue, lebendeRadiolarien des Mittelmeeres," Monatsberichte der Kdniglichen. 794-817.
- **HALTERNORTH T. et DILLER H., 1985** – Mammifères d'Afrique et de Madagascar. Ed. Delachaux et Niestlé, Neuchâtel, 397 p.
- **HAMDINE W., 1991** – Ecologie de la genette (*Genetta genetta* LINNE 1758). Dans le parc national du Djurdjura – Station de Tala – Guilef. Thèse Magister. Inst. Nat. Agro., El Harrach (Alger), 152 p.
- **HARBI E., 1991** – Répartition des mammifères d'Algérie. Thèse Ingénieur. Agro., Institut. Nat. Agro., El Harrach (Alger), 97 p.
- **HARBI E., 1991** – Répartition des mammifères d'Algérie. Thèse Ingénieur. Agro., Institut. Nat. Agro., El Harrach (Alger), 97 p.
- **HAUBOLD B., 2000.** Practical bioinformatics. Phylogenetics—ML. Max Planck InstitutfürchemischeÖkologie und Friedrich Schiller. Ins. Informatik. Univ. Jena. 55p.
- **HUELSENBECK J. P., 1995.** The Robustness of Two Phylogenetic Methods: Four-Taxon Simulations Reveal a Slight Superiority of Maximum Likelihood over Neighbor Joining. *Mol. Biol. Evol.* 12(5): 843-849.
- **HULO N., BAIROCH A., BULLIARD V., CERUTTI L., DE CASTRO E., LANGENDIJK-GENEVAUX PS., PAGNI M. et SIGRIST C. J., 2006.**The PROSITE database. *Nucleic Acids Res*, 34(Database issue), D227–30
- **KHIREDDINE A., 1977** – Etude écologique pour un aménagement cynégétique dans le Massif, Senabla Chergui à Djelfa. Thèse Ingénieur. Agro., Inst. Nat. Agro., El Harrach (Alger) , 113 p.
- **KIMURA M., 1980.** A simple method for estimating evolutionary rate of base substitutions through comparative studies of nucleotide sequences. *J. Mol. Evoi*, 16, 111-120.
- **KINGMAN J. F., 2000.** Origins of the coalescent. 1974-1982. *Genetics* 156:1461-1463.

- **KLEPPE K., OHTSUKA E., KLEPPE R., MOLINEUX I., KHORANA H. G., 1971.** Studies on polynucleotides. XCVI. Repair replications of short synthetic DNA's as catalyzed by DNA polymerases. *J. Mol. Biol.*, 56: 341-361.
- **KOWALSKI K. et RZEBIK – KOWALSKA B., 1991** – Mammals of Algeria. Ed . Polish, Acad. Sci. Inst. Syst. and Evol. Mammal, p: 48 – 52.
- **KOWALSKI K., RZEBIK- KOWALSKA B., 1991-** Mammals of Algeria. Ed. Polish, Acad. Sci. Inst. Syst. and Evol. Mamma., 48 – 52.
- **KUHNER M. K., FELSEINSTEIN J., 1994.** A Simulation Comparison of Phylogeny Algorithms under Equal and Unequal Evolutionary Rates. *Mol. Biol. Evol.* 11(3): 459-468.
- **KUMAR S., TAMURA K., NEI M., 1994.**MEGA: Molecular Evolutionary Genetics Analysis software for microcomputers, Vol.10. P.189-191.
- **LAAMARI M., 1985** – Aperçu sur la faune à Oued El Biod dans la région d'Arris en milieu agricole et naturel. Thèse Ingénieur. Agro., Inst. Nat. Agro., El Harrach (Alger), 60 p.
- **LE BERRE M., 1990** – Faune du Sahara : Mammifères. Ed. Le chevalier R. Chabaud, Paris, T. 2, 359 p.
- **LEMEY P., SALEMI M., VANDAMME A. M., 2009.**The phylogenetic handbook. Cambridge: Cambridge University Press.
- **LEMEY P., SALEMI M., VANDAMME A. M., 2009.**The phylogenetic handbook. Cambridge: Cambridge University Press.
- **LI K.B., 2003.** ClustalW-MPI: ClustalW analysis using distributed and parallel computing 19: 1585–1586.
- **LI W. H., 1997.** Molecular evolution. Sunderland, Massachussets: SinauerAssac. 487.
- **LIBRADOP., AND ROZAS J., 2009.**DnaSP v5: A software for comprehensive analysis of DNA polymorphism data. *Bioinformatics* 25: 1451-1452.
- **LOBRY J. R.,1996.** Asymmetric substitution patterns in the two DNA strands of bacteria. *Molecular biology and evolution* 13, 660-665
- **MADKOUR G., 1982** – Comparative ostéologique studies on *Paraechinus aethiopicus* of Qatar. *Zool. Anz. Jene*, 209 n° 1 – 2 p. p: 120 – 136.
- **MATTHIEWS L.H., 1972** – La vie des mammifères. « La grande encyclopédie de la nature ». Ed. Bordas, Paris , Vol. 15, T. I, 383 p.
- **MAYRE., 1965.** Classification and phylogeny. *Amer. zool.*, 5: 165-174.
- **MAZARI G., 1988** – Premières notes sur l'inventaire de la faune du parc national de Chréa. *Ann. Inst. Nat. Agro., El Harrach (Alger)*, vol. 12 n° spéciaux, p : 325–354.
- **MEINERTZHAGEN R., 1934** – The biogeography status of Ahaggar plateau in the central Sahara, with special reference to birds. *Ibis, loners*, 4PP: 528-571.

