



République Algérienne Démocratique et Populaire

Ministère de l'Enseignement Supérieur et de la Recherche Scientifique

Université de Blida 1

Faculté de SNV

Département de Biologie et Physiologie Cellulaire

Option : Bioinformatique

Mémoire présenté pour l'obtention
Du diplôme de Master Académique
Intitulé

**L'utilisation des outils de bioinformatique dans la
phylogénétique, Modèle : le genre *Gazella*.**

Présenté par :

M^{elle} KAHOUL Fairouz

&

M^{elle} DJELLAB Assia

Soutenu devant le jury composé de :

Présidente : *M^{me} BOUKRIT A* *MAA* *USDB*

Examinatrice : *M^{me} EDDAKRA A* *MCB* *USDB*

Promotrice : *M^{me} DEROUICHE L* *MCB* *ESSAIA*

Co-promotrice : *M^{me} ZATRA Y* *MAA* *USDB*

Année universitaire : 2016 /2017

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 gratitudee notre
promotrice M^{me} DEROUICHE .L*

*Maitre de conférences à L' École Supérieure des Sciences de l'Aliment et Industries
Agroalimentaires– Alger, pour sa patience, et sur tout pour sa confiance, ses remarques
et ses conseils,
sa disponibilité et sa bienveillance.*

Et

*Notre Co-promotrice M^{me} ZATRA .Y Maitre assistante au département de biologie
Je voudrais également remercier les membres des jurys :*


A la présidente Mme BOUKRITA Maitre assistante au département de biologie

D'avoir bien voulu nous faire l'honneur de présider ce jury

*L'examinatrice Mme EDDAKRA ,pour avoir accepté d'évaluer ce travail et pour
toutes leurs remarques et critiques.*

*Nos remerciements s'adressent également à tous nos enseignants, qui nous ont donné les
bases de la science.*

Merci



Je dédie ce mémoire à
A MAIS CHERS PARENTS

*Je vous remercie pour tout le soutien et l'amour que vous me portez depuis mon enfance et
j'espère que votre bénédiction m'accompagne toujours.*

Un vraiment spécial dédicace pour

Mes frères et Mes sœurs

Mes cousins et cousines

Ma tante AICHA

*Puisse Allah te donner longue vie et bonne santé,
et son mari MOHEMED.*


Ainsi qu'à Mon binôme FAIROUZ et sa famille.

*Toutes mes amies en particulier RAZKALAH IMENE, FAIZA, NOUR EL HOUDA, et
LAMIA.*

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

Assia

*Thank
you*



Je dédie ce mémoire à :

Mes parents :

Ma mère, qui a œuvré pour ma réussite, de par son amour, son soutien, tous les sacrifices consentis et ses précieux conseils, pour toute son assistance et sa présence dans ma vie, reçois à travers ce travail aussi modeste soit-il, l'expression de mes sentiments et de mon éternelle gratitude.

Mon père, qui peut être fier et trouver ici le résultat de longues années de sacrifices et de privations pour m'aider à avancer dans la vie. Puisse Dieu faire en sorte que ce travail porte son fruit ; Merci pour les valeurs nobles, l'éducation et le soutien permanent venu de toi.

Mes frères Abderrahmane, Rachid, Nabil et Oussama qui n'ont cessé d'être pour moi des exemples de persévérance, de courage et de générosité.

Mon binôme Assia,

Mes amies Zahra, wided, Faiza, Nour el Houda, Lamia, Marwa.

Et un cher ami Ilyas pour son encouragement tout le temps.

Fairouz

*Thank
you*



ABREVIATIONS

- **ADN:** Acide D soxyriboNucl ique.
- **ADNmt :** ADN mitochondrial.
- **ARN:** Acide RiboNucl ique.
- **BLAST:** Basic Local Aligement Search Tool.
- **cyt b :** cytochrome b.
- **DAMBE:** Data Analysis in Molecular Biology and Evolution.
- **DDBJ:** DNA Data Bank of Japan.
- **DnaSP:** DNA Sequence 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 / R action de polym risation en cha ne.
- **PDB:** Protein Data Bank.
- **PHYLIP:** PHYLogeny Inference Package.
- **PIR:** Protein Information Resource.
- **UE:** Unit s Evolutives.
- **UPGMA:** Unweighted Pair-Group Method using Arithmetic Averages.

Liste des figures

N° de figure	Titre	Page
01	<i>Gazella dorcas</i> , Hamma 2017.	03
02	<i>Gazella cuvieri</i> , Hamma 2017.	03
03	Répartition de genre <i>Gazella</i> dans le monde (Kowalski et al., 1991).	04
04	Répartition géographique de genre <i>Gazella</i> en Algérie (Kowalski et al., 1991).	05
05	Exemple d'un arbre phylogénétique.	11
06	Les différents arbres phylogénétiques : a) arbre enraciné, b) arbre non enraciné.	12
07	Diverses catégories de groupes taxonomiques (Tourasse, 1992).	13
08	Les étapes initiales pour récupérer les séquences de la GenBank.	22
09	Les dix-neuf espèces de genre « <i>Gazella</i> » obtenues dans la GenBank, ces espèces sont indiquées par des flèches en rouge.	23
10	Exemple des séquences nucléotidiques trouvées pour l'espèce <i>Gazella arabica</i> .	23
11	Exemple d'une séquence nucléotidique au format FASTA.	24
12	La page principale du programme Fabox.	25
13	La conversation des séquences sous forme fasta par l'utilisation de Fabox.	26
14	Alignment trimmer (la forme fasta).	26
15	Barre d'outils de MEGA5.	27
16	Le programme ClustalW utilisé pour l'alignement multiple des séquences du marqueur moléculaire choisi.	27
17	La méthode UPGMA utilisée pour la construction de l'arbre phylogénétique.	28
18	Les paramètres utilisés pour la reconstruction phylogénétique.	29
19	Barre d'outils de DnaSP.	30

20	Interface utilisateur de DnaSP.	30
21	Barre d'outil du programme DAMBE.	31
22	Barre d'outil qui fait les graphes de C/G et A/T.	31
23	La page d'accueil de Network 5.0.	33
24	Ouverture du fichier de données de rdf pour le calcul de MJ Network.	34
25	Un fichier sous forme out a été sauvegardé.	35
26	Reconstruction de réseau haplotypique.	35
27	Structure de l'ADN mitochondrial et la position de cytb et D-loop (Rustin et al., 1995).	38
28	Les séquences de cytb ont été convertir sous forme fasta à partir de bloc note grâce à fabox.	40
29	Arbre phylogénétique reconstruit par Méthodes UPGMA basées sur 70 séquences de nucléotides de cytochrome-b de genre <i>Gazella</i> en utilisant MEGA5.	42
30	L'arbre final de genre <i>Gazella</i> basé sur 66 séquences de cytb en utilisant la méthode UPGMA « Kimura 2-paramater » par la suite informatique MEGA5.	44
31	Comparaison entre les séquences de <i>G. arabica</i> et <i>G. gazella</i> par le programme BLAST.	46
32	Comparaison entre les sequences de <i>G. spkie</i> , <i>G. arabica</i> et <i>G. gazella</i> à l'aide du programme BLAST.	46
33	Comparaison entre la séquence de <i>G. saudiya</i> et les séquences de <i>G. dorcas</i> par logiciel Blast.	47
34	Arbre phylogénétique des séquences D-loop du genre <i>Gazella</i> en utilisant la méthode UPGMA et le logiciel MEGA5.	48
35	Comparaison entre les séquences de Clade A par le programme BLAST.	49
36	Comparaison entre les séquences de <i>G. arabica</i> et <i>G. gazella</i> à l'aide du programme BLAST.	50
37	Les haplotypes ont été construit à l'aide du logiciel DnaSP de cytb.	52
38	Résultat d'alignement des séquences D-loop du genre <i>Gazella</i> en utilisant le logiciel FABOX.	53

39	Les haplotypes des séquences D-loop du genre <i>Gazella</i> en utilisant le logiciel DNASP.	54
40	Variations des fréquences GC (A) et de l'AT (B) de Cytb produites à partir de DAMBE.	56
41	Variations des fréquences GC (C) et de l'AT (D) de D-loop produites à partir de DAMBE.	57
42	Comparaison des fréquences entre cytb et D-loop réalisés par logiciel DAMBE.	58
43	Réseau haplotypique réalisé par le logiciel Network à partir de 66 séquences de Cytb de genre <i>Gazella</i> .	59
44	La croissance des haplotypes des séquences D-loop du genre <i>Gazella</i> en utilisant le logiciel Network.	60

Liste des tableaux

N° de tableau	Titre	Page
I	La description morphologique de genre <i>Gazella</i> .	6
II	Les formats de données utilisés par logiciel DAMBE (Xia, 2001).	32
III	Les séquences de cytb et D-loop extraites de la GenBank de genre <i>Gazella</i> .	37
IV	Tableaux qui représentent le pourcentage de similarité entre les séquences par l'utilisation de serveur Blast.	50
V	Comparaison des indices de diversités calculés avec DnaSP sur nos échantillons étudiés.	56

VI	Exemple d'une séquence au format GenBank.	Annexe
VII	Exemple d'une séquence nucléotidique au format EMBL.	Annexe
VIII	Exemple d'une séquence nucléotidique au format DDBJ.	Annexe
IX	Les séquences extraites à partir de la GenBank (types des gènes et leur origine géographique).	Annexe

GLOSSAIRE

- **Ancêtre** : spécifique ou infra-spécifique (sinon, il est paraphylétique et équivaut au « groupe ancestral » classique, qui n'est pas un groupe naturel).
- **Bootstrap** est une méthode statistique utilisée pour évaluer la robustesse d'un arbre phylogénétique. Elle part de l'hypothèse que les sites évoluent de manière indépendante.
- **Caryologie** : une branche basée sur le caryotype de la cellule nucléaire.
- **FASTA** : algorithme proposé par Pearson basé sur l'identification rapide de fragments strictement identiques entre la séquence cible et les séquences de la banque (**Comet, 1998**).
- **Homologie** est une similitude héritée d'un ancêtre commun, contrairement à l'homoplasie (**Lankaster, 1870**).
- **Homoplasie** est divisée en convergence (apparition indépendante d'un caractère chez deux espèces) et réversion (apparition d'un caractère ayant l'apparence de la morphologie ancestrale).
- **IUCN** (la liste rouge) est un outil important pour surveiller la biodiversité et la gestion des risques d'extinction (**Bärman et al., 2012**).
- **Jackknife** Cette méthode a été appliquée aux problèmes de phylogénie par **Mueller et Ayala (1982)**. Supposons une matrice de données constituée de K caractères. Le Jackknife consiste à effectuer K reconstructions phylogénétiques différentes, chacune d'elles ayant été obtenue en supprimant un caractère différent.
- **Lignée** est le regroupement de toutes les espèces issues du même ancêtre.
- **Matrice de dissimilarités ou matrice de distances** est un tableau à double entrée comprenant horizontalement comme verticalement la même série d'espèces. Chacune des cases de la matrice contient la distance de dissimilarité qui sépare les deux espèces concernées. La valeur de dissimilarité est égale à zéro pour deux espèces identiques.
- **Maximum de parcimonie** est une approche statistique non-paramétrique très utilisée, notamment pour l'inférence phylogénétique. Sous l'hypothèse du maximum de

parcimonie, l'arbre phylogénétique inféré est celui qui requiert le plus petit nombre de changements évolutifs.

- **Maximum de vraisemblance** est une approche statistique courante utilisée pour inférer les paramètres de la distribution de probabilité d'un échantillon donné. Elle est très employée dans le domaine de la bioinformatique.
- **Taxon (Lam, 1950 ; Mayr, 1953)** : groupe d'organismes reconnu en tant qu'unité formelle à chacun des niveaux de la classification (**Simpson, 1961**) biologique hiérarchisée . Unité de la taxinomie,
- **Taxonomie (Candolle, 1813)**, devenu Taxinomie (Littré) : Théorie et pratique de la classification des organismes (Code International de Nomenclature Zoologique, 4^{ème} édition, 2000)
- **Transition** est une mutation au cours de laquelle une base purique est remplacée par une autre base purique ou par une base pyrimidique.
- **Transversion** est une mutation au cours de laquelle une base purique est remplacée par une base pyrimidique ou vice-versa.
- **Unité Evolutive (UE)** ou **Taxon terminal** : taxon dont les données sont observées a priori,
- **UPGMA** : une méthode utilisée pour construire des arbres phylogénétiques si les séquences ne sont pas trop divergentes (**Golding et al, 2003**).

RÉSUMÉ

Ce mémoire expose l'utilisation des outils de bioinformatique dans les analyses phylogénétique du genre *Gazella*, tel que la reconstruction de l'arbre phylogénétique, l'identifier des haplotypes, et avoir l'évolution des espèces. Nous avons tout d'abord présenté le genre *Gazella*, 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 genre *Gazella*.

Les analyses phylogénétique 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 l'identification des haplotypes, **Network** pour avoir l'évolution des espèce, et **DAMBE** pour faire des calcule statistique, avec l'utilisation de différentes méthodes comme UPGMA, Kimura- 2 parameters ; ces analyses appliquée sur 66 séquences de cytochrome b et 16 séquences de D-loop de l'ADN mitochondrial de 19 espèces appartiennent au genre *Gazella* prises dans différentes régions géographiques démontrent la monophylie du genre *Gazella*.

Les méthodes d'analyses phylogénétiques, que ce soit en bifurcation, les haplotypes ou en réseau par les logiciels MEGA, DnaSP, Network ont montré pour le genre *Gazella* l'existence de deux clades distincts.

Mots clés : *Gazella*, cytochrome b, D-loop, phylogénétique, logiciel, monophylie.

Summary

This memory exposes the use of useful of bioinformatic in the analyses phylogenetic of the genus *Gazella*, such as the reconstruction of the phylogenetic tree, to identify haplotypes, and have the evolution of the species. First of all we have presented the genus *Gazella*, their morphology, their distribution in the world and Algeria in particular; we have outlined the methods of phylogenetic reconstruction to infer phylogenies of kind *Gazella*.

The phylogenetic analysis obtained in this work have been developed using the different software such as, MEGA5 for the reconstruction of the phylogenetic tree, DnaSP for the identification of haplotypes, Network for having the evolution of species, and DAMBE to make calculates statistics, with the use of different methods as, UPGMA, Kimura- 2 parameters; these analysis applied on 66 sequences of cytochrome b and 16 sequences of D-loop of mitochondrial DNA of 19 species belong to the genus *Gazella* taken in different geographical regions demonstrate the monophyly of the genus *Gazella*.

The methods of phylogenetic analyzes, either in Junction, the haplotypes or in network by the MEGA, DnaSP, Network softwares have shown for the kind *Gazella* the existence of two distinct clades.

Key words: *Gazella*, cytochrome b, d-loop, phylogenetic, software, monophyly.

ملخص

تتضمن هذه المذكرة استخدام مختلف وسائل الاعلام الالي الحيوي في تحليل دراسة نشوء جنس الغزال , مثل إعادة بناء شجرة النشوء , التطور , تحديد les haplotypes و تطور الأنواع. قمنا بدراسة جنس الغزال, مورفولوجيا و توزيعه على المستوى العالمي و المحلي بصفة خاصة, قدمنا أساليب إعادة الإعمار النشوء لاستنتاج نشوء جنس غزال .

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 الانواع, **DAMBE** يقوم بعمل حسابات احصائية. كما قمنا باستخدام مختلفة الطرق (UPGMA, كيمورا - 2 parameters). قمنا بمعالجة 66 سلسلة من مورثة cytb و 16 سلسلة من مورثة D-loop لـ 19 نوع من جنس الغزال في مختلف المناطق الجغرافية. وفي النهاية تحصلنا على النتائج التي تدل على la monophylie التي تخص نوع الغزال.

طرق التحليل الفيلوجيني, سواء في haplotypes , التشعيب او شبكة. البرامج MEGA, DnaSP, Network اظهرت في النوع الغزال بوجود تشعيب مترابطين.

المصطلحات: الغزال , D-loop, Cytb , شجرة النشوء, البرامج, la monophylie.

SOMMAIRE

INTRODUCTION 1

CHAPITRE I : RAPPELS BIBLIOGRAPHIQUES

I.1. Terminologie et modèle d'étude 3

I.1.1. Présentation de l'espèce3

I.1.2. Distribution géographique de genre *Gazella*4

I.1.3. Description morphologique5

I.2. Etude phylogénétique 9

I.2.1. Définition de la phylogénie9

I.2.2. Phylogénie moléculaire9

I.2.3. Représentation phylogénétique10

I.2.4. Arbre phylogénétique11

I.2.4.1. Caractéristiques générales des arbres11

I.2.4.2. Différentes représentations graphiques pour les arbres13

I.2.5. Reconstruction d'arbres phylogénétiques14

I.2.6. Critères de choix entre les méthodes15

I.3. Stockage de données biologiques 15

I.3.1. Banques des données15

I.3.1.1. Les banques généralistes16

I.3.1.2. Les banques spécialistes17

I.3.2. Différentes étapes et logiciels d'une étude phylogénétique17

I.3.2.1. Logiciels de nettoyage17

I.3.2.2. Logiciels d'alignement18

I.3.2.3. Calcul des distances19

I.3.2.4. Logiciels de reconstruction de l'arbre phylogénétique19

I.3.2.5. Logiciels de visualisation d'arbres phylogénétiques20

CHAPITRE II : MATERIEL ET METHODES

II.1. Analyses moléculaires et phylogénétiques 21

II.1.1. Etapes d'Analyses Moléculaires21

II.1.1.1. Echantillonnage.....21

II.1.1.2. Extraction et quantification de l'ADN21

II.1.1.3. Amplification de l'ADN par PCR.....22

II.1.1.4. Séquençage22

II.2. Extraction des données 22

II.2.1. Définition de la GenBank22

II.2.2. Traitement des séquences de la GenBank24

CHAPITRE III : RESULTATS ET DISCUSSION

III.1. Résultats de la GenBank 36

III.2. Analyses phylogénétiques 38

III.2.1. Construction des arbres phylogénétiques38

III.2.2. Résultat de DnaSP52

III.2.3. Analyses statistique56

III.2.4. Résultat de network59

CONCLUSION 62

RÉFÉRENCES BIBLIOGRAPHIQUES 64

BIBLIO-WEB

ANNEXE

Introduction

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**).

Pendant la dernière décennie, plusieurs études ont conduit sur les gazelles (**Bärmann et al., 2013**), gazelle désigne un mammifère ruminant voisin de l'antilope, mais qui a une plus petite taille. La gazelle est un animal très rapide qui vit dans les steppes d'Afrique et d'Asie. La gazelle fait partie de la famille des Bovidés.

L'examen des gazelles sur longue période a permis de mieux comprendre la façon dont elles pouvaient croître. En particulier, il apparaît que, pour une grande majorité d'entre elles, cette croissance a lieu sur une période relativement restreinte.

La sélection de gazelles sur la base d'une période plus restreinte permettrait à la fois de mieux coller à ce constat statistique et d'examiner le devenir des gazelles ainsi sélectionnées sur le reste de la période d'étude.

Notre travail est basé sur les analyses phylogénétiques. Le but de ce mémoire repose sur l'utilisation de plusieurs logiciels de bioinformatique et utiliser une base des données de biologie moléculaire pour traiter les séquences afin de trouver l'arbre phylogénétique et la position de différentes espèces de genre *Gazella*.

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

- ❖ Présenter notre modèle d'étude Genre *Gazella*, leur classification et leur distribution géographique dans le monde et particulièrement en Algérie ainsi que leur morphologie.
- ❖ Définir la phylogénie, un 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, rapporter 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 ainsi que les perspectives ouvertes par ce travail.

I.1. Terminologie et modèle d'étude

I.1.1. Présentation de l'espèce

Les gazelles sont des mammifères, de la famille des bovidés, comportent quatre genres monophyletiques : *Antilope*, *Nanger*, *Eudorcas* et *Gazella* (Lerp et al., 2013). La tribu Antilopiné comporte aujourd'hui 13 genres : *Raphiceros*, *Ourebia*, *Madoqua*, *Dorcatragus*, *Saiga*, *Litocranius*, *Ammodorcas*, *Antidorcas*, *Procapra*, *Eudorcas*, *Nanger*, *Antilope* et *Gazella*, (Efron et al., 1976; Gentry, 1992; Rebholz et al., 1999; Groves, 2000; Grubb, 2005; Groves et al., 2011; Hassanin et al., 2012). Notre travail sera basé sur le genre *Gazella*, exemple : *G.dorcas* (Figure1), *G.cuvier* (Figure2).

La classification systématique de ces mammifères est la suivante :

Classification	
Règne	Animalia
Embranchement	Chordata
Classe	Mammalia
Ordre	Artiodactyla
Sous/Ordre	Ruminantia
Famille	Bovidae
Genre	<i>Gazella</i> (Blainville, 1816)



Espèces :	Sous-espèces :
-----------	----------------

G.arabica

G.bennettii

G. cuvieri

G. dorcas *G. dorcas pelzelni*

G. dorcas x leptoceros

G. erlangeri

G.gazella *G.gazella gazella*

G.gazella muscatensis

G. leptoceros

G. saudiya

G. spekei

G. subgutturosa

G. subgutturosa marica

G. subgutturosa sairenses

G. subgutturosa subgutturosa

G. subgutturosa yarkandensis

G. bilkis

G. dama

G. granti

G. rufifrons

G. rufina

G. soemmerringii

G. thomsonii

Figure 1 : *Gazella dorcas*, Hamma
2017

Figure2 : *Gazella cuvieri*, Hamma
2017

I.1.2. Distribution géographique de genre *Gazella*



Le genre *Gazella* (*bovidae*, *Artiodactyla*) est l'un des taxons les plus communs dans la faune des « Hipparion », largement distribuées en Eurasie et en Afrique. (Arambourg, 1959 ; Gentry., 1970, 2010).

✚ Distribution dans le monde

Les gazelles sont distribuées à travers l'Afrique et l'Asie et sont adaptées aux environnements arides et semi-arides (Lerp et al., 2013), leurs gamme géographique enjambe l'Afrique du nord, Arabie, Moyen-Orient, Inde, la Chine, et la Mongolie (Bärmann et al., 2013) (Figure 3).

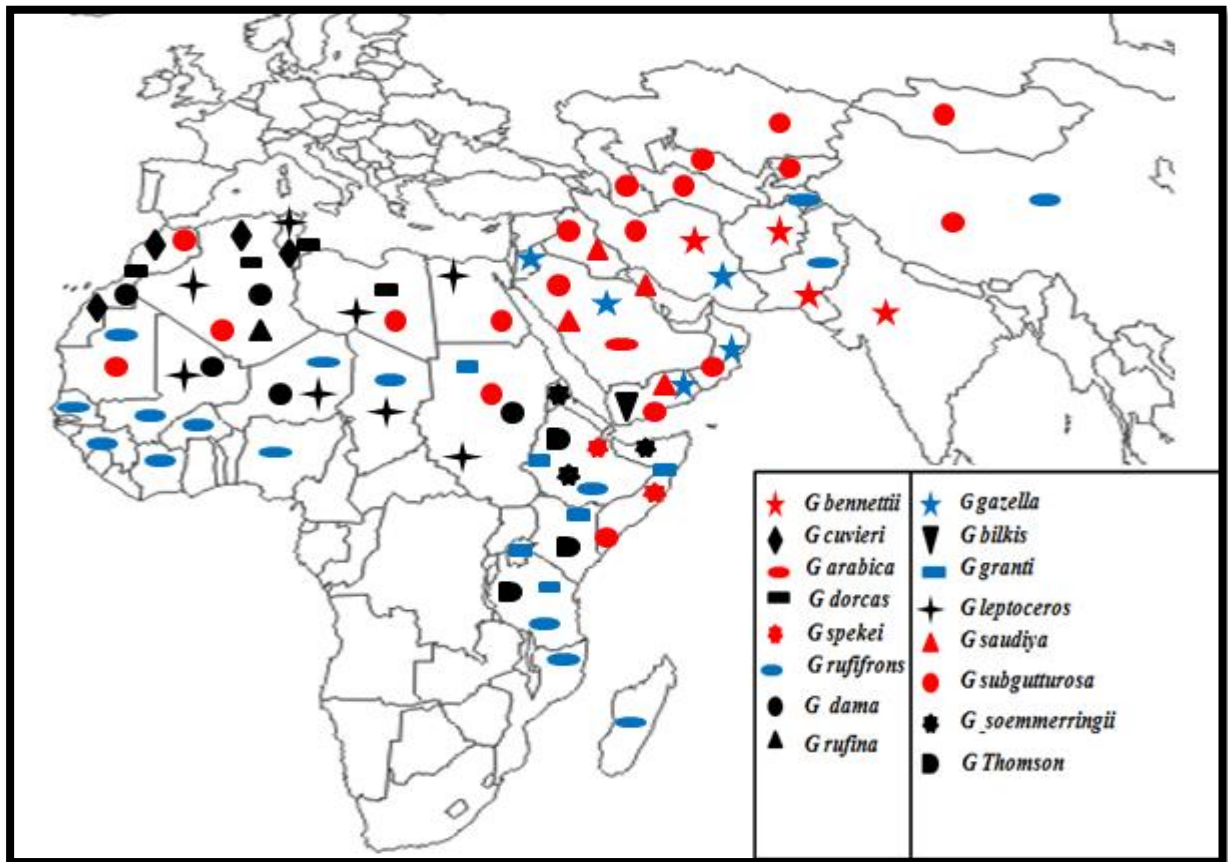


Figure 3 : Répartition de genre *Gazella* dans le monde (Original).

✚ Distribution en Algérie

En Algérie, les travaux ont permis de dénombrer 107 espèces de mammifères, selon le livre « les mammifères d'Algérie, des origines à nos jours », Kowalski a mentionné la distribution de genre *Gazella* en différents régions en Algérie(Figure 4).