- **METREF S., 1994** – Contribution à l'étude bioécologique de l'avifaune (Aves) d'une oliveraie à Boumlih (Cap – Djinet) – Relations trophiques de quelques espèces de vertébrés. Mémoire, Ingénieur. Agro., Inst. Nat. Agro., El Harrach (Alger), 233p.
- **MORDJI D., 1988** – Etude faunistique dans la réserve naturelle du Mont Babor. These Ingénieur. Agro., Inst. Nat. Agro., El Harrach (Alger), 100 p.
- **MORET B. M. E., WARNOW T., 2002-** Reconstructing Optimal Phylogenetic Trees: A Challenge in Experimental Algorithmics. R. Fleischer et al. (Eds.): Experimental Algorithmics, LNCS 2547. 163–180.
- **MOSTEFAI N. ED., 1990**– Contribution à l'étude de la faune (Oiseaux et Mammifères) du Parc national de Taza. Etude particulière de la Sittelle Kabylie et possibilité de réintroduction du Cerf de Barbarie. Thèse Ingénieur. Agro., Inst. Nat. Agro., El Harrach (Alger), 86 p.
- **NEI M., LI W. H., 1979-** Mathematical model for studying genetic variation in terms of restriction endonucleases. ProcNatlAcadSci USA 76 : 5269-5273
- **NEI M., 1987-** Molecular Evolutionary Genetics. Columbia University Press, New York.
- **NEI M., 1996-** Phylogenetic Analysis in Molecular Evolutionary Genetics. Annu. Rev. Genet. 30: 371-401.
- **PAGE R. D. M., 1996-** TREEVIEW: An application to display phylogenetic trees on personalcomputers. Computer Applications in the Biosciences 12: 357-358.
- **PERNA N. T., KOCHER T. D., 1995-** Patterns of nucleotide composition at fourfold degenerate sites of animal mitochondrial genomes. J. Mol. Evol. 41, 353e358.
- **PETTER F., 1955** – Nouvelle note de biologie sur le hérisson du désert. Mammalia, T. n° 18 p. p : 220 – 221.
- **RASMONT R., 1997-** Evolution Biologique. Traduction de la 2ème édition anglaise. Ed. Départ. Deboeck Univ. Paris. Bruxelles. 371-507.
- **RASMONT R., 1997-** Evolution Biologique. Traduction de la 2ème édition anglaise. Ed. Départ. Deboeck Univ. Paris. Bruxelles. 371-507.
- **REEVE N., 1994** – Hedgehogs. Pauser. Nat. Hist., London, 313 p.
- **RISCHKOWSKY B., PILLING D., 2008** -L'Etat des Ressources Zoogénétiques pour l'Alimentation et l'Agriculture dans le Monde, pp393-410.
- **ROBINSON M., 1997-**Diversite des Modes d'évolution des Génomes de Rongeurs. Thèse de doctorat en Biologie. Dir. Thèse : M. Dominique. Univ. Claude Bernard - LYON I. 29-43. 158p.
- **RON A. M FOUCHIER., VINCENT M., ANDERS W., THEO M BESTEBROER., SANDER H., DEREK S., GUUS F RIMMELZWAAN., BJÖRN O., AND ALBERT D. M. E OSTERHAUS ., 2005.** Characterization of a Novel Influenza A Virus

- Hemagglutinin Subtype (H16) Obtained from Black- Headed Gulls. *J. Virol.* 79(5): 2814–2822.
- **SAINT – GIRON M. C., 1973** – Les mammifères de France et de Benelux. Ed. Doin, Paris, 481 p.
 - **SAITOU N., NEI M., 1987**- The neighbor-joining method: a new method for reconstructing phylogenetic trees. *MolBiolEvol.* 4:406–425.
 - **SALEMI M., VANDAMME A. M., 2003**-The phylogenetic handbook: a practical approach to DNA and protein phylogeny. Cambridge: Cambridge University Press.
 - **SANGER F., NICKLEN S. AND COULSON A., 1977**- DNA sequencing with chain-terminatinginhibitors. *Proc. Natl. Acad. Sci. USA* 74, 5463–5467.
 - **SAYAH C., 1988** – Comparaison faunistiques entre quatre stations dans le parc national du Djurdjura (Tikjda). Thèse Ingénieur. Agro., Inst. Nat. Agro., El Harrach (Alger), 130 p.
 - **SCHAAL B. A., OLSEN K. M., 2000**- Gene genealogies and population variation in plants. *ProcNatlAcadSci U S A* 97: 7024–7029.
 - **SCHILLING D., SINGER D. & DILLER H., 1986** – Guide des Mammifère d’Europe. Ed. Delachaux et Niestlé, Neuchâtel, Paris, 280 p.
 - **SCHMIDT H. A., 2003**- Phylogenetic Trees from Large Datasets Inaugural. Thèse de Doctorat en Mathématique. Dir. Thèse : Von Haeseler A. Univ. Heinrich–Heine–Düsseldorf vorgelegt von Heiko. 123p.
 - **SERVANT F., BRU C., CARRERE S., COURCELLE E., GOUZY J., PEYRUC D. & KAHN D., 2002**-ProDom : automated clustering of homologous domains. *Brief Bioinform*, 3(3), 246–51.
 - **SEURAT L. G., 1924** – Zoologique forestière de l’Algérie. Gouv. Gén. Algérie, Ecole des brigadiers des eaux et des forêts, 54 p.
 - **SI BACHIR, 1991** – Etude bioécologique de la faune du lac de Boulhilet ou petit Ank Djamel (Oum Bouaghi). Mémoire Magister. Univ. Sétif, Algérie, 139 p.
 - **SUNNUCKS P., 2001**- Efficient genetic markers for population biology. *Tree*, 15: 199–203
 - **SWOFFORD D. L., OLSEN G. J., WADDELL P. J., & HILLIS D. M., 1996**- Phylogenetic inference. In D. M Hillis, C. Moritz, and B. Mable (eds.) *Molecular Systematics* (2nd ed.), Sinauer Associates, Sunderland, Massachusetts: 407-514.
 - **SWOFFORD D.L., 1998-PAUP***. Phylogenetic Analysis Using Parsimony and other methods. Version 4.0 (beta version). Laboratory of Molecular Systematics Smithsonian.
 - **TALBI K. H., 1985** – Inventaire du gibier en vue d’un aménagement cynégétique de la forêt d’Oum – Graf – Saïda. Thèse Ingénieur. Agro., Inst. Nat. Agro., El Harrach (Alger), 80 p.

- **TAMURA K., PETERSON D., PETERSON N., STECHER G., NEI M. & KUMARS., 2011-** MEGA5: Molecular Evolutionary Genetics Analysis using Likelihood, Distance, and Parsimony methods [logiciel]. Molecular Biology and Evolution.
- **TELAILIAS S., 1990** – Bioécologie de la faune de différents milieux de la zone du lac Tonga (parc national d'El – Kala). Thèse Ingénieur. Agro., Inst. Nat. Agro., El Harrach (Alger), 111 p.
- **THOMPSON J. D., HIGGINS D. G., GIBSON T. J., 1994-** CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. Nucleic Acids Research 22 (22): 4673-4680.
- **TILLIER E. R., et COLLINS R. A., 2000-** The contributions of replication orientation, gene direction, and signal sequences to base-composition asymmetries in bacterial genomes. Journal of molecular evolution 50, 249-257.
- **TOURASSE N. J., 1992-** Développement d'une distance évolutive entre séquences prenant en compte la variabilité du taux de substitution entre sites et application à la reconstruction de phylogénies moléculaires anciennes. Thèse Doctorat en Génétique et Biologie des Populations. Univ. Claude Bernard - Lyon 1. 1-61. 186p.
- **VESMANIS I. E., 1979** – Bemerkungen zur Verbreitung und Taxonomie von Erinaceus a. algirus Lereboullet 1842 nud Paraechinus aethiopicus deserti Loche 1858. Afr. Small Mammal News, Special 1: 1 – 14.
- **VICENTE M. C., et FULTON T., 2003-** Utilisation des marqueurs moléculaires pour l'étude de la diversité génétique végétale. International Plant Genetic Institute for Genomic Diversity Resources Institute.
- **WU CH., APWEILER R., BAIROCH A., NATALE D A., BARKER WC., BOECKMANN B., FERRO S., GASTEIGER E., HUANG H., LOPEZ R., MAGRANE M., MARTIN M. J., MAZUMDER R., O'DONOVAN C., REDASCHI N. et SUZEK B., 2006-** The Universal Protein Resource (UniProt): an expanding universe of protein information. Nucleic Acids Res, 34(Database issue), D187–91.
- **WU CH., YEH LS., HUANG H., ARMINSKI L., CASTRO-ALVEAR J., CHEN Y., HU Z., KOURTESIS P., LEDLEY R. S., SUZEK B. E., VINAYAKA C. R., ZHANG J. & BARKER W. C. 2003-** The Protein Information Resource. Nucleic Acids Res, 31(1), 345–7.
- **XIA X., 2001-** Data analysis in molecular biology and evolution. Boston: Kluwer Academic Publishers.