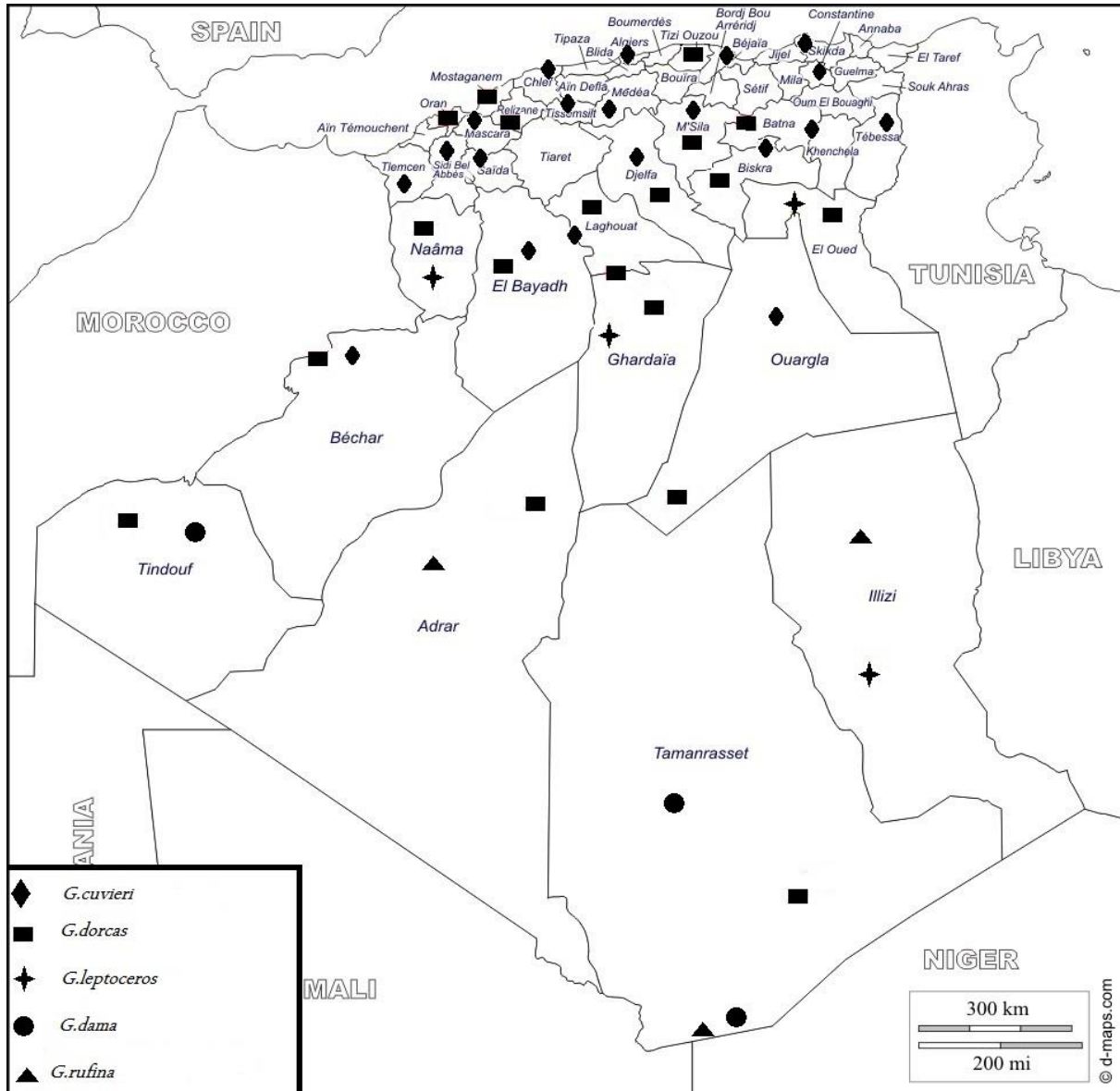


Figure 4 : Répartition géographique de genre *Gazella* en Algérie (Original).

I.1.3. Description morphologique

La gazelle est l'un des genres les plus riches en espèces dans les ruminants à cornes. En dépit de la similitude globale dans la taille du corps et la morphologie (Bärmann et al., 2013) (Tableau I).

Tableau I : La description morphologique de genre *Gazella* (Original).

<i>Espèce</i>	<i>Morphologie</i>
<i>G. bennettii</i>	<ul style="list-style-type: none"> • Mesurant entre 80 et 95 cm de long, de 50 à 65 cm de haut pour un poids allant de 14 à 25 kg. • Les femelles font généralement la moitié de la taille des mâles, elles sont plus minces et les anneaux sont moins prononcés. • Le pelage d'été de cette gazelle est de couleur rougeâtre - chamois sur le dessus et blanc pâle sur le ventre, avec un poil lisse et brillant. Le front est marron foncé ou noir. • Du coin de l'œil jusqu'au museau on peut voir une bande blanche (Loder et al., 1894).
<i>G. cuvieri</i>	<ul style="list-style-type: none"> • Est une gazelle de taille moyenne, son poids est d'environ de 35kg. • Le mâle se différencie de la femelle par son poids légèrement supérieur. • Les deux sexes possèdent sensiblement les mêmes mensurations (70 cm au garrot).
<i>G. arabica</i>	<ul style="list-style-type: none"> • Les caractéristiques de crâne distinguent le spécimen de toutes autres gazelles. • Il est très improbable que le spécimen représente une ancienne population sauvage des gazelles (Barmann et al., 2013).
<i>G. dorcas</i>	<ul style="list-style-type: none"> • C'est un ongulé de petite taille. • Le mâle possède des cornes en forme de lyres recourbées vers l'arrière, puis vers le haut (18 à 28 anneaux peuvent être dénombrés). • Le pelage ras et lisse présente une coloration générale fauve pâle avec une bande plus foncée sur les flancs que accentue le contraste avec le blanc du ventre, des culottes et de la face interne des membres, le dessus du museau et le front brun roux sont marqués par une bande blanchâtre que s'étend de la base des cornes au museau, encerclant l'œil, la queue de longueur moyenne, se termine par une touffe de poils noirs. (Kowalski et al., 1991).
<i>G. erlangeri</i>	<ul style="list-style-type: none"> • Est une petite gazelle sombre avec un corps robuste et des jambes courtes (Barmann et al., 2013).
<i>G. gazella</i>	<ul style="list-style-type: none"> • Est une petite gazelle mesurant de 98 à 115 cm de long pour un poids allant de 16 à 30 kg. La queue mesure de 8 à 13 cm. • Les deux sexes sont dotés de cornes qui mesurent entre 5,8 et 29 cm de long (Barmann et al., 2013).
<i>G. bilkis</i>	<ul style="list-style-type: none"> • Une altitude de 1,230 – 2,150 m (Barmann et al., 2013).
<i>G. granti</i>	<ul style="list-style-type: none"> • Mesurant entre 1,40 et 1,66 m de long, de 87 à 97 cm de haut, pour un poids allant de 38 à 80 kg.

	<ul style="list-style-type: none"> • Le dimorphisme sexuel est présent chez cette espèce, les mâles étant plus grands et plus lourds que les femelles. • La queue mesure entre 20 et 28 cm de long. • Les deux sexes disposent de cornes qui mesurent entre 60 et 80 cm chez le mâle contre 30 à 40 cm pour celles des femelles. • Le pelage est de couleur sable sur la partie supérieure du corps et blanc sur la partie intérieure des jambes et le ventre. • La caractéristique la plus distinctive de cette gazelle est la bande verticale noire distincte qui descend de chaque côté des fesses blanches <p>(Legendre, 1997).</p>
<i>G. leptoceros</i>	<ul style="list-style-type: none"> • Cornes du mâle lyrées et faiblement incurvées en forme de S, assez droites, pointes dressées, écartement variables, section à la base ovale allongé : 20 à 25 anneaux, quart supérieur lisse, cornes de la femelle plus courtes, plus fines, plus arrondies, mais recourbées en S plus droites et plus parallèles avec 20 à 26 faibles anneaux, souvent inégales ou irréguliers. Oreilles et sabots très longs et étroits, sabots latéraux courts et larges. (Kowalski k et al., 1991).
<i>G. saudiya</i>	<ul style="list-style-type: none"> • Est une petite gazelle. • Les jambes plus courtes que la gazelle dorcas. • Le pelage est également plus clair. • Les oreilles sont de couleur chamois clair et elle n'a pas de tache sur le nez. • Les bandes lumineuses du visage sont blanc-chamois (Barmann et al., 2013).
<i>G. spekei</i>	<ul style="list-style-type: none"> • Mesurant entre 95 et 105 cm de long, de 50 à 60 cm haut, pour un poids allant de 15 à 25 kg. • La longueur de la queue varie de 15 à 20 cm. • C'est la plus petite des espèces de gazelles actuelles (Farian, 2009).
<i>G. subgutturosa</i>	<ul style="list-style-type: none"> • Gazelle subgutturosa est plus grande et plus lourd comparé aux dorcas. • Poids est d'environ de 23,2 kg (femelle) et 27,4kg (mâles) (Barmann et al., 2013).
<i>G. rufifrons</i>	<ul style="list-style-type: none"> • Mesurant entre 90 et 110 cm de long, de 65 à 70 cm de haut, pour un poids allant de 25 à 30 kg. • La taille de la queue varie de 20 à 30 cm. • Mâles et femelles sont tous deux de taille similaire. • Le pelage est généralement de couleur rouge fauve, sauf le ventre et le croupion qui sont blancs. Il dispose d'une mince bande noire distincte de 2 à 4 cm de haut qui va de coude au grasset (patte arrière). • Les cornes sont présentes chez les deux sexes. Elles forment une légère courbe en S. Celles des femelles sont lisses et minces (Groves et al., 2011).

<p><i>G. dama</i></p>	<ul style="list-style-type: none"> • Cornes du mâle lyrées, recourbées en arrière, redressées à l'extrémité et parfois de nouveau incurvées en avant, section de la base circulaire ovale : 18 à 23 anneaux parfois lisses. • Cornes de la femelle semblables mais plus courtes et plus fine. • Anneaux plus plats. • Oreilles assez longues, pelage court et lisse. • Face supérieure de la queue portant une faible touffe de poils. • Coloration variant selon l'âge, la saison et la région. • Le cou, le dos, les flancs, la face externe des membres est fauve, marron vif, mais la partie postérieure du corps et blanche, la tête est de teinte clair à l'exception du front de couleur brunâtre (Kowalski et al., 1991).
<p><i>G. rufina (G Rouge)</i></p>	<ul style="list-style-type: none"> • Cornes du mâle plus ou moins lyrées et faiblement incurvées en S. • Pointers dressées ou convergentes, section de la base circulaire, 9 à 18 anneaux, quart supérieur lisse. • Cornes de la femelle plus courtes et minces, plus droites, presque lisse avec seulement 13 anneaux, dessus (tête, cou, épaules, dos, cuisses, face externe des pattes). • Une tâche foncée sur le museau et, le plus souvent, sourcil blanchâtre de la base des cornes au mufler (Kowalski et al., 1991).
<p><i>G. soemmerringii</i></p>	<ul style="list-style-type: none"> • Mesurant entre 1,25 et 1,50 m de long, de 85 à 92 cm de haut, pour un poids allant de 38 à 46 kg. • La taille de la queue varie de 18 à 28 cm. La taille de la queue varie de 18 à 28 cm. • Les cornes des mâles sont en forme de lyrées et fortement striées. Elles mesurent environ 58 cm de long. • Le pelage est de couleur fauve rouge uniforme sur la partie supérieure du corps avec une grande tache blanche sur la croupe. La partie inférieure, l'intérieur des jambes et la queue sont également blancs. • Les principales caractéristiques qui la distinguent des autres gazelles sont la tache blanche sur le croupion qui s'étend sur les côtés, les couleurs de la partie dorsale et ventrale qui se rencontrent le long des flancs avec une ligne nette, et les marques faciales présentes sur le visage. Des bandes sombres descendent du nez et des yeux au nez, séparés par des bandes blanches. • La queue est courte et conique et se termine par une touffe noire (Bardelli et al., 2014).
<p><i>G. Thomson</i></p>	<ul style="list-style-type: none"> • Elle mesure entre 80 et 120 cm de long, de 55 à 82 cm de haut, pour un poids allant de 15 à 35 kg. • Le dimorphisme sexuel est présent chez cette espèce, les mâles étant plus grands et plus lourds que les femelles. • Le pelage est court et lisse. La partie supérieure est brun clair. Une bande plus claire marque les flancs, soulignée d'une bande brune ou noire très visible. La partie ventrale ainsi que les fesses sont blanches. • Le visage est marqué de blanc, de fauve, de brun foncé et de noir. Ces

	<p>couleurs varient selon les individus.</p> <ul style="list-style-type: none"> • Les cornes sont présentes chez les deux sexes (<i>Kingdon et al., 2015</i>).
--	---

Beaucoup d'espèces de gazelle sont mises en danger ou menacées en raison de la chasse, de la perte d'habitat, et de la concurrence avec bétail domestique (**Cloudsley-Thompson ; 1992; East, 1988, 1989, 1990, 1992, 1993 ; IUCN, 1988; Loggers et al., 1992; Ryder, 1987; Saleh, 1987**). l'IUCN a classé les espèces de *G. arabica*, *G. bennetti*, *G. cuvieri*, *G. dorcas*, *G. gazella*, *G. leptoceros*, *G. marica*, *G. saudiya*, *G. spekie*, *G. subguttorsa*, *G. bilkis*, *G. dama*, *G. granti*, *G. rufina*, *G. soemmerringii*, *G. thomsonii* comme des espèces « éteinte a l'état sauvage » et leur chasse est interdite.

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 (**Haeckel, 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 utilise 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 adapté. Par exemple, si le but 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 0.8 Kb 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 régulatrice (**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 :

✚ **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.

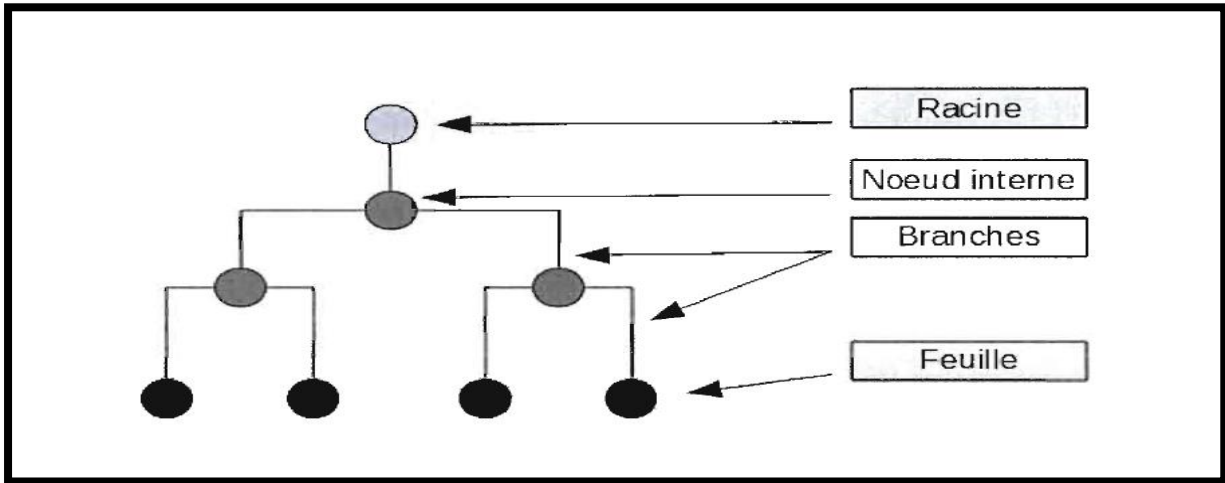


Figure 5 : Exemple d'un arbre phylogénétique.

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.

- a) **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 6- a).
- b) **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 6- b).

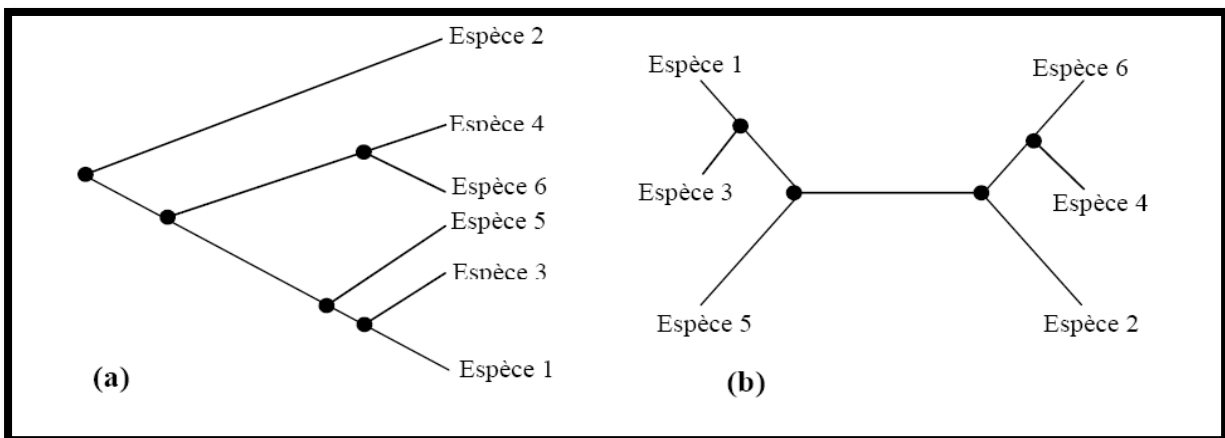


Figure 6 : Les différents arbres phylogénétiques : a) arbre enraciné, b) arbre non enraciné.

c) **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 issu d'un même ancêtre commun, c'est le cas de groupe (X, Y, Z) de la Figure 7
- ❖ **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 7)
- ❖ **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 7).

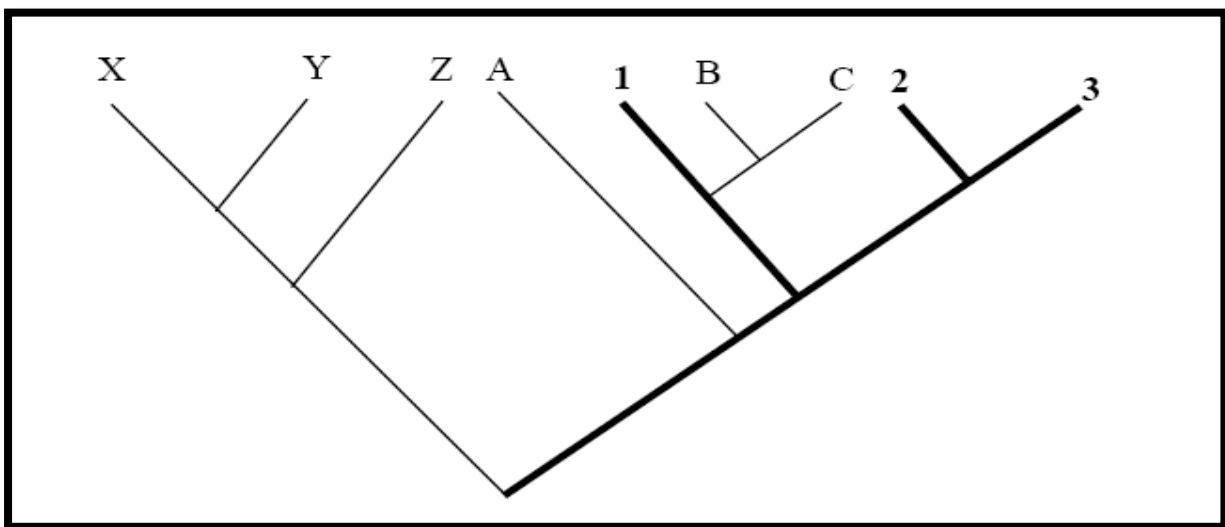


Figure 7 : Diverses catégories de groupes taxonomiques. Les taxa X, Y, Z forment un groupe monophylétique. Les taxa 1, 2 et 3 forment un groupe paraphylétique. Les taxa A, B et C forment un groupe polyphylétique (**Tourasse, 1992**).

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 et al., 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 et al., 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 phénétique** : 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 de données 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 données 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 contiguës relativement courtes afin de créer des séquences plus longues.

❖ Bioedit

Beaucoup de chercheurs dans le domaine de la biologie moléculaire ont utilisé des modules de BioEdit pendant leur recherche originale. BioEdit 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 (**Bérard, 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 ont été parallélisés pour réduire le temps d'exécution (**Li, 2003**).

❖ Blast

Basic Local Alignment Search Tool ; permis d'identifier des régions de similarité locale entre 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 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 D.L 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 telles 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 DnaML. 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**).

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 phylogénétique du genre *Gazella*, tel que la reconstruction de l'arbre phylogénétique, l'identifier des haplotypes, et avoir l'évolution des espèce.

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 genre *Gazella*. 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 phylogénique.

II.1.1. Etapes d'Analyses Moléculaires

II.1.1.1. Echantillonnage

Les échantillons employés pour extraire l'ADN se 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 commercialisé. 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 Polymerase Chain Reaction (PCR) (Kleppé 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. Extraction des données

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 du genre *Gazella* 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 sur la source NCBI.
3. Ecrire le nom de notre genre « *Gazella* » (Figure 8).

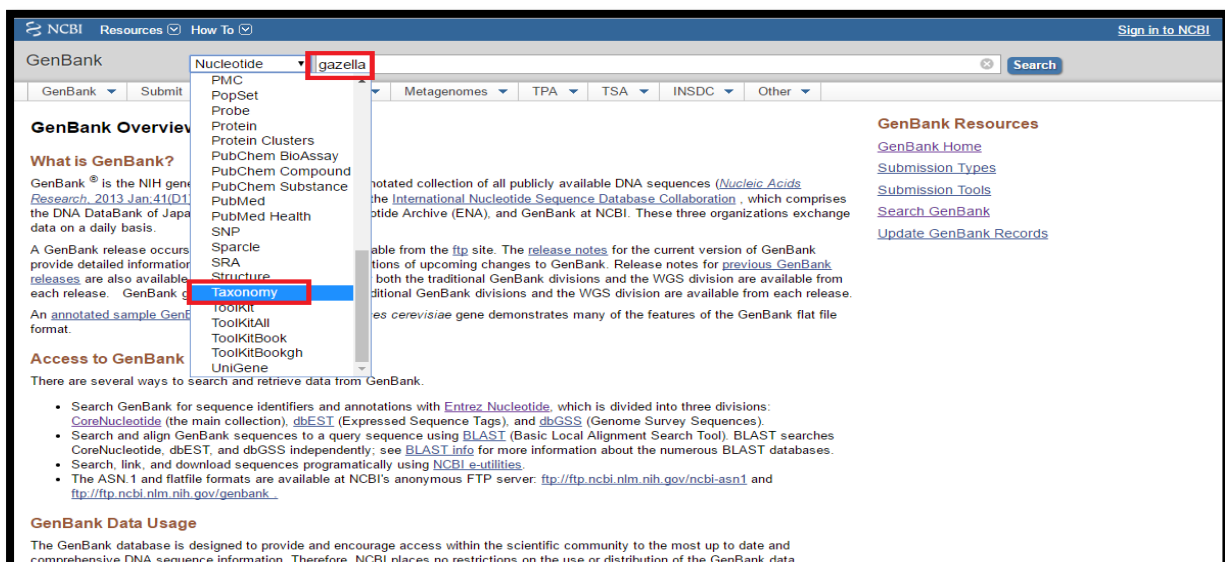
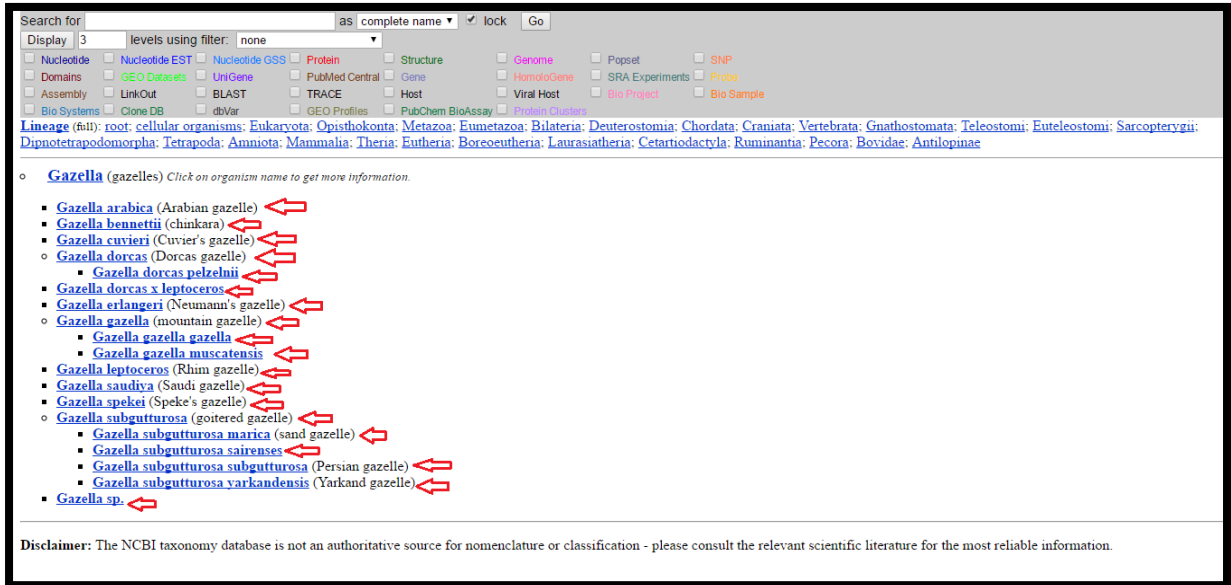


Figure 8 : Les étapes initiales pour récupérer les séquences de la GenBank.

4. En entrant sur *Gazella*, nous obtenons toutes les espèces qui appartiennent à ce genre



(Figure 9).

Figure 9 : Les dix-neuf espèces de genre « *Gazella* » obtenues dans la GenBank ; ces espèces sont indiquées par des flèches en rouge.

5. Nous avons analysé toutes les espèces une par une pour identifier le nombre des séquences nucléotidiques trouvées dans cette banque (Figure 10).

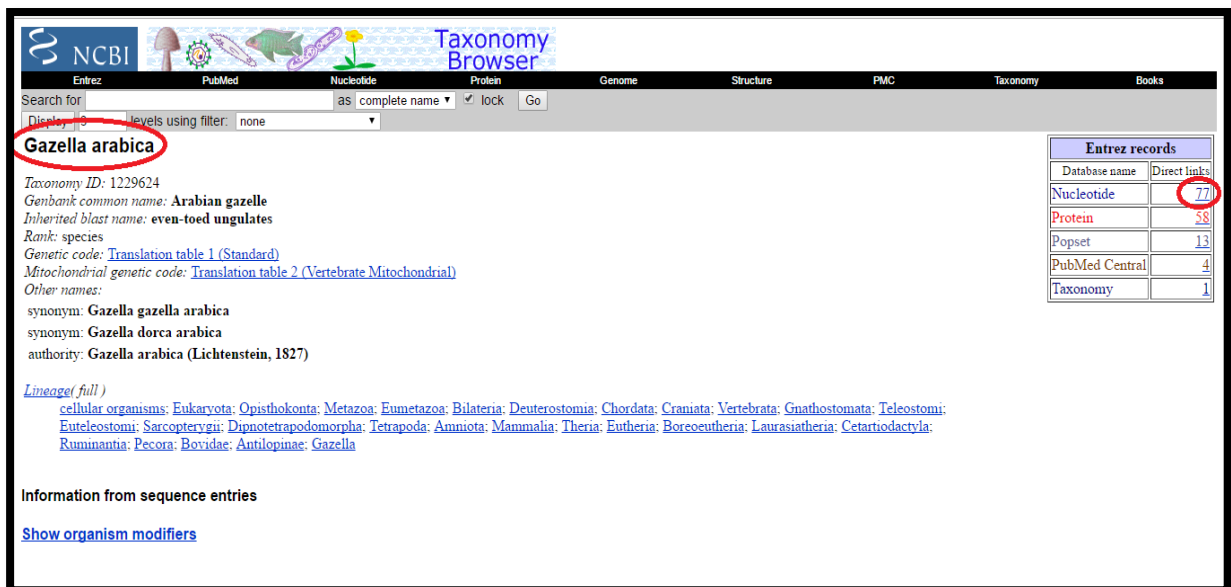


Figure 10 : Exemple des séquences nucléotidiques trouvées pour l'espèce *Gazella arabica*.