- **YOSHIKAWA H., DOGRUMAN-AL F., DOGRUMAN-AI F., TURK S., KUSTIMUR S., BALABAN N., ET AL. 2011.**Evaluation of DNA extraction kits for molecular diagnosis of human Blastocystis subtypes from fecal samples. *ParasitolRes.*;109:10

Annexes

TabV: Les séquences extraites à partir de la GenBank (types des gènes et leur origine géographique).

Espèce	Région	Gène	Taillepb	Locus
<i>Atelerix albiventris</i> (663)	Indéfini	Cytb (1) Cytb (1) Cytb (1) Cytb (1)	470 435 402 369	KY777648 KY777649 KJ193305 KY777650
	Sénégal	BRCA1 (1) BRCA1 (1) Cytb (1)	834 823 1140	KF783081 KF783066 KF783143
<i>Atelerix algirus</i> (265)	Espagne	TTR (1) Cytb (1) ARN ribosomal 12S (1) VWF (1) BRCA1 (1) BRCA1 (1) GHR (1) RAG1 (1) BFIBR (3) RAG1 (1) BFIBR (3) Cytb (8)	1502 1130 966 935 834 823 804 662 639 636 633 356	KF783163 KF783144 KF783176 KF783097 KF783082 KF783067 KF783111 KF783035 KP279821 KF783052 KP279820 KP279811
	Indéfini	APOB (2) ARN ribosomal 12S (2) Cytb (1) D-loop (3) D-loop (4) D-loop (8) Cytb (1) ARN ribosomal 12S (1) ARN ribosomal 12S (1) E.algirusry (1)	727 594 470 409 407 405 367 329 259 158	KX959478 KX959469 KY777651 KU179783 KU179779 KU179780 KY777652 KY777658 KY777659 X90866
	Algérie	BRCA1 (2) BFIBR (5) BFIBR (14) RAG1 (1) BFIBR (1) BFIBR (20) RAG1 (1) D-loop (1) D-loop (1)	683 641 639 639 637 633 609 461 455	KY558570 KP279827 KP279826 MK510335 KP279863 KP279828 MK510384 MK510278 MK510232

		Cytb (44) Cytb (3)	356 743	KP279765 KY558558
	Maroc	RAG1(1) BFIBR(2) RAG1(36) RAG1(1) BFIBR(3) RAG1(1) RAG1(1) RAG1(1) RAG1(1) RAG1(2) RAG1(2) RAG1(2) RAG1(1) RAG1(1) RAG1(1) RAG1 (1) RAG1 (1) D-loop (6) D-loop (11) D-loop (20) D-loop (7) D-loop (2) D-loop (2) D-loop (1) Cytb(2)	640 639 639 637 633 629 619 615 612 611 610 609 608 602 597 580 572 461 640 459 458 454 450 431 356	MK510407 KP279866 MK510334 MK510333 KP279867 MK510403 MK510354 MK510393 MK510341 MK510368 MK510362 MK510385 MK510331 MK510336 MK510424 MK510361 MK510329 MK510286 MK510231 MK510228 MK510226 MK510303 MK510230 MK510281 KP279809
	Tunisie	BFIBR(1) RAG1(5) RAG1(1) D-loop (4) D-loop (1) Cytb(1)	639 639 632 459 403 356	KP279823 MK510328 MK510383 MK510242 MK510274 KP279818
	Libye	RAG1(1) D-loop (1)	639 459	MK510321 MK510219
<i>Atelerixfront alis (8)</i>	Afrique du Sud	TTT(1) Cytb(1) ARN ribosomal 12 (1) BRCA1 (1) BCAR1(1) GHR(1) RAG(1) RAG(1)	1502 1140 967 834 823 804 715 636	KF783164 KF783145 KF783177 KF783083 KF783068 KF783112 KF783036 KF783053