6. Les séquences représentatives de chacune des 19 espèces ont été téléchargées sous format

```
FASTA Send: ▾

Gazella arabica isolate 182 SPARC-related modular calcium binding protein 1 (SMOC1) gene, partial cds

GenBank: KU560887.1
GenBank Graphics PopSet

>KU560887.1 Gazella arabica isolate 182 SPARC-related modular calcium binding protein 1 (SMOC1) gene, partial cds
GTCCGCTGCCGGGACCTCCACACGGTAAGCCCTGGAGCCTGTCCTGCCCTCAGAGGGACCTGGCCC
AGGAGAACCAGCTTCCTGTTGGGCTGGCACCTCGCGTCTCCCAGATTCAGTCTCCTGGGCAGGTGGGAAAC
TGAGACACTCGGGGGCAGCATCTCTGCTCGTGGGCAGATGAGGTCCTGGTGGAGAGACTGGATCTTGG
ATTCTGGTACAGGGCTGGCCCTTCTAGCTGGGTGGCCTTGGGAAAGGAGGTAATCTCTGAACCTC
AGTTCACCTCAGGCAGGTATTGCCCTGCCTGAGTCAGTAGGTGCTATGGGTCAAAAGATCCTGTGTGCA
GACATAAGGGGACACCACCTACCTGAGTAACTCAGACAATGTGGCAACATGCCTGAACATGTCCTGTGG
GGATGCTGATGGTGGGATTGCCCTGTAGCTAGTGGACGGTGGGGAGCTATTTGCCACCAGCTCTGGGT
GCTCTAGGCTTAGAGACCACAAAATCATTGCTAGTGGTGGGGCTGAAGTCGGCTCAGGTTCCAGCCCC
TCCTTACCCCTGCTTTCTCTTCTTAGCTACGTGATGCCAGTTGTGAGAGTGATGCCAGGGCTAAG
AGTGGGGAGGTGG
```

FASTA à partir de la "GenBank" (Figure 11).

Figure 11 : 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 nombre logiciels de traitement et d'analyse de séquences.

Au total, la GenBank contient 1261 séquences de 19 espèces du genre «*Gazella* » :

- 214 séquences d'ADN nucléaire.
- 1026 séquences d'ADN mitochondrial.

II.2.2. Traitement des séquences de la GenBank

Pour réaliser une phylogénie, un bon nombre de programmes informatiques sont disponible, 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.

Le MEGA est conçu pour réaliser de diverses analyses statistiques dans une 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 du genre « *Gazella* » par différentes étapes :

Le programme FaBox a été employé pour convertir les fichiers, de la page principale (Figure 12) l'utilisateur peut choisir une liste de services, chacune 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 de Fasta.

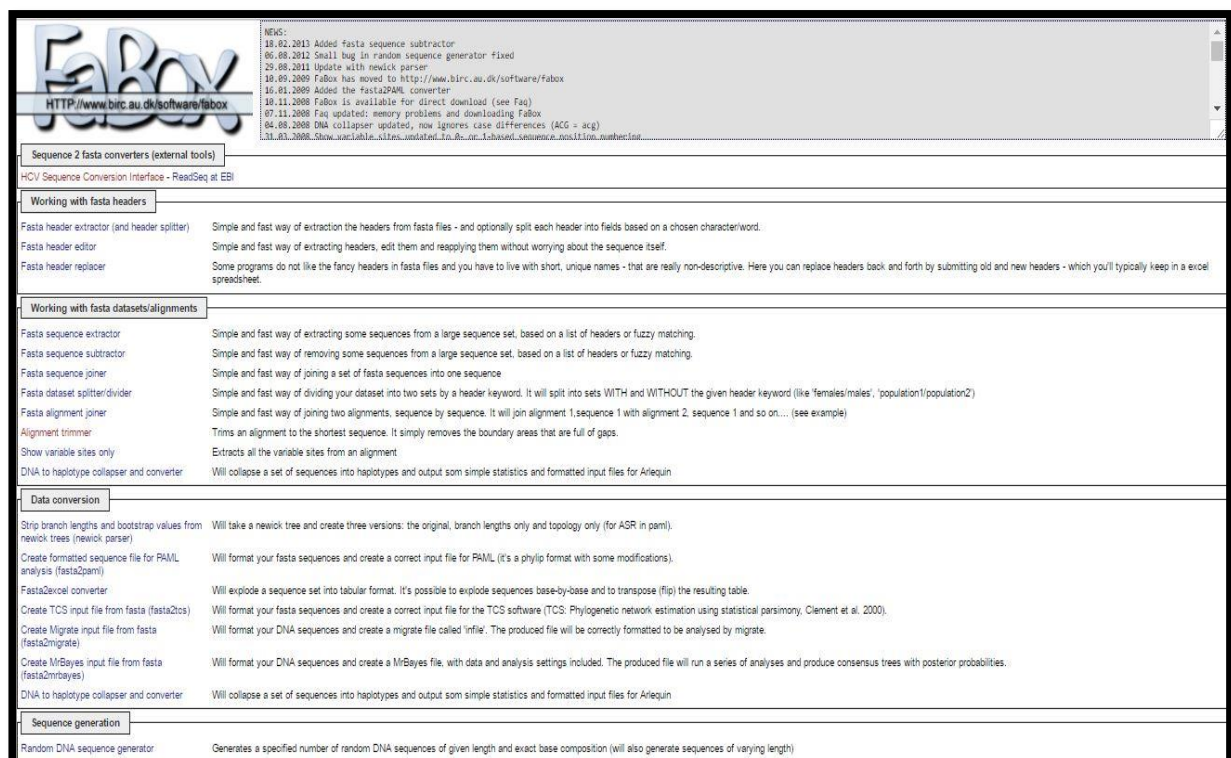


Figure 12 : La page principale du programme FaBox.

L'utilisateur peut 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 (Figure13).

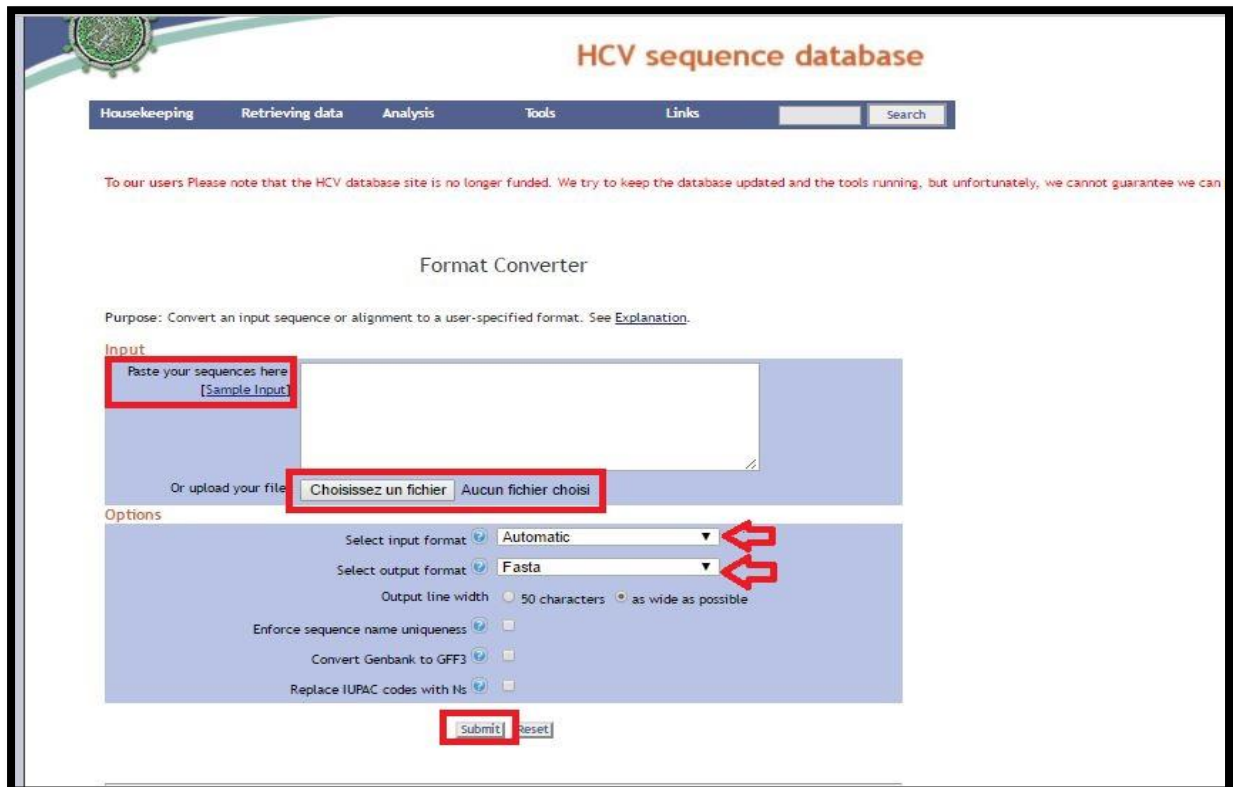


Figure 13 : La conversion des séquences sous forme Fasta par l'utilisation de Fabox.

Alignment cropper va réduire l'alignement (en format Fasta) à la longueur de la séquence la plus courte (suppression les gaps dans le début et la fin de l'alignement). Appuyez sur " show example" pour l'essayer sans soumettre votre propre séquence (Figure14).

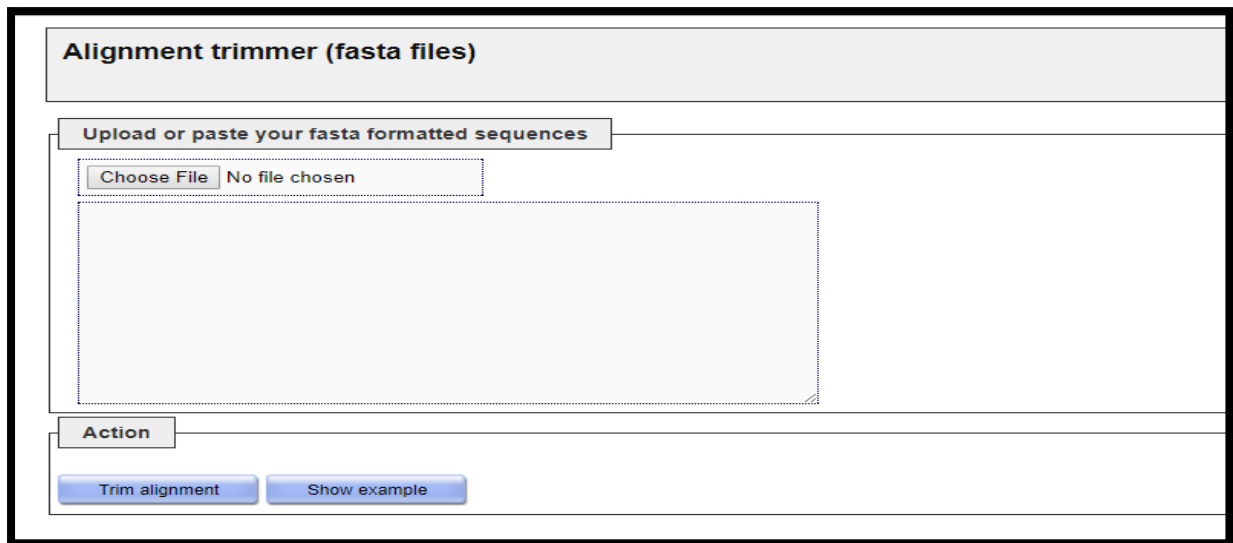


Figure 14 : Alignment trimmer (la forme Fasta).

Les alignements multiples des séquences nucléotidiques et protéiques ont été réalisés au moyen du logiciel MEGA version 5 (Tamura et al., 2011) (Figure 15) par la méthode d'alignement progressif ClustalW (Thompson et al., 1994) (Figure 16).

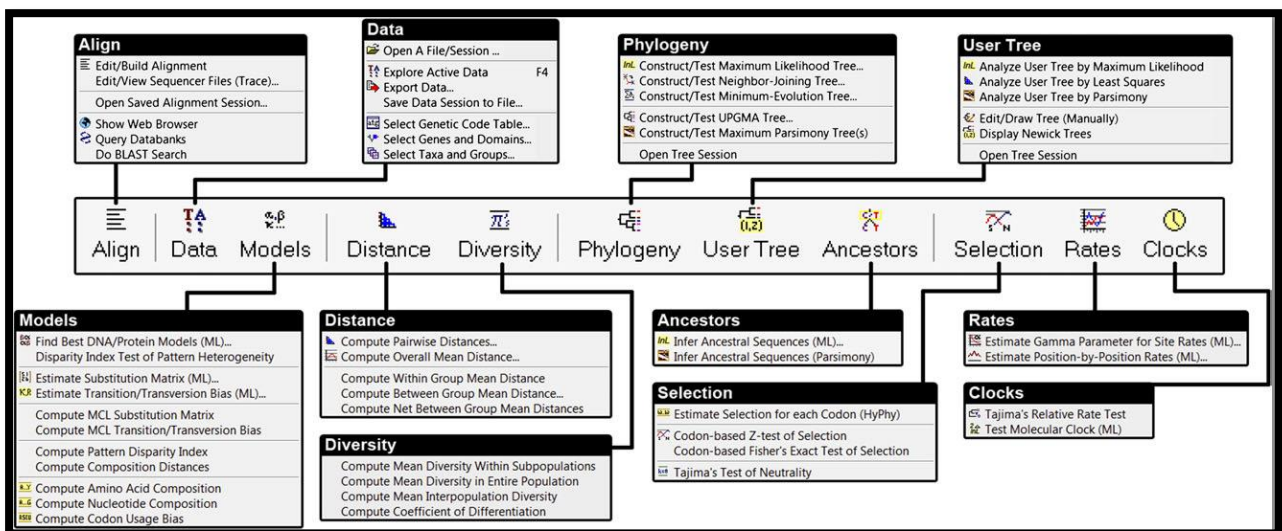


Figure 15 : Barre d'outils de MEGA 5.

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.

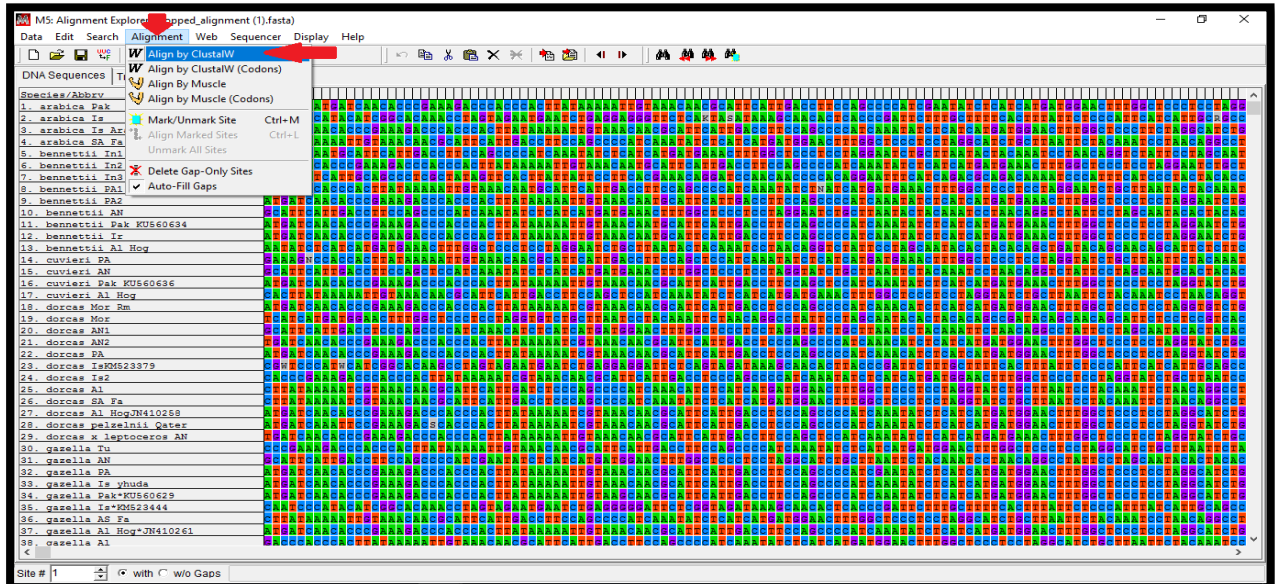


Figure 16 : Le programme ClustalW utilisé pour l'alignement multiple des séquences du marqueur moléculaire choisi.

Une fois les séquences alignées 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 17), 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 ».

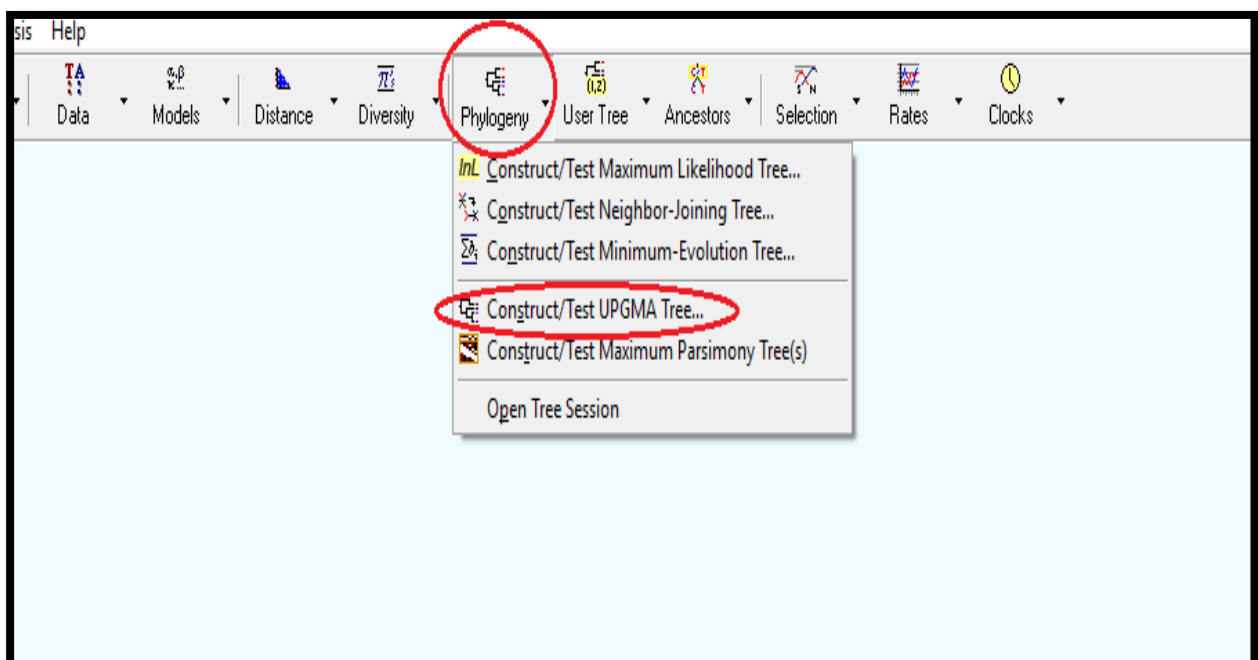


Figure 17 : La méthode UPGMA utilisée pour la construction de l'arbre phylogénétique.

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 18). Les réplifications sont ramenées en pourcentages et ils sont indiqués au niveau des nœuds.

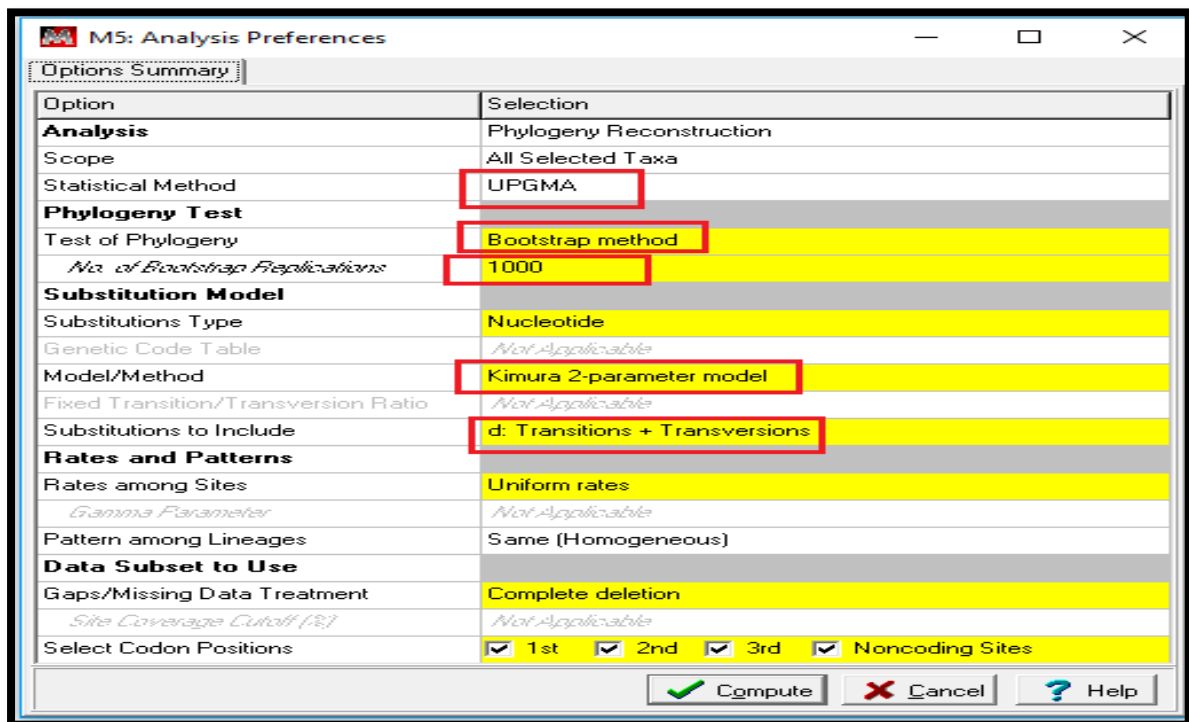


Figure 18 : Les paramètres utilisés pour la reconstruction phylogénétique.

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 mitochondriques pour l'ensemble de totalité des séquences et à chaque population séparément ont été évalués pour plusieurs paramètres comprenant le nombre de haplotypes (h), et les diversités des haplotypes (Hd) 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 19).

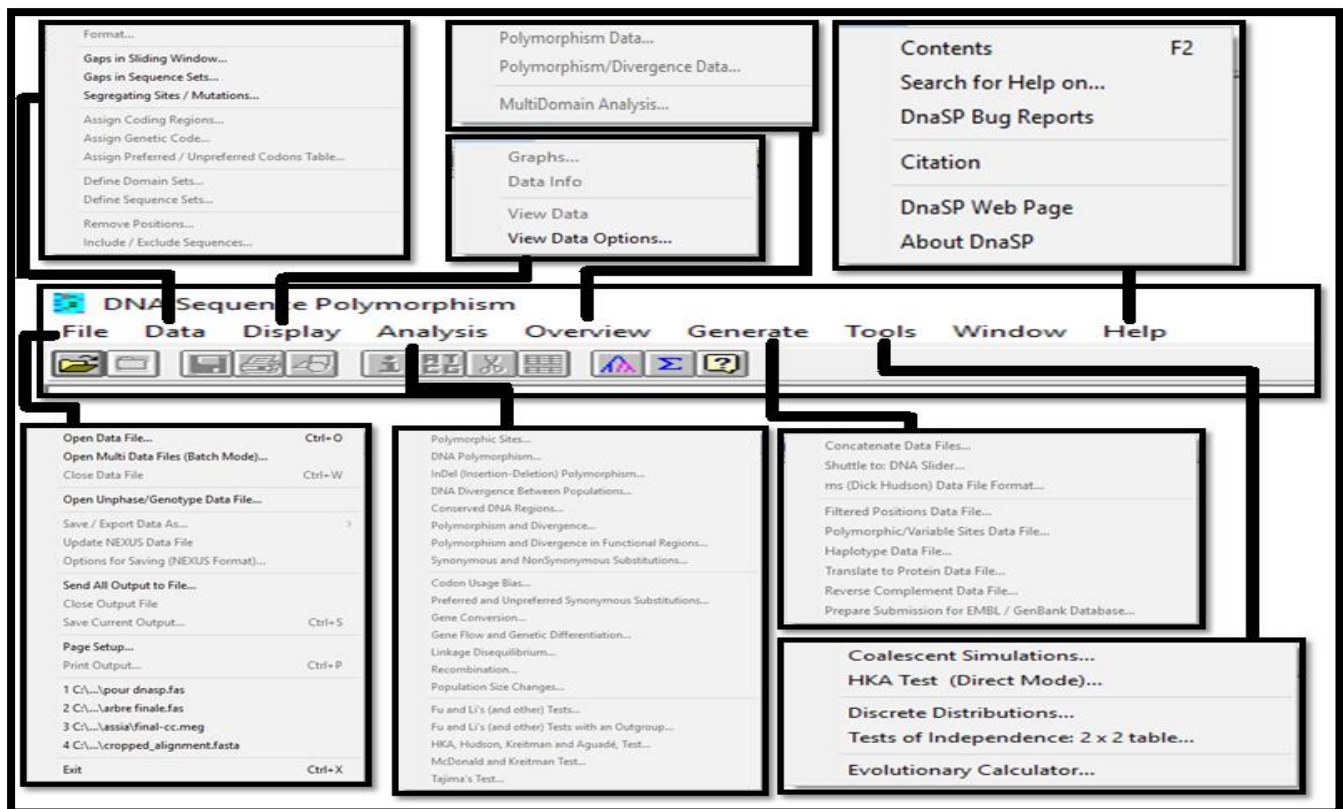


Figure 19 : Barre d'outils de DnaSP.

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 20).

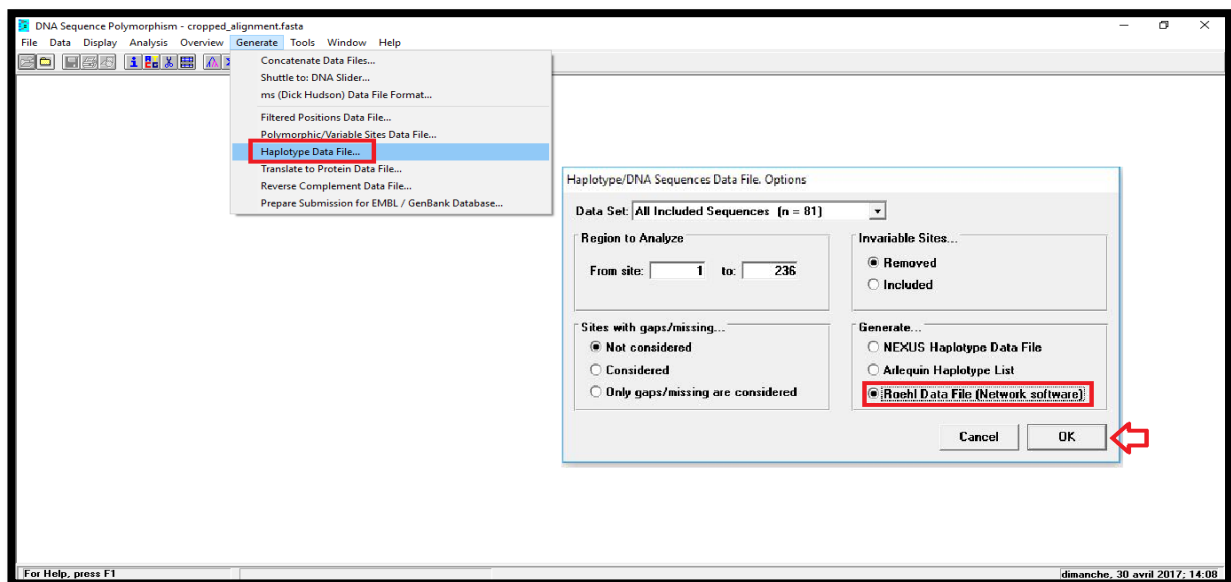


Figure 20 : Interface utilisateur de DnaSP.