<i>Erinaceus amurensis</i> (41)	Japon	APOB(1) RAG1(2) BRCA1(2) APOB(1) ATP7A(2) BDNF(2) D-loop (1)	1154 1095 1032 1008 673 561 394	LC124833 LC124956 LC124922 LC124834 LC124860 LC124889 LC094446
	Russie	TTR(1) Cytb(1) ARN ribosomal 12S (1) Vwf(1) BRCA1(1) BRCA1(1) GHR(1) RAG1(1) RAG1(1)	1492 1134 966 935 837 823 804 715 636	KF783156 KF783125 KF783169 KF783091 KF783075 KF783060 KF783104 KF783028 KF783045
	Corée du sud	D-loop(4)	452	AB557981
	Indéfini	COX1(10) NADH(2) ARN ribosomal 12S(1) ARN ribosomal 12S(1) Génomecomplet(1)	516 1044 969 968 16941	KP992997 HQ857501 HQ857483 HQ857482 KX964606
<i>Erinaceus concolor</i> (132)	Israël	BRCA1(1) Cytb(1)	837 555	KF783073 AY011447
	Abkhazia	Cytb(1)	316	KF783028
	Indéfini	BRCA1(1) Cytb(2) Cytb(15)	823 1140 883	KF783058 KF783120 KF783120
<i>Erinaceusecaudatus</i> (3)	Indéfini	CryaA(1) Cryaa(1) Cryaa(1)	506 426 39	AJ299584 AJ272233 AJ270465
<i>Erinaceus europaeus</i> (227868)	Nouvelle Zélande	BRCA1(1)	5556	XM_016194170
	Indéfini	BRCA1(1) BRCA1(1) BRCA1(1) Cytb(2) Cytb(5) Cytb(2) Cytb(1) Cytb(1) Cytb(1)	2776 2742 792 743 476 474 471 392 320	DQ354459 AF284008 DQ630279 KY558565 KY777640 KY777647 KY777646 KY777638 KY777639
	Grande Bretagne	BRCA1(1)	856	KF783057

	Portugal	BRCA1(1)	683	KY558573
	Russie	BRCA1(1)	823	KF783056
		BRCA1(1) Cytb(2)	804 1140	KF783071 KF783116
	Royaume Uni	BRCA1(1) Cytb(1)	804 1140	KF783072 KF783118
	Italie	Cytb(1)	1140	KF783119
	Allemagne	Cytb(1)	627	JF318975
<i>Erinaceusrou manicus</i> (293)	Russie	TTR(1)	1495	KF783155
		Cytb(1)	1140	KF783122
Cytb(1)		1128	KF783123	
Cytb(1)		1127	KF783124	
BRCA1(1)		823	KF783059	
ARN ribosomal 12S (1)		962	KF783168	
VWF(1)		935	KF783090	
BRCA1(1)		837	KF783074	
GHR(1)		804	KF783103	
RAG1(1)		725	KF783027	
RAG1(1)		636	KF783044	
	Indéfini	Microsatellite:E32(1)	559	MH683423
		Microsatellite:E25(1)	526	MH683253
		Microsatellite:E19(2)	522	MH683324
		Microsatellite:E19(1)	517	MH683323
		Microsatellite:E25(1)	515	MH683254
		Microsatellite:E36(1)	512	MH683299
		Microsatellite:E3(1)	508	MH683189
		Microsatellite:E19(1)	507	MH683320
		Microsatellite:E20(1)	504	MH683278
		Microsatellite:E23(1)	501	MH683443
		Microsatellite:E20(1)	499	MH683269
		Microsatellite:E26(2)	499	MH683382
		Microsatellite:E3(1)	498	MH683191
		Microsatellite:E8(1)	498	MH683240
		Microsatellite:E3(2)	496	MH683182
		Microsatellite:E26(1)	496	MH683383
		Microsatellite:E8(1)	495	MH683241
		Microsatellite:E3(2)	494	MH683184
		Microsatellite:E22(1)	494	MH683225
		Microsatellite:E20(1)	494	MH683276
		Microsatellite:E8(1)	492	MH683242
	Microsatellite:E3(1)	490	MH683181	
	Microsatellite:E29(1)	490	MH683538	
	Microsatellite:E8(1)	489	MH683238	
	Microsatellite:E20(1)	489	MH683267	
	Microsatellite:E22(1)	488	MH683222	
	Microsatellite:E20(1)	487	MH683275	
	Microsatellite:E36(2)	487	MH683301	

	Microsatellite:E26(1)	487	MH683386
	Microsatellite:E22(1)	486	MH683226
	Microsatellite:E29(3)	486	MH683535
	Microsatellite:E3(1)	484	MH683188
	Microsatellite:E22(3)	484	MH683223
	Microsatellite:E3(2)	482	MH683193
	Microsatellite:E5(1)	472	MH683338
	Microsatellite:E35(3)	477	MH683203
	Microsatellite:E31(1)	476	MH683229
	Microsatellite:E10(1)	473	MH683520
	Microsatellite:E2(3)	472	MH683313
	Microsatellite:E27(1)	471	MH683389
	Microsatellite:E2(2)	470	MH683318
	Microsatellite:E2(4)	468	MH683312
	Microsatellite:E2(3)	466	MH683315
	Microsatellite:E18(3)	465	MH683196
	Microsatellite:E35(1)	464	MH683206
	Microsatellite:E36(3)	462	MH683296
	Microsatellite:E16(5)	460	MH683265
	Microsatellite:E5(6)	458	MH683336
	Microsatellite:E7(2)	457	MH683449
	Microsatellite:E11(4)	456	MH683371
	Microsatellite:E18(1)	455	MH683197
	D-loop(3)	455	MK510247
	Microsatellite:E30(6)	454	MH683305
	Microsatellite:E24(9)	452	MH683266
	Microsatellite:E37(3)	451	MH683281
	Microsatellite:E16(3)	450	MH683257
	Microsatellite:W32(1)	449	MH683422
	Microsatellite:E11(4)	448	MH683422
	Microsatellite:W32(1)	447	MH683421
	Microsatellite:E6(6)	446	MH683220
	Microsatellite:E18(2)	445	MH683199
	Microsatellite:E9(1)	444	MH683518
	Microsatellite:W12(1)	442	MH683476
	Microsatellite:W33(1)	440	MH683505
	Microsatellite:W30(1)	439	MH683363
	Microsatellite:E10(1)	437	MH683526
	Microsatellite:E16(2)	435	MH683258
	Microsatellite:E9(1)	432	MH683515
	Microsatellite:E9(1)	431	MH683517
	Microsatellite:E9(1)	430	MH683516
	Microsatellite:W3(3)	425	MH683360
	Microsatellite:W12(1)	424	MH683472
	Microsatellite:EW11(1)	421	MH683499
	D-Loop(1)	420	KY366249
	D-Loop(12)	419	KY366248
	Microsatellite:w19(1)	419	MH683289
	Microsatellite:W10(1)	418	MH683512
	Microsatellite:W30(2)	417	MH683359

		Microsatellite:W28(2)	416	MH683349
		Microsatellite:W16(2)	415	MH683546
		Microsatellite:W19(3)	414	MH683288
		Microsatellite:W30(3)	413	MH683356
		Microsatellite:W29(3)	412	MH683352
		Microsatellite:W16(1)	411	MH683545
		Région de contrôle(2)	409	HM462028
		Microsatellite:W7(2)	409	MH683327
		Région de contrôle(4)	407	HM462024
		Microsatellite:W1(5)	407	MH683174
		Microsatellite:W31(1)	406	MH683453
		Région de contrôle(1)	405	HM462027
		Région de contrôle(2)	403	KY489945
		Microsatellite:W18(2)	403	MH683461
		Région de contrôle(4)	401	KY489931
		Région de contrôle(1)	400	KY489953
		Microsatellite:W14(1)	400	MH683396
		Région de contrôle(39)	399	KY489901
		Microsatellite:W7(3)	399	MH683333
		Région de contrôle(1)	398	KY489952
		Microsatellite:W23(1)	398	MH683209
		Région de contrôle(6)	397	KY489940
		Microsatellite:W1(2)	397	MH683172
		Microsatellite:W14(2)	396	MH683391
		Microsatellite:W26(4)	395	MH683202
		Microsatellite:W23(5)	394	MH683213
		Microsatellite:W1(2)	392	MH683173
		Microsatellite:W12(1)	391	MH683215
		Microsatellite:W23(3)	390	MH683212
		Microsatellite:W24(2)	389	MH683231
		Microsatellite:W5(1)	388	MH683246
		Microsatellite:W13(2)	386	MH683217
		Microsatellite:W7(2)	384	MH683332
		Microsatellite:W5(2)	383	MH683245
		Microsatellite:W5(1)	378	MH683252
		Microsatellite:W15(1)	367	MH683454
	Autriche	COI(1)	743	KY754506
	Grèce	RAG1(1)	639	MK510350
		RAG1(1)	612	MK510351
		RAG1(1)	610	MK510352
		D-loop(3)	455	MK510247
	Ukraine	COI(1)	286	MH329869

<i>Hemiechinus auritus</i> (43)	Turkménistan	Cytb(3) BRCA1(1) BRCA1(1) RAG1(1) RAG1(1) GHR(1) TTR(1) ARN ribosomal 12S (1) Vwf(1)	1140 837 823 715 524 804 1500 970 935	KF783128 KF783077 KF783062 KF783030 KF783047 KF783106 KF783158 KF783171 KF783093
	Mongolie	Cytb(2) Cytb(1) Cytb(1) COI(1)	1140 1.131 1.119 688	KF783129 KF783131 KF783134 KX859285
	Ouzbékistan	Cytb(1) Cytb(1)	1140 1.136	KF783138 KF783133
	Kazakhstan	Cytb(1)	1.136	KF783136
	Russie	Cytb(1) Cytb(1)	1.113 1082	KF783139 KF783135
	Egypte	Cytb(1) BRCA1(1) BRCA1(1) RAG1(1) RAG1(1) GHR(1) TTR(1) ARN ribosomal 12S (1) VWF(1)	1090 837 823 724 636 804 1500 990 935	KF783140 KF783078 KF783063 KF783031 KF783048 KF783107 KF783159 KF783172 KF783094
	Jordan	Cytb(1) BRCA1(1)	743 683	KY558561 KY558572
	Koweït	Cytb(1)	743	KY558562
	Indéfini	Cytb(1) Cyt(2) ND2(1) APOB(1) ARN ribosomal 12S (1) ARN ribosomal 12S (1) ARN ribosomal 12S (1) Mitochondrial genome complet(2)	1140 743 1047 727 973 595 594 17283	HQ857522 KY558564 KY558561 KX959480 HQ857484 KX959471 KX959472 NC_005033