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, est un logiciel qui intègre des paquet pour manipulation et analyse des séquences comportant une interface conviviale et un grand choix de fonctions analytiques dedans bio-informatique, phylogenetics, et génomique descriptive et comparative (Salemi et al., 2003; Felsenstein 2004; Lemey et al., 2009). Permet de convertir, manœuvrant, statistiquement et graphiquement décrivant (Figure21).

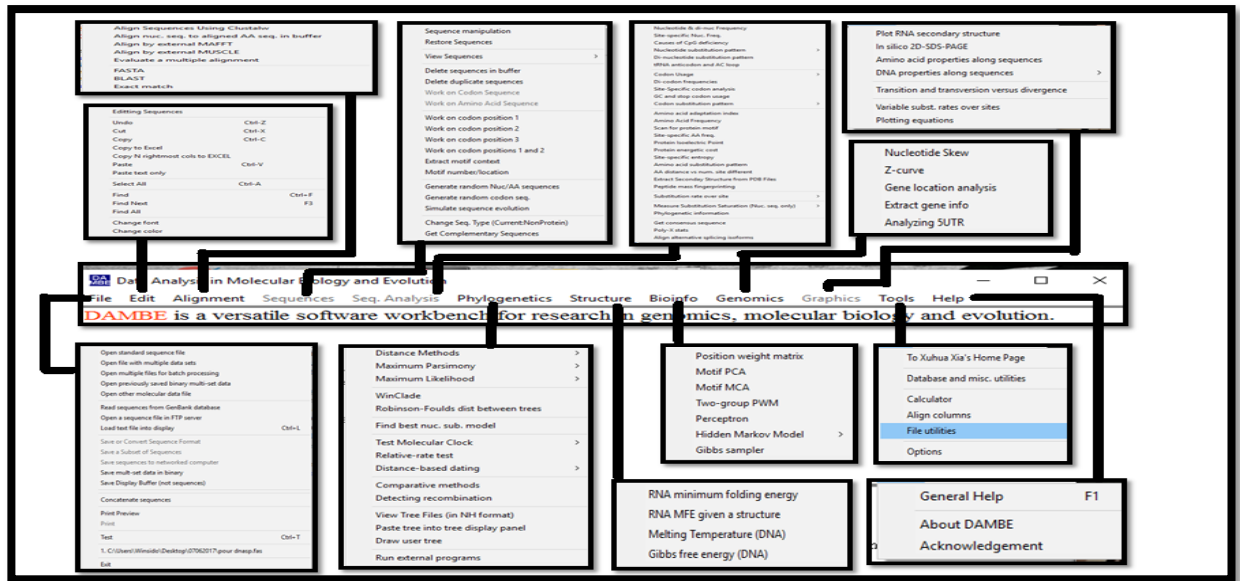


Figure 21 : Barre d'outil du programme DAMBE

DAMBE peut représenter graphiquement le taux de C/G et A/T (Figure 22).

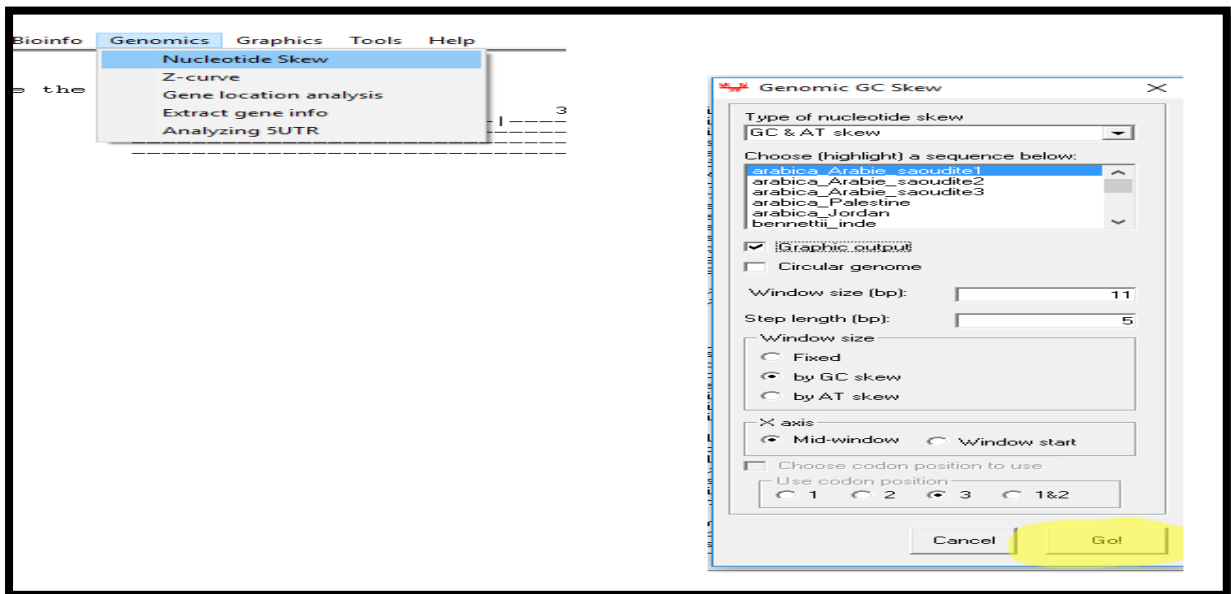


Figure 22 : Barre d'outil qui fait les graphes de C/G et A/T.

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

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

La format des séquences	Lire à partir de	Convertir à
PHYLIP	+	+
PAUP	+	+
MEGA	+	+
CLUSTAL	+	-
FASTA	+	+
GenBank	+	+
GCG	+	+
MSF	+	+
DNA strider	+	+
PAML	+	-
RST, MP _a	-	+
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 reconstruits 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 de routine (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 23). Les méthodes de network sont maintenant disponibles et sont très utilisés par les chercheurs dans des domaines aussi différents que la phylogéographie (**Schaal et al., 2000**).

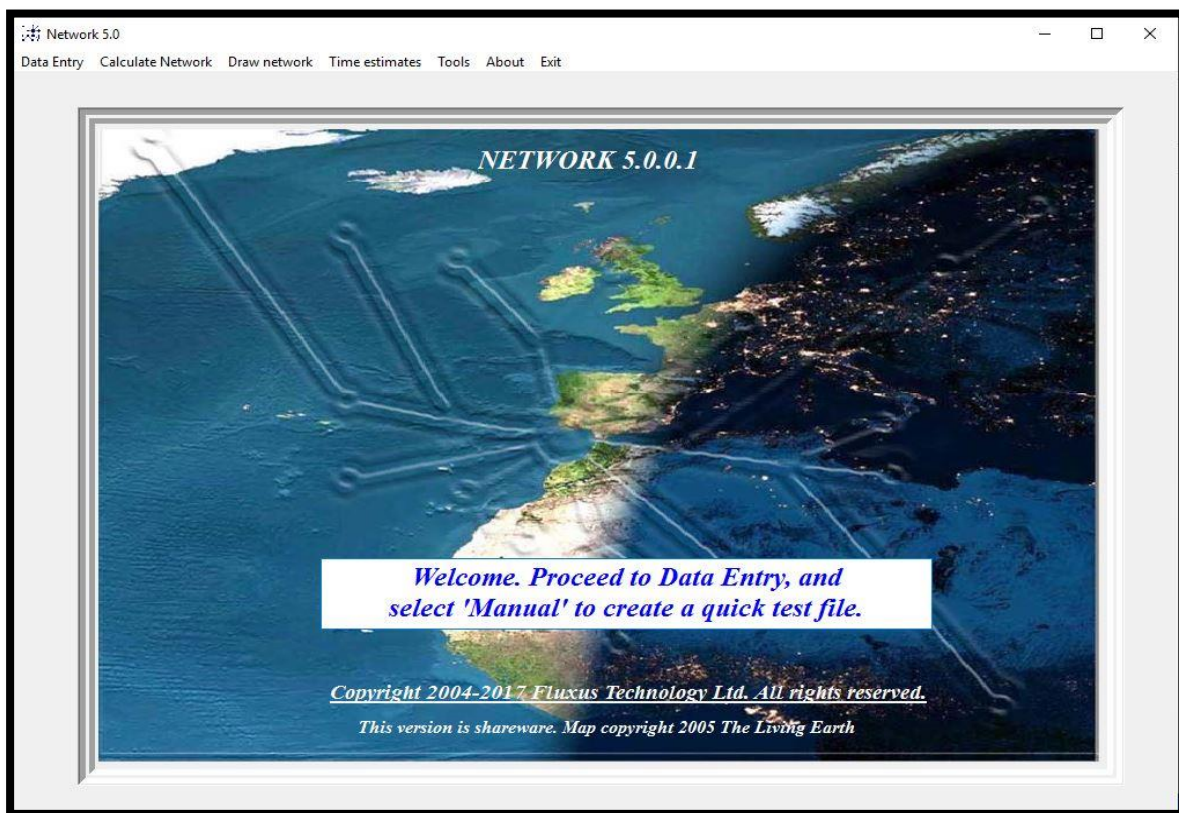


Figure 23 : La page d'accueil de Network 5.0.

Pour le calcul du réseau initial, nous passons par différents étapes.

1. Calculate Network menu.
2. Network calculations.
3. Median Joining.
4. File.
5. Open.
6. DNA data file (rdf) (Figure24).

Sélectionnez le dossier de rdf dont vous avez manuellement créé dans le rédacteur de données de Network ou exporté du programme DnaSP.

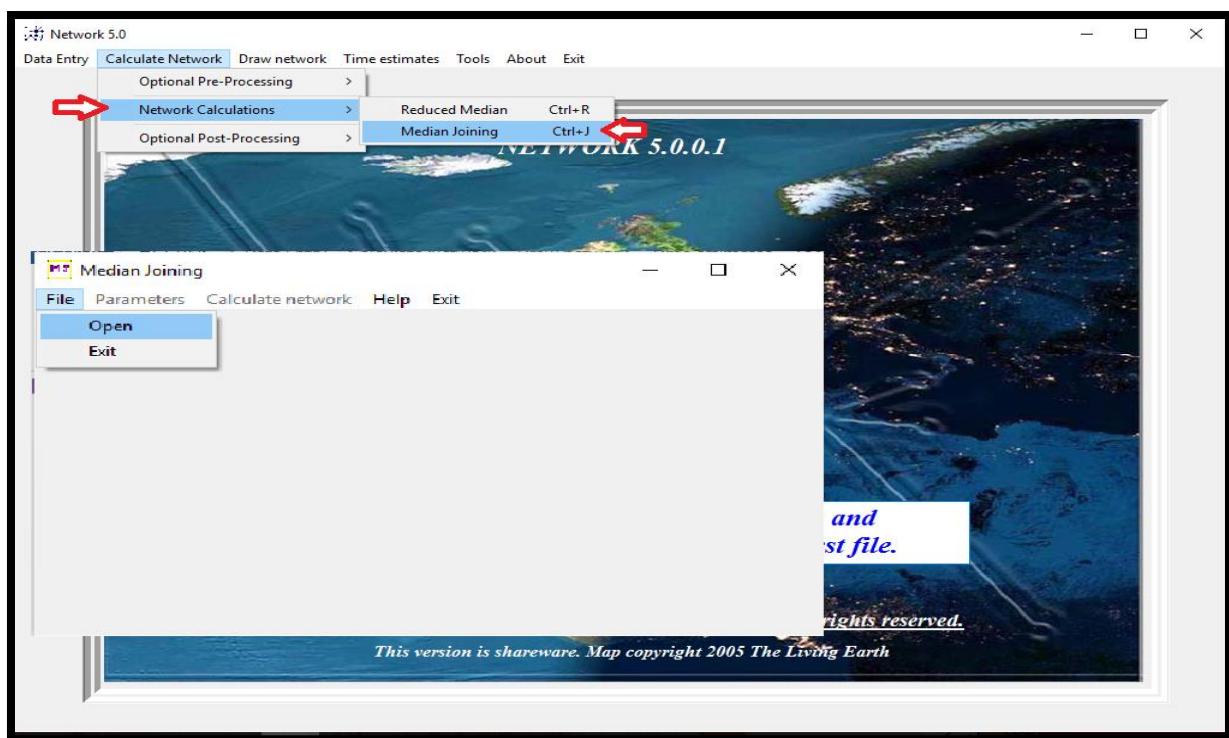


Figure 24 : Ouverture du fichier de données de rdf pour le calcul de MJ Network.

Après l'ouverture du dossier, le Network peut produire des fichiers sous forme *.out afin que le calcul network peut reconstruire le réseau (Figure 25).

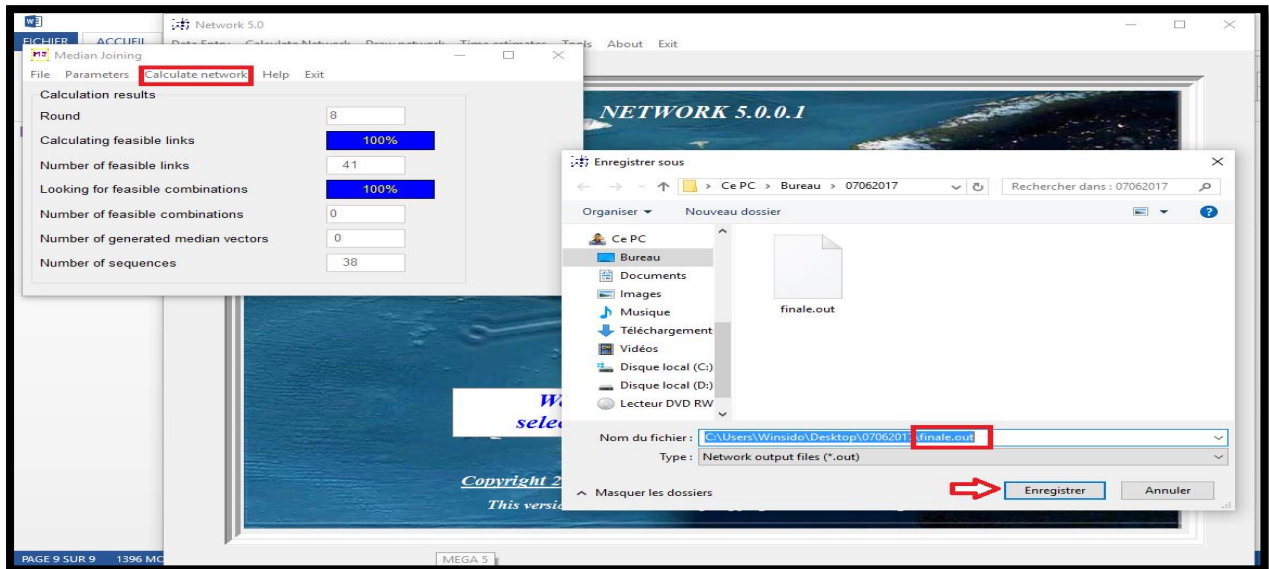


Figure 25 : Un fichier sous forme *.out a été sauvegardé.

Le Network a présenté la phylogénétique des espèces avec les cycles graphiques sous forme d'un réseau par la fonction Draw network (Figure 26).

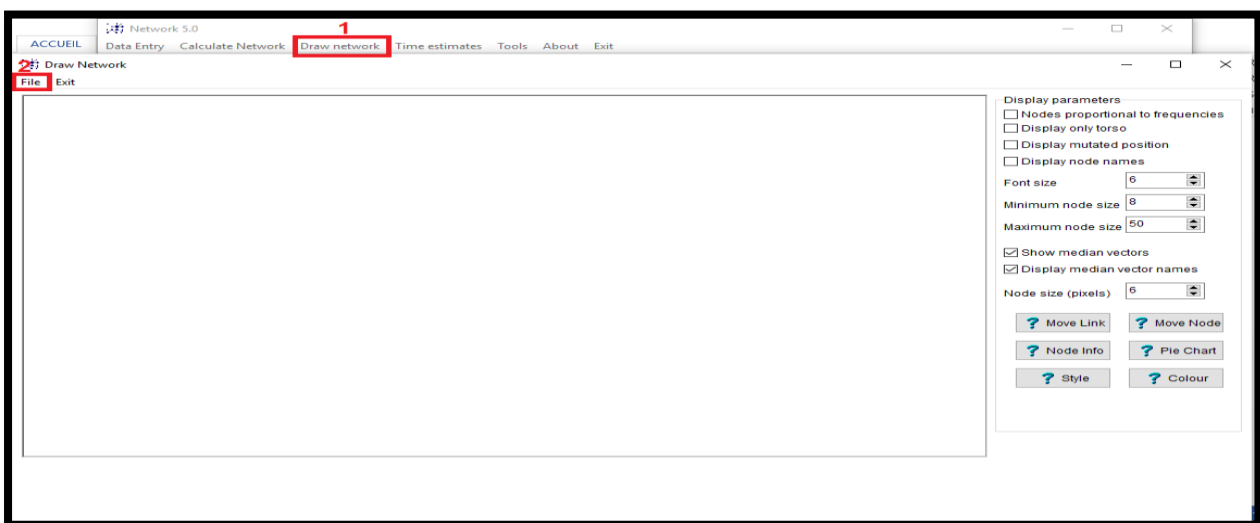


Figure 26 : reconstruction de réseau haplotypique.

Nos résultats consistent sur la comparaison entre les espèces de genre *Gazella* sur le plan moléculaire à partir des données de la GenBank, afin de déterminer la position de ces espèces dans l'arbre phylogénétique, et avoir l'évolution des espèces de genre *Gazella* par des analyses statistiques des gènes étudiés.

III.1. Résultats de la GenBank

Notre étude est basée sur l'utilisation plusieurs logiciels de bioinformatique afin de traiter une base des données de biologie moléculaire pour reconstruire l'arbre phylogénétique et la position de différentes espèces de genre *Gazella* à partir de 82 séquences (gène : *cytb*, D-loop).

19 espèces appartenant au genre *Gazella*, récupérées auprès de la GenBank. Les espèces qui ont été utilisé dans ces analyses phylogénétiques sont : *G. arabica*, *G. bennetti*, *G. cuvieri*, *G. dorcas*, *G. leptoceros*, *G. marica*, *G. saudiya*, *G. spekie*, *G. subguttorsa*, *G. gazella*, *G. dorcas pelzelinii*, *G. dorcas x leptoceros*, *G. erlangeri*, *G. gazella muscatensis*, *G. subgutturosa marica*, *G. subgutturosa sairenses*, *G. subgutturosa subgutturosa*, *G. subgutturosa yarkandensis*, *G. sp.* localisées dans différentes régions géographiques.

Les séquences de chaque espèce extraite à partir de la GenBank, leur type de gène et leur origine géographique sont mentionnées dans le tableau en annexe. (Voir tableau IX).

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 genre *Gazella* extraites de la GenBank.

Espèce	Nombre total des séquences	Séquences d'ADN nucléaire	Séquences d'ADN mitochondrial	Nombre total de cytb	Nombre total de D-loop	Les séquences non identifiées
<i>G. arabica</i>	77	34	43	26	3	/
<i>G. bennetti</i>	73	28	45	34	0	/
<i>G. cuvieri</i>	24	12	12	7	2	/
<i>G. Dorcas</i>	299	49	249	155	3	1
<i>G. dorcas pelzelinii</i>	1	0	1	1	0	/
<i>G. dorcas x leptoceros</i>	21	1	20	20	0	/
<i>G. erlangeri</i>	2	0	2	0	0	/
<i>G. gazelle</i>	336	19	317	162	91	/
<i>G. gazella gazelle</i>	3	0	3	1	1	/
<i>G. gazella muscatensis</i>	1	0	1	1	0	/
<i>G. leptoceros</i>	30	13	13	10	0	4
<i>G. saudiya</i>	10	0	4	9	0	/
<i>G. spekei</i>	29	21	8	5	0	/
<i>G. subgutturosa</i>	143	19	116	69	17	8
<i>G. subgutturosa marica</i>	94	18	76	33	21	/
<i>G. subgutturosa sairenses</i>	4	0	4	0	4	/
<i>G. subgutturosa subgutturosa</i>	76	0	76	0	76	/
<i>G. subgutturosa yarkandensis</i>	20	0	20	0	20	/
<i>G. sp.</i>	18	0	10	10	0	8
Total :	1261	214	1026	543	634	21

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

🚩 214 séquences d'ADN génomique : NLRP2, SMOC1, PANK4...ect

🚩 1026 séquences d'ADN mitochondriale majoritairement cytb, cox iii, D-loop ...ect

Notre étude est basée sur l'ADN mitochondrial (le cytochrome b et les régions de contrôle D-loop) pour établir des liens de parenté entre les espèces du genre *Gazella* (Figure 27). En général les taux de mutation de l'ADN mitochondrial sont plus grands que ceux de l'ADN nucléaire (Ballard et al., 2004).

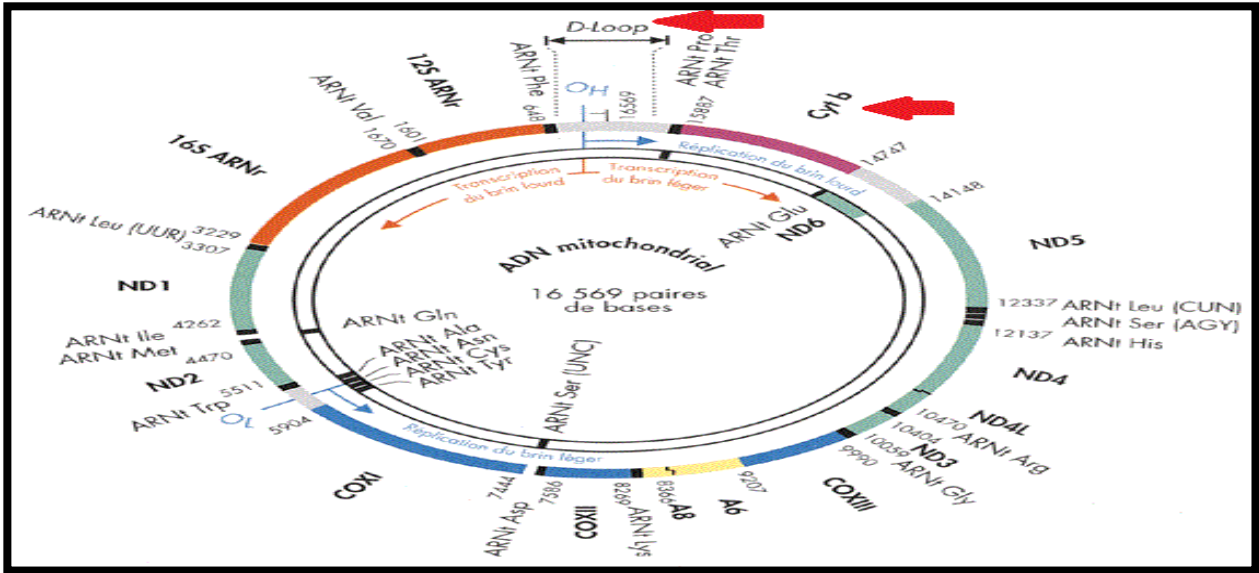


Figure 27: Structure de l'ADN mitochondrial et la position de cytb et D-loop (Rustin et al., 1995).

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 D-loop en se basant sur leurs origines géographiques :

Cyt b:

- ***G. arabica*** : Au total, il y'a 6 séquences, 3 séquences sont de l'Arabie saoudite, 2 séquences de Palestine, et une séquence de Jordanie, nous avons choisi 3 séquences d'Arabie saoudite de différentes tailles (entre 336pb à 1149pb), une séquence de Palestine (1132pb) et une de Jordanie (1195pb).
- ***G. bennettii*** : 12 séquences au total, 5 séquences sont de l'Inde de tailles (204pb à 751), 2 séquences sont de Pakistan (300pb et 1149pb), 2 séquences de la péninsule arabique (333pb et 375pb), 2 séquences de l'Arabie saoudite (1149pb et 1018pb) et une seule séquence d'Iran (410pb). Nous avons choisi une seule séquence d'Inde et une du Pakistan de

taille (415pb) et (1149pb) respectivement, les 2 séquences de péninsule arabique (333pb et 375pb), 2 séquences de l'Arabie saoudite (1149pb et 1018pb) et la séquence d'Iran de (410pb).

- ***G. cuvieri*** : nous prendrons toutes les séquences de ce genre, une seule séquence de la péninsule arabique (333pb) et une de l'Afrique du nord (300pb) et 2 séquences sont de l'Espagne (1095pb et 1149pb).

- ***G. dorcas*** : 17 séquences ont été trouvées dont 2 séquences du Maroc (375pb et 716pb), 2 séquences d'Afrique du nord (300pb et 450pb), 2 séquences d'Algérie (1152pb) une seule séquence pour : Mali (1152pb), Soudan (1152pb), Qatar (1152pb), Arabie saoudite (336pb), péninsule arabique (336pb), Allemagne (336pb), Tchad (1152pb), Tunisie (1152pb) et 3 séquences de Palestine (790pb à 1152pb), on a éliminé seulement 2 séquences, un du Maroc de 375pb et l'autre de Palestine de 790pb.

- ***G. dorcas pelzelni*** : une seule séquence de Qatar de taille de 1140pb.

- ***G. dorcas x leptoceros*** : une seule séquence d'Afrique du nord de taille de 450pb.

- ***G. gazella*** : Nous avons trouvé 12 séquences dont 3 séquences de Palestine de différents taille (790pb à 1149pb) on a choisi une seule séquence de 1149pb, 2 séquences d'Arabie saoudite (336pb et 1152pb) on a choisi la séquence de 1152pb, une seule séquence pour les pays suivants : Turquie de 400pb, l'Afrique du nord (300pb), Péninsule Arabique (375pb), Dubaï (1011pb), Allemagne (360pb), Qatar (981pb) et l'Eurasie (1134pb).

- ***G. gazella gazella*** : une seule séquence d'Allemand de taille 336pb.

- ***G. gazella muscatensis*** : une seule séquence de la péninsule arabique de la taille 375pb.

- ***G. leptoceros*** : nous prendrons toutes les séquences de cette espèce. Au total nous avons 7 séquences dont 2 séquences de la péninsule arabique (333pb et 375pb), une seule séquence de chacun des pays suivants : l'Afrique du nord (300pb), Tunisie (1025pb), Egypte (1104pb), l'Algérie (1152pb), et République de chèque (1140pb).

- ***G. saudiya*** : nous avons au total 3 séquences dont une séquence d'Afrique du nord de 300pb, et 2 séquences d'Arabie saoudite de 375pb et 404pb. On prend la séquence d'Afrique du nord et une seule séquence d'Arabie saoudite de 404pb.

- ***G. spekei*** : Nous avons 2 séquences l'un d'Afrique du nord (300pb) et l'autre de Dubaï (1149pb).

- ***G. subgutturosa*** : Au total nous avons 14 séquences dont ; 2 séquences de la Mongolie (375pb et 1083pb), une seul séquence pour les pays suivants : l’Eurasie (1143pb), Azerbaïdjan (333pb), Turkestan (332pb), péninsule arabique (375pb), Afrique du nord (300pb), Qatar (1128pb), la Chine (1143pb). 4 séquences d’Iran de différentes tailles (258pb à 470pb). On a éliminé la Mongolie (375pb) et Iran (333pb).

- ***G. subgutturosa marica*** : nous avons au total 6 séquences, on prend une séquence d’Arabie saoudite (1116pb), une séquence d’Oman (333pb), une séquence d’UAE (333pb) et une de la péninsule arabique (375pb).

- ***G. sp.*** : On a 2 séquences, l’un de la Péninsule Arabique (375pb) et l’autre d’Iran (383pb).

Nous allons étudier ici 70 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 azotique qui forme une séquence d’une seule ligne (Figure 28).

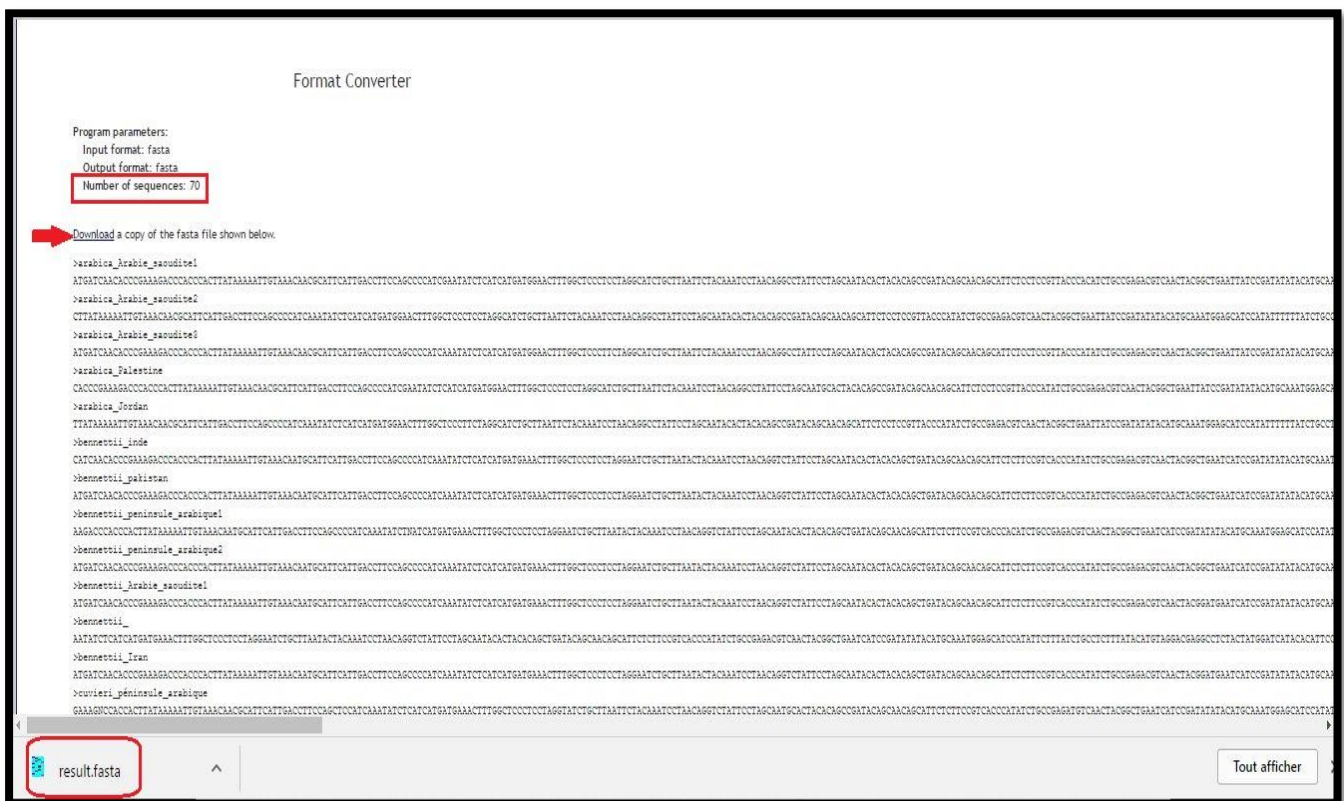


Figure 28 : Les séquences de cytb ont été convertir sous forme Fasta à partir de bloc note grâce à Fabox.

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 genre *Gazella* ont été représentées à l'aide d'un arbre réalisé grâce au programme MEGA 5 (Figure 29).

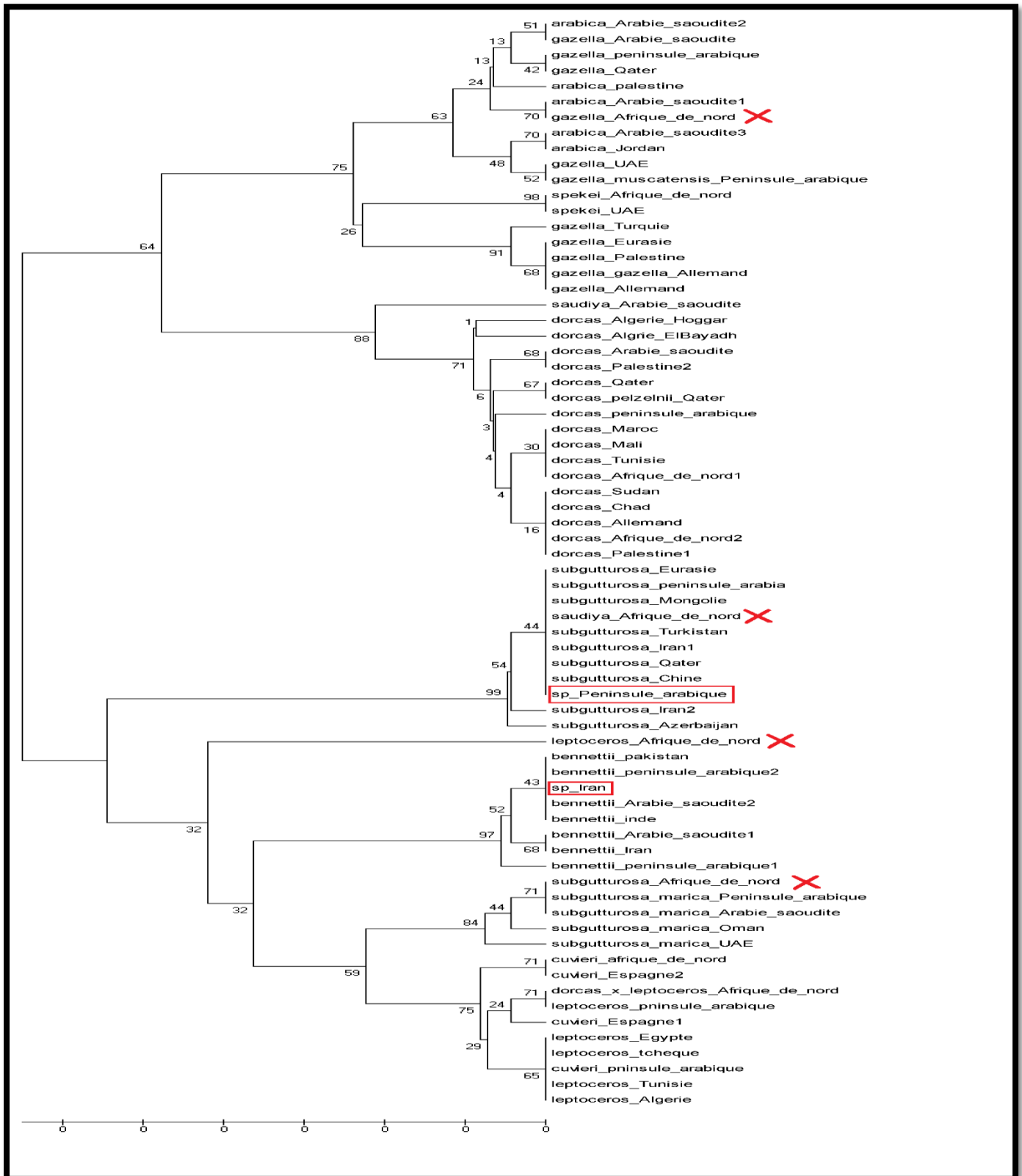


Figure 29 : Arbre phylogénétique reconstruit par Méthodes UPGMA basées sur 70 séquences de nucléotides de cytochrome-b de genre *Gazella* en utilisant MEGA5.

Le but de visualiser graphiquement est de voir si une proximité génétique existait entre les individus au sein d'un même groupe. L'ADN mitochondrial a été largement utilisé comme un outil pour déchiffrer l'histoire évolutive et démographique des populations et des espèces (Ballard et al., 2004).

Correction de l'arbre brut obtenu

Nous avons remarqué dans l'arbre obtenu qu'ils y'a certaines espèces qui ne sont pas dans ces réel clade : *G. gazella*, *G. saudiya*, *G. leptoceros*, *G. subguttorosa*. Et nous avons remarqué aussi que ces quatre séquences de l'article (« Phylogenetic Relationships in the Bovid Subfamily Antilopinae Based on Mitochondrial DNA Sequences, 1998 » des auteurs Wilhelmus Rebholz et Eric Harley), repose essentiellement sur la morphologie des espèces, donc on a les éliminé complètement.

Nous avons identifié les deux séquences de *Gazella sp.* Selon l'arbre précédent (Figure 29) *G. sp.* de la péninsule arabique AF187721 fait partie de l'espèce de *G. subguttorosa* et *G. sp.* d'Iran KR873132 fait partie de l'espèce *G. bennettii*.

Après l'élimination et l'identification des sp, nous avons obtenu au total 66 séquences de cytb afin de reconstruire l'arbre final.

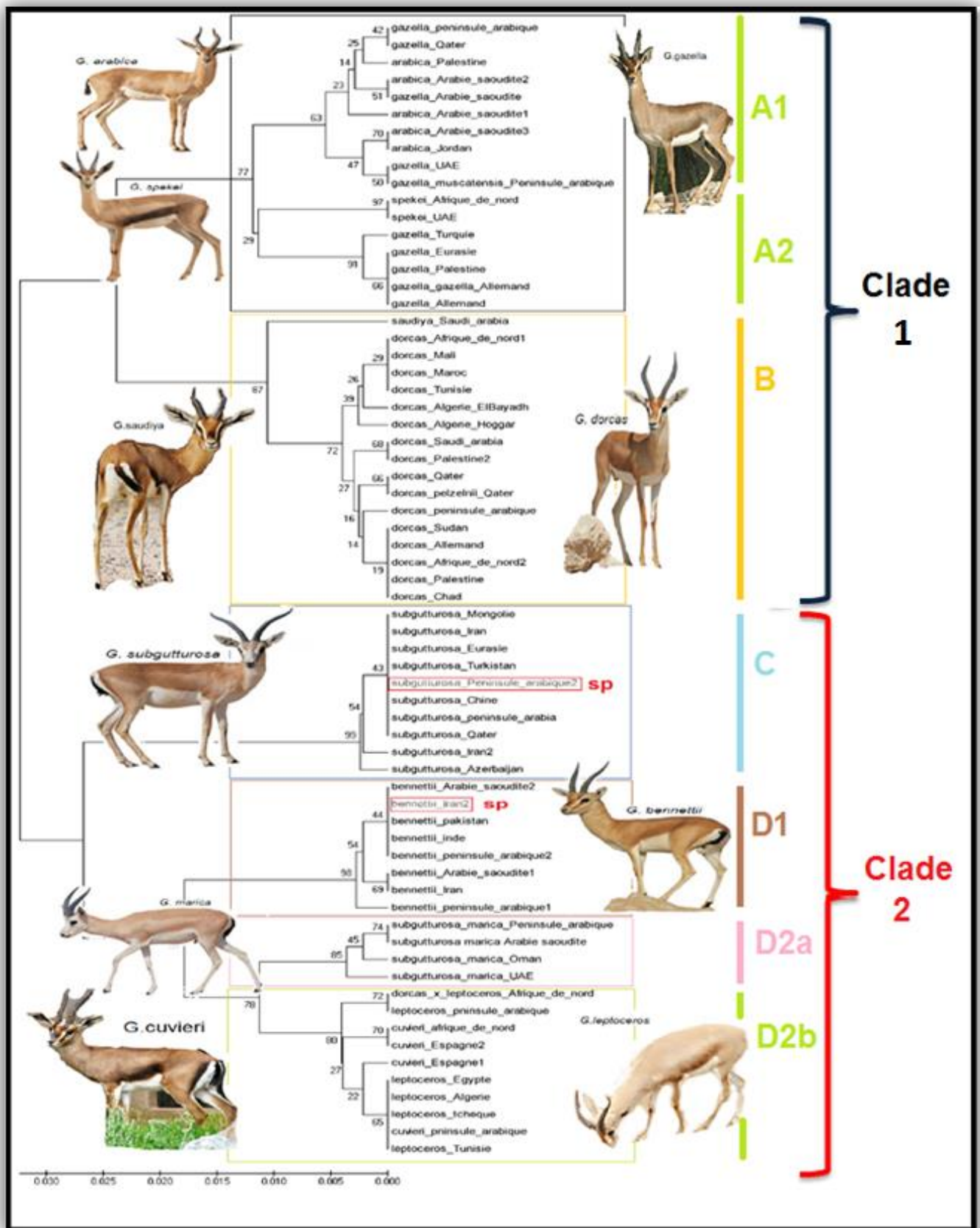


Figure 30 : L'arbre final de genre *Gazella* basé sur 66 séquences de cytb en utilisant la méthode UPGMA « Kimura 2-paramater » par la suite informatique MEGA5.

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

Cytb : D'après l'arbre, les séquences de cytb se répartissent en 2 clades principaux :

Clade 1: Il comprend *G. arabica*, *G. spekie*, *G. gazella*, *G. dorcas* et *G. saudiya*

Ce clade est subdivisé en deux sous-clades :

- ✓ Sous Clade A (A1 et A2) : Il renferme les espèces *G. arabica*, *G. spekie*, *G. saudiya* et les sous espèces *G. gazella gazella* et *G. gazella muscatensis*.
- ✓ Sous Clade B : Il correspond à *G. dorcas*, *G. saudiya* et le sous espèce *G. dorcas pelzelinii*.

Clade 2 : Il comprend *G. subguttorosa*, *G. bennettii*, *G. leptoceros*, *G. cuvieri* et *G. subguttorosa marica*.

Ce clade est subdivisé en 2 sous-clades :

- ✓ Sous Clade C : Il n'est représenté que par la seule espèce correspond à *G. subguttorosa*.
- ✓ Sous Clade D : Est subdivisé aussi en :
 - D1 : Il n'est représenté que par la seule espèce correspond à *G. bennettii*.
 - D2a : Il correspond à *G. subguttorosa marica*.
 - D2b : Il renferme les espèces *G. leptoceros*, *G. cuvieri* et *G. dorcas x leptoceros*.

De manière générale, l'arbre obtenu dans cette partie montre que *Gazella* est monophylétique (Wronski et al., 2010), absence des nœuds signifie qu'il n'y a pas une séparation nette des espèces, donc elles sont des espèces très proches, à partir d'une matrice de séquences traitées (cytb) avec la méthode UPGMA.

Les résultats trouvés par Blast sont les mêmes, ce logiciel identifie des régions de similarité locale entre séquences. Le programme compare des séquences nucléotidiques ou protéiques et calcule la significativité des résultats.

Les séquences sont classées en fonction d'un "score" qui dépend de l'homologie avec la séquence requête, de la taille de la banque et de la valeur du "E-Value". Plus celui-ci est petit, plus l'homologie entre la séquence requête et celle de la banque est grande.

Nous avons également vérifié si ces séquences sont proches. Nous avons comparé les séquences de *G. arabica* et *G. gazella* de A1 de clade 1. L'homologie entre ces séquences est de 99%.

	Description	Max score	Total score	Query cover	E value	Ident	Accession
<input type="checkbox"/>	Gazella arabica isolate Wadi Tari 3 cytochrome b (cytb) gene, partial cds, mitochondrial	693	693	100%	0.0	100%	KC188762.1
<input type="checkbox"/>	Gazella gazella isolate AWWP mitochondrion, complete genome	693	693	100%	0.0	100%	JN632640.1
<input type="checkbox"/>	Gazella dorcas isolate North-east 9 cytochrome b (cytb) gene, complete cds, mitochondrial	693	693	100%	0.0	100%	JN410224.1
<input checked="" type="checkbox"/>	Gazella gazella NHM 40.316 cytochrome b (cytb) gene, partial cds, mitochondrial gene for mitochondrial product	693	693	100%	0.0	100%	AF187713.1
<input checked="" type="checkbox"/>	Gazella arabica isolate GGF41 cytochrome b (cytb) gene, complete cds, mitochondrial	688	688	100%	0.0	99%	KU560630.1
<input type="checkbox"/>	Gazella gazella isolate South 10 cytochrome b (cytb) gene, complete cds, mitochondrial	688	688	100%	0.0	99%	JN410260.1
<input type="checkbox"/>	Gazella gazella cytochrome b (cytb) gene, partial cds, mitochondrial gene for mitochondrial product	688	688	100%	0.0	99%	AF187701.1
<input type="checkbox"/>	Gazella erlangeri isolate AWWP mitochondrion, complete genome	676	676	100%	0.0	99%	JN632639.1
<input checked="" type="checkbox"/>	Gazella gazella isolate South 11 cytochrome b (cytb) gene, complete cds, mitochondrial	676	676	100%	0.0	99%	JN410261.1
<input checked="" type="checkbox"/>	Gazella arabica voucher TAUM11048 cytochrome b (cytb) gene, partial cds, mitochondrial	673	673	97%	0.0	99%	KC188769.1
<input type="checkbox"/>	Gazella arabica isolate Omani cytochrome b (cytb) gene, partial cds, mitochondrial	671	671	100%	0.0	99%	KU560648.1
<input type="checkbox"/>	Gazella arabica isolate Wadi Tari 1 cytochrome b (cytb) gene, complete cds, mitochondrial	665	665	100%	0.0	99%	KC188765.1
<input checked="" type="checkbox"/>	Gazella gazella muscatensis HZM 26.4534 cytochrome b (cytb) gene, partial cds, mitochondrial gene for mitochondrial product	665	665	100%	0.0	99%	AF187897.1

Figure 31 : Comparaison entre les séquences de *G. arabica* et *G. gazella* par le programme BLAST.

Nous avons comparé aussi les deux séquences de *G. sepkie* sous les numéros d'accèsion (AF030608, KU560632) avec les séquences de *G. gazella* et de A2 de clade1. Ces dernières ont une grande similarité de 97%. Ce résultat montre que ces séquences sont très homologues.

<input checked="" type="checkbox"/>	Gazella speikai cytochrome b (cyt b) gene, mitochondrial gene encoding mitochondrial protein, partial cds	555	555	100%	3e-154	100%	AF030608.1
<input checked="" type="checkbox"/>	Gazella speikai isolate 3467 cytochrome b (cytb) gene, complete cds, mitochondrial	549	549	100%	1e-152	99%	KU560632.1
<input type="checkbox"/>	Gazella speikai isolate 3463 cytochrome b (cytb) gene, complete cds, mitochondrial	549	549	100%	1e-152	99%	KU560631.1
<input type="checkbox"/>	Gazella speikai isolate AWWP mitochondrion, complete genome	549	549	100%	1e-152	99%	JN632642.1
<input type="checkbox"/>	Gazella arabica isolate Wadi Tari 3 cytochrome b (cytb) gene, partial cds, mitochondrial	521	521	100%	3e-144	98%	KC188762.1
<input type="checkbox"/>	Gazella gazella voucher AWWP 9923 cytochrome b (cytb) gene, partial cds, mitochondrial	521	521	100%	3e-144	98%	JX274674.1
<input type="checkbox"/>	Gazella gazella isolate AWWP mitochondrion, complete genome	521	521	100%	3e-144	98%	JN632640.1
<input type="checkbox"/>	Gazella dorcas isolate North-east 9 cytochrome b (cytb) gene, complete cds, mitochondrial	521	521	100%	3e-144	98%	JN410224.1
<input type="checkbox"/>	Gazella gazella isolate KKWRC 255 cytochrome b (cytb) gene, partial cds, mitochondrial	521	521	100%	3e-144	98%	GU384841.1
<input type="checkbox"/>	Gazella gazella isolate KKWRC 1343 cytochrome b (cytb) gene, partial cds, mitochondrial	521	521	100%	3e-144	98%	GU384827.1
<input type="checkbox"/>	Gazella gazella NHM 40.316 cytochrome b (cytb) gene, partial cds, mitochondrial gene for mitochondrial product	521	521	100%	3e-144	98%	AF187713.1
<input type="checkbox"/>	Gazella arabica isolate GGF41 cytochrome b (cytb) gene, complete cds, mitochondrial	516	516	100%	1e-142	98%	KU560630.1
<input checked="" type="checkbox"/>	Gazella gazella isolate GH1 cytochrome b (cytb) gene, complete cds, mitochondrial	516	516	100%	1e-142	98%	KU560629.1

Figure 32 : Comparaison entre les séquences de *G. sepkie*, *G. arabica* et *G. gazella* à l'aide du programme BLAST.

Nous avons comparé aussi la séquence d'Arabie saoudite sous le numéro JN410323 de *G. saudiya* par rapport aux autres séquences de *G. dorcas*. Ces dernières ont une grande similarité de 98% avec la séquence de *G. dorcas*. Ce résultat montre que ces séquences sont très homologues aussi.

Description	Max score	Total score	Query cover	E value	Ident	Accession
<input checked="" type="checkbox"/> Gazella saudiya voucher NHM 40.310 cytochrome b (cytb) gene, partial cds, mitochondrial	747	747	100%	0.0	100%	JN410323.1
<input type="checkbox"/> Gazella dorcas isolate Unknown 3 cytochrome b (cytb) gene, complete cds, mitochondrial	725	725	100%	0.0	99%	JN410258.1
<input type="checkbox"/> Gazella dorcas haplotype h55 cytochrome b (cytb) gene, partial cds, mitochondrial	719	719	100%	0.0	99%	KM582071.1
<input checked="" type="checkbox"/> Gazella dorcas voucher TAUM11238 cytochrome b (cytb) gene, partial cds, mitochondrial	719	719	100%	0.0	99%	KC188752.1
<input type="checkbox"/> Gazella dorcas isolate AWWP mitochondrion, complete genome	719	719	100%	0.0	99%	JN632638.1
<input type="checkbox"/> Gazella dorcas isolate South-east 11 voucher Powell Cotton Museum PCGDL50 cytochrome b (cytb) gene, partial cds, mitochondrial	719	719	100%	0.0	99%	JN410335.1
<input type="checkbox"/> Gazella dorcas isolate South-east 10 voucher Powell Cotton Museum PCGDL49 cytochrome b (cytb) gene, partial cds, mitochondrial	719	719	100%	0.0	99%	JN410334.1
<input type="checkbox"/> Gazella dorcas isolate North-east 6 cytochrome b (cytb) gene, partial cds, mitochondrial	719	719	100%	0.0	99%	JN410316.1
<input type="checkbox"/> Gazella dorcas isolate North-east 12 cytochrome b (cytb) gene, complete cds, mitochondrial	719	719	100%	0.0	99%	JN410227.1
<input type="checkbox"/> Gazella dorcas isolate North-east 7 cytochrome b (cytb) gene, complete cds, mitochondrial	719	719	100%	0.0	99%	JN410222.1
<input type="checkbox"/> Gazella dorcas isolate North-east 3 cytochrome b (cytb) gene, complete cds, mitochondrial	719	719	100%	0.0	99%	JN410219.1
<input type="checkbox"/> Gazella dorcas voucher NHM 36.5.20.19 cytochrome b (cytb) gene, partial cds, mitochondrial	717	717	99%	0.0	99%	JN410320.1
<input type="checkbox"/> Gazella dorcas isolate South-east 9 voucher Powell Cotton Museum PCGDL34 cytochrome b (cytb) gene, partial cds, mitochondrial	715	715	100%	0.0	99%	JN410333.1
<input type="checkbox"/> Gazella dorcas haplotype h77 cytochrome b (cytb) gene, partial cds, mitochondrial	713	713	100%	0.0	99%	KM582093.1
<input type="checkbox"/> Gazella dorcas haplotype h54 cytochrome b (cytb) gene, partial cds, mitochondrial	713	713	100%	0.0	99%	KM582080.1
<input type="checkbox"/> Gazella dorcas haplotype h54 cytochrome b (cytb) gene, partial cds, mitochondrial	713	713	100%	0.0	99%	KM582070.1
<input type="checkbox"/> Gazella dorcas isolate South-central 15 cytochrome b (cytb) gene, partial cds, mitochondrial	713	713	100%	0.0	99%	JN410330.1
<input type="checkbox"/> Gazella dorcas isolate South-central 13 cytochrome b (cytb) gene, partial cds, mitochondrial	713	713	100%	0.0	99%	JN410328.1
<input type="checkbox"/> Gazella dorcas isolate South-central 8 cytochrome b (cytb) gene, partial cds, mitochondrial	713	713	100%	0.0	99%	JN410326.1
<input type="checkbox"/> Gazella dorcas isolate South-central 9 cytochrome b (cytb) gene, complete cds, mitochondrial	713	713	100%	0.0	99%	JN410240.1
<input type="checkbox"/> Gazella dorcas isolate South-central 7 cytochrome b (cytb) gene, complete cds, mitochondrial	713	713	100%	0.0	99%	JN410239.1
<input type="checkbox"/> Gazella dorcas isolate South-central 6 cytochrome b (cytb) gene, complete cds, mitochondrial	713	713	100%	0.0	99%	JN410238.1
<input type="checkbox"/> Gazella dorcas pelzelinii isolate TS089 cytochrome b (cytb) gene, complete cds, mitochondrial	712	712	99%	0.0	99%	JF728788.1
<input checked="" type="checkbox"/> Gazella dorcas isolate 4 cytochrome b (cytb) gene, complete cds, mitochondrial	710	710	99%	0.0	99%	JN410233.1
<input checked="" type="checkbox"/> Gazella dorcas haplotype h78 cytochrome b (cytb) gene, partial cds, mitochondrial	708	708	100%	0.0	98%	KM582084.1
<input type="checkbox"/> Gazella dorcas haplotype h78 cytochrome b (cytb) gene, partial cds, mitochondrial	708	708	100%	0.0	98%	KM582092.1

Figure 33 : Comparaison entre la séquence de *G. saudiya* et les séquences de *G. dorcas* par logiciel BLAST.