<i>Mesechinusd auricus</i> (39)	Russie	TTR(1) Cytb(2) ARN ribosomal 12S(1) Vwf(1) BRCA1(1) BRCA1(1) GHR(1) RAG1(1) RAG1(1)	1501 1140 976 935 837 823 804 725 609	KF783157 KF783126 KF783170 KF783092 KF783076 KF783061 KF783105 KF783029 KF783046
	Chine	Cytb(5)	1140	HQ857525
	Indéfini	Cytb(5) ND2(5) ARN ribosomal 12S(2) ARN ribosomal 12S(1) ARN ribosomal 12S(2) SINE ERI-1(1) SINE ERI-1(1) SINE ERI-1(1) SINE ERI-1(1) SINE ERI-1(1) SINE ERI-1(1) SINE ERI-1(1) SINE ERI-2(1) SINE ERI-1(1) SINE ERI-2(1) SINE ERI-2(1) SINE ERI-2(1) SINE ERI-2(1) SINE ERI-2(1) SINE ERI-2(1) SINE ERI-2(1)	1140 1047 982 980 978 512 487 383 365 364 352 318 315 307 303 269 254 249 233	HQ857525 HQ857506 HQ857489 HQ857491 HQ857487 AF195909 AF195906 AF195904 AF195907 AF195903 AF195908 AF195916 AF195905 AF195911 AF195912 AF195910 AF195914 AF195915 AF195913
<i>Mesechinush ughi</i> (9)	Chine	Cytb(2)	1140	MK881608
	Indéfini	Cytb (2) ND2(2) ARN ribosomal 12S (2) SCN9A(1) SCN9A(1)	1140 1047 974 592 525	HQ857530 HQ857511 HQ857492 KF052687 JX17125
<i>Paraechinus aethiopicus</i> (77)	ArabieSaoudite	TTR(1) Cytb(1) ARN ribosomal 12S(1) VWF(1) BRCA1(1) BRCA1(1) GHR(1) RAG1(1)	1495 1140 972 935 837 820 804 636	KF783161 KF783142 KF783174 KF783096 KF783080 KF783065 KF783109 KF783050

	Qatar	APOB(2) RAG1(2) BRCA1(2) ATP7A(1) ATP7A(1) BDNF (2) Cytb(2)	1154 1095 1032 673 667 561 1140	LC124835 LC124958 LC124924 LC124863 LC124862 LC124891 HQ857537
	Indéfini	Cytb(2) ND2(2) ARN ribosomal 12S(2) APOB(3) ARN ribosomal 12S(4) BDNF(1) D-loop (4) D-loop (7) BMI1 (1) PLCB4(1)	1140 1047 975 727 596 523 409 408 330 324	HQ857537 HQ857518 HQ857499 KX959474 KX959463 AY986748 KU179772 KU179768 AY986739 AY451989
	Mauritanie	RAG1(1) RAG1(4) D-loop (1) D-loop (4)	640 639 457 456	MK510371 MK510343 MK510240 MK510243
	Maroc	RAG1(6) D-loop (1) D-loop (1) D-loop (5)	639 474 458 456	MK510319 MT090965 MK510267 MK510217
	Libye	RAG1(1) D-loop (1)	639 457	MK510320 MK510218
	Sénégal	RAG1(1) D-loop (1)	639 456	MK510327 MK510225
	Algérie	BRCA1(1) Cytb(2)	683 743	KY558567 KY558553
	UAE	BRCA1(1) Cytb(1)	683 773	KY558568 KY558554
<i>Paraechinus hypomelas</i> (18)	Turkménistan	TTR(1) Cytb(1) ARN ribosomal 12S(1) Vwf(1) BRCA1(1) BRCA1(1) GHR(1) RAG1(1) RAG1(1)	1494 1140 971 935 837 823 804 715 603	KF783160 KF783141 KF783173 KF783095 KF783079 KF783064 KF783108 KF783032 KF783049
	Arabie Saoudite	RAG1(1)	715	KF783033