D-loop :

- G. arabica* : Nous avons trouvé 2 séquences, une séquence d'Arabie saoudite de taille de 337pb, et une séquence de l'Allemagne de taille 307pb.
- G. cuvieri* : Une seule séquence d'Espagne de taille de 784pb.
- G. dorcas* : On a 2 séquences, une séquence d'Arabie saoudite (250pb) et l'autre de l'Allemagne (268pb).
- G. gazella* : Nous avons trouvé 3 séquences, 2 séquences de l'Allemagne (234pb et 211pb), une seule séquence d'Arabie saoudite (366pb).
- G. gazella gazella* : Une seule séquence de l'Allemagne de taille 362pb.
- G. subgutturosa* : Une seule séquence de la Chine de taille 984pb.
- G. subgutturosa marica* : On a 2 séquences, une séquence d'Arabie saoudite de taille 358pb et l'autre d'Arabie Saoudite de taille 1106pb.
- G. subgutturosa sairenses* : Une seule séquence de la Chine de taille de 785pb.

- i) *G. subgutturosa subgutturosa* : Nous avons trouvé 2 séquences, une séquence d'Azerbaïdjan de taille 980pb et l'autre d'Uzbekistan de taille 973pb.
- j) *G. subgutturosa yarkandensis* : Une seule séquence de la chine de taille 785pb.

L'arbre phylogénétique de genre *Gazella* a été réalisé à partir des séquences de la région de contrôle (D-loop) de l'ADN mitochondrial :

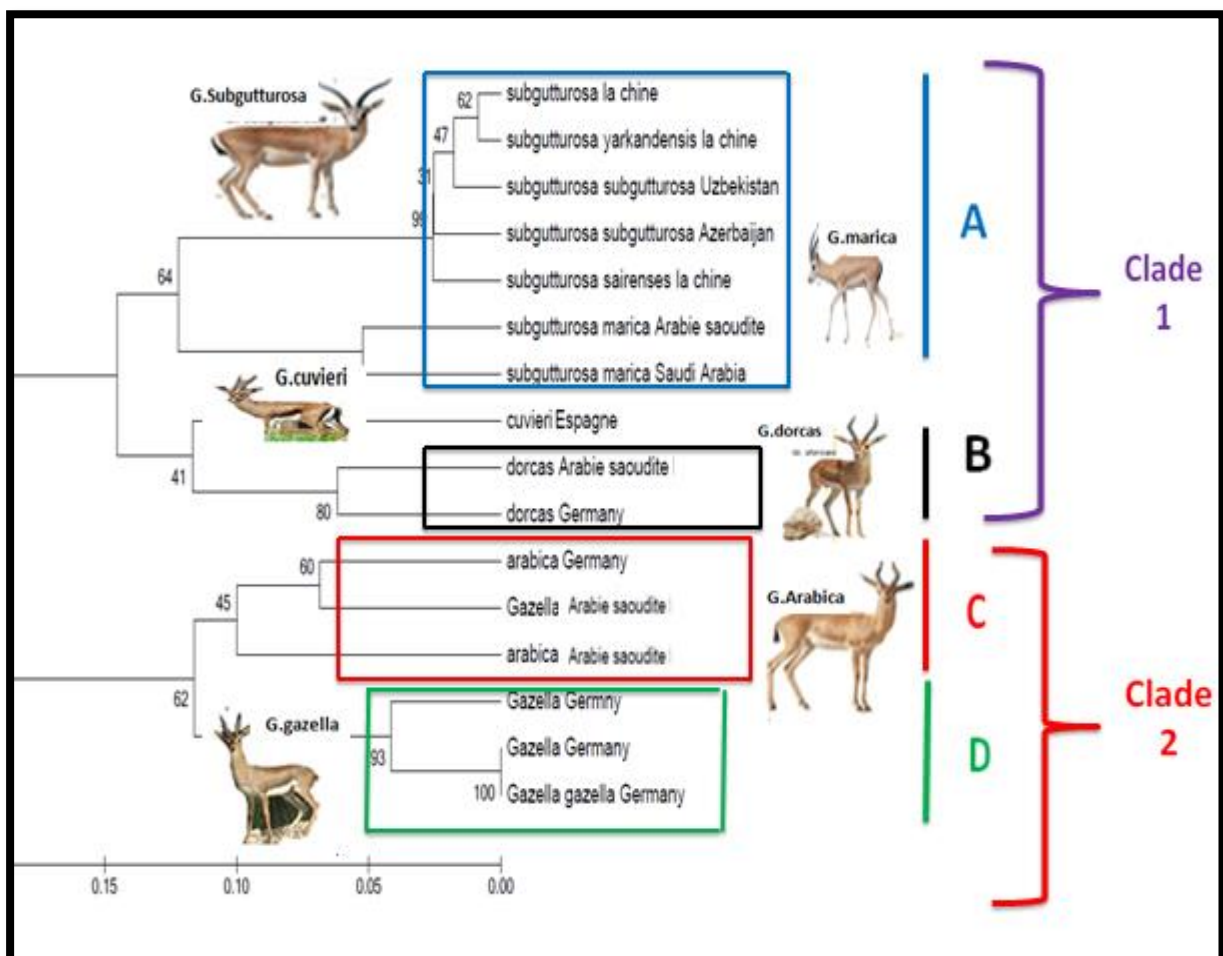


Figure 34 : Arbre phylogénétique des séquences D-loop du genre *Gazella* en utilisant la méthode UPGMA par logiciel MEGA5.

D'après l'arbre, les séquences de D-loop se répartissent en 2 clades principaux :

Clade 1 : Il comprend la majorité des espèces : *G. subgutturosa*, *G. subgutturosa yarkandensis*, *G. subgutturosa subgutturosa*, *G. subgutturosa sairenses*, *G. subgutturosa marica*, *G. cuvieri*, *G. dorcas*.

Correspondrait donc au genre «*Gazella*», ce clade est subdivisé en 2 sous-clades :

- ✓ **Sous-clade A**: Il comprend deux espèces de *G. subgutturosa marica* d'Arabie saoudite ; une espèce de *G. subgutturosa sairenses* ; deux espèces de *G. subgutturosa subgutturosa* et *G. subgutturosa yarkandensis*.

Nous avons utilisé le serveur BLAST afin de pouvoir comparer les séquences de clade A1 à celles de la base de données GenBank.

Description	Max score	Total score	Query cover	E value	Ident	Accession
Gazella gazella voucher AMWP 9673 D-loop, partial sequence, mitochondrial	673	673	100%	0.0	100%	JX274993.1
Gazella erianaei isolate AMWP mitochondrial, complete genome	488	488	100%	4e-134	91%	JX274992.1
Gazella gazella isolate AMWP mitochondrial, complete genome	486	486	100%	1e-133	91%	JX274991.1
Gazella gazella gazella voucher ZMR-MAM-66104 D-loop, partial sequence, mitochondrial	468	468	100%	5e-128	90%	JX274990.1
Gazella gazella voucher AMWP 9679 D-loop, partial sequence, mitochondrial	448	448	100%	7e-122	89%	JX274989.1
Gazella gazella arabica voucher ZSM AM1063 D-loop, partial sequence, mitochondrial	427	427	63%	9e-116	99%	JX274988.1
Gazella gazella arabica voucher ZMR-MAM-2109 D-loop, partial sequence, mitochondrial	422	422	82%	4e-114	92%	JX274987.1
Gazella gazella mitochondrial rRNA-Thr, rRNA-Pro, rRNA-Phe genes and D-loop	422	422	99%	4e-114	88%	AJ235320.1
Gazella dorcas isolate North-east 1 control region, partial sequence, mitochondrial	399	399	92%	2e-107	88%	JN410262.1
Gazella subgutturosa haplotype H5 control region, partial sequence, mitochondrial	381	381	95%	7e-102	86%	GU993274.1
Gazella subgutturosa subgutturosa haplotype 03b D-loop, complete sequence, mitochondrial	374	374	85%	1e-99	88%	HQ185425.1
Gazella arabica voucher ZMR-MAM-2115 D-loop, partial sequence, mitochondrial	372	372	83%	4e-99	89%	JX274986.1
Gazella subgutturosa haplotype H6 control region, partial sequence, mitochondrial	372	372	95%	4e-99	86%	GU993275.1
Gazella subgutturosa haplotype H1 control region, partial sequence, mitochondrial	372	372	95%	4e-99	86%	GU993270.1
Gazella subgutturosa isolate 31 D-loop, partial sequence, mitochondrial	370	370	86%	1e-98	88%	KJ499638.1
Gazella subgutturosa isolate 29 D-loop, partial sequence, mitochondrial	370	370	86%	1e-98	88%	KJ499636.1

Figure 35 : Comparaison entre les séquences de clade A par le programme BLAST.

Cette méthode nous a permis d'obtenir le pourcentage de similarité entre les séquences, la probabilité que cette similarité soit due au hasard, ainsi que le pourcentage de recouvrement des deux séquences.

Les résultats de BLAST montre que il y'a une similarité entre les sous-espèces *G. subgutturosa yarkandensis*, *G. subgutturosa subgutturosa*, *G. subgutturosa sairenses*, *G. subgutturosa marica*. Les pourcentages de la similarité (Tableau IX):

Tableau IV: Tableaux qui représente le pourcentage de similarité entre les séquences par l'utilisation de serveur Blast.

	<i>G. subgutturosa yarkandensis</i>	<i>G. subgutturosa subgutturosa</i>	<i>G. subgutturosa sairenses</i>	<i>G. subgutturosa marica</i>
<i>G. subgutturosa yarkandensis</i>	100 %	99 %	100 %	100 %
<i>G. subgutturosa subgutturosa</i>	99 %	100 %	100 %	100 %
<i>G. subgutturosa sairenses</i>	100 %	100 %	100 %	70 %
<i>G. subgutturosa marica</i>	100 %	100 %	70 %	100 %

- ✓ **Sous-clade B :** Au total, nous avons 2 séquences de l'espèce dorcas d'Arabie saoudite et l'Allemagne.

Clade 2 : Il comprend la majorité des espèces (*Gazella arabica*, *Gazella gazella*, *Gazella gazella*). Ce clade est subdivisé en 2 sous-clades :

- ✓ **Sous-clade C :** Au total, il y'a 2 séquences de l'espèce de *G. arabica* et une séquence de *G. gazella*.
- ✓ **Sous-clade D :** Nous avons 2 séquences de l'espèce *G. gazella* et une séquence de *G gazella gazella*.

Nous avons utilisé le serveur BLAST qui compare une séquence à celles de la base de données GenBank.

Select All None Selected 0

Alignments Download GenBank Graphics Distance tree of results

Description	Max score	Total score	Query cover	E value	Ident	Accession
Gazella arabica voucher ZMB-MAM-2115 D-loop, partial sequence, mitochondrial	623	623	100%	1e-174	100%	JX274964.1
Gazella erlangeri isolate ANWP mitochondrion, complete genome	422	422	89%	4e-114	92%	JN632639.1
Gazella gazella voucher ANWP 9679 D-loop, partial sequence, mitochondrial	399	469	100%	2e-107	91%	JX274969.1
Gazella gazella voucher ANWP 9923 D-loop, partial sequence, mitochondrial	372	372	89%	4e-99	89%	JX274968.1
Gazella gazella isolate ANWP mitochondrion, complete genome	372	372	89%	4e-99	89%	JN632640.1
Gazella gazella arabica voucher ZMB-MAM-2109 D-loop, partial sequence, mitochondrial	344	344	87%	8e-91	88%	JX274976.1
Gazella gazella gazella voucher ZMB-MAM-66104 D-loop, partial sequence, mitochondrial	342	342	88%	3e-90	87%	JX274973.1
Gazella gazella isolate KKWRIC 134 D-loop, partial sequence, mitochondrial	333	333	60%	2e-87	96%	GU384743.1
Gazella dorcas isolate North-east 1 control region, partial sequence, mitochondrial	331	401	91%	7e-87	88%	JN410262.1
Gazella gazella isolate KKWRIC 724 D-loop, partial sequence, mitochondrial	331	331	60%	7e-87	95%	GU384735.1
Gazella gazella isolate KKWRIC 848 D-loop, partial sequence, mitochondrial	327	327	60%	8e-86	95%	GU384736.1

Figure 36 : Comparaison entre les séquences de *G. arabica* et *G. gazella* à l'aide du programme BLAST.

Nous avons comparé les séquences de *G. gazella* avec une séquence de *G. arabica*. Ces dernières ont une grande similarité de 90% avec la séquence de *G. gazella*. Ce résultat montre que ces séquences sont très homologues.

✚ Interprétation des relations phylogénétiques entre les différentes espèces du genre *Gazella*

Les gazelles sont identifiées comme étant le groupes taxonomiquement le plus complexes des bovidés (Groves et al., 1967 ; Groves, 1997). Ce problème est provoqué en grande partie par la variation intraspécifique considérable étant confondue avec la similitude interspécifique, particulièrement en caractères tels que la forme des cornes, taille du corps et la coloration de pelage (Groves, 1996).

D'après les arbres globaux (de cyt b et D-loop) du genre *Gazella*, nous avons deux clades importants : (1) un clade Asiatique (*G. bennettii*, *G. subgutturosa*, *G. marica*, *G. leptoceros* et *G. cuvieri*) et (2) un clade Africain (*G. saudiya*, *G. dorcas*, *G. spekei*, *G. gazella*, et *G. arabica*) (Lerp et al., 2013).

- *G. marica* (Thomas, 1897), a été englobé dans *G. leptoceros* par Ellerman et al., 1951. Plus tard, le *G. marica* a été considéré comme une sous-espèce de *G. subgutturosa* a la base des similitudes morphologique (Groves et al., 1967; Kingswood et al., 1996, 1997).
- Les classifications tât placent *G. bennettii* comme une sous-espèce de *G. gazella* (Haltenorth et al., 1977; Roberts, 1977) ou comme sous-espèce de *G. dorcas* (Gentry, 1964; Groves, 1969; Lange, 1972).
- Les données de caryologie, ont trouvé que *G. bennettii* indépendantes du *G. gazella* (Furley et al., 1988; Kumamoto et al., 1995).
- La classification taxonomique de *G. cuvieri* et du *G. leptoceros* demeure embrouillant. Lange (1972) a classifié *G. cuvieri* sous *G. gazella*, et *G. leptoceros* étaient considérés comme une sous-espèce de *G. subgutturosa*.
- En revanche Hassanin (2012), propose récemment que *G. marica* et *G. leptoceros* sous-espèce de *G. cuvieri* en raison de leur divergence mitochondriques.
- A propos de *G. dorcas*, plusieurs sous-espèces sont décrites sur la base de la variation phénotypique, telle que la coloration de pelage et la forme et la longueur des cornes (Groves, 1969, 1981; Alados, 1987; Yom-Tov et al., 1995; Groves et al., 2011).
- Une étude phylogeographique basée sur la variation de séquence mitochondrial de cytochrome b et de contrôle région indique récemment que *G. dorcas* comprenant « *G.*

saudiya» et « *G. pelzelni*» représente un groupe réciproquement monophyletique avec des relations de groupe-sœur au *G. gazella* et à *G. arabica* (Lerp et al., 2011).

III.2.2. Résultat de 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 substitution de nucléotides.

- **Cytb** : L'analyse de Cytb de l'ADN mitochondrial révèle une diversité génétique très importante ; les 66 individus étudiés ont permis de caractériser 30 haplotypes différents. 13 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 ($Hd = 0,9622$) (Figure 37).

```

Haplotype/DNA Sequences Data File
Input Data File: C:\...\Apour dnasp.fas
Number of sequences: 66 Number of sequences used: 66
Selected region: 1-1186 Number of sites: 1186
Total number of sites (excluding sites with gaps / missing data): 233
Sites with alignment gaps: not considered
Number of variable sites: 41
===== Haplotype Distribution =====
Haplotype diversity, Hd: 0.9622
Hap_1: 1 [1]
Hap_2: 2 [2-37]
Hap_3: 2 [3-5]
Hap_4: 1 [4]
Hap_5: 5 [6-7 9 11 51]
Hap_6: 1 [8]
Hap_7: 2 [10 12]
Hap_8: 5 [13 45-48]
Hap_9: 2 [14 16]
Hap_10: 1 [15]
Hap_11: 4 [17-18 20 30]
Hap_12: 5 [19 21 25 27 31]
Hap_13: 2 [22 32]
Hap_14: 2 [23 26]
Hap_15: 1 [24]
Hap_16: 1 [28]
Hap_17: 1 [29]
Hap_18: 2 [33 44]
Hap_19: 4 [34 36 40 42]
Hap_20: 1 [35]
Hap_21: 2 [38 41]
Hap_22: 2 [39 43]
Hap_23: 1 [49]
Hap_24: 8 [50 54-55 57-60 62]
Hap_25: 2 [52-53]
Hap_26: 1 [56]
Hap_27: 1 [61]
Hap_28: 1 [63]
Hap_29: 1 [64]
Hap_30: 2 [65-66]

Hap_1: 1 [arabica_AS_1]
Hap_2: 2 [arabica_AS_2 gazella_AS]
Hap_3: 2 [arabica_AS_3 arabica_Jor]
Hap_4: 1 [arabica_Palestine]
Hap_5: 5 [bennettii_Inde bennettii_pak bennettii_PA_2 bennettii_AS_2 bennettii_Iran2]
Hap_6: 1 [bennettii_PA_1]
Hap_7: 2 [bennettii_AS_1 bennettii_Iran]
Hap_8: 5 [cuvieri_PA leptoceros_Tun leptoceros_Egy leptoceros_Alger leptoceros_tch]
Hap_9: 2 [cuvieri_AN cuvieri_Esp2]
Hap_10: 1 [cuvieri_Esp1]
Hap_11: 4 [dorcas_Maroc dorcas_AN_1 dorcas_Mali dorcas_Tun]
Hap_12: 5 [dorcas_AN_2 dorcas_Sud dorcas_Pal dorcas_All dorcas_Chad]
Hap_13: 2 [dorcas_Qat pelzelni_Qat]
Hap_14: 2 [dorcas_AS dorcas_Pal_2]
Hap_15: 1 [dorcas_PA]
Hap_16: 1 [dorcas_Alg_Hoggar]
Hap_17: 1 [dorcas_Alg_EIBayadh]
Hap_18: 2 [dorcas_x_leptoceros_leptoceros_PA]
Hap_19: 4 [gazella_All gazella_Pal gazella_All gazella_Eur]
Hap_20: 1 [gazella_Tur]
Hap_21: 2 [gazella_PA gazella_Qat]
Hap_22: 2 [gazella_UAE g_muscataensis_PA]
Hap_23: 1 [saudiya_AS]
Hap_24: 8 [subgutturosa_PA_2 subgutturosa_Mongoli subgutturosa_Eurasie
subgutturosa_Turkist subgutturosa_PA subgutturosa_Qat subgutturosa_Iran
subgutturosa_Chine]
Hap_25: 2 [spekei_AN spekei_UAE]
Hap_26: 1 [subgutturosa_Aze]
Hap_27: 1 [subgutturosa_Iran_2]
Hap_28: 1 [marica_Oman]
Hap_29: 1 [marica_UAE]
Hap_30: 2 [marica_PA marica_AS]

```

Figure 37: Les haplotypes ont été construits à l'aide du logiciel DnaSP de cytb.

- **D-loop** : 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.



Figure 38 : Résultat d'alignement des séquences D-loop du genre *Gazella* 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.

Sur les 16 séquences de D-loop de genre *Gazella* obtenues, 15 haplotypes ont pu être identifiés, 14 haplotypes privés ont été détectés. Chaque haplotype contient une seule

séquence sauf l'haplotype 7 contient 2 séquences. Le nombre d'haplotypes ont été construit à l'aide du logiciel DnaSP (Figure 39).

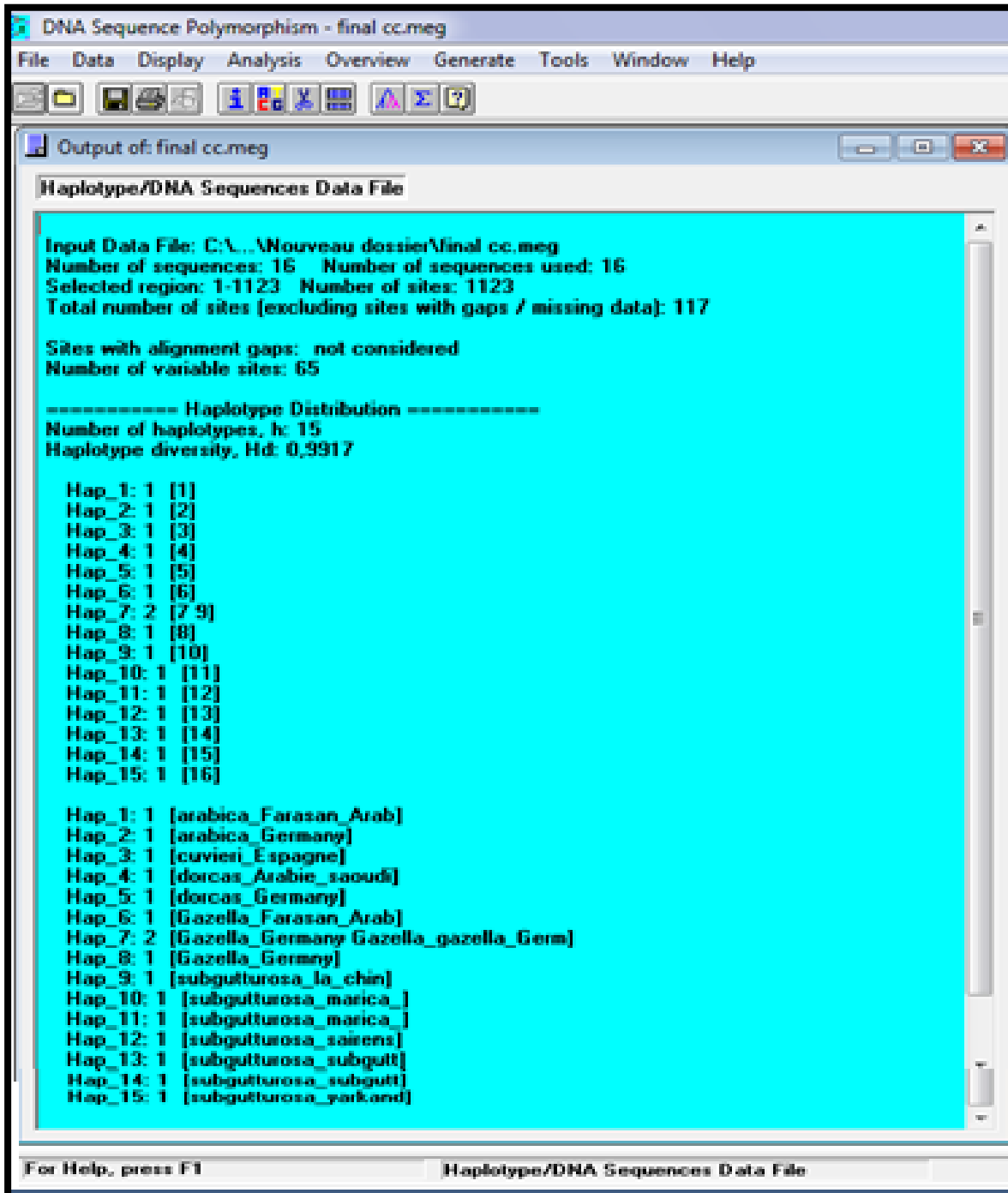


Figure 39 : Les haplotypes des séquences D-loop du genre *Gazella* en utilisant le logiciel DnaSP.

Tableau V : 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	66	17	0,9622	1186	41
	2		13			
D-loop	1	16	10	0.9917	1123	65
	2		4			

Interprétation des résultats de DnaSP

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 (Salemi 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 40-41). 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) :

$$\text{Skew} = (\text{nucleotide1} - \text{nucleotide2}) / (\text{nucleotide1} + \text{nucleotide2})$$

- 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.

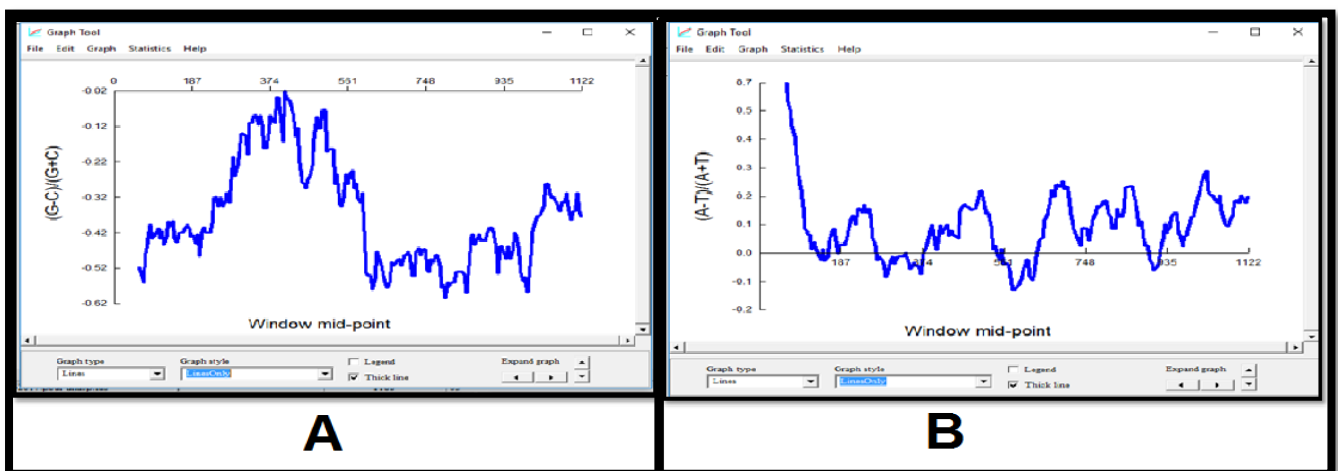


Figure 40 : Variations des fréquences GC (A) et de AT (B) de Cytb produites à partir de DAMBE.

Nous avons remarqué que le GC Skew est négatif (Figure 40 A) cependant l'AT Skew est positif (Figure 40 B), Par conséquent, 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).

Selon Saccone (1999) dans tous les DNAsmt des vertébrés on observe remarquablement un GC skew négatif, plus élevé pour les reptiles, l'aves et les mammifères.