	Indéfini	ARN ribosomal 12S(2) APOB(1)	594 727	KX959467 KX959476
	Yemen	BRCA1(1) Cytb(1)	683 743	KY558569 KY558557
<i>Echinosorex gymnura</i> (15)	Malaisie	Cytb(6)	634	MK112014
	Indéfini	NADH2(1) ARN ribosomal 12S(1) GHR(1) BDNF(1) CREM(1) BMI1(1) Génomecomplet(2) PLCB4(1)	1044 966 904 563 435 324 17088 285	AF434827 AF434820 AF392887 AY986746 AY451978 AY986735 AF348079 AY451986
<i>Hylomysmeg alotis</i> (10)	Laos	Cytb(1) Cytb(1) ARN ribosomal 12S (3) VWF(1) BRCA1(1) GHR(1) RAG1(1) RAG1(1)	1140 885 955 880 839 775 728 636	KF783147 KF783146 KF783178 KF783098 KF783084 KF783113 KF783037 KF783054
<i>Hylomys parvus</i> (12)	Indéfini	Cytb(1) Cytb(1) ARN ribosomal 16S(1) ARN ribosomal 16S(1) Cytb(1) BRCA1(1) Cytb(2) APOB(3) ARN ribosomal 16S(1)	1,123 1,116 1,001 971 936 801 639 505 314	DQ630430 DQ630427 DQ630370 DQ630367 DQ630429 DQ630282 AH009817 DQ630203 DQ630369
<i>Hylomyssuill us</i> (31)	Indonésie	Génome mitochondrial complet(1) COI(1)	17290 481	AM905041 JF459627
	Malaisie	Génome mitochondrialcomplet (2)	15353	AM905042
	Indéfini	Elément ultra-conservé(1) BRCA1(1) ARN ribosomal 12S(1) Adra2b(1) Cytb(2) Cytb(1)	2820 2738 2605 1173 1140 1116	BRB01000000 AY121754 AY121770 AJ505819 HQ857523 DQ630428

		NADH(1) ND2(2) ARN ribosomal 16S(1) ARN ribosomal 12S(2) ARN ribosomal 12S(1) BRCA1(1) PRNP(1) COI(4) BDNF(1) APOB(1) Vwf(1) PLCB4(1)	1050 1047 982 964 962 801 701 690 522 505 403 296	AF434828 HQ857504 DQ630368 HQ857485 AF434822 DQ630283 AY133044 KY605332 AY986747 DQ630204 AY121760 AY451987
	Cambodge	l. d'élément ultra-conservé(1)	1818	KBVR0000000 0
	Thaïlande	BRCA1(1) Cytb(1)	810 402	KF783069 AB125602
<i>Hylomys cf. suillus</i> JJS-2016 (8)	Viet Nam	APOB(2) RAG1(1) RAG1(1) BRCA1(2) BDNF(2)	1166 1095 1083 607 561	LC12483 LC124961 LC124960 LC124926 LC124893
<i>Neohylomys hainanensis</i> (13)	Chine	Cytb(1)	302	AB368766
	Indéfini	Cytb(3) ND2(3) ARN ribosomal 12S(3) COI(1) COI(1) COI(1)	1144 966 648 647 645	HQ857534 HQ857496 HM031765 HM031766 HM031764
<i>Neotetracus inensis</i> (23)	Viet Nam	Cytb(1) ARN ribosomal 12S(1) BRCA1(1) BRCA1(1) GHR(1) RAG1(1) RAG1(1) VWF(1)	1140 961 849 810 804 718 636 315	KF783148 KF783181 KF783086 KF783070 KF783115 KF783038 KF783055 KF783099
	Indéfini	Cytb(2) ND2(2) ARN ribosomal 12S(2) COI(4) SCN9A(1) SCN9A(1)	1140 1044 963 690 592 525	HQ857532 HQ857513 HQ857494 KY605339 KF052692 JX171260
	Chine	COI(1) mitochondrion, complete genome(2)	657 16982	JQ600048 JX519466

<i>Podogymnura truei</i> (32)	Indéfini	TTN(1)	4428	JN632930
		BRCA2(1)	3881	JN414310
		BRCA1(1)	2838	JN414177
		ENAM(1)	2773	JN414597
		APOB(1)	2331	JN414025
		IRBP(1)	1273	JN414802
		DMP1(1)	1246	JN414473
		Adra2b(1)	1196	JN413881
		NADH2(1)	1044	AF434829
		ARN ribosomal 12S(1)	967	AF434823
		RAG(1)	931	JN414935
		CNR1(1)	918	JN633187
		GHR(1)	903	JN414740
		EDG1(1)	852	JN633256
		ADRB2(1)	810	JN633682
		RAG1(1)	692	JN633620
		ATP7A(1)	678	JN633737
		APP(1)	660	JN633548
		Cytb(2)	639	AH009818
		Vwf(1)	620	JN415064
		BDNF(1)	555	JN633375
		BDNF(1)	527	AY986745
		CREM(1)	474	AY451980
		TYR(1)	426	JN633851
		BMI1(1)	390	AY986742
		CREM(1)	364	JN633496
		RAG2(1)	331	JN633319
		ADORA3(1)	321	JN633426
		PLCB4(1)	316	JN633106
		BMI1(1)	315	JN632757
		PLCB4(1)	256	AY541991