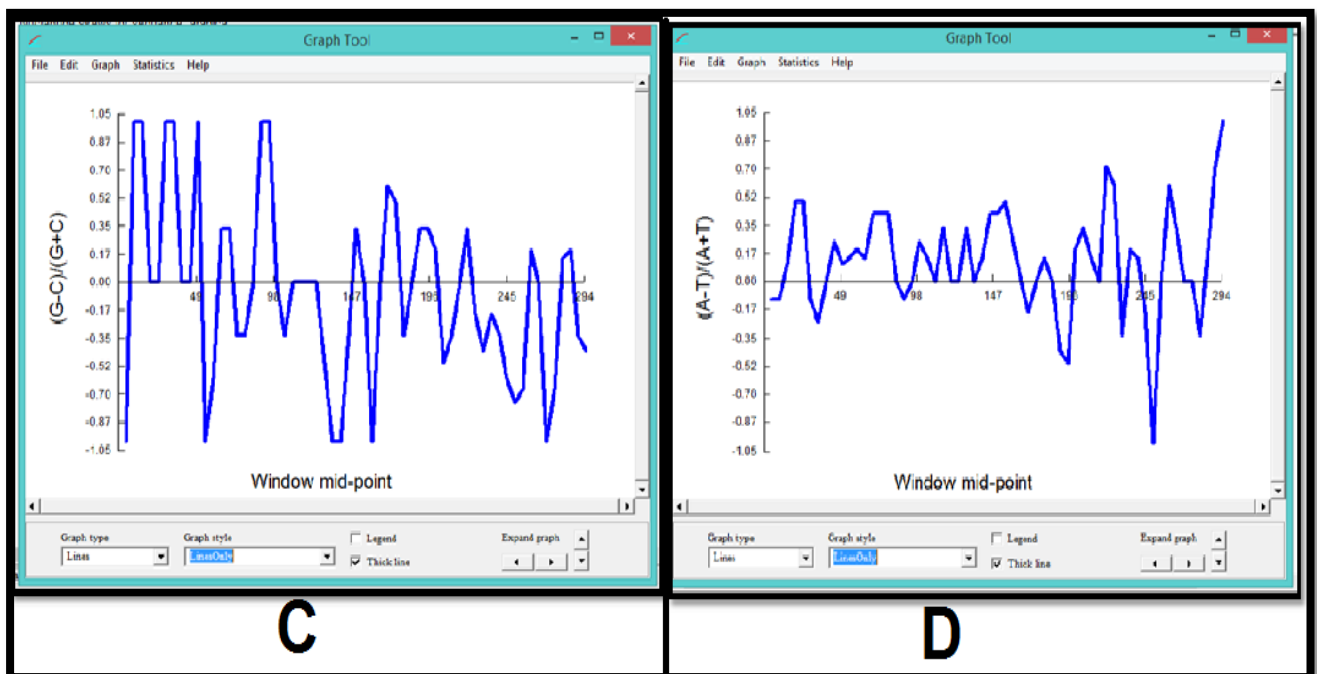


Figure 41 : Variations des fréquences GC (C) et de AT (D) de D-loop produites à partir de DAMBE

Nous avons remarqué que la fréquence GC Skew est généralement négatif (Figure 41 C) cependant l'AT Skew est généralement positif (Figure 41 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 D-loop (Figure 42).

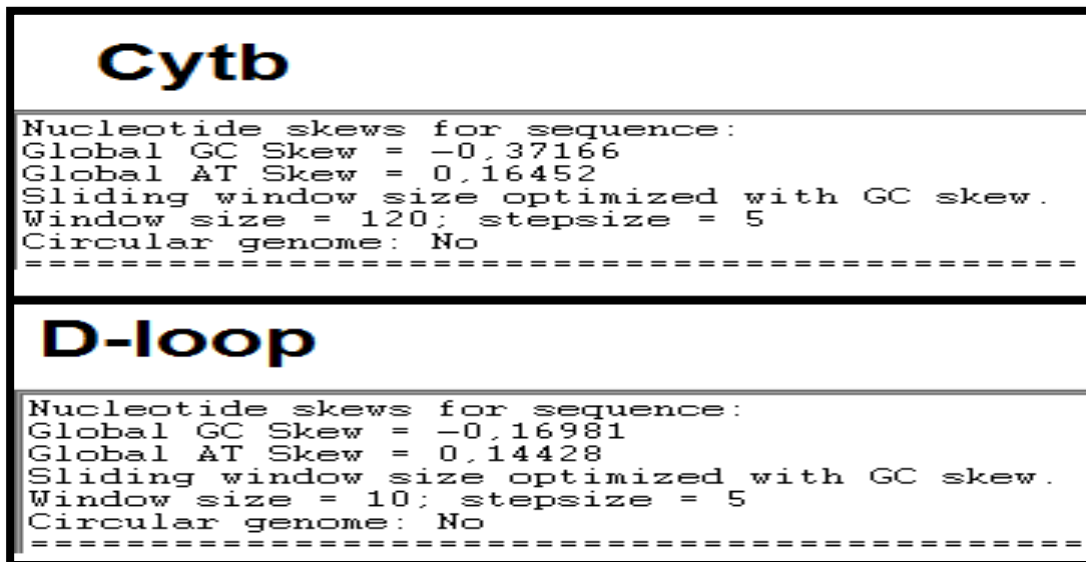


Figure 42 : Comparaison des fréquences entre cytb et D-loop réalisés par le logiciel DAMBE.

Nous avons remarqué qu'il y a une richesse de la cytosine sur la guanine et de l'adénine sur la thymine pour cytb et D-loop, mais la fréquence des D-loop est plus élevée par rapport au cytb (Figure 42) c.-à-d. le nombre de CG présente dans cytb est plus élevé par rapport au D-loop, parce que cette dernière est moins conservée que le cytb, ça est due de leurs rôles dans l'organisme.

Le gène du cytb code pour une protéine membranaire intégrale alors il est conservé cependant D-loop joue le rôle d'une région de contrôle, il ne code pas pour une protéine donc il est moins protégé.

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.

Sur les 66 séquences de Cytb de genre *Gazella* obtenues, 30 haplotypes ont pu être identifiés. La structure du réseau (Figure 43) apparaît plus complexe et avec de nombreux haplotypes divergents.

Haplogroupe 1 correspond au groupe Asiatique : *G. bennettii*, *G. subgutturosa*, *G. marica*, *G. leptoceros* et *G. cuvieri*. pour les haplotypes suivants : H5, H6, H7, H8, H9, H10, H18, H24, H26, H27, H28, H29, H30.

❖ **Haplogroupe 2** correspond au groupe d'Afrique : *G. saudiya*, *G. dorcas*, *G. spekei*, *G. gazella*, et *G.* pour les haplotypes suivants : H1, H2, H3, H4, H11, H12, H13, H14, H15, H16, H17, H19, H20, 21, H22, H23, H25.

Conclusion

Les gazelles sont des mammifères, de la famille des bovidés, comportent quatre genres monophyletiques (*Antilope*, *Nanger*, *Eudorcas* et *Gazella*). La tribu Antilopiné comporte aujourd'hui 13 genres : *Raphiceros*, *Ourebia*, *Madoqua*, *Dorcatragus*, *Saiga*, *Litocranius*, *Ammodorcas*, *Antidorcas*, *Procapra*, *Eudorcas*, *Nanger*, *Antilope* et *Gazella*, La situation de genre *Gazella* est plus complexe, avec des espèces existant en Afrique, en Moyen-Orient, et en Asie.

Généralement les gazelles sont caractérisées par une variabilité phénotypique prononcée qui n'est pas toujours reflétée par une divergence moléculaire des séquences d'ADN, une partie de cette variation peut être due à la plasticité phénotypique.

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 *D-loop* qui représentent le clade d'Afrique 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 le genre *Gazella* l'existence de deux clades distincts qui sont le clade Asiatique (*G. bennettii*, *G. subgutturosa*, *G. marica*, *G. leptoceros* et *G. cuvieri*) et le clade Africain (*G. saudiya*, *G. dorcas*, *G. spekei*, *G. gazella*, et *G. arabica*).

La baisse dans le nombre des Gazelle est mondiale, pour cela il y a un besoin urgent pour que les chercheurs déterminent leurs unités de conservation et décrivent des relations phylogénétiques entre les espèces de ce genre. .

-
- **ADJRAD S., 2005.** Les Méthodes de reconstruction phylogénétique: Modèles, Algorithmes, Exploitation des Logiciels et Applications à *Vicia*L. (Fabaceae), *Hyoseris*L. (Asteraceae) et *Medicago*L. (Fabaceae). Mémoire. Université Abderrahmane Mira de Béjaïa. Algérie.10-75. p137
 - **Alados C. L., 1987.** A Cladistic Approach to the Taxonomy of the Dorcas Gazelles. *Israel Journal of Zoology*, 34(1–2), 33–49.
 - **Arambourg C., 1959.** Vertébrés continentaux du Miocène supérieur de l’Afrique du Nord. *Publ Serv Carte Géol Algérie*, n s, Paléont, Mém, 4: 1–159.
 - **Arigon A. M., 2006.** Développements d’outils pour l’aide à l’identification dans les grandes banques de familles de gènes. Thèse de doctorat : Université claudes bernard-lyon 1. 12-13. 167p.
 - **Avise J. C. S., Arnold J., Ball MR., Bermingham E., Lamb T., Neigel JE., Reeb CA., and Saundres NC., 1987.** “Intraspecific phylogeography: The mitochondrial DNA Bridge Between population genetics and systematics.” *Annual review of ecology and systematics* 18: 489-522.
 - **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.
 - **Bardelli G., Ricci M., De Marchi G., Cardini A., 2014.** "Just another island dwarf? Phenotypic distinctiveness in the poorly known Soemmerring's Gazelle, *Nanger soemmerringii* (Cetartiodactyla: Bovidae), of Dahlak Kebir Island". *Biological Journal of the Linnean Society*. **111** (3): 603–620.
 - **Bärmann E. V., Wronski T., Lerp H., Azanza B., Börner S., Erpenbeck D., Rössner G. E., Wörheide G., 2013.** Data from: A morphometric and genetic framework for the genus *Gazella* de Blainville, 1816 (Ruminantia: Bovidae) with special focus on Arabian and Levantine mountain gazelles.
 - **Barrowclough G. F., and Flesness N. R., 1996.** Species, subspecies and races: the problem of units of management in conservation. Pages 247-254(Phylogenetic Reanalysis of the Saudi Gazelle and Its Implications).
 - **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. & Wheeler D. L., 2006.** GenBank. Nucleic Acids Res, 34(Database issue), D16–20.
- **Bérard 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 Protein Data Bank. Nucleic Acids Res, 28(1), 235–42.
- **Boore J. L., 1999.** “Animal mitochondrial genome”. Nucleic acids Res 27(8): 1767-1780.
- **Camin J. H., and Sokal R. R., 1965.** A method for deducing branching sequences in phylogeny. Evolution, 19: 311-326.
- **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 Compiegne : 5-38. 216p.

-
- **Candolle A. P. de., 1813.** Théorie élémentaire de la Botanique ou exposition du principe de classification élémentaire naturelle et de l'art de décrire et d'étudier les végétaux. Déterville (Paris).
 - **De Carvalho Junior S. A., 2003.** Sequence Alignment Algorithms. Thesis of Master of Sciences in Advanced Computing. King's College London. Univ. London. 4-18.
 - **Diallo A. B., 2009.** Bioinformatique avancée. *Université du Québec à Montréal. Canada.*
 - **Diallo A. B., 2007.** Développement et parallélisation d'algorithmes bioinformatiques pour la reconstruction d'arbres phylogénétiques et de réseaux réticulés. Mémoire. *Université du Québec à Montréal. Canada.* 7-10. 194p
 - **Effron M., Bogart M. H., Kumanoto A. T., Benirschke K., 1976.** Chromosome studies in the mammalian subfamily Antilopinae. *Genetica* 46:419–444.
 - **Efron, B., 1979.** Bootstrap methods : another look at the jackknife. *Ann. Statist.*, 7: 1-26.
 - **Ellerman J. R., Morrison-Scott T. C. S., 1951.** Checklist of palearctic and Indian mammals, 1758 to 1946. In: Trustees of the British Museum. London.
 - **Felsenstein J., 2004.** Inferring phylogenies. Sunderland (MA): Sinauer Associates: 664.
 - **Felsenstein J., 1993.** PHYLIP (PHYLogeny Inference Package) version 3.6a2, Distributed by the author, Department of Genetics, University of Washington, Seattle, WA.
 - **Felsenstein J., 1985a.** Confidence limits on phylogenies with a molecular clock. *Syst. Zool.*, 34(2): 152-161.
 - **Felsenstein J., 1985b.** Confidence limits on phylogenies: an approach using the bootstrap. *Evolution*, 39: 783-791.
 - **Furley C. W., Tichy H., Uerpman H. P., 1988.** Systematics and chromosomes of the Indian gazelle, *Gazella bennetti* (Sykes, 1831). *Zeitschrift für Säugetierkunde* 53:48–54.
 - **Gattiker A., Michoud K., Rivoire C., Auchincloss A. H., Coudert E., Lima T., Kersey P., Pagni M., Sigrist C. J., Lachaize C., Veuthey A. L., Gasteiger E., & Bairoch A., 2003.** Automated annotation of microbial proteomes in SWISS-PROT. *Comput Biol Chem*, 27(1), 49–58.

-
- **Gattolliat 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., Jochen Graw., 2006.** 4. Auflage, Springer Verlag Berlin Heidelberg.
 - **Gentry A. W., 2010.** Bovidae. In: Werdelin L, Sanders W J eds. Cenozoic Mammals of Africa. Berkeley and Los Angeles: University of California Press. 741–796
 - **Gentry A. W., 1992.** The subfamilies and tribes of the family Bovidae. Mammal Review 22:1–32.
 - **Gentry A. W., 1970.** The Bovidae (Mammalia) of the Fort Ternan fossil fauna. In: Leakey L S B, Savage R J G eds. Fossil Vertebrates of Africa. London and New York: Academic Press. 243–323.
 - **Gentry A. W., 1964.** Skull characteristics of African gazelles. Annales and Magazine of Natural History 13:353–382.
 - **Golding B., Morton D., 2003.** Elementary Sequence Analysis. Department of Biology. McMaster. Univ. Hamilton, Ontario. 190 p.
 - **Groves C. P., Grubb P., 2011.** Ungulate taxonomy. John Hopkins University Press, Baltimore.
 - **Groves C. P., 2000.** Phylogenetic relationships within recent Antilopinae. In: Vrba ES, Schaller GB, editors. Antelopes, deer and relatives – fossil record, behavioural ecology, systematics and conservation. Yale University Press. p. 223–233.
 - **Groves C. P., 1997.** Taxonomy of Arabian gazelles. In: Habibi K, Abu- Zinada A, Nader I (eds) The gazelles of Arabia. National Commission for Wildlife Conservation and Development, Riyadh, pp 24–51.
 - **Groves C. P., 1996.** Taxonomic diversity in Arabian gazelles: the state of the art. In: Greth, A., Magin, C., Ancrenaz, M. (Eds.), Conservation of Arabian Gazelles. National Commission for Wildlife Conservation, Development, Riyadh, pp. 8–39.
 - **Groves C. P., 1981.** Notes on gazelles. 3. The Dorcas gazelles of north Africa. Anali Museo Civico di Storia Naturale di Genova, 83, 455–471.
 - **Groves C. P., 1969.** On the smaller gazelles of the Genus Gazella de Blainville, 1816. Zeitschrift für Säugetierkunde 37:38–60.
 - **Groves C. P., 1967.** On the gazelles of the genus Procapha Hodgson, 1846. Z. Säugetierk. 32: 144–149.

-
- **Grubb P., 2005.** Order Artiodactyla. In: Wilson DE, Reeder DM, editors. Mammal species of the world, Vol. 1. Baltimore: John Hopkins University Press. p. 629–722.
 - **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.
 - **Guindon S., et Gascuel O., 2003.** A simple, fast and accurate algorithm to estimate large phylogenies by maximum likelihood. *Systematic Biology*.52: 696-704.
 - **Haeckel E., 1860.** "Über neue, lebende Radiolarien des Mittelmeeres," Monatsberichte der Kdniglichen. 794-817.
 - **Holdenworth T., & Diller H., 1977.** Säugetiere Afrikas. Munich: BLV.
 - **Hassanin A., Delsuc F., Ropiquet A., et al. (11 co-authors) 2012.** Pattern and timing of diversification of Cetartiodactyla (Mammalia, Laurasiatheria), as revealed by a comprehensive analysis of mitochondrial genomes. *Comptes Rendus Biologies* 335:32–50.
 - **Haubold B., 2000.** Practical bioinformatics. Phylogenetics—ML. Max Planck Institut für chemische Ökologie 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. & Sigrist C. J., 2006.** The PROSITE database. *Nucleic Acids Res*, 34(Database issue), D227–30.
 - **IUCN., 1988.** "Red List of Threatened Animals," Unwin, Surrey. Jermin, L., Graur, D., Lowe, R., and Crozier, R. (1994). Analysis of directional mutation pressure and nucleotide content in mitochondrial cytochrome b genes. *J. Mol. Evol.*39: 160–173
 - **Kawalski K., Rzebick-Kowalska B., 1991.** Mammals of Algeria
 - **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.
 - **Kingdon J., 2015.** The Kingdon Field Guide to African Mammals ,London, UK: Bloomsbury Publishing. pp.
 - **Kingdon J., 2013.** Mammals of Africa. London: Bloomsbury.

-
- **Kingman J. F., 2000.** Origins of the coalescent. 1974-1982. *Genetics* 156:1461-1463.
 - **Kingswood S. C., Rebholz W. E. R., Vassart M., Kumamoto A. T., 1997.** Cytogenetics and molecular genetics of Arabian gazelles: chromosomes, proteins and nucleic acids. In: Habibi K, Abu-Zinada AH, Nader IA, editors. *The Gazelles of Arabia*. Riyadh: National Commission for Wildlife Conservation and Development. p. 159–191.
 - **Kingswood S. C., Vassart M., Williamson D., 1996.** Systematics of *Gazella subgutturosa* : implications for conservation in the Arabian Peninsula. In: Greth A, Magin C, Ancrenaz M, editors. *Conservation of Arabian gazelles*. Riyadh: National Commission for Wildlife Conservation and Development. p. 104–111.
 - **Kingswood S. C., Blank D. A., 1996.** *Gazella subgutturosa* . *Mammalian Species* 518:1– 10.
 - **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.
 - **Kuhner M. K., Felsenstein J., 1994.** A Simulation Comparison of Phylogeny Algorithms under Equal and Unequal Evolutionary Rates. *Mol. Biol. Evol.* 11(3): 459-468.
 - **Kumamoto A. T., Kingswood S. C., Rebholz W. E. R., Houck M. L., 1995.** The chromosomes of *Gazella bennetti* and *Gazella saudiya* . *Zeitschrift für Säugetierkunde* 60:159–169.
 - **Kumar S., Tamura K., Nei M., 1994.** MEGA:Molecular Evolutionary Genetics Analysis software for microcomputers, Vol.10. P.189-191.
 - **Lam H. J., 1950.** Proposal to indicate a taxonomic group of any rank with the term taxon (plural taxa). In: J. Lanjouw (Ed.), *Botanical nomenclature and taxonomy*. Union inter. Sc. biol., Colloquia, Ser. B., vol. 2 (Paris). pp. 1-88.
 - **Lange J., 1972.** Studien an Gazellenschädeln. Ein Beitrag zur Systematik der kleinen Gazellen, *Gazella* (de Blainville, 1816). *Säugetierkundliche Mitteilungen* 20:193–249.
 - **Lankester E. R., 1870.** On the use of the term homology in modern zoology, and the distinction between homogenetic and homoplastic agreements. *Ann. Mag. Nat. Hist.*,4(6): 34-43.

-
- **Legendre L., & Legendre P., 1979.** Écologie numérique. 1. Le traitement multiple des données écologiques. Masson, Paris & Les Presses de l'Université du Québec, Montréal.
 - **Lemey P., Salemi M., Vandamme A. M., 2009.** The phylogenetic handbook. Cambridge: Cambridge University Press.
 - **Lerp H., Wronski T., Butynski T. M., & Plath M., 2013.** Speciation of Arabian gazelles. In: *Speciation: Natural Processes, Genetics and Biodiversity*. P. Michalak, ed. Nova Science, Hauppauge, New York. Pp. 59-82
 - **Lerp H., Wronski T., Plath M., Schröter A., Pfenninger M., 2013.** Phylogenetic and population genetic analyses confirm the existence of a distinct species boundary between Mountain (Gazella gazella) and Arabian Gazelles (G. arabica) in the Levant. *Mammalian Biology*. In press.
 - **Lerp H., Wronski T., Pfenninger M., Plath M., 2011.** A phylogeographic framework for the conservation of Saharan and Arabian Dorcas gazelles (Artiodactyla: Bovidae). *Organisms Diversity and Evolution* 11:317–329.
 - **Li W. H., 1997.** Molecular evolution. Sunderland, Massachusetts: Sinauer Assoc. 487.
 - **Li K.B., 2003.** ClustalW-MPI: ClustalW analysis using distributed and parallel computing 19: 1585–1586.
 - **Librado P., 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
 - **Loggers C., Thevenot M., and Aulagnier S., 1992.** Status and distribution of Moroccan wild ungulates. *Biol. Cons.* 59: 9–18.
 - **Mallon D. P., & Cuzin F., 2008.** "Gazella cuvieri". IUCN Red List of Threatened Species. Version 2008. International Union for Conservation of Nature. Retrieved 18 August 2015.
 - **Mayr E., 1969.** Principles of systematic zoology. McGraw Hill (New York).
 - **Mayr E., 1965.** Classification and phylogeny. *Amer. zool.*, 5: 165-174.
 - **Mayr E., Linsley E. G., and Usinger R., 1953.** Methods and principles of systematic zoology. McGraw-Hill (New York).

-
- **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.
 - **Mueller, L. D. et Ayala, F. J., 1982.** Estimation and interpretation of genetic distance in empirical studies. *Genet. Res.*,40: 127-137.
 - **Nei M., 1996.** Phylogenetic Analysis in Molecular Evolutionary Genetics. *Annu. Rev. Genet.* 30: 371-401.
 - **Nei M., 1987.** Molecular Evolutionary Genetics. Columbia University Press, New York.
 - **Nei M., Li W. H., 1979** Mathematical model for studying genetic variation in terms of restriction endonucleases. *Proc Natl Acad Sci USA* 76 : 5269-5273
 - **Okubo K., Sugawara H., Gojobori T., & Tateno Y., 2006.** DDBJ in preparation for overview of research activities behind data submissions. *Nucleic Acids Res*, 34(Database issue), D6–9.
 - **Page R. D. M., 1996.** TREEVIEW: An application to display phylogenetic trees on personal computers. *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.
 - **Peter Yakovchuk., Ekaterina Protozanova., and Maxim D., Frank-Kamenetskii., 2006.** Base-stacking and base-pairing contributions into thermal stability of the DNA double helix. *Nucleic Acids Research*.
 - **Rasmont R., 1997.** Evolution Biologique. Traduction de la 2ème édition anglaise. Ed. Départ. Deboeck Univ. Paris. Bruxelles. 371-507.
 - **Rebholz W. E. R., Harley E. H., 1999.** Phylogenetic relationships in the bovid subfamily Antilopinae based on mitochondrial DNA sequences. *Molecular Phylogenetics and Evolution* 12:87–94.
 - **Rischkowsky B., Pilling D., 2008** -L'Etat des Ressources Zoogénétiques pour l'Alimentation et l'Agriculture dans le Monde, pp393-410.
 - **Roberts T. J., 1977.** The mammals of Pakistan. Ernest Benn, London.
 - **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.
 - **Rustin P., Rötig A., 1995.** Mécanismes moléculaires des cytopathies mitochondriales. *Med Therap* 1: 435-447.
 - **Ryder O., 1987.** Conservation action for gazelles: An urgent need. *Trends Ecol. Evol.* 6: 143–144.
 - **Saccone C., De Giorgi C., Gissi C., Pesole G., Reyes A., 1999.** Evolutionary genomics in Metazoa: the mitochondrial DNA as a model system. *Gene* 238, 195–209.
 - **Saitou N., Nei M., 1987.** The neighbor-joining method: a new method for reconstructing phylogenetic trees. *Mol Biol Evol.* 4:406–425.
 - **Saitou N., and Nei M., 1997.** The evolutionary history was inferred using the Neighbour-Joining method.
 - **Saleh M., 1987.** The decline of gazelles in Egypt. *Biol. Cons.* 39:83-95.
 - **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-terminating inhibitors. *Proc. Natl. Acad. Sci. USA* 74, 5463–5467.
 - **Schaal B. A., Olsen K. M., 2000.** Gene genealogies and population variation in plants. *Proc Natl Acad Sci U S A* 97: 7024–7029.
 - **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.
 - **Simpson G. G., 1961.** Principles of animal taxonomy. Columbia University Press (New York).
 - **Sunnucks P., 2001.** Efficient genetic markers for population biology. *Tree*, 15: 199–203.

-
- **Swofford D.L., 1998-PAUP***. Phylogenetic Analysis Using Parsimony and other methods. Version 4.0 (beta version). Laboratory of Molecular Systematics Smithsonian.
 - **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.
 - **Tamura K., Peterson D., Peterson N., Stecher G., Nei M. & Kumar S., 2011**. MEGA5: Molecular Evolutionary Genetics Analysis using Likelihood, Distance, and Parsimony methods [logiciel]. *Molecular Biology and Evolution*. (logiciel libre,téléchargeable sur <http://www.megasoftware.net/>)
 - **Thomas O., 1897**. On a new gazelle from Central Arabia. *Annals and Magazine of Natural History* 6, 19, 162–163.
 - **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., & 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.
 - **Vicente M. C., & Fulton T., 2003**. Utilisation des marqueurs moléculaires pour l'étude de la diversité génétique végétale. International Plant Genetic Institute for Genomic Diversity Resources Institute.
 - **Wronski T., Wacher T., Hammond R. L., Winney, B., Hundertmark K. J., Blacket MJ., Mohammed OB., Flores B., Omer SA., Macasero W., Plath M., Tiedemann R., Bleidorn C., 2010**. Two reciprocally monophyletic mtDNA lineages elucidate the taxonomic status of Mountain gazelles (*Gazella gazella*). *Syst. Biodivers.* 8, 119–129, <http://dx.doi.org/10.1080/14772001003613192>.

-
- **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. & 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 L S., 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.
 - **Yom-Tov Y., Mendelssohn H., & Groves C. P., 1995.** *Gazella dorcas*. *Mammalian Species*, 491,1–3.
 - **Yoshikawa H., Dogruman-AI 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. *Parasitol Res.*;109:1045-50

Biblio-web

Bioedit: <http://www.mbio.ncsu.edu/BioEdit/bioedit.html>

Blast: <https://blast.ncbi.nlm.nih.gov/Blast.cgi>

Clustalw: <http://www.ebi.ac.uk/Tools/msa/clustalw2/>

DNAdist: <http://mycor.nancy.inra.fr/egglib/v2/installation.html>

DnaMLK: <http://fr.freedownloadmanager.org/Windows-PC/T-Rex-GRATUIT.html>

DAMBE: <http://dambe.bio.uottawa.ca/DAMBE/dambe.aspx>

Dendroscope : <http://ab.inf.uni-tuebingen.de/software/dendroscope/>

Exonerate: <http://ftp.ebi.ac.uk/pub/software/vertebrategenomics/exonerate/exonerate-2.2.0-cygwin.tar.gz>

MEGA: <http://www.mybiosoftware.com/mega-5-05-molecular-evolutionary-genetics-analysis.html>

PAUP: <http://fr.freedownloadmanager.org/Windows-PC/PAUP-GRATUIT.html>

PHYML: <http://www.atgc-montpellier.fr/phyml/versions.php>

Sequencher: <http://www.commentcamarche.net/download/telecharger-34084043-sequencher>

TreeView: http://www.01net.com/telecharger/windows/Internet/editeur_de_site/fiches/12283.html

Exemples des Formats des Banques de données :

Tab VI: Exemple d'une séquence au format GenBank

```

GenBank: KU560630.1
FASTA  Graphics  PopSet
LOCUS      KU560630          1149 bp    DNA     linear   MAM 10-FEB-2016
DEFINITION Gazella arabica isolate GGF41 cytochrome b (cytb) gene, complete
           cds; mitochondrial.
ACCESSION  KU560630
VERSION    KU560630.1
KEYWORDS   .
SOURCE     mitochondrion Gazella arabica (Arabian gazelle)
ORGANISM   Gazella arabica
           Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
           Mammalia; Eutheria; Laurasiatheria; Cetartiodactyla; Ruminantia;
           Pecora; Bovidae; Antilopinae; Gazella.
REFERENCE  1 (bases 1 to 1149)
AUTHORS    Lerp,H., Klaus,S., Allgower,S., Wronski,T., Pfenninger,M. and
           Plath,M.
TITLE      Phylogenetic analyses of gazelles reveal repeated transitions of
           key ecological traits and provide novel insights into the origin of
           the genus Gazella
JOURNAL    Mol. Phylogenet. Evol. (2016) In press
PUBMED     26826603
REMARK     Publication Status: Available-Online prior to print
REFERENCE  2 (bases 1 to 1149)
AUTHORS    Lerp,H., Klaus,S., Allgoewer,S., Wronski,T., Pfenninger,M. and
           Plath,M.
TITLE      Direct Submission
JOURNAL    Submitted (07-JAN-2016) Natural History Collections, Museum
           Wiesbaden, Friedrich-Ebert-Allee 2, Wiesbaden, Hessen 65185,
           Germany
COMMENT    ##Assembly-Data-START##
           Sequencing Technology :: Sanger dideoxy sequencing
           ##Assembly-Data-END##
FEATURES   Location/Qualifiers
           source          1..1149
                           /organism="Gazella arabica"
                           /organelle="mitochondrion"
                           /mol_type="genomic DNA"
                           /isolate="GGF41"
                           /db_xref="taxon:1229624"
                           /country="Saudi Arabia: Farasan Islands"
                           /collected_by="T. Wronski"
           gene                7..1146
                           /gene="cytb"
           CDS              7..1146
                           /gene="cytb"
                           /codon_start=1
                           /transl_table=2
                           /product="cytochrome b"
                           /protein_id="AM081621.1"
                           /translation="MINTRKTHPLMKIVNNAFIDLPA PSNISSWNNFGSLLGICLILQ
                           ILTGLFLAMHYTADTATAFSSVTHICRDVNYGWIIRYMHANGASMFFICLFMHVGRGL
                           YYGSYTFLETHNIGVILLFATMATAFMGYVLPWQMSFWGATVITNLLSAIPYIGTNL
                           VEWIWGGFSDKATLTRFFAFHFILPFI AALAMVHLLFLHETGSNNPTGISSDADKI
                           PFHPYYTIKDILGALLLILVLM LVLVFS PDL LGDPDNYTPANLNTPPHIKPEWYFLF
                           AYAILRSIPNKLGGVLALVLSILILIMPL LHTSKQRSMMFRPISQCLFWILVADLLT
                           LTWIGGQPVEHPYIIIGQLASIMYFLLILVLM PAASTIENSLLKW"
ORIGIN
1  acactaatga tcaacaccgg aaagaccac ccacttataa aaattgtaaa caacgcattc
61  attgaccttc cagccccatc gaatatctca tcatgatgga actttggctc cctcctaggc
121 atctgcttaa ttctacaaat cctaacaggc ctattcctag caatacacta cacagccgat
181 acagcaacag cattctcctc cgttaccac atctgccgag acgtcaacta cggctgaatt
241 atccgatata tacatgcaaa tggagcatcc atatttttta tctgcctctt tatacacgta
301 ggacgaggcc tctactatgg atcatataca ttcttagaaa catggaatat cggagtaatt
361 cttctatttg caacaatagc tacagcattt ataggatacg tcctaccatg aggacaaata
421 tctttctgag gagcaacagt tatcacgaac ctctctcag caatcccata catcggcaca
481 aacctagtag aatgaatctg aggaggggtc tcagtagata aagcaacact caccggattc
541 tttgcttttc actttattct cccattcatc attgcagccc tcgctatagt ccacttacta
601 ttcttccacg aaacaggatc caacaacccc acaggaattt catcagacgc agacaaaatc
661 ccatttcacc cctactacac catcaaggac atcctaggag cactactact aatcctagtt
721 cttatactcc tagttctatt ctaccggac ctactcggag acccagacaa ctatacaca
781 gcaaatccac tcaacacacc cccacacatc aaacctgaat ggtacttctt attcgcata
841 gcaattctcc gatcaattcc caataaacta ggaggggttc tagccttagt cctatcaatt
901 ctaatcctaa tccttatacc cctactacac acatccaaac aacgaagcat aatatttcga
961 ccaatcagcc aatgcctatt ctgaatcctt gtagcagacc tgctaacct cacatgaatc
1021 ggaggacaac cagtcgaaca cccatacatc attatcggac aactagcatc aatcatalat
1081 ttcctgctca ccttagtact gataccagca gccagcacca ttgaaaacag tctcctaaaa
1141 tgaagataa
//

```

Tab VII: Exemple d'une séquence nucléotidique au format EMBL.

```

ID KUS60630; SV 1; linear; genomic DNA; STD; MAM; 1149 BP.
XX
AC KUS60630;
XX
DT 12-FEB-2016 (Rel. 127, Created)
DT 23-JUN-2016 (Rel. 129, Last updated, Version 2)
XX
DE Gazella arabica isolate GGF41 cytochrome b (cytb) gene, complete cds;
DE mitochondrial.
XX
KW .
XX
OS Gazella arabica (Arabian gazelle)
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia;
OC Eutheria; Laurasiatheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
OC Antilopinae; Gazella.
OG Mitochondrion
XX
RN [1]
RC Publication Status: Available-Online prior to print
RP 1-1149
RX DOI; 10.1016/j.ympev.2016.01.012.
RX PUBMED; 26826603.
RA Lerp H., Klaus S., Allgower S., Wronski T., Pfenninger M., Plath M.;
RT "Phylogenetic analyses of gazelles reveal repeated transitions of key
RT ecological traits and provide novel insights into the origin of the genus
RT Gazella";
RL Mol. Phylogenet. Evol. 98:1-10(2016).
XX
RN [2]
RP 1-1149
RA Lerp H., Klaus S., Allgoewer S., Wronski T., Pfenninger M., Plath M.;
RT ;
RL Submitted (07-JAN-2016) to the INSDC.
RL Natural History Collections, Museum Wiesbaden, Friedrich-Ebert-Allee 2,
RL Wiesbaden, Hessen 65185, Germany
XX
DR MD5; cf46524a94702a3e44d4ac4be82b2f82.
XX
CC ##Assembly-Data-START##
CC Sequencing Technology :: Sanger dideoxy sequencing
CC ##Assembly-Data-END##
XX
FH Key Location/Qualifiers
FH
FT source 1..1149
FT /organism="Gazella arabica"
FT /organelle="mitochondrion"
FT /isolate="GGF41"
FT /mol_type="genomic DNA"
FT /country="Saudi Arabia:Farasan Islands"
FT /collected_by="T. Wronski"
FT /db_xref="taxon:1229624"
FT gene 7..1146
FT /gene="cytb"
FT CDS 7..1146
FT /codon_start=1
FT /transl_table=2
FT /gene="cytb"
FT /product="cytochrome b"
FT /db_xref="GOA:A0A125RWS3"
FT /db_xref="InterPro:IPR005797"
FT /db_xref="InterPro:IPR005798"
FT /db_xref="InterPro:IPR016174"
FT /db_xref="InterPro:IPR027387"
FT /db_xref="InterPro:IPR030689"
FT /db_xref="UniProtKB/TrEMBL:A0A125RWS3"
FT /protein_id="AMD81621.1"
FT /translation="MINTRKTHPLMKIVNNAFIDLPAISNISSWVNFGLSLGICLILQI
FT LTGLFLAMHYTADTATAFSSVTHICRDVNYGWIIRYMHANGASMFIFCLFMHVGRGLYY
FT GSYTFLETWNIIGVILLFATMATAFMGYVLPWQMSFWGATVITNLLSAIPYIGTNLVEW
FT IWGGFSDVKATLTRFFAFHFILPFIIAALAMVHLLFLHETGSNNPTGISSDADKIPFHP
FT YTIKIDILGALLLILVLMMLLVFSPDLLGDPDNYTPANPLNTPPHIKPEWYFLFAYAIL
FT RSIPNKLGGVLALVLSILILMLPLLHTSKQRSMMFRPISQCLFWILVADLLTLTWIGG
FT QVPEHPYIIIGQLASIMYFLILVLMPPAASTIENSLKWK"
XX
SQ Sequence 1149 BP; 366 A; 336 C; 153 G; 294 T; 0 other;
acactaatga tcaacaccgg aaagaccacc ccacttataa aaattgtaaa caacgcattc 60
attgaccttc cagccccatc gaatatctca tcatgatgga accttggctc cctcctaggc 120
atctgcttaa tctacaaat cctaacaggc ctattcctag caatacacta cacagccgat 180
acagcaacag cattctcctc cgttaccacc atctgcccag acgtcaacta cggctgaatt 240
atccgatata tacatgcaaa tggagcatcc atatttttta ctgacctctt tatacacgta 300
ggacgaggcc tctactatgg atcatataca ttcttagaaa catggaatat cggagtaatt 360
ctcttatttg caacaatagc tacagcattt ataggatagc tctaccatg aggacaaata 420
tctttctgag gagcaacagt taccacgaac ctctctcag caatcccata catcggcaca 480
aacctagtag aatgaatctg aggagggttc tcagtagata aagcaaacct caccggattc 540
tttgcttttc actttattct cccattcacc attgcagccc tcgctatagt ccacttacta 600
ttccttcacg aaacaggatc caacaacccc acaggaattt catcagacgc agacaaaatc 660
ccatttcacc cctactacac catcaaggac atcctaggag cactactact aatcctagtt 720
cttactactc tagttctatt ctaccgggac ctactcggag acccagacaa ctatacacca 780
gcaaatccac tcaacacacc cccacacatc aaacctgaat ggtacttctt attcgcatat 840
gcaattctcc gatcaattcc caataaacta ggaggggttc tagccttagt cctatcaatt 900
ctaatcctaa tccttatacc cctactacac acatccaaac aacgaagcat aatatttcga 960
ccaatcagcc aatgcttatt ctgaatcctt gtagcagacc tgctaacact cacatgaatc 1020
ggaggacaac cagtccaaca cccatacacc attatcggac aactagcacc aatcatalat 1080
ttcctgctca tcctagtagt gataccagca gccagcacca ttgaaaacag tctcctaaaa 1140
tgaagataa 1149
//

```

Tab VIII: Exemple d'une séquence nucléotidique au format DDBJ.

LOCUS	KU560630	1149 bp	DNA	linear	MAM 10-FEB-2016
DEFINITION	Gazella arabica isolate GGF41 cytochrome b (cytb) gene, complete cds; mitochondrial.				
ACCESSION	KU560630				
VERSION	KU560630.1				
KEYWORDS	.				
SOURCE	mitochondrion Gazella arabica (Arabian gazelle)				
ORGANISM	Gazella arabica Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Laurasiatheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae; Antilopinae; Gazella.				
REFERENCE	1 (bases 1 to 1149)				
AUTHORS	Lerp,H., Klaus,S., Allgower,S., Wronski,T., Pfenninger,M. and Plath,M.				
TITLE	Phylogenetic analyses of gazelles reveal repeated transitions of key ecological traits and provide novel insights into the origin of the genus Gazella				
JOURNAL	Mol. Phylogenet. Evol. (2016) In press				
PUBMED	26826603				
REMARK	Publication Status: Available-Online prior to print				
REFERENCE	2 (bases 1 to 1149)				
AUTHORS	Lerp,H., Klaus,S., Allgoewer,S., Wronski,T., Pfenninger,M. and Plath,M.				
TITLE	Direct Submission				
JOURNAL	Submitted (07-JAN-2016) Natural History Collections, Museum Wiesbaden, Friedrich-Ebert-Allee 2, Wiesbaden, Hessen 65185, Germany				
COMMENT	##Assembly-Data-START## Sequencing Technology :: Sanger dideoxy sequencing ##Assembly-Data-END##				
FEATURES	Location/Qualifiers				
source	1..1149 /organism="Gazella arabica" /organelle="mitochondrion" /mol_type="genomic DNA" /isolate="GGF41" /db_xref="taxon:1229624" /country="Saudi Arabia: Farasan Islands" /collected_by="T. Wronski"				
gene	7..1146 /gene="cytb"				
CDS	7..1146 /gene="cytb" /codon_start=1 /transl_table=2 /product="cytochrome b" /protein_id="AMD81621.1" /translation="MINTRKTHPLMKIVNNAFIDLPA PNSSWVNFSGLLGICLIQ ILTGLFLAMHYTADTATAFSSVTHICRDVNYGWIIRYMHANGASMFICLFMHVGRGL YYSYTFLETWNIGVILLFATHATAFMGYVLPWQMSFWGATVITNLLSAIPYIGTNL VEWIWGGFSVDKATLTRFFAFHFILPFIIAALAMVHLLFLHETGSNNPTGISDADKI PFPYYTIKDI LGALLLILVLMVLFSPDLGDPDNYTPANPLNTPPHIKPEWYFLF AYAILRSIPNKLGGVLALVLSILILIMPLLHTSKQRSMFRPISQCLFWILVADLLT LTWIGGPVEHPYIIIGQLASIMYFLLILVLMPAASTIENSLLKW"				
BASE COUNT	366 a	336 c	153 g	294 t	
ORIGIN	1 acactaatga tcaacaccg aaagaccac ccacttataa aaattgtaaa caagcattc 61 attgaccttc cagccccatc gaatatctca tcatgatgga accttggtc cctcctaggc 121 atctgcttaa ttctacaat cctaaccaggc ctattcctag caatacacta cacagccgat 181 acagcaacag cattctctc cgttaccac atctgccgag acgtcaacta cggctgaatt 241 atccgatata tacatgcaaa tggagcatcc atatttttta tctgcctctt tatacacgta 301 ggacgaggcc tctactatgg atcatataca ttccagaaa catggaatat cggagtaatt 361 ctctatttg caacaatagc tacagcattt ataggatacg tcctaccatg aggacaaata 421 tctttctgag gagcaacagt tatcacgaac ctctctcag caatcccata catcggcaca 481 aacctagtag aatgaatctg aggagggttc ttagtagata aagcaacact caccgattc 541 tttgcttttc accttattct cccattcatc attgcagccc tgcctatagt ccacttacta 601 ttccttcacg aaacaggatc caacaacccc acaggaattt catcagacgc agacaaaatc 661 ccatttcacc cctactacac catcaaggac atcctaggag cactactact aatcctagtt 721 ctataactcc tagttctatt ctaccggag ctactcggag acccagacaa ctatacacca 781 gcaaatccac tcaacacacc cccacacatc aaacctgaat ggtacttctt attcgcatat 841 gcaattctcc gatcaattcc caataaacta ggagggttc tagccttagt cctatcaatt 901 ctaatcctaa tccttatacc cctactacac acatccaaac aacgaagcat aatatttcca 961 ccaatcagcc aatgcctatt ctgaatcctt gtagcagacc tgctaacact cacatgaatc 1021 ggaggacaa cagtcgaaca cccatacatc attatcggac aactagcatc aatcatatat 1081 ttctgctca tcttagtact gataccagca gccagcacca ttgaaacag tctcctaaaa 1141 tgaagataa				
	//				

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

Espèce	Région	Gène	Taille pb	Locus	
<i>G. Arabica</i> (77)	Oman	SMOC1 (5)	671	KU560886	
		PANK4(6)	458	KU560796	
	Arabie saoudite	NLRP2(5)	549	KU560751	
		CHD2(5)	669	KU560709	
		EPS15L1(5)	412	KU560666	
		cytb(3)	1149	KU560630	
		D-loop(1)	337	JX274664	
		cytb(1)	336	JX274663	
		cytb(5)	1140	KC188765	
	Palestine	ZNF618(5)	686	KU560840	
		control region (12)	324	KM523455	
		cytb(12)	790	KM523446	
		cytb(4)	1132	KC188759	
	Jordan	cytb(1)	1095	KC188747	
	Allemagne	D-loop(2)	307	JX647826	
		12S ribosomal RNA(2)	991	JX647803	
	Qatar	SPTBN1(2)	829	JX647831	
		KCAS(1)	725	JX647815	
	<i>G. bennettii</i> (73)	L'inde	COI(1)	1545	KT372099
			cytb(5)	378	KT353977
cytb(3)			415	KT369291	
cytb(2)			204	KP142700	
cytb(2)			314	DQ919166	
16S ribosomal RNA(4)			535	DQ989297	
12S ribosomal RNA (1)			386	EF133853	
misc_feature(1)			376	EF079832	
cytb(2)			751	EU434731	
L'afrique de nord		cox iii(2)	783	AF030480	
Pakistan		cytb(2)	300	AF030614	
		SMOC1(5)	671	KU560899	
		cytb(1)	1149	KU560633	
Péninsule Arabique		cytb(2)	333	HQ316151	
		cytb(1)	375	AF187698	
Arabie saoudite		ZNF618(5)	694	KU560856	
		PANK4(5)	458	KU560815	
		NLRP2(5)	549	KU560768	

		CHD2(4)	669	KU560727
		EPS15L1(4)	396	KU560683
		cytb(1)	1149	KU560634
		cytb(3)	1018	JN410357
	Iran	cytb(10)	410	KP729619
	Soudaine	mitochondrion genome(2)	16435	JN632635
<i>G. cuvieri</i> (24)	Péninsule arabique	cytb(2)	333	HQ316155
	L'afrique de nord	cox iii(1)	783	AF030475
		cytb(1)	300	AF030609
	Espagne	SMOC1(2)	671	KU560905
		ZNF618(2)	688	KU560860
		PANK4(2)	458	KU560819
		NLRP2(2)	550	KU560772
		CHD2(2)	669	KU560730
		EPS15L1(2)	396	KU560686
		cytb(2)	1149	KU560636
		cytb(2)	1095	JN410343
	D-loop(2)	784	JQ065722	
	Soudaine	mitochondrion genome(2)	16427	JN632636
	<i>G. dorcas</i> (299)	Maroc	cytb(3)	375
cytb(12)			716	JQ676952
L'afrique de nord		cox iii(3)	783	AF030484
		cytb(3)	300	AF030618
		cytb(33)	450	KM582094
		KCAS(1)	367	KM582053
Mali		mitochondrion genome(3)	16439	JN632637
		cytb(1)	1152	JN410256
Sudan		PANK4(8)	458	KU560808
		cytb(9)	1152	JN410251
Arabie saoudite		cytb(2)	336	JX274673
		D-loop(2)	250	JX274666
peninsule arabique		cytb(8)	375	AF187719
Arabie saoudite		CHD2(7)	669	KU560720
		EPS15L1(8)	396	KU560675
Qatar		cytb(27)	1152	JN410233
Palestine		ZNF618(6)	688	KU560849
		control region(31)	319	KM523547
		cytb(28)	790	KM523379
		cytb(1)	1132	KC188752
	cytb(13)	1152	JN410231	

	Allemagne	D-loop(1)	268	JX647824
		cytb(1)	336	JX647822
		12SRibosomal RNA(2)	992	JX647802
	Algerie	NLRP2(6)	550	KU560763
		SMOC1(8)	671	KU560891
		cytb(3)	1152	JN410252
		control region(52)	451	JN410262
		cytb(1)	1152	JN410255
	Tunisie	cytb(1)	1152	JN410257
	Chad	cytb(9)	1152	JN410242
	Qatar	SPTBN1(2)	829	JX647830
		KCAS(2)	707	JX647814
	Espagne	sry(1)	606	AJ003127
seqX (1)		372	GAZEXONS	
<i>G. dorcas pelzelni</i> (1)	Qatar	cytb(1)	1140	JF728768
<i>G. dorcas x leptoceros</i> (21)	L'Afrique de nord	cytb(20)	450	KM582114
		KCAS(1)	367	KM582054
<i>G. erlangeri</i> (2)	Soudaine	mitochondrion genome(2)	16434	JN632639
<i>G. gazella</i> (336)	Turquie	cytb(1)	400	JF719321
		L'afrique de nord	cox iii(5)	783
	cytb(5)		300	AF030623
	Palestine	SMOC1(3)	670	KU560882
		ZNF618(3)	685	KU560837
		PANK4(3)	458	KU560792
		NLRP2(3)	550	KU560746
		CHD2(3)	668	KU560705
		EPS15L1(3)	396	KU560661
		cytb(1)	1149	KU560629
		cytb(27)	1140	KC188776
		control region(56)	324	KM523490
		cytb(65)	790	KM523444
	TP53(1)	1021	KM233187	
	Arabie saoudite	cytb(2)	336	JX274675
		D-loop(3)	366	JX274668
		cytb(10)	1152	JN410261
	Peninsule arabique	cytb(3)	375	AF187713
	United Arab Emirates: Dubai	cytb(1)	1011	JN410353
	Allemagne	D-loop(1)	234	JX647825
cytb(45)		360	GU384870	

		D-loop(87)	211	GU384825
	Qatar	cytb(1)	981	JF728766
	Eurasie	cytb(1)	1134	AJ222682
	France	tRNA-Thr, tRNA-Pro, tRNA-Phe genes and D- loop(1)	1212	AJ235320
	Soudaine	mitochondrion genome(2)	16447	JN632640
<i>G. gazella gazelle</i> (3)	Allemand	D-loop(1)	362	JX647823
		cytb(1)	336	JX647821
		12S ribosomal RNA(1)	536	JX647811
<i>G. gazella muscatensis</i> (1)	Peninsula arabique	cytb(1)	375	AF187697
		cytb(2)	333	HQ316153
<i>G. leptoceros</i> (30)		cytb(1)	375	AF187699
	USA	MC1R(1)	936	FJ773324
	L'afrique de nord	cox iii(1)	783	AF030476
		cytb(1)	300	AF030610
	Tunisie	SMOC1(2)	671	KU560906
		ZNF618(2)	684	KU560862
		PANK4(2)	458	KU560821
		NLRP2(2)	550	KU560774
		CHD2(2)	669	KU560732
		EPS15L1(2)	420	KU560688
		cytb(2)	1025	JN410345
	Egypt	cytb(2)	1118	JN410347
	Algerie	cytb(1)	1152	JN410259
	Czech Republic	cytb(1)	1140	JF728767
	X	seqX(4)	783	KF787942
Soudaine	Mitochondrion genome(2)	16439	JN632641	
<i>G. saudiya</i> (10)	L'afrique de nord	cox iii(1)	783	AF030481
		cytb(1)	300	AF030615
	Arabie saoudite	cytb(6)	375	AF187722
		cytb(2)	404	JN410323
<i>G. spekei</i> (29)	USA	ytochrome c oxidase II(1)	684	GSU18824
	L'Afrique de nord	cox iii(1)	783	AF030474
		cytb(1)	300	AF030608
	UAE: Dubai	SMOC1(1)	669	KU560888
		ZNF618(3)	688	KU560846
		PANK4(4)	458	KU560802
		NLRP2(4)	544	KU560754
	CHD2(4)	669	KU560714	

		EPS15L1(4)	397	KU560670
		cytb(4)	1149	KU560632
	Soudaine	mitochondrion genome(2)	16440	JN632642
<i>G. subgutturosa</i> (143)	Mongolie	COI(4)	664	KX859293
		COI(1)	605	KU527897
		cytb(1)	375	DQ269164
		cytb(1)	1083	KU560652
		NLRP2(2)	540	KU560764
		EPS15L1(2)	396	KU560679
	Eurasie	cytb(1)	1143	AF036282
	Allemagne	PrP(1)	679	AF117313
	Azerbaïdjan	cytb(1)	333	HQ316158
	Turkistan	cytb(2)	332	HQ316159
	Peninsule arabique	cytb(3)	375	AF187716
	L'afrique de nord	cox iii(2)	783	AF030478
		cytb(2)	300	AF030612
	Qatar	SMOC1(2)	665	KU560898
		ZNF618(1)	684	KU560854
		PANK4(2)	458	KU560812
		CHD2(2)	669	KU560724
		cytb(1)	1128	KU560653
	Soudaine	mitochondrion genome(3)	16436	JN632643
	Iran	cytb(1)	333	HQ316157
		cytb(8)	410	KM387306
		COXIII(1)	384	KM948600
		cytb(1)	258	KM983326
		cytb(15)	470	KF790616
		control region(7)	480	GU993273
	la Chine	multiple endocrine neoplasia type I(1)	366	KF027381
		COI(9)	644	KC679027
		cytb(32)	1143	KM978991
		D-loop(17)	984	KJ499836
		pklr(1)	305	JN167349
	Turquie	misc_feature	400	JF719320
		DRB(3)	262	KC309407
Russia	16S ribosomal RNA(1)	1423	AF538359	
	12S ribosomal RNA(1)	955	AY141134	
USA	12S ribosomal (1)	241	GSU87007	
	16S ribosomal (1)	340	GSU87057	
X	seqX (8)	512	AX084559	

<i>G. subgutturosa marica</i> (94)	Oman	cytb(3)	333	HQ316160
	UAE	cytb(3)	333	HQ316161
	Peninsule arabique	cytb(2)	375	AF187718
	Arabie saoudite	SMOC1(3)	670	KU560910
		ZNF618(3)	689	KU560865
		PANK4(3)	458	KU560824
		NLRP2(2)	548	KU560775
		CHD2(2)	669	KU560734
		EPS15L1(2)	396	KU560689
		cytb(3)	1116	KU560654
		cytb(2)	336	JX274677
		D-loop(2)	358	JX274670
		cytb(20)	390	JN376064
		16S ribosomal RNA(20)	879	JN376044
	D-loop(19)	1106	JN376024	
	Qatar	12S ribosomal RNA(2)	992	JX647804
SPTBN1(2)		782	JX647833	
KCAS(1)		725	JX647816	
<i>G. subgutturosa sairenses</i> (4)	la Chine	D-loop(4)	785	KC934198
<i>G. subgutturosa subgutturosa</i> (76)	Azerbaïdjan	D-loop(10)	980	HQ185411
	Ouzbékistan	D-loop(66)	973	HQ185457
<i>G. subgutturosa yarkandensis</i> (20)	la Chine	D-loop (20)	785	KC934194
<i>G. sp.</i> (18)	Peninsule arabique	cytb(6)	375	AF187721
	Iran	cytb(4)	383	KR873132
	X	seqX(8)	512	BD261